2020

Essays on Employment Growth, Wage Discrimination, and Marijuana Legalization

Candon Johnson
West Virginia University, cjohns77@mix.wvu.edu

Essays on Employment Growth, Wage Discrimination, and Marijuana Legalization

Candon R. Johnson

Dissertation submitted to the
College of Business and Economics
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Economics

Brad R. Humphreys, Ph.D., Chair
Jane E. Ruseski, Ph.D.
Bryan C. McCannon, Ph.D.
Ann Marie Hibbert, Ph.D.

Department of Economics

Morgantown, West Virginia
2020

Keywords: Olympic Games, Employment Growth, Synthetic Control, Wage Discrimination, National Basketball Association, Marijuana Legalization, Substitute

Abstract

Essays on Employment Growth, Wage Discrimination, and Marijuana Legalization

Candon R. Johnson

The opening chapter covers the impact of the Olympic Games on employment growth. The Olympics Games stand as the largest sporting event in the world. The Games include approximately 200 countries during the Summer Olympic Games and 90 countries competing in the Winter Games, each occurring once every four years. Potential host cities fiercely compete to host the games under the guise of economic prosperity. Event promoters claim substantial economic benefits, such as employment growth, to be had from hosting these costly games. This paper examines the impacts of the Olympic Games on employment growth rates using a synthetic control approach. Results show transitory increases in employment growth rates following a county being awarded the Olympic Games in Fulton County, GA and Salt Lake County, UT. A decrease in employment growth rate appears in Los Angeles County, CA due to being awarded the 1984 Summer Olympic Games. Results suggest that potential hosts should proceed with caution when considering hosting the Olympic Games.

Chapter two investigates the prominence of wage discrimination in the National Basketball Association (NBA) using free agent signings from 2011-2017. Free agent signings allow us to better capture the determinants of players' wages, a limitation of the previous NBA wage discrimination literature. Using the Oaxaca-Blinder decomposition and weighted linear regression models, we find that black athletes are paid significantly less than their counterparts. In addition, weighted quantile regressions show evidence of the presence of consumer discrimination in the league. This is observed through the result that black players with high audience visibility experience a larger racial wage gap; moreover, this gap is positively related to the share of white population of the MSA where the player is employed.

In the final chapter, I explore the impact of legalization of marijuana on risky consumption of alcohol and tobacco. Utilizing BRFSS data and a differences-in-differences approach with entropy balancing, results indicate that individuals in states that introduce legal recreational marijuana experience a decrease in risky behaviors. Legal states experience a decrease in the overall use of alcohol, drinking and driving, and smokeless tobacco use. Legalization can weed out risky behaviors involving alcohol and tobacco, indicating that marijuana represents a substitute for alcohol and smokeless tobacco. No significant changes in cigarette smoking occurs following legalization.

Acknowledgements

I would like to thank the people who have supported me during my time at West Virginia University. My advisor and dissertation committee chair, Dr. Brad Humphreys, provided many valuable lessons to navigate both my career and life. Thank you for all of the time and effort you put in as a mentor. Having you in my corner gave me the confidence I needed to complete my Ph.D.

I also wish to thank each of my dissertation committee members Dr. Jane Ruseski, Dr. Bryan McCannon, and Dr. Ann Marie Hibbert. I am forever grateful for their time and advice. Dr. Ruseski never failed to provide me with thoroughly marked up copies of early drafts of papers. Dr. McCannon provided me with advice numerous times when I randomly stopped by his office during my time on the job market. Thank you Dr. Hibbert for believing in me as a young undergraduate student in your finance course. In addition, I would like to thank Dr. Joshua Hall for his endless support and guidance.

I am grateful to have had the opportunity to interact with many phenomenal graduate students while at WVU; in particular, Eduardo Minuci, Alex Scarcioffolo, Amir Neto, Alex Cardazzi, Zach Rodriguez, and Hyunwoong Pyun. Thank you for your support, allowing me to brainstorm ideas, and being willing to provide beneficial feedback.

To my parents, Penny and Danny, and the remainder of my family, thank you for the endless love, motivation, and patience during this process. I never could've made it through without the strong support system my family collectively provided. I love you all.

Contents

List of Figures

List of Tables

Chapter 1

The Impact of the Olympic Games on Employment Growth: A Synthetic Control Approach

1.1 Introduction

Potential hosts fiercely compete to host the Olympic Games, the largest sporting event in the world, in part because of expected economic growth generated by hosting the mega-event. Hosting the Olympic Games costs billions of dollars, a portion of which taxpayers subsidize. For example, public funds accounted for $115 million, in 2018 dollars, of the cost of the 1984 Summer Olympics in Los Angeles, $920 million for the 1996 Summer Olympics in Atlanta, and nearly $2 billion for the 2002 Winter Olympics in Salt Lake City (US General Accounting Office, 2000). The Games cost $761 million in Los Angeles, $4.3 billion in Atlanta, and $2.6 billion in Salt Lake City overall (Flyvbjerg et al., 2016). Host cities use claimed economic benefits resulting from the Olympic Games to justify subsidizing the cost of hosting the Olympic Games. The exorbitant cost of hosting the Olympic Games, and taxpayer subsidization of these costs, makes assessing the tangible economic benefits generated by hosting the Games an important topic.

