An Investigation Into The Economic Impact of a Professional Sports Franchise

by

Albi Bilali

An essay submitted to the Department of Economics
in partial fulfillment of the requirements for
the degree of Master of Arts

Queen's University
Kingston, Ontario, Canada

August 2019

Acknowledgments

This essay would not have been possible without the help of my supervisor, Alper Arslan, who was the first to introduce me to the synthetic control methodology. I would also like to thank my fellow MA classmates, from whom I have learned a great deal in the past year.

Contents

1 Introduction

On July 22nd, 2016 the U.S. Bank Stadium, new home of the NFL team the Minnesota Vikings officially opened its doors after years of tumultuous back and forth between the Vikings, the Minneapolis local government, and the Minnesota state government. The roughly $1.06 billion dollar stadium is a perfect example of the complex relationship sports organizations have with the municipalities that house them. The stadium is home to the Vikings, but it is jointly owned by the city of Minneapolis and the state of Minnesota, who together provided nearly half of the construction costs for the stadium. While such a large public investment into a luxury sport stadium may seem odd, this type of relationship is fairly common in all of the major sports leagues in the United States. State and local governments have a desire to attract major sport franchises, and are often willing to provide incentives to teams to get them to move in/stay. Sports teams, on the other hand, enjoy a form of monopoly power, since leagues limit the number of teams that can participate within them. This allows them to make demands from cities, who are typically willing to meet them to keep the teams in town. The most obvious reason why cities want to attract/retain teams is simply because their constituents demand it. This isn't much of a shock since professional sports is one of the largest industries world wide, with several of the top teams being worth billions of dollars (Badenhausen, 2019). A second reason might be that housing a large sport franchise is seen as a status symbol for cities. This too isn't surprising since sports franchises often form a part of a city's history and add to its prestige, two such examples are the Chicago Bulls NBA team and the New York Yankees MLB team, both of whom have worldwide recognition. However these aren't the benefits legislators cite when they propose funding a new stadium, or providing tax-exempt status to a local sports franchise. They instead claim that such investments are sound and that they benefit the local economy. This third reason, which will be the focus of this paper, is less obvious and far more controversial.

Advocates of sports franchises claim that while the costs of stadiums can be high, the direct and indirect benefits they provide will exceed the costs in the long run (Baade and Dye, 1990). The direct benefits include revenue from ticket sales, luxury seating, and in-stadium concession stands. Indirect benefits are more vague, but can generally be summarized as increased economic activity in the area surrounding a sports stadium, or increased economic activity as a result of a sports franchise. One recent example of the latter is a Statistics Canada report crediting the playoff success of the Toronto Raptors for providing a boost to the Canadian economy (Evans, 2019). On the other hand, critics of sports investments claim that while the revenue brought in by stadiums may cover operating costs, they will never cover the initial construction costs. Additionally, stadiums located in the heart of large cities have indirect costs, like the typically forsaken property taxes that could be brought in were the land used for a different venture, and increased traffic and congestion. Despite these concerns, it is clear that the municipal governments seem to be in favour of housing sport franchises. Seattle, Chicago, Atlanta, and Dallas are just a few of the cities that have been more than willing to offer teams hundreds of millions of dollars to help pay for a stadium in hopes of attracting or retaining an existing franchise (Gius and Johnson, 2001). Cities provide this funding either through the sale of bonds or often by increasing or imposing a local tax, justifying the costs by citing the aforementioned benefits. While governments, and perhaps even citizens, believe a sports team is a profitable investment, economists remain more skeptical.

Most of the ex-post analysis conducted by economists and independent parties has found that sports franchises and stadiums do not provide any significant economic benefit, and stadiums construction costs typically run far higher than projections. Unfortunately, the methodology these researchers use has significant holes and gaps, which this paper will seek to remedy. Most of the research has employed a multivariate regression analysis to identify whether constructing a new sports stadium or obtaining a new franchise has a significant impact on a region's local economy. Unfortunately, this type of analysis is vulnerable to omitted variable bias as large cities have complex economies, and all the contributors to economic growth cannot possibly be captured by the limited data sets that exist. Additionally, sports franchises do not relocate very often, due to the high cost of moving; and new franchises rarely open, due to league regulations that limit the number of existing teams. This leads to a lack of general variation over shorter periods of time, but unfortunately looking at a longer period of time only increases the bias that omitted variables may have on the analysis. This study will attempt to circumvent this issue by using the synthetic control technique from Abadie and Gardeazabal (2003) to investigate whether Oklahoma City benefited from acquiring the Seattle SuperSonics NBA team (now the Oklahoma City Thunder) from the city of Seattle in 2008.

The city of Seattle had been home to the SuperSonics from their inception in 1967 until their eventual relocation to Oklahoma city in 2008. After a period of declining ticket sales and financial losses, the Seattle SuperSonics sought roughly $220 million of funding from the state of Washington to renovate their then home KeyArena. After they were unsuccessful in obtaining the funding, the team was sold to a business group composed mostly of Oklahoma City businessmen. Though it was stipulated that the new ownership would attempt to find a suitable home in the city of Seattle, they relocated to Oklahoma City where they received $120 million dollars to renovate the existing Ford Center venue, tax-breaks and other incentives in the form of stadium revenue from the city. This move provides the perfect opportunity to investigate the whether public funding into a sports team and venue provides a positive economic return for a local municipality. The purpose of this article is to provide an answer to this question using data from the American Community Survey (from years 2005 to 2017) and a synthetic control methodology.

Synthetic control studies were first popularized by Abadie and Gardeazabal in a 2003 article where they investigated the impact of terrorism on the economic growth of the Basque region in Spain (Abadie and Gardeazabal, 2003). They use a weighted average of other Spanish regions to construct a counter-factual to the Basque region, and thus observe the economic impact of terrorism. This study will similarly use other US Metropolitan Statistical Areas (MSAs) to construct a counter-factual for Oklahoma City, which can then be used as a control to determine whether the Seattle SuperSonics had a positive impact on the Oklahoma City economy.

The rest of this paper will continue as follows. Section 2 will review some of the existing literature on the subject of the economic impact of sport investments. Section 3 will provide a short explanation of the synthetic control methodology, and the data set that will be used for the study. Section 4 will showcase the results of the analysis, and lastly section 5 will provide the conclusions and some potential next steps.

2 Literature Review

Though somewhat limited, the existing literature investigating the economic impact of sports franchises and investments can be generally divided into an ex-ante and ex-post category. With the former being popular until roughly the late 1980s, and the latter being popular from the 1990s onwards. The ex-ante studies are impact studies commissioned by sports organizations or local governments who are investigating the effect of building a sports venue. Baade and Dye (1988) provide an excellent review of the impact study literature, highlighting the short-comings of the methodology. They note that due to the prospective nature of impact studies, a large number of assumptions must be made. These assumptions range from the prospective attendance of games and events to an estimate of the multiplier effect that the stadium will have on local spending. These studies also fail to mention that construction costs typically end up surpassing projected costs by quite a large margin. Additionally, impact studies implicitly assume that all spending following the construction of the venue will be new spending, rather than spending redirected from other forms of entertainment (Baade and Dye, 1988). While impact studies constitute much of the economic investigation prior to the 1990s, their prospective nature and potential conflicts of interest diminish much of their credibility. It is for this reason that much of the recent literature has been of a retrospective nature, and has relied on regression analysis to discern whether teams and stadiums have a significant economic impact.

Baade and Dye (1990) were some of the first researchers to utilize a retrospective multivariate regression approach when investigating the impact of stadiums and teams. They use census data on nine metropolitan areas over a number of years to investigate the effect of stadium development and team relocation on the local economy and its development. Their findings vary between different locations, however they conclude that generally obtaining a sports team does not have a significant impact on the local economy. Constructing or investing in a stadium does not either, and in fact it sometimes has a negative impact on the economy. They theorize that stadium development possibly leads to an economy developing more low-pay service jobs, in a prime city location (Baade and Dye, 1990). Unfortunately, Baade and Dye's analysis suffers from a small sample size, and it possibly speaks more to those selected communities than sports teams in general.

Gius and Johnson (2001) seek to remedy the small sample size issue by conducting a large cross-sectional analysis on all US cities with a population greater than 25 000 to determine if being host to a sports team has a significant impact on the income of city citizens. They find that while being host to one franchise does not have a statistically significant impact on per capita income, being host to two or more sports teams does (Gius and Johnson, 2001). The benefit of their cross sectional approach is the size of their sample and the fact that they can control for many variables by choosing a year where data is readily available. However, the downside is there are inevitable fixed city factors that influence local income that the researchers cannot account for, and that the causality is unclear. Do cities with many professional sports teams have a higher per-capita income because of said teams? Or do professional teams choose to locate in cities where individuals have a higher income?

Santo (2005) on the other hand, is critical of the prior analysis conducted by Baade and Dye and offers some evidence that modern stadiums may be a benefit to the local economy. Santo notes that Baade and Dye and other similar articles typically used data on stadiums and teams from the 1960s to the late 1980s. He believes that the type of stadium constructed during that time is incomparable to today's modern stadium, both in structure and location. Stadiums constructed from the 1960s to the 1980s tended to be bare bones, utilitarian venues that were constructed far away from the city center. Where as today's modern stadiums are luxurious and typically tend to be located in the city's downtown area. Santo believes this allows venues to serve as tourist attractions, and brings in income from non-residents who come for the stadium but stay to enjoy other forms of entertainment the city has to offer (Santo, 2005). He conducts a similar analysis to Baade and Dye (1990) but uses a a larger number of cities and more recent data. His findings contradict the conclusions of Baade and Dye. Santo finds that the construction of new baseball stadiums is positively correlated with per capita income, although obtaining a baseball team is not statistically significant. Santo hypothesizes that this may be due to the larger number of games in an MLB season as opposted to other sports leagues. He concludes that this is evidence that the common consensus among economists may be misguided, and that the context in which a stadium is constructed may be important in determining its economic impact.

A more contextual approach is taken by Austrian and Rosentraub in their 2002 paper. Rather than look at stadium constructions on a larger, cross-country scale, they choose to look closely at how four different US cities attempted to revive their downtown cores. From this viewpoint the goal of a stadium or a sports franchise is not to raise the per capita income of a city, but to slow the decentralization of a city's economy, a trend that is observed in all metropolitan areas as suburbs become more and more popular. The authors compare the cities of Cleveland and Indianapolis to Cincinnati and Columbus. All four municipalities were concerned with the outflow of jobs from the city core to suburbs, and took steps to combat it. Cleveland and Indianapolis invested heavily in sports entertainment and tourism while Cincinnati and Columbus did not. The author's find that the efforts of Cleveland and Indianapolis did seem to slow the outflow of jobs to suburban areas in the above cities (Austrian and Rosentraub, 2002). However, they do note that the four cities faced different challenges, and that the evidence was far from conclusive, even for Indianapolis and Cleveland.

Though other articles exist that investigate this question, they typically use one of the techniques highlighted above and reach similar conclusions as the above authors. Recent articles have rekindled the debate on whether stadiums and teams are a good investment, but additional research is also needed due to the methodological flaws of previous research. To truly answer the question of whether a sports investment is a net benefit, we need to use techniques that can provide causal inference, or at the very least a technique that can provide a robust correlation result. A simple cross-sectional study will always be vulnerable to omitted variable bias, as comparing cities to one another will undoubtedly fail to capture some of the variables that influence per capita income growth in cities. A time-series analysis on the other hand will fail to capture the fact that teams often consider the future projections of a city's economy when they choose to relocate. This article will contribute to the existing literature by using the synthetic control technique to control for these two factors. By creating a synthetic control, we can simulate a counter-factual that closely resembles our original city (Oklahoma City), and hopefully obtain an unbiased estimate of the impact of obtaining a major sports franchise.

3 Methodology and Data Source

3.1 A Brief Formal Introduction to Synthetic Control

The methodology outlined in this section will be a somewhat condensed version of the model outlined in Abadie et al. (2010), where the authors showcase how a synthetic control can be used for a comparative case study. The authors provide much of the theoretical justification for the methodology, a great deal of which won't be listed in this article.

Suppose we observe $J+1$ regions over a period of time $\{t = 1, \ldots, T\}$, and one of these regions is exposed to a treatment or intervention of interest beginning at time $t = T_0 + 1$. In this study, the regions are Metropolitan Statistical Areas and the intervention of interest is the relocation of the Oklahoma City Thunder to Oklahoma City, the region of interest. We will order the regions such that the region of interest is the 1st region, and the remaining J regions serve as our potential controls. Let Y_{it}^N be the outcome that would be observed for region i at time t in the absence of the intervention. In our study, this outcome will be the relative median household income (RMHI) of the MSA. Additionally, let Y_{it}^I be the outcome observed for region i at time t if the region is exposed to the treatment. If we let T_0 be the number of pre-intervention periods, with $1 \leq T_0 < T$ then we clearly have $Y_{it}^I = Y_{it}^N$ for all $t \epsilon \{1, \ldots, T_0\}$ and all observed regions $\{1, \ldots, J+1\}$. Lastly, let $\alpha_{it} = Y_{it}^I - Y_{it}^N$ be the effect of the intervention for unit i at time t. Then, because only the region of interest (Oklahoma City) is exposed to the intervention, and only after time period T_0 our goal is to estimate $(\alpha_{1T_{0+1}}, \ldots, \alpha_{1T})$ where $\alpha_{1t} = Y_{1t} - Y_{1t}^N$ since the treated outcome is observed. Essentially, we want to estimate the impact of the treatment, which is equal to the observed RMHI minus the RMHI of the unobserved non-treatment outcome (i.e. the counterfactual). Thus, in order to estimate α_{1t} we simply need to estimate Y_{1t}^N.