The claimed benefits from hosting the Olympic Games includes long-term employment growth. Employment growth potentially occurs due to construction associated with the

Olympic Games venues and other Olympic-related construction like new hotels and transportation infrastructure. The International Olympic Committe (IOC) requires host areas to have more than 40,000 hotel rooms for the Summer Olympics and nearly 24,000 hotel rooms for the Winter Olympics, an Olympic village capable of housing all participating athletes, and for sport venues to meet their requirements (Baade and Matheson, 2016). Construction projects undertaken to meet these requirements potentially generate increases in local employment. A persistent increase in tourism as a result of hosting the Games represents an additional mechanism for sustained employment growth. If an influx of tourism occurs after the Games, the local labor force will expand to accommodate the increase in tourism. This potential local increase in employment growth serves as the focus of this study.

Previous literature assessed the impact of the Olympic Games on employment growth in the host area, finding inconclusive results. Hotchkiss et al. (2003) and Hotchkiss et al. (2015) found a large, persistent increase in employment growth in Atlanta, GA due to hosting the 1996 Summer Olympic Games. In contrast, Feddersen and Maennig (2013a) revisited the studies and found no impact on overall employment growth in a reply to Hotchkiss et al. (2003). Feddersen and Maennig (2013b) found only an increase of 29,000 jobs in July 1996, when the Olympic Games were being held, in only Fulton County (the county in which Atlanta is located). Games promoters estimated that the 2002 Winter Olympic Games in Salt Lake City would generate 36,000 job-years of employment. However, the Games increased employment by 4,000–7,000, with this increase dissipating quickly (Baumann et al., 2012b).

Prior research on the economic impact of the games primarily uses either an event study framework comparing outcomes in the host city before and after the Games or a difference-in-differences approach with a relatively small control group. Hotchkiss et al. (2003) use counties within Georgia that did not hold an Olympic event as their control group, while Baumann et al. (2012b) use states adjacent to Utah for example. The synthetic control method (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015) represents a reasonable alternative approach to analyzing the economic impact of the Games.

The methodology utilized in this paper follows the approach used by Islam (2019) to examine the impact of introducing an National Football League (NFL) team to a metropolitan statistical area (MSA) on local employment growth. The methodology differs from Islam

(2019) by analyzing county-level employment growth, a smaller geographic impact area, and focusing on the Olympic Games. The counties analyzed include Los Angeles County (1984 Summer Games), Fulton County (1996 Summer Games), and Salt Lake County(2002 Winter Games). Essex County, the host of the 1980 Winter Olympic Games, is excluded due to its small size and limited data in the pre-treatment period. The fixed boundaries of counties makes county level analysis preferable to MSAs due to changes in the boundaries of MSAs over time.

Like Feddersen and Maennig (2013b) and Baumann et al. (2012b), the synthetic control results in this paper show that the 1996 Summer Olympic Games in Atlanta and the 2002 Winter Olympic Games in Salt Lake City caused transitory increases in employment growth. Atlanta experienced the largest impact, experiencing an increase in employment growth in several years between being selected to host the Games and hosting the Games. Salt Lake City experienced an increase in only the year following selection to host the Games. The Summer Olympic Games, a much larger event than the Winter Olympic Games, partially explains the smaller impact in Salt Lake County relative to Fulton County. The synthetic control results also show evidence of a negative economic impact from the 1984 Summer Olympic Games in Los Angeles after being awarded hosting rights but two years prior to hosting the Games.[1]

This paper makes several contributions to the literature. The results show that hosting the Olympics can either increase or decrease employment growth, but these effects are transitory. This paper contains the first evidence of a decrease in employment growth due to hosting the Olympics. Previous literature finds either positive or no impact. The causal evidence of a transitory employment impact due to the 1996 Olympics developed here provides clarity to the debate between Hotchkiss et al. (2003, 2015) and Feddersen and Maennig (2013a). The synthetic control method represents a casual inference method not previously used to examine the impact of the Olympic Games. The introduction of this casual inference method advances the literature.

[1]Multiple studies utilize the synthetic control method to show transitory impacts (Eren and Ozbeklik, 2016; Kreif et al., 2016; Tirunillai and Tellis, 2017).

1.2 Hosting the Olympic Games: Process and Impacts

The process of hosting the Olympic Games begins nearly a decade before the Games occur in a specific area. The host city selection process involves many steps. Consider, for example, the selection process for the Games of the XXVI Olympiad, informally known as the 1996 Summer Olympic Games, in Atlanta, Georgia. Atlanta submitted their bid as a potential US candidate city for the 1996 Games to the United States Olympic Committee (USOC) in September 1987. 13 other US cities submitted bids to the USOC. The USOC reduced the field from 14 to two, Atlanta and Minneapolis-St.Paul, with Atlanta being selected as the US candidate city in April 1988.

Atlanta then competed with cities selected by National Olympic Committees (NOCs) around the world, including Athens, Greece; Toronto, Canada; Melbourne, Australia; Manchester, Great Britain; and Belgrade, SFR Yugoslavia, for the rights to host the 1996 Games. The following year, IOC members visited each candidate city before holding a vote to select the host city in 1990. Voting consisted of five rounds, with the city receiving the lowest number of votes in each round being eliminated from consideration. Atlanta defeated Athens, Greece 51-35 in round five of voting to become the host of the 1996 Olympic Games (Atlanta Committee for the Olympic Games, 1997). The Winter Olympics follows a similar selection process.

After the awarding of hosting rights, the NOC forms a local Organizing Committee for the Olympic Games (OCOG), and dissolves the OCOG after the Games occur. OCOGs receive local, state, and federal government subsidies in order to put on the Games. The size of the subsidies depend on the amount of funding the OCOG receives from the IOC and the availability of private funding. The budget of the OCOG primarily includes operating costs of the Games, while the host city is largely responsible for infrastructure (Humphreys and Howard, 2008).