Next, consider a vector of weights $W = (w_2, \ldots, w_{J+1})'$, such that each weight is greater than or equal to 0 and the sum of all weights is equal to 1. Each particular W represents a potential synthetic control for our region of interest. In other words, a synthetic control is essentially just a weighted average of the other control regions. Suppose that our untreated outcome Y_{it}^N can be modeled by the following equation:

$$Y_{it}^N = \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it} \tag{1}$$

Where Z_i and μ_i represent observable and non-observable covariates, the former of which are assumed to be unaffected by the treatment. Then the outcome variable for our synthetic control can be modeled by:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \theta_t \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt} \tag{2}$$

The main assumption then is that there exists a set of weights $(w_2^*, \ldots, w_{J+1}^*)$ such that:

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_1 1 \tag{3}$$

$$\sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_2 2 \tag{4}$$

$$\ldots \tag{5}$$

$$\sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0} \tag{6}$$

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1 \tag{7}$$

In words and for our specific context, this states that there exists a set of weights for which the weighted average of control MSA region median household income will sum to the pre-treatment Oklahoma City median household income. Additionally, the weighted average of the observed covariates will also be equal to the Oklahoma City covariates. Abadie et al. (2010) show that for this vector of weights W^*, $Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}$ approaches 0 if the number of preintervention periods is large relative to the scale of the shocks (Abadie et al., 2010). With this we arrive at our estimator for α_{1t} which is:

$$\hat{\alpha_{1t}} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \tag{8}$$

The impact of the treatment can be estimated by subtracting the observed synthetic control post-treatment dependent variable outcome from the observed post-treatment outcome. This type of comparison between the treated region and the synthetic control is very similar to a Difference-in-Differences or Two-Way Fixed Effects methodology. However, the synthetic control method additionally allows for the region specific fixed effects (λ_t) to vary over time, where as a DID framework requires the region specific fixed effect to be time invariant. Often it is the case that no W^* which perfectly satisfies equations 3-7 exists. In this case Abadie et al. (2010) show that the results may still hold if the synthetic control approximately fits the pretreatment outcome. They ultimately leave it to the researcher to decide if the characteristics of the treated units are sufficiently matched by the synthetic control. They also advise choosing control regions that are similar to the region of interest. If the linear model specified in equation 1 does not hold over the entire set of regions, then interpolating across regions may lead to biased results.

With the theoretical justification for the methodology covered, we are now left with the task of determining how exactly to choose our set of weights, W^*. Let the $T_0 x1$ vector $K = (k_1, \ldots, k_{T_0})'$ define a combination of pre-intervention outcomes such that $\bar{Y}_i^K = \sum_{s=1}^{T_0} k_s Y_{is}$. Consider M such linear combinations defined by $K_1, \ldots, K_M$ and let $X_1 = (Z_1', \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})'$ be an (kx1) vector of characteristics for our region of interest (Oklahoma City), $k = M + z$ where z is the number of observed covariates. This formulation looks puzzling, but an obvious choice for each $\bar{Y}_1^{K_j}$ is simply Y_{1j} for $j = 1, \ldots, T_0$. In other words, we can simply use the observed outcome values as our pre-treatment characteristics. However, this is not necessary and it can be computationally preferable to use a smaller set of characteristics. Lastly, let X_0 be a similar (k x J) matrix that contains the same variables as X_1 but for the unaffected J regions. Then our goal is to find a W^* to minimize the distance between X_1 and $X_0 W$. Abadie et al. (2010) employ the following measure:

$$||X_1 - X_0 W||_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)} \tag{9}$$

Where V is some symmetric, positive semidefinite (k x k) matrix. While the inference is correct for any V, Abadie et al. recommend a data-driven method for choosing V. They recommend choosing the V that minimizes the mean square prediction error of the outcome variable during the pre-treatment period (Abadie et al., 2011). This article will follow a similar approach. Thus, the synthetic control procedure can be summarized as follows:

1. Using both the treated and control regions, construct the vector X_0 and X_1
2. Using the data-driven approach, find the optimal value of weights W^* to construct the synthetic control
3. Using our new synthetic control, estimate $\hat{\alpha}_{1t}$ for $t = T_0 + 1, \ldots, T$

Fortunately, Abadie et al. (2011) have developed a convenient R-package, named **Synth** that carries out this exact procedure. The synthetic control analysis in this article will then use a number of MSA regions similar to Oklahoma City to simulate the counter-factual Oklahoma City that never obtained an NBA team. By comparing the actual outcome to the synthetic outcome, we can see whether obtaining the major sports franchise, and consequently funding the Ford Center renovations, had an impact on the local Oklahoma City economy.

3.2 Data & Summary Statistics

For this study, data was gathered from the American Community Survey on the largest 64 Metropolitan Statistical Areas in the United States from the years 2005-2008 and 2010-2017. The American Community Survey is an annual survey conducted by the US Census Bureau. Every year, roughly 3.5 million households are surveyed by the Census Bureau, who then use the information to publish projections and estimates at different aggregate levels (e.g. county, state, MSA). Because the information published by the Census Bureau is an estimate, there is a margin of error for all independent and dependent variables. However, as this study will focus on the most populated MSAs, the margin of error has a smaller impact as it is more likely that households in said MSAs are surveyed.

Tables 1 and 2 display the average statistics for the variables used in the analysis for the years 2005 and 2017 respectively. In the tables, the Bachelor Degree variable refers to the percentage of individuals whose highest completed level of education is an undergraduate degree. Similarly, the High School Diploma variable refers to the percentage of individuals whose highest completed level of education is High School. Thus, a lower level in the High School Diploma variable could indicate

	NFL Franchise (30)	NBA Team (27)	MLB Team (25)	None (26)
Asian/PI	0.049	0.058	0.058	0.059
Black	0.163	0.153	0.143	0.112
Bachelor Degree	0.196	0.200	0.202	0.175
High School Diploma	0.274	0.263	0.267	0.284
Median Household Income	$63, 996	$64, 579	$66, 028	$60, 398
Population	3, 295, 217	3, 317, 632	3, 369, 081	1, 175, 355
Poverty Rate	0.114	0.118	0.112	0.129
State Income	$58, 207	$58, 243	$59, 653	$58, 063
Unemployment Rate	0.054	0.055	0.055	0.055
MHI/SI	1.102	1.110	1.108	1.043

Table 1: Summary statistics of different metropolitan areas in 2005. All dollar values are in 2017 dollars

	NFL Franchise (27)	NBA Team (27)	MLB Team (25)	None (26)
Asian/PI	0.062	0.065	0.069	0.069
Black	0.162	0.160	0.148	0.126
Bachelor Degree	0.224	0.225	0.227	0.198
High School Diploma	0.250	0.244	0.243	0.261
Median Household Income	$67, 535	$66, 937	$69, 548	$64, 744
Population	4, 635, 696	4, 691, 402	5, 190, 385	1, 323, 996
Poverty Rate	0.120	0.125	0.116	0.132
State Income	$62, 797	$62, 437	$64, 817	$61, 966
Unemployment Rate	0.032	0.033	0.033	0.035
MHI/SI	1.07	1.069	1.069	1.056

Table 2: Summary statistics of different metropolitan areas in 2017. All dollar values are in 2017 dollars

either a larger proportion of High School drop outs or a larger proportion of college graduates. State Income refers to the median household income of the state the MSA is located in. This can be problematic as some MSAs are on the border of two states and thus have households in both of them. In this case, the more applicable state was chosen for the MSA. For example, the Kansas City MSA includes residents in both Missouri and Kansas, but for that observation only Missouri's state income is used. As can be seen in the tables, MSAs that have a professional sports team tend to be wealthier on average, more populated, more diverse, have lower levels of poverty and have more educated citizens. This could be due to the impact of the sports team on the MSA, but is more than likely due to teams being selective in which MSA they choose to reside.