IOC voting on the host city of the Games generally follows the format discussed above, with one notable exception. Only one city, Los Angeles, placed a bid to host the 1984 Olympic Games. The lack of interest in hosting the 1984 Olympics stemmed from events surrounding Games prior to 1984, including violence and financial losses. Mexico City experienced violence and protests in 1968. Eleven Israeli Olympic athletes were killed by terrorists in Munich in 1972. The 1976 Summer Olympic Games in Montreal cost nearly 10 times more that budgeted leading to a debt that took thirty years to eventually pay down. Denver won the rights to host the 1976 Winter games in 1970 but a 1972 referendum on public subsidization of the games failed and the games moved to Innsbruck, Austria. Los Angeles agreed to host following the IOC guaranteeing any losses and confirming the adequacy of the city's existing sports infrastructure for Olympic events (Zimbalist, 2016, pp.1).

From 1960-2016, sports-related costs averaged $5.213 billion for the Summer Olympics, and $3.112 billion for the Winter Olympics, in 2015 US dollars (Flyvbjerg et al., 2016). Non-sport infrastructure, security, opening ceremonies, and other spending add to the total cost of hosting the Games. An extravagant opening ceremony alone cost nearly $350 million at the 2008 Summer Olympic Games in Beijing, China. Security costs soared following terrorist attacks throughout the US on September 11, 2001. Athens estimated security costs at $400 million in their initial bid to host the 2004 Summer Olympic Games, submitted before 9/11. The final cost ballooned to approximately $1.5 billion (Zimbalist, 2016, pp.42–43). Of the more than $13 billion spent hosting the 2016 Summer Olympic Games in Rio de Janeiro, non-sport related infrastructure accounted for $8.2 billion (Associated Press, 2017). Total expenditures to host the Olympics reached as high as $40 billion for 2008 Summer Olympics in Beijing, and $50 billion for the 2014 Winter Olympics in Sochi (Zimbalist, 2016, pp.2).

The Olympic Games represent a major investment undertaken by host cities. Proponents of hosting the Games claim that the events will generate an array of positive outcomes, both socially and economically, in the host area. Opponents claim that there can be negative outcomes, and the positive impacts that do exist are not large enough to warrant the high cost of hosting these events.[2]

Pride and prestige associated with hosting the Olympics potentially generates an uplifted mood in the host area. Smith (2009) argues for the presence of a connection between hosting mega sporting events and an increase in mental health in the local community. Hosts often believe that hosting the Olympics also generates an increase in physical activity, but Bauman et al. (2013) suggest that physical activity increases much less than projected, or not at all.

[2]Potential impacts and an assessment of literature are discussed in Scandizzo and Pierleoni (2018).

Atkinson et al. (2008) conduct a willingness to pay (WTP) study to estimate the value of intangible benefits of hosting the London Olympic Games, finding an aggregate household WTP of nearly $2 billion. Atkinson et al. (2008) state that, given that economic studies generally show negligible or negative impacts, this WTP represent a credible approach to assessing the public choice problem of hosting the Olympics. This WTP pales in comparison to the actual cost of the London Games to taxpayers. Of the $14.6 billion it cost to host the 2012 Olympic Games in London, $4.4 billion came from taxpayers (Schwarz, 2015).

Prestige associated with hosting the Olympic Games potentially makes the host a more desirable destination for tourists. Kang and Perdue (1994) found an increase in tourism in South Korea following the 1988 Olympics. The increase peaked in the year following the Games and dissipated in the following years. Giesecke and Madden (2011) found no induced tourism impacts as a result of the 2000 Olympic Games in Sydney, Australia. Induced tourism represents a mechanism for a persistent increase in employment.

Negative impacts such as increases in crime in the host area arise as well. Baumann et al. (2012a) found that the Olympic Games led to a 10% increase in property crime. Hosting the Olympics or other mega-events, such as the World Cup, can cause political unrest due to hosting being unpopular among local residents. This occurred in Brazil prior to hosting the 2014 World Cup; widespread political unrest occurred in Brazil during the Confederations Cup. The Confederations Cup, an international soccer competition held the year prior to the World Cup (in 2013 in the case of Brazil), drew over a million Brazilian protesters to the streets. Protesters disproved of the government spending $15-20 billion for hosting the 2014 World Cup. The protests continued as the World Cup approached. Many Brazilian cities experienced strikes by police and teachers, among other workers, in the run-up to the World Cup (Zimbalist, 2016, pp.2).

Hosting the Olympic Games requires large infrastructure investments. In addition to the construction of new sport facilities, the Games also require investment in the surrounding area on non-sport related infrastructure. The Olympic Games potentially draw substantial tourism activity, and hosts must be equipped to handle the increased inflow of visitors. This infrastructure requirement could be beneficial, potentially boosting employment growth due to construction. Additionally, the claimed increases in tourist activity could increase in employment in tourism related industries.

Employment growth represents the economic outcome of interest in this study for two primary reasons. First, mixed results on the impact of the Olympic Games on employment in the literature makes this study necessary to add clarity. Second, the strict infrastructure requirements for hosting the Olympic Games dictated by of the IOC makes an increase in local construction activity almost certain to occur. This increase in construction activity potentially leads to increased employment growth, although it could simply crowd out other local construction projects.

1.3 Literature Review: Olympic Economic Impact

A substantial literature exists studying the impact of the Olympic Games on employment, yielding inconsistent results. Studies focusing on the 1996 Olympic Games in Atlanta provide an interesting set of conflicting results. Hotchkiss et al. (2003) compare counties near Olympic venues to those not near Olympic venues in Georgia finding a persistent increase in employment due to hosting the 1996 Olympic Games. Feddersen and Maennig (2013a) questioned this positive impact on multiple grounds, with a focus on accounting for pre-treatment trends and the treatment period used. Maennig and Fedderson find no significant increase in employment associated with hosting the 1996 Games after accounting for local time trends. They also perform numerous nonparametric tests in lieu of the standard differences-in-differences model tests, again finding no effect.