Working with data at the MSA level has its benefits and drawbacks. One of the benefits is that a Metropolitan Statistical Area is designed with economic activity in mind. The Census Bureau aims to capture the characteristics all individuals who participate in a certain economic area. This means that there is a lower chance of the treatment spilling over to neighbouring MSAs. Unfortunately, because of this definition, MSAs are also frequently redefined. For this dataset, care has been taken to avoid using MSAs that have undergone significant revisions over the 2005-2017 time period. A second drawback is that the American Community Survey data only goes as far back as 2005. Aggregate data for MSAs prior to 2005 does exist, but yearly entries do not exist, and typically different characteristics are calculated for different years. For example, some surveys contain population measures for the years of 2000 and 2001, while others contain unemployment rates for year of 2003. This fragmentation of data leads to a need to extrapolate values for variables which are not present in the available data sets. Additionally, as these pre-2005 surveys are conducted by different parties, the definition of an MSA is not consistent throughout them. Rather than extrapolate data, I have chosen to simply omit the prior years, but this comes at the cost of the synthetic control estimator being potentially less accurate.

All of the MSAs will be used to conduct a preliminary analysis to check that the 2005-2017 dataset matches some of the results presented in the literature. For the synthetic control portion of the analysis, a subset of 24 MSA regions will be used as the control pool. A full list of regions will be listed in section 4 of the paper, along with the weights placed on each MSA region. The pool of control regions is restricted to MSAs that share somewhat similar characterstics to Oklahoma City. This is both in an effort to avoid over-fitting, and to avoid the possible interpolation biases highlighted in section 3.1. Table 3 provides summary statistics for Oklahoma City in comparison to the pool of controls for the year 2005. As can be seen, the characteristics are similar but the average control region tends to be more populated, and wealtheir than Oklahoma City. Figures 5 and 6 in appendix A also provide a brief look at the changes in median household income and relative median household income in Oklahoma City, Seattle, and Tulsa (another city in Oklahoma), the vertical line indicates the first year after the Seattle SuperSonics became the Oklahoma City Thunder. All 3 cities seem to follow somewhat similar patterns, with Tulsa matching Oklahoma City very closely. In fact, were it not for the spillover of the treatment, Tulsa would be a potentially viable DID control region for Oklahoma City. While alone the figures do not provide any key insight, they do show that all 3 MSAs seem to follow similar trends, despite the fact that one lost an NBA franchise, one gained an NBA franchise, and the third had no treatment.

	Oklahoma City	Control MSA Regions (24)
Asian/PI	0.030	0.087
Black	0.100	0.068
Bachelor Degree	0.181	0.197
High School Diploma	0.271	0.263
Median Household Income	\$50, 473	\$67, 790
Population	1, 156, 578	2, 151, 195
Poverty Rate	0.155	0.108
Median State Household Income	\$47, 368	\$63, 042
Unemployment Rate	0.046	0.053
Median Household Income/State Income	1.07	1.077

Table 3: Summary statistics of Oklahoma City and the pool of control regions for the year 2005

4 Analysis & Results

Following the methodology of Baade and Dye (1990) and Gius and Johnson (2001) the variables of interest will be both the Median Household Income and the MHI/SI or RMHI variable. If a public investment in a sports team results in an increase in wealth to the local area, then we should see the MSA Median Household Income increase relative to the Median State Household Income, thus RMHI should be postively correlated with such an investment. The analysis will be broken up into two sections. The first section will attempt to replicate Baade and Dye (1990) and Gius and Johnson (2001) to determine if the modern data set displays similar characteristics that the existing literature describes. The second section will contain the results of the synthetic control methodology to determine if acquiring the Seattle SuperSonics had an economic impact on Oklahoma City.

4.1 Replication and Naive Regressions

The first part of the analysis will be replicating Baade and Dye (1990) using the Oklahoma City and Seattle MSA regions. This entails using the 2005-2017 data to estimate the following linear regression equations:

$$\mathrm{MHI}_t = X_t\beta + \varepsilon_t \tag{10}$$

$$\mathrm{MHI}_t/\mathrm{SI}_t = \mathrm{RMHI}_t = X_t\beta + \varepsilon_t \tag{11}$$

Where MHI is the Median Household Income for the MSA, and SI is the State Median Household Income. Two versions of equation 10 were estimated, with different controls. The first controlled for the proportion of Asian/Pacific Islander residents in the MSA, the proportion of Black residents in the MSA, the proportion of individuals whose highest level of education was an undergraduate degree, the proportion of individuals whose highest level of education is a high school diploma, an indicator variable identifying whether the MSA possesses an NBA team, population, poverty rate, unemployment and a constant term. The second version of equation 10 used the same covariates but omitted the High School Graduate and the Poverty Rate variables. The reason for estimating this second version is that due to the low number of observations including a large number of covariates increases the size of the standard errors, thus reducing the chance of finding a significant result. Of course, reducing the number of controls increases the chance of biased estimates, so there is a trade-off that has to be accepted.

The results for this analysis can be seen in Table 4 and Table 6 (in Appendix A). Table 4 contains the results for Oklahoma City. For equations 10 version 1 and 11 the coefficient for an NBA Franchise is not found to be statistically significant at an acceptable confidence level. This is consistent with the literature and with the conclusions of Baade and Dye (1990). However, the fact that so many of the coefficients are negative has some troubling implications about the analysis. While both the R^2 and Adjusted R^2 are high for the models, I believe the financial recession of 2007-2008 could be impacting the dataset. This would explain why things that are typically positively correlated with income, like the proportion of college graduates, have a negative coefficient value. The coefficient estimate for an NBA Franchise in equation 10 version 2, model number 3 in Table 4, is statistically significant at the 5% level, and again is negative. This is likely due to the aforementioned trade-off that comes with controlling for fewer covariates. The significant and negative coefficient is likely biased by the recession, which occurred at a similar time as the SuperSonics move to Oklahoma City. This also explains why the NBA Franchise estimate for Seattle **(Table 6 model 3)** is positive and statistically significant. This variable is really just comparing the pre-2008 Median Household Income of the cities to the post-2008 Median Household Income, while not controlling for the Great Recession of 2006-2008.

A second analysis that may be of interest is a pooled fixed-effects regression analysis across all 64 MSA regions. This would be a similar, though somewhat more reliable version of the model estimated by Gius and Johnson (2001). The equation that was estimated was a dummy variable regression equation, which on top of the fixed effect covariates also included a indicator variables for an MLB team, NBA Team, NFL Team, and more than 1 professional sports team. Table 7 in Appendix A contains the results for this regression, which are consistent with the common findings

	Dependent variable:		
	Median Household Income	MHI/SI	Median Household Income
	(1)	(2)	(3)
Asian/PI	−490,473.100	−11.682	−309,229.500
	(494,128.400)	(9.148)	(337,729.300)
Black	49,146.920	-21.039^{***}	137,559.600
	(188,020.100)	(3.481)	(110,569.700)
College Graduate	−44,083.460	-9.186^{**}	−4,775.570
	(96,905.190)	(1.794)	(58,115.060)
High School Graduate	−51,547.840	-6.016^{**}	
	(84,631.260)	(1.567)	
NBA Franchise	−4,639.556	−0.023	$-4{,}261.917^{**}$
	(2,774.227)	(0.051)	(1,562.028)
Population	0.073	0.00000	0.067^{**}
	(0.036)	(0.00000)	(0.023)
Poverty Rate	−29,831.080	-6.016^{**}	
	(80,713.620)	(1.494)	
Unemployment	69,474.160	1.370	67,426.160
	(101,410.700)	(1.877)	(84,060.480)
Constant	−695.037	6.391^{***}	−33,585.400
	(55,934.110)	(1.036)	(22,177.360)
Observations	12	12	12
R^2	0.935	0.956	0.925
Adjusted R^2	0.763	0.837	0.836
Residual Std. Error	932.894 (df = 3)	0.017 (df = 3)	776.928 (df = 5)
F Statistic	5.423^{*} (df = 8; 3)	8.063^{*} (df = 8; 3)	10.313^{**} (df = 6; 5)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Naive Regression results for Oklahoma City

in the literature. Interestingly, the MLB Franchise coefficient is positive and statistically significant at the 1% level in both models. This is perhaps more evidence to show that the number of games in a season does play a role in the economic success of the stadium, as the MLB has far more games per season than the NFL and NBA. It should be noted that there is not a great deal of variation for MLB teams in this data set. In fact no MLB team relocated during the 2005-2017 time period. Overall the findings so far suggest that our data set exhibits similar characteristics to the data sets used by previous researchers. This will serve as a robustness check for the synthetic control findings.

4.2 Synthetic Control Analysis

For the synthetic control analysis the data-driven methodology explored in section 3.1 was implemented. A vector V was chosen to minimize the Mean Square Prediction Error (MSPE) of the outcome variable during the pretreatment (2005-2008) period. That is, V was selected to minimize $||X_1 - X_0W||_V = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)}$ for an optimal W. The pool of control variable and their respective weights are outlined in Table 5. As can be seen, despite including a large number of control regions, the synthetic control is composed mainly of two regions, Austin and Syracuse, whose respective weights add up to 0.96 of a possible 1. This suggests that the pool of regions is either too large, or it consists of regions that are too different from Oklahoma City to be of any use. However, filtering the pool after running a preliminary analysis should be avoided, as it is somewhat akin to p-hacking in a regular regression framework.

	Weight		Weight
Albany-Schenectady-Troy	0	New Haven-Bridegport-Stamford	0
Austin	0.314	Phoenix	0
Boston Metro Area	0.001	Portland-Vancouver	0.002
Cincinnati	0	Providence (RI)	0.001
Denver	0.002	Riverside-San Bernardino	0.002
Grand Rapids-Muskegon-Holland	0	Rochester	0.001
Hartford	0	Sacramento	0.002
Honolulu	0.001	Salt Lake City	0
Kansas City	0.001	San Diego	0.003
Las Vegas	0.001	San Francisco	0
Los Angeles-Long Beach	0.022	San Jose	0.001
Minneapolis	0.001	Syracuse	0.645

Table 5: Weight composition of Synthetic Control for Oklahoma City

Figure 1 shows the synthetic control vs the actual observed outcome for the Median Household Income of Oklahoma City. Unfortunately the synthetic control does not do an adequate job simulating

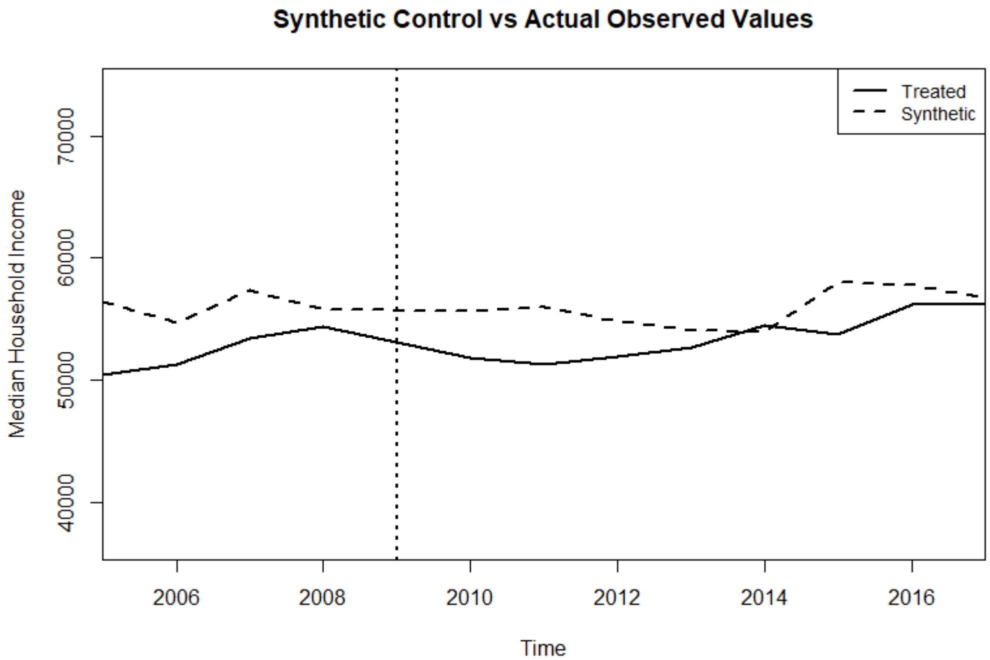

Figure 1: Synthetic Control vs Actual Outcome for Median Household Income

the pre-treatment outcomes, and thus drawing conclusions from this form of analysis for this variable would be unwise. We will instead focus on the MSA Median Household Income/State Median Household Income variable, referred to as the Relative Median Household Income (RMHI) from here onwards. Figure 2, displays the synthetic control RMHI vs the observed RMHI for Oklahoma City. Immediately it is obvious this is a much better fit in the pre-2009 period than the synthetic control in Figure 1, though it's not quite a fit to the same level that Abadie et al. (2010) find in their paper. The synthetic control reinforces the general consensus of the literature, and the regression analysis in section 4.1. The economic impact of acquiring the Seattle SuperSonics seems to be marginal, if it exists at all. Figure 3 shows the gap between the synthetic control RMHI and the observed Oklahoma City RMHI. If the Oklahoma City Thunder did have any impact on the Oklahoma City economy, it seems to be a negative one rather than a positive one. This could be evidence that sport franchises do lead to the creation of more low-wage local jobs, however it's more likely that the team simply did not have a statistically significant impact.