Hotchkiss et al. (2015) revisited the topic of their initial paper in response to Feddersen and Maennig (2013a). Hotchkiss et al. (2015) again found evidence that employment growth in Georgia counties near Olympic venues outpaced growth in other Georgia counties. Hotchkiss et al. (2015) reported positive impacts from hosting the Olympics, but at a lower magnitude than their original paper. In this study they find a smaller impact, 11%. Their comparison of MSAs in Georgia that hosted the Olympics to similar MSAs throughout the southern United States provides their most convincing evidence. Results indicate that MSAs hosting the Olympics outpaced employment gain in other southern states by 5%.

Baade et al. (2002) highlight the importance of the time period studied on results, finding an employment increase of approximately 3,500 to 43,000 from the 1996 Olympics in Atlanta depending on the period examined. Baade et al. (2002) found that much of the expenditures on the Games occurred in 1994 and 1995. Their estimate coincides with the increase of 37,000 jobs projected in Atlanta by Humphreys and Plummer (1995).[3]

Feddersen and Maennig (2013b) conducted an additional study examining mega-events and sectoral employment using the 1996 Olympic Games. They analyzed monthly data for 16 different sectors using a nonparametric approach to isolate any employment effects. Their results show a slight boost in employment, but no evidence of a persistent shift in employment growth. Fulton County (the county in which Atlanta is located) experienced an increase of 29,000 jobs in July 1996 when the Games took place. Three sectors of the economy accounted for the increase retail trade; accommodation and food services; and arts, entertainment, and recreation.

Baumann et al. (2012b) further studied the impact the Olympics on employment growth by analyzing the 2002 Winter Olympics in Salt Lake City, UT relative to outcomes in adjacent states. They found an increase in employment substantially lower than estimated by promoters. Promoters estimated an increase of 35,000 job-years while Baumann et al. (2012b) find an increase of 4,000-7,000 jobs using a control group of states adjacent to Utah. Like Feddersen and Maennig (2013b), the leisure industry accounted for the increase in employment and the effect dissipated after a year. Considering the mixed results on the impact of the Games on local employment growth, slight job growth appears associated with hosting the Olympic Games, but at a magnitude much less than claimed ex ante and dissipating quickly.

Research on the economic impacts of hosting the Games extends beyond employment growth. Rose and Spiegel (2011) find a "robust, permanent and large" increase in exports as a result of hosting the Olympics. Results indicate that countries placing an unsuccessful

[3]Humphreys and Plummer (1995) estimated an increase of 77,000 jobs in all of Georgia. With 48% of Georgia's population residing in Atlanta, Humphreys and Plummer (1995) claims that 48% of the employment growth would occur in Atlanta. This translates to an increase in approximately 37,000 jobs occurring in Atlanta. With arguably more than 48% of Olympic spending occurring in Atlanta this forecast is likely understated.

bid experienced a similar impact. Maennig and Richter (2012) reexamined this peculiar result, finding no impact on exports when using an appropriate matching and treatment methodology, suggesting that results in Rose and Spiegel (2011) may suffer from selection bias.

Baade et al. (2010) assessed the impact of the 2002 Winter Olympic Games in Salt Lake City on local taxable sales. They used quarterly taxable sales data from 1982 through 2006 and estimated an auto-regressive-moving-average (ARMA) model. The overall impact of hosting Olympics, based on impacts estimated for several different local sectors, showed a net negative effect on taxable sales. While hotels, and eating and drinking establishments experienced gains, losses elsewhere outweighed these gains leading to a net loss of $167.4 million. Similarly, in a study analyzing the impact of the 2000 Summer Olympic Games in Sydney Australia, Giesecke and Madden (2011) found that the Olympics generated a loss in real consumption of $2.1 billion.

The research on the economic impact of another mega-event warrants discussion: the World Cup, the largest soccer tournament in the world. Baade and Matheson (2004) study the impact of the 1994 World Cup hosted by the United States using income data from 1970-2000 to estimate the effect of hosting the Games on income growth. Baade and Matheson (2004) compared predicted growth to actual growth in each city that hosted a match. Nine of the thirteen US cities that hosted World Cup matches experienced growth lower than the predicted value, indicating an economic loss from hosting the event. The combined losses total up to $9.26 billion compared to the ex ante estimate of $4 billion in benefits.

Mega-events such as the Olympic Games and World Cup are high cost/low reward investments. Positive economic impacts are generally low, if existing at all. Matheson (2012) discuss that the economic impacts of hosting the mega-events may be even lower for developing countries. Hosting mega-events can allow politicians to clear political hurdles to invest in infrastructure, but this comes with paying a large price for unproductive sports infrastructure. Further reviews of the literature can be found in Scandizzo and Pierleoni (2018) and Baade and Matheson (2016).

1.4 Empirical Analysis

1.4.1 Data and Methodology

Data come from the Bureau of Economic Analysis (BEA) Regional Economic Accounts CAINC30 dataset. CAINC30 data includes variables reflecting annual population, per capita income, and employment at the county level over the 1969-2016 period. Population estimates come from the Census Bureau's annual (July 1) midyear population estimates. The BEA uses this population estimate to calculate per capita income. BEA compiles data on the county employment level including full-time and part time jobs. Conversion of data from levels to growth rates, as in Islam (2019), leaves an analysis data set covering 1970-2016.