While the economic literature and regression analysis point to the Thunder not having an impact on the OKC economy, it is important to consider some other potential explanations for this finding. One aspect that might bias the results is that due to Hurricane Katrina in 2005, the New Orleans

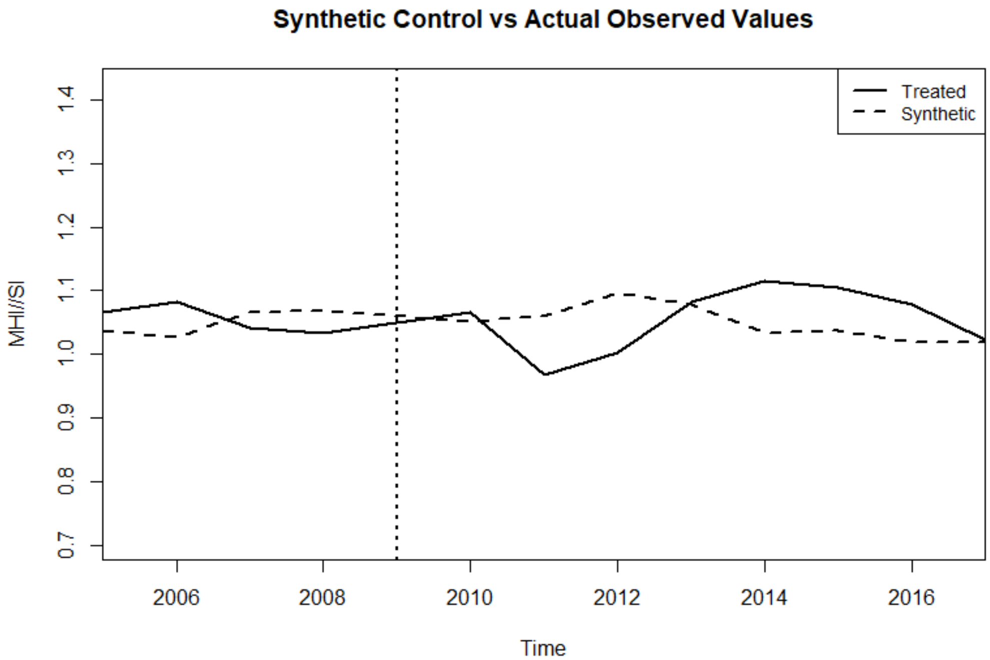

Figure 2: Synthetic Control vs Actual Outcome for Relative Median Household Income

Hornets NBA team had to temporary relocate to Oklahoma City during the 2005-2007 time period. This means that while Oklahoma City did not officially possess a major sports franchise in the 2005-2008 period, they did de facto possess an NBA team. This could bias the results as it is possible that once the SuperSonics moved into OKC in 2009, they simply took over where the Hornets left off. Thus we would not expect to see a jump in relative household income, even if acquiring an NBA franchise does in fact have positive economic benefits. In other words, instead of bringing in new business from basketball lovers outside of Oklahoma City, it's possible the SuperSonics simply capture the market that the Hornets had already created. Secondly, due to a standstill in negotiations between NBA players and owners, the NBA experienced a lockout at the start of the 2011-2012 NBA. Besides delaying the start of the season, the lockout also led to a shortened season, as teams only played 66 regular season games as opposed to the 82 games normally played. As Santo (2005) hypothesizes, this reduced number of games may have reduced the economic benefits of renovating the Ford Center and housing the Oklahoma City Thunder.

While the equivalent of a t-statistic may not exist for the synthetic control method, we can check the robustness of our analysis using a placebo test. A placebo test consists of conducting the analysis on an MSA region that in reality had no treatment applied to it. If the synthetic control is

Figure 3: Gap Between Observed RMHI and Synthetic Control RMHI for multiple MSA regions

sound, then there should be no gap between the control and the region after the treatment period. Abadie and Gardeazabal (2003) use a placebo test on the Catalonia region in their investigation to validate their findings for the Basque region in Spain. Abadie et al. (2010) propose an even more intensive methodology, where the analysis is performed on every control region in the control pool, and the gap between the region of interest (Oklahoma City in our case) and it's synthetic control is compared to the gap between other control regions and their respective synthetic controls. If the analysis is robust, then we would expect the gap for the region of interest to be larger than the gap for other test regions. The authors call this form of robustness check a form of permutation inference Abadie et al. (2010). Figure 4 showcases the permutation inference for the Oklahoma City analysis. The black line is the gap in predicted and actual RMHI for Oklahoma city, and the gray lines are the gaps for the other control regions. It should be noted that following Abadie et al. (2010), all MSA regions with an MSPE greater than 2 times the MSPE of the Oklahoma City MSA have been omitted from the chart. As can be seen, the methodology does a decent job in predicting the RMHI for all regions prior to the treatment period. And similarly to the prior regression analysis results, the gap for Oklahoma City following the initial treatment period does not seem to be an larger than the average placebo gap. This suggests that the treatment had little impact on the RMHI of

Oklahoma City, as the gaps during the treatment period are comparable to those in regions where no treatment occurred.

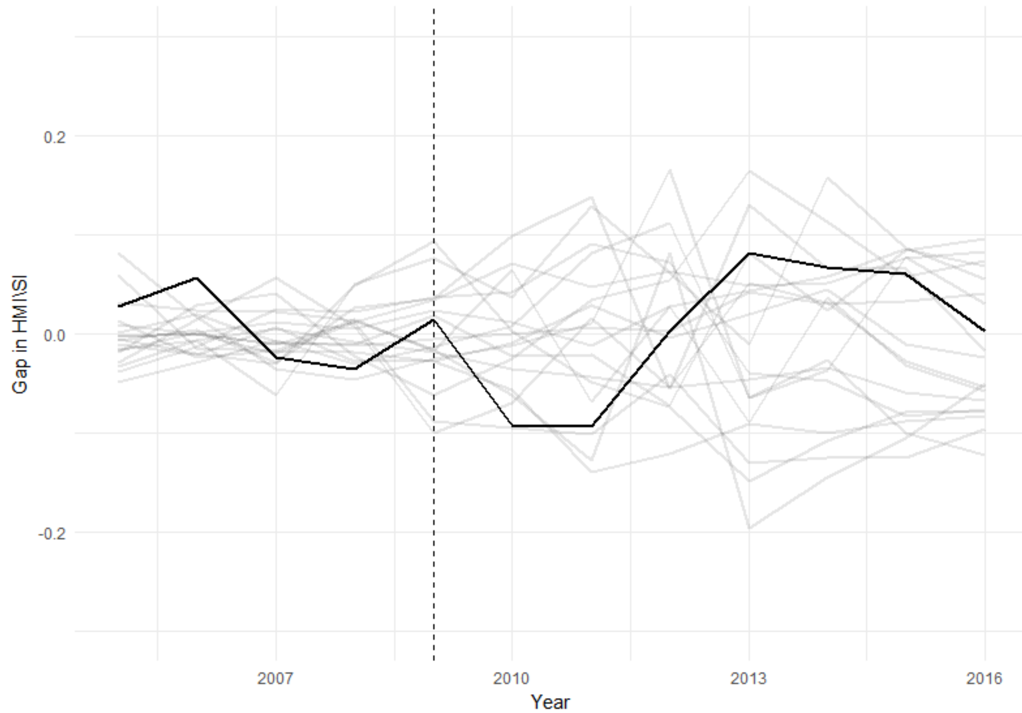

Figure 4: Permutation Inference for Oklahoma City synthetic control

5 Conclusion & Possible Follow Ups

This article sought to investigate whether the relocation of the Seattle SuperSonics to Oklahoma City in 2009 had a significant impact on the relative median household income of Oklahoma City residents. Using data from the American Community Survey and multiple analytical methodologies I found that we cannot reject the null hypothesis that despite the claims of sports advocates, obtaining the Seattle SuperSonics had no effect on the RMHI of Oklahoma City residents. These findings are consistent with much of the literature, which comes to a similar conclusion regarding all major sports teams and stadium investments. This begs the question of why local and state governments are still willing to fund these teams, and offer them incentives to stay. The only possible solution may be that when it comes to professional sports, the intangibles simply cannot be ignored. Despite the consensus amongst economists, the Superbowl is still one of the most viewed yearly television events, and most major sports teams are still able to sell out stadiums for regular seasons games, let alone playoffs. It's possible that despite there being no direct or indirect economic benefit, city residents prefer to consume professional sports, and are willing to be taxed at a higher rate to ensure they have a hometown team. Unfortunately, even if this is the case, this is not the claim teams and government officials make when they request or approve additional funding. They instead claim that an investment into a state of the art stadium, or tax-breaks for sports franchises will benefit the local economy, and that claim has yet to be demonstrated.

This study could be improved on in a number of ways. Firstly, having a larger number of pre-treatment periods would not only improve the accuracy of the synthetic control methodology, but it would lead to a larger sample size for the multivariate regression analysis. Unfortunately, yearly data at the MSA aggregate level has only recently been made readily available. But, perhaps in a number of years this study could be replicated for a team such as the LA Rams, who moved to LA from St. Louis in 2016. Secondly, the impact of the Great Recession likely had a role in biasing some of the results of this analysis. If this study is replicated in the future, care should be taken to choose a time period that has relatively few exogenous shocks on the MSA regions. Lastly, an interesting point of inquiry could be to identify the effect of being a competitive or winning sports franchise as opposed to a non-competitive one. Since stadium ticket revenue is one of the main sources of revenue for teams and local governments, how competitive a newly acquired franchise is could play a role in their economic impact. It is for this reasons that newly formed teams are typically given favourable drafting conditions, allowing them to form a competitive roster without a heavy investment. Ultimately, given the consensus amongst most economists, it may just be that despite the love that the general public has for professional sports, it simply is not a great economic investment for a city.

6 Works Cited

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association 105*(490), 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2011). Synth: An r package for synthetic control methods in comparative case studies. *Journal of Statistical Software 42*(13).

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American economic review 93*(1), 113–132.

Austrian, Z. and M. S. Rosentraub (2002). Cities, sports, and economic change: A retrospective assessment. *Journal of urban affairs 24*(5), 549–563.

Baade, R. A. and R. F. Dye (1988). Sports stadiums and area development: A critical review. *Economic development quarterly 2*(3), 265–275.

Baade, R. A. and R. F. Dye (1990). The impact of stadium and professional sports on metropolitan area development. *Growth and change 21*(2), 1–14.

Badenhausen, K. (2019, Jul). The world's 50 most valuable sports teams 2019.

Evans, P. (2019, Aug). Toronto raptors helped spike canada's economy in may, statscan says — cbc news.

Gius, M. and D. Johnson (2001). An empirical estimation of the economic impact of major league sports teams on cities. *The Journal of Business and Economic Studies 7*(1), 32.

Hlavac, M. (2018). stargazer: Well-formatted regression and summary statistics tables. *R Package Version 5.2.2*.

Santo, C. (2005). The economic impact of sports stadiums: Recasting the analysis in context. *Journal of Urban Affairs 27*(2), 177–192.

7 Appendix A: Figures

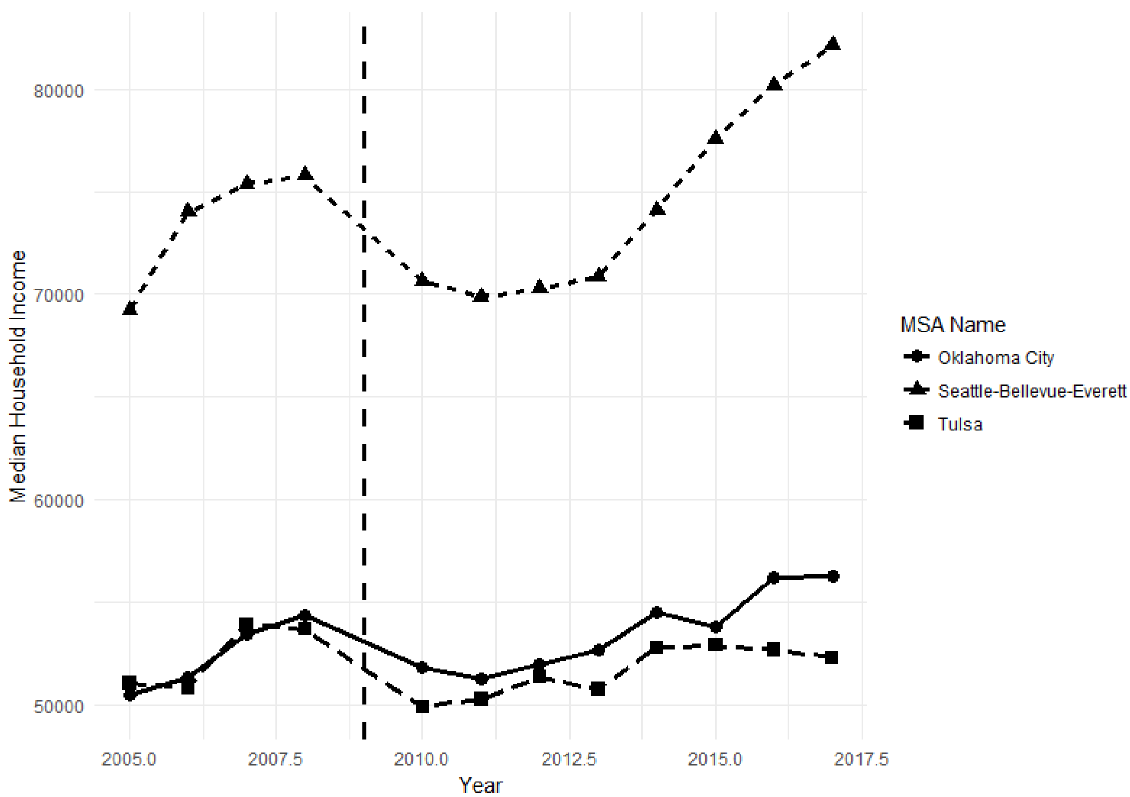

Figure 5: Median Household Income over time of Oklahoma City, Tulsa, and Seattle MSAs