To examine the impact of the Olympic Games on county employment growth, I use the synthetic control. The synthetic control method appears throughout the economic literature analyzing local employment growth (Munasib and Rickman, 2015; Peri and Yasenov, 2015), as well as in sports economics (Islam, 2019; Pyun, 2018), and in research analyzing overall economic conditions (Grier and Maynard, 2016). Synthetic control creates a synthetic version of the treatment area to provide a counterfactual. The control group provides a comparison to assess the impact of an event or policy. The synthetic version of the treated counties in this study are constructed using a weighted average of other U.S. counties in a donor pool of counties with observable characteristics similar to treated counties that hosted the Games. Donor pools exclude counties contiguous to treated counties and counties that also competed to host the Olympic Games.

The data contains observations for a total of T years. $1, ..., (T_0 - 1)$ constitutes the period before treatment occurs and $T_0, ..., T$ the post-treatment period. The treatment occurs in year T_0. The donor pool consists of $J+1$ counties, $j = 1, 2, 3, ...J+1$ defined so that county 1 is treated. The synthetic control method chooses a vector of optimal weights, W^*, for each county in the donor pool that minimizes

$$\sum_{m=1}^{k} v_m (X_{1m} - X_{0m}W)^2 \tag{1.1}$$

where X_{1m} represents a vector of predictor variables for the treated county (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015; Pyun, 2018). X_{0m} represents a (kxj) vector of predictor variables for counties in the donor pool and j indexes the number of counties in the donor pool. v_m reflects the weight, showing the relative importance assigned to the mth variable when measuring the difference between X_1 and X_0. Each W^* is bounded between 0 and 1, and the total weights must sum to 1.

The synthetic control method selects a weight v_m that minimizes the root mean square prediction error (RMSPE). The RMSPE for Olympic hosting counties is defined as

$$RMSPE = \left(\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2 \right)^{1/2} \tag{1.2}$$

where Y represents the outcome variable. RMSPE measures statistical fit between outcomes in the treated and synthetic county, with a lower RMSPE indicating a better fit. A high post-treatment RMSPE indicates a lack of fit in the post-treatment period, suggesting important impacts in treated counties. Comparison of post-treatment and pre-treatment RMSPE shows the credibility of any impacts found. A large post-treatment RMSPE does not indicate a large impact of treatment the pre-treatment period also has a large RMSPE, as no discernible difference between the pre-treatment and post-treatment periods exist. Therefore, a high post-treatment to pre-treatment ratio indicates a potentially larger impact from treatment (Abadie et al., 2015).

The synthetic control method requires identifying a donor pool of counties that did not receive the treatment. The counties included in the donor pool generate the synthetic control group based on pre-treatment data. Using all US counties as the donor pool poses a problem as the donor pool will contain many counties with little similarity to the treated county. To correct for this, the donor pool excludes counties with large differences in population compared to the treated county. The donor pool includes counties with populations larger than 1,000,000 for Los Angeles County, populations between 500,000 and 1,000,000 for Fulton County, and populations between 500,000 and 1,250,000 for Salt Lake County. Robustness checks show results are not sensitive to using alternative population criterion for identifying the donor pool.

The construction of a synthetic county follows Islam (2019) who analyzed the impact of National Football League teams appearing in US cities. Islam (2019) found no evidence of positive employment growth effects from new NFL teams. The overall average population growth and per capita income growth during the pre-treatment period, as well as select years of the outcome variable employment growth, Y, construct the synthetic control county. Employment growth every five years before treatment is used for construction when able to do so. Kaul et al. (2015) warn against using all past values of the outcome variable to construct the synthetic control group, as this results in all other predictors having no contributing weight. Kaul et al. (2015) recommends using one lag for the outcome variable and from the year prior to treatment. Although three lags of employment growth are used here, results remain similar when using only one year of employment growth (see Appendix 4.1.1).

I define the treatment year as the year in which the IOC awards the rights to host the Olympic Games, not the year when the Games occur. Construction of infrastructure related to hosting the Olympic Games takes place between the awarding of the games and the staging of the event and could potentially generate employment impacts. Construction reasonably begins shortly after the awarding of the rights to host the Games. Feddersen and Maennig (2013a) consider the second quarter of 1990 as the beginning of the treatment period when analyzing impact of the 1996 Olympic Games in Atlanta. Baade et al. (2002) found that much of the impact from the 1996 Olympic Games occurred in 1994 and 1995, with smaller impacts occurring in years prior. This highlights the importance of including the entire period after gaining the rights to host the Games in the treatment period. Using this treatment year, the treatment period begins approximately 6-7 years before the Games occur. The specific treatment years are 1978 for Los Angeles County, 1990 for Fulton County, and 1995 for Salt Lake County.

Placebo tests act as sensitivity tests to identify significant impacts on employment growth due to the Olympic Games. In this approach, every county in the donor pool receives a placebo “treatment" as if the county hosted the Olympic Games. Placebo tests compare placebo counties to the county that actually hosted the Olympics. When employing placebo tests, the time path of the outcome variable for the placebo treatments should not significantly deviate from their synthetic counterpart. A large portion of the donor units exhibiting similar impacts to the treated unit in the placebo test calls any initial synthetic control results into question (Abadie et al., 2010).

Cunningham (2018) discusses constructing p-values based on the placebo tests. After assigning placebo treatments to counties that did not host the Olympic Games, post-treatment to pre-treatment RMSPE ratios are calculated. Where the treated county's post-treatment to pre-treatment RMSPE ratio ranks among the placebo counties is used to calculate a p-value. Consider Los Angeles County as an example. The post-treatment to pre-treatment RMSPE ratios for Los Angeles County and the 19 other counties donor counties are calculated following the placebo test. Los Angeles County's ratio of 3.4614 ranks third out of 20 counties, yielding a p-value of 0.15 (i.e. 3/20).

1.4.2 Results: The Olympic Games and Employment Growth

Figure 1.1 presents results for the synthetic control method applied to county employment growth generated by hosting the three Olympic Games. The upper panel shows results for the 1984 Summer Olympics in Los Angeles County, the middle panel the 1996 Summer Olympics in Fulton County, and the bottom panel the 2002 Winter Olympics in Salt Lake County. In each panel a dashed vertical line identifies the treatment year, the year in which the IOC awarded the host the Olympic Games, and a solid vertical line identifies the year in which the Olympic Games occurred. The solid vertical line highlights any impacts experienced around the year the Games occurred, either through construction taking place close to the Games, or job creation due to increases in tourism.

Table 1.1 shows the counties that contribute to the synthetic Los Angeles County. Middlesex County, MA, a county in Boston containing Cambridge, MA, represents the largest contributor at 0.266. Bronx County, NY follows at 0.239, and Santa Clara County, CA at 0.212. Table 1.2 shows predictor balance and root mean square prediction error for this case. Average population growth, average income growth, and three select years of employment growth select the synthetic Los Angeles County.[4] Treatment occurred in 1978, so the three years of employment growth rates used include 1970, 1974, and 1977. From Table 1.2, Los Angeles County and Synthetic Los Angeles County exhibit good predictor balance. Each predictor variable utilized shows little or no difference between real and synthetic Los Ange-

[4]While Islam (2019) selects the three years in five year increments. Here a shorter time span between years is used due to data beginning in 1970.

les County. While a ratio of post-treatment to pre-treatment RMSPE well above 1 suggests a potentially significant impact from hosting the Olympics, a p-value of 0.15 suggests no persistent shift in the post-treatment time path of employment growth.

Results in Figure 1.1 indicate a potential negative impact on employment growth rate in Los Angeles County after being awarded the Olympic Games. The gap between actual and synthetic Los Angeles appears negative from 1979-1985 with the largest gap occurring in 1982. In 1982 employment in synthetic Los Angeles County grew at a rate of 0.7% while Los Angeles County experienced a decline in employment growth of -2%. A p-value of 0.15 indicates no persistent impact on employment growth. However, a high post-treatment to pre-treatment RMSPE ratio, in addition to results shown in Figure 1.1 suggest a negative transitory effect on employment growth in Los Angeles County.

Reduced employment growth is a potentially surprising result given the legacy of the 1984 Summer Olympic Games, which were generally regarded as a success. Prior to Los Angeles being awarded the 1984 Summer Olympic Games, no city wanted to be the host, following a series of tumultuous Olympic Games. With the IOC offering a guarantee to cover any losses, and Los Angeles having some appropriate infrastructure in place to host the Games, the city agreed to host. The 1984 Games proved to be one of the most financially successful Games in history, turning a modest profit of $215 million (Zimbalist, 2016, pp.1). The financial success of the 1984 Games spurred renewed competition to host the Olympics in the following years.

The synthetic control results indicate that this financial success came with high costs economically, in terms of a loss in employment growth. An excess demand for building materials and construction labor induced from hosting the Olympic Games can explain lower output in a tight labor market. In a tight labor market, induced labor demand will not lead to additional output, but instead cause reallocation of scarce resources towards the Olympic Games-related economic activity.

Next, consider the results from the 1996 Summer Olympic Games in Atlanta (Fulton County). Table 1.3 shows the synthetic control weights following the methods described in Section 4.1. Essex County, MA, which lies adjacent to Boston, represents the largest contributor at 0.42. Montgomery County, MD, a county adjacent to Washington D.C. and the most populous county in Maryland, follows at 0.355.

Predictor balance and RMSPE are presented in Table 1.4. Predictor balance indicates a good fit with nearly identical population growth, per capita income growth, and employment growth in treated and synthetic Fulton County. A high p-value of 0.5384 indicates no persistent shift in employment growth, but a post-treatment to pre-treatment RMSPE ratio greater than one suggests the potential for transitory impacts.

The middle panel of Figure 1.1 shows the time path of actual employment growth in Fulton County and synthetic Fulton County. Overall actual employment growth lies above synthetic employment growth following treatment through 1997, with the exception of 1991 and 1992. The largest gaps between the actual and synthetic employment growth occur between 1993 and 1997. This increase coincides with the construction of Centennial Olympic Stadium. Construction of Centennial Olympic Stadium started in 1993, on July 10th, with completion and opening of the stadium occurring on May 18th, 1996 (Atlanta Committee for the Olympic Games, 1997). Hosting the Olympics seems to have had a temporary positive impact on employment growth in Fulton County, particularly in the lead up to, and hosting of, the 1996 Summer Games. The largest gap occurs the year Fulton County hosted the Games, when Fulton County outpaced synthetic Fulton County by 3.2 percentage points. Synthetic Fulton grew at a rate of 2% and Fulton County grew at 5.2%. Fulton County experienced large impacts in 1993, and 1994 as well, with employment growth more than doubling in comparison to synthetic Fulton County. While Fulton County experiences significant short-term employment growth, the impact appears to dissipate by 1998.

The 2002 Winter Olympic Games in Salt Lake City serves as the final US Olympics analyzed studied. Like Atlanta, Salt Lake City faced competition to become the host of the Olympic games, prevailing over bids from Quebec City, Quebec, Canada; Sion, Switzerland; and Östersund, Sweden (Baade et al., 2010). Table 1.5 shows the synthetic control weights. Pima County, AZ, which contains Tuscon, represents the largest contributor at 0.485. DuPage County, IL, a county adjacent to Chicago, follows at 0.364. Oklahoma County, OK, location of the state capital Oklahoma City, contributes 0.146.