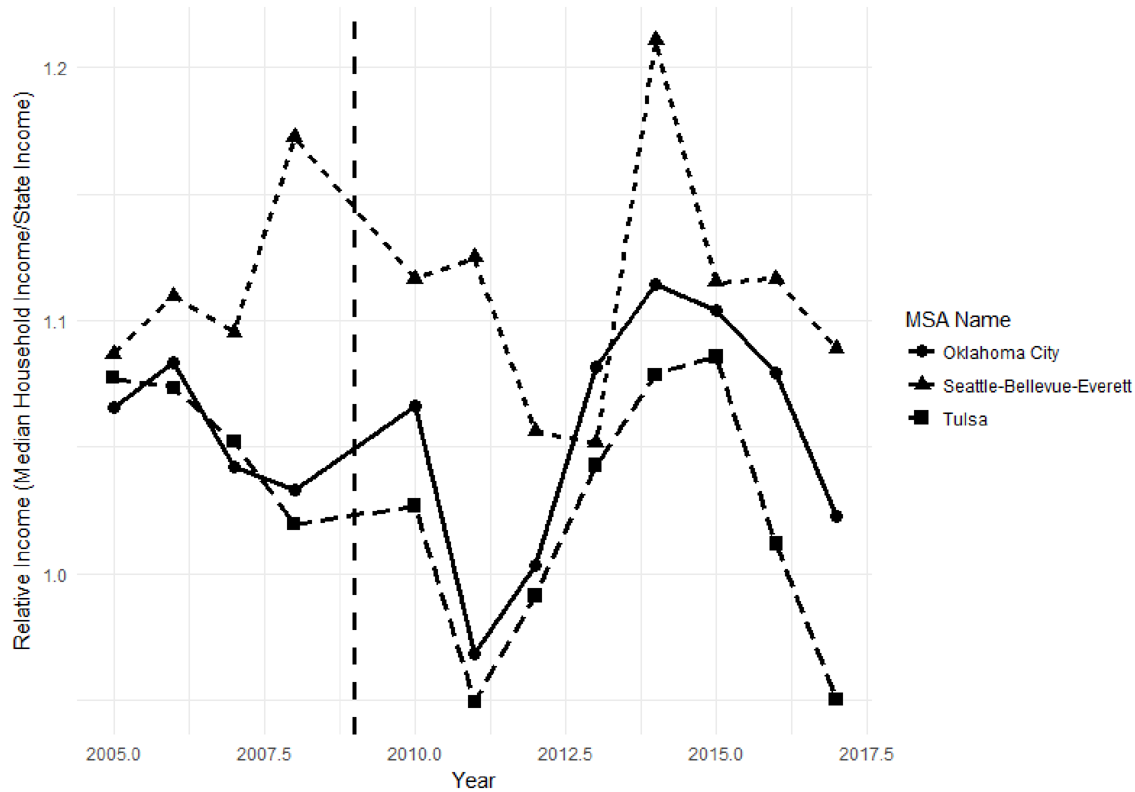

Figure 6: Median Household Income divided by Median State Household Income of Oklahoma City, Tulsa, and Seattle MSAs

8 Appendix B: Tables

	Dependent variable:		
	Median Household Income	MHI/SI	Median Household Income
	(1)	(2)	(3)
Asian/PI	−179,958.000	−0.498	−87,195.760
	(248,181.500)	(11.264)	(155,429.600)
Black	−1,026,165.000	34.396	−568,069.600
	(1,373,482.000)	(62.337)	(1,446,293.000)
College Graduate	−22,581.350	−7.861	6,352.866
	(160,181.800)	(7.270)	(142,567.500)
High School Graduate	−61,951.480	−5.630	
	(182,565.600)	(8.286)	
NBA Franchise	21,733.870	−0.117	49,772.870**
	(21,414.800)	(0.972)	(15,861.700)
Population	0.030	−0.00000	0.054**
	(0.020)	(0.00000)	(0.016)
Poverty Rate	−191,054.600	−2.926	
	(115,000.200)	(5.219)	
Unemployment	−41,045.860	−0.710	50,259.900
	(124,207.700)	(5.637)	(124,450.200)
Constant	90,427.300	3.325	−75,317.430
	(140,634.400)	(6.383)	(101,011.500)
Observations	12	12	12
R^2	0.968	0.405	0.934
Adjusted R^2	0.883	−1.181	0.854
Residual Std. Error	1,447.853 (df = 3)	0.066 (df = 3)	1,617.198 (df = 5)
F Statistic	11.355** (df = 8; 3)	0.256 (df = 8; 3)	11.702*** (df = 6; 5)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Regression results for the Seattle MSA area

	Dependent variable:	
	MHI	MHI/SI
	(1)	(2)
Asian/PI	176,300.300***	0.319
	(37,822.060)	(0.805)
Black	2,607.748	−0.058
	(10,115.510)	(0.215)
College	1,241.649	−1.319**
	(25,353.790)	(0.540)
High School	−21,617.110	−1.304***
	(20,173.310)	(0.429)
NBA Franchise	329.134	0.005
	(1,778.135)	(0.038)
MLB Franchise	19,242.510***	0.278***
	(4,521.951)	(0.096)
NFL Franchise	1,428.514	0.012
	(1,419.034)	(0.030)
More Than 1	−2,381.926	0.016
	(2,308.346)	(0.049)
Population	−0.001***	−0.000
	(0.0002)	(0.000)
Poverty Rate	−115,987.000***	−0.391*
	(9,962.075)	(0.212)
Unemployment	−22,322.410***	0.187
	(8,047.687)	(0.171)
Constant	79,002.200***	1.789***
	(10,080.320)	(0.215)
Observations	768	768
R^2	0.909	0.833
Adjusted R^2	0.899	0.816
Residual Std. Error (df = 694)	4,075.163	0.087
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 7: Fixed-Effects Regression Results for Median Household Income and Relative Median Household Income

	Treated Region	Synthetic Control	Sample Mean
Asian/PI	0.03	0.032	0.088
Black	0.102	0.073	0.071
College	0.178	0.189	0.197
High School	0.281	0.281	0.262
Population	1, 182, 911.5	1, 182, 696.1	2, 254, 844.6
Poverty Rate	0.149	0.133	0.107
Unemployment	0.041	0.054	0.056

Table 8: Observed Oklahoma City mean covariate values compared to synthetic control and sample mean covariate values