Table 1.6 reports predictor balance and RMSPE. The closeness of predictor variables between Salt Lake County and synthetic Salt Lake County indicates a good fit. As in Fulton

County, a high p-value of 0.4737 indicates no persistent shift in employment growth, but a post-treatment to pre-treatment RMSPE ratio greater than one suggest the potential for transitory impacts.

The the bottom panel on Figure 1.1 shows the time path of actual employment growth in Salt Lake County and synthetic Salt Lake County. Employment growth rates in Salt Lake County and synthetic Salt Lake County lie close to one another following treatment, with the exception of 1996 and 1997. In 1996 and 1997 Salt Lake County grew at a rate of 5.4% and 3.7% while synthetic Salt Lake County grew at 2.4% and 2.2%, respectively. This spike correlates with hotel expansion in Salt Lake City that occurred from 1994 to 2002. In that span of time, the number of hotel rooms in Salt Lake County increased by 63%, an increase that led the director of sales and marketing for the first five-star hotel in Salt Lake City to state: "There's no doubt we're overbuilt, a 63 percent growth is tough to support no matter where you are. Las Vegas, whatever"(Isidore, 2002). Hosting the Olympics appeared to generate a positive shock on Salt Lake County in the two years following treatment. Similar to Fulton County, the positive employment growth dissipates quickly.

To assess the ability of synthetic control to capture economic impacts, placebo tests act as significance tests. Placebo tests apply the synthetic control method to every unit in the donor pool. This approach indicates whether treatment or randomness drives the results. When employing placebo tests, the time path of the outcome variable for the placebo "treatments" should not significantly deviate from their synthetic counterpart. Figure 1.2 reports placebo tests for each of the three Olympic Games with the gap in employment growth between the county tested and the county's synthetic counterpart graphed on the Y axis in each year. The bold black line represents the county that hosted the Olympic Games, while the light gray lines each represent a county in the donor pool. Impacts experienced by the host county compared to placebo counties determines the significance of the impacts, based on the percentage of donor units that deviate from the treated county. A large percentage of placebo counties experiencing larger changes in employment growth than the treatment county calls into question the validity of the synthetic control results.

The top panel of Figure 1.2 shows placebo test results for the 1984 Los Angeles Games. A large portion of the post-treatment period appears to no have a significant impact on Los

Angeles County with the exception of 1982. In 1982 only one placebo county experienced an effect larger than Los Angeles. This placebo county is dropped in Appendix 4.1.2 when placebo counties with high mean squared prediction errors (MSPE) are removed from the donor pool (Abadie et al., 2010).[5]

Placebo test results for the 1996 Atlanta Games appear in the middle panel of Figure 1.2. Fulton county experienced significant increases on employment growth in 1993, 1994, and 1996, in line with the years containing significant employment increases reported in previous studies (Feddersen and Maennig, 2013b; Baade et al., 2002). The bottom panel of Figure 1.2 shows placebo test results for the 2002 Salt Lake Games. Placebo test results indicate that Salt Lake experienced increased employment growth in only one year, 1996. An increase in employment appears in 2007 as well, however the amount of time passing between treatment, hosting, and this increase coupled with the lack of impact prior to 2007 calls into question attributing this increase to the Olympic Games.

Results in Figures 1.1 and 1.2 indicate that each county experienced some transitory impacts to employment growth after acquiring the rights to host the Games. Los Angeles County saw a decrease in employment growth two years prior to the Games being held. Fulton County and Salt Lake County each experienced transitory increases in employment growth. Salt Lake County's growth coincides with a documented increase in hotel construction, while Fulton County's growth matches the time period of construction of the Centennial Olympic Stadium. The lack of a persistent increase in employment growth calls into question the claimed benefit of persistent increases in tourism caused by hosting the Games.[6]

[5] Abadie et al. (2010) present placebo tests dropping donor states that have MSPE two times, five times, and twenty times higher than the treated state.

[6] The counties studied represent the focal point of each Olympic Games, but few events were held in other counties throughout the hosting state. Considering the size of the event that the Olympics represents and events being held throughout Olympic hosting states, analyzing spillover effects becomes important. Appendix 4.1.3 presents results analyzing state-level impacts.

1.4.3 Robustness Checks

Alternate Donor Pools

Due to the subjective nature of selecting counties to include in the donor pool, using alternative population limits to identify donor counties checks the robustness of results. Alternative selection criterion use both a wider and narrower range of county populations to identify donor counties. Figures 1.3, 1.4, and 1.5 present results using alternative donor pool criteria.

For the 1984 Games, donor pools using counties with populations greater than 750,000 and 1,250,000 test for sensitivity, compared to the 1,000,000 population threshold utilized for results above. Figure 1.3 indicates that results are not sensitive to the donor pool composition, finding a decrease in employment growth in each alternative donor pool. The alternative population criterion for the 1996 Games includes counties with population of 250,000 to 1,250,000, and 500,000 to 850,000 instead of 500,000 to 1,000,000 used above. Figure 1.4 indicates that Fulton County experienced employment growth in 1993 through 1997 for each alternative donor pool.

Figure 1.5 presents results for the 2002 Games. While the original donor pool includes counties with populations from 500,000 to 1,000,000, alternative ranges of 250,000 to 1,500,000 and 650,000 to 1,000,000 constitute the alternative donor pools. As with Los Angeles County and Fulton County, initial results for Salt Lake County persist when using these alternative donor pools.

Failed Olympic Bids

The three Olympics Games analyzed above generated transitory impacts on employment growth. As a robustness check, I consider unsuccessful bids to host the Games by Minneapolis, MN (Hennepin County) and Chicago, IL (Cook County). Minneapolis unsuccessfully bid against Atlanta to represent the USOC in the competition for the 1996 Olympics. Chicago advanced further into the bidding process, being selected by the USOC to compete with Rio De Janeiro, Brazil; Madrid, Spain; and Tokyo, Japan to host the 2016 Summer Olympic Games (Baade and Sanderson, 2012). Chicago spent more than $100 million on the failed bid attempt (Zimbalist, 2016, pp.42).

Counties may select into bidding to host the Olympic Games based on a belief that the county will experience substantial economic growth in the future. Analyzing counties with unsuccessful bids to host the Olympic Games mitigates concerns of selection bias by counties that bid to host the Olympics. Issues with selection bias appears throughout the literature on the economic impact of the Olympics; for example Maennig and Richter (2012) refute results in Rose and Spiegel (2011) on these grounds.

The analysis of outcomes Cook County, IL and Hennepin County, MN follows the same approach as Los Angeles County, Fulton County, and Salt Lake County, using population growth, income growth, and three select years of employment growth. Cook County, the second largest county in the US, uses a donor pool consisting of counties with populations larger than 1 million. Hennepin County's donor pool includes counties with a population between 500,000 and 1,250,000, the same range used for Salt Lake County. 1988 represents the treatment year for Hennepin County, the year in which the USOC selected Atlanta over Minneapolis, MN.

Table 1.7 shows synthetic control weights and reports RMSPE for Hennepin County. The largest contributor to synthetic Hennepin County is Hartford County, CT at 0.365, followed by Prince George's County, MD (0.263) and Oakland County, MI (0.203). Table 1.8 presents predictor balance and pre-treatment model fit. A high p-value and post-treatment to pre-treatment ratio less than one suggest no post-treatment change in Hennepin County. Synthetic control results are presented in Figure 1.6. Since the USOC did not select Hennepin County's bid to host the Olympic Games, there should be no discernible effects seen after the bid failed. From Figure 1.6, decreases in employment growth can be seen between actual and synthetic Hennepin County. However, Figure 1.7 presents placebo tests, highlighting the absence of any significant gap.

Synthetic Cook County provides an arguably more telling examination of the role played by selection bias in this setting. Chicago made it to the final phase of IOC voting to determine the host the 2016 Summer Olympic Games, costing Chicago $100 million in bid preparation costs in the process. Allegheny County, PA (0.566) and Cuyahoga County, OH (0.223) constitute most of synthetic Cook County, as shown in Table 1.9. Allegheny County, PA includes the city of Pitsburgh, and Cuyahoga County, OH includes Cleveland.

As with Hennepin County, results on Table 1.10 show a high p-value of 0.6 and a post-treatment to pre-treatment ratio less than one, indicating no significant impact on employment growth. There are no discernible differences between actual and synthetic Cook County found in either the synthetic control results nor placebo tests shown in Figures 1.8 and 1.9. Overall, synthetic control results from these two counties that made unsuccessful bids show no evidence of an economic impact, mitigating concerns that selection bias drives the results in actual host counties. This further validates the robustness of results for Los Angeles County, Fulton County, and Salt Lake County.

1.5 Conclusion

This paper analyzes the impact of three separate Olympic Games held in the United States between 1984 and 2002 on the employment growth rates in the counties that hosted the Games. The synthetic control method assesses this impact by constructing synthetic Fulton, Salt Lake, and Los Angeles counties to provide valid comparison groups for each Games. The Games examined include one Winter and two Summer Games. The results show transitory changes in employment growth following the awarding of the rights to host Olympic Games; positive transitory effects on two counties and a negative transatory impact in one.

A decrease in employment growth occurred in Los Angeles County in 1982, caused by hosting the Games. In contrast, Fulton County and Salt Lake County each experienced transitory increases in employment growth. Fulton County experienced increased employment growth in 1993, 1994, and 1996. Salt Lake County experienced a smaller increase in employment growth in a single year, 1996. The smaller size of the Winter Olympics compared to the Summer Olympics partially explains why Fulton County experienced a larger impact than Salt Lake County.

Back of the envelope calculations reveal the magnitude of the impact of hosting the Olympic Games by calculating the difference between Olympic host counties and their synthetic counterparts in significantly different years. For Los Angeles County in 1982, the only year of significant impact, the Olympic Games resulted in a decrease about 118,000 jobs relative to synthetic Los Angeles. In 1982 Los Angeles County actually lost over 86,000 jobs, experiencing an employment growth rate of nearly -2%. While Los Angeles County lost employment, synthetic Los Angeles County grew at a rate of 0.7%, accounting for an increase of nearly 32,000 jobs. The increase in jobs in synthetic Los Angeles County, coupled with Los Angeles County experiencing a decrease in over 86,000, leads to a net difference of about 118,000 jobs.

The same approach applies to increase employment growth experienced in Fulton and Salt Lake County. Significant differences between Fulton County and synthetic Fulton County in 1993, 1994, and 1996 led to an increase of about 63,000 jobs. The 63,000 increase resembles the forecasts in Humphreys and Plummer (1995). Job creation of over 24,000 in 1996 due to the Olympic Games resembles results in Feddersen and Maennig (2013b). Feddersen and Maennig (2013b) estimated an increase of around 29,000 jobs in 1996, the year the Games took place. Salt Lake County experienced significant positive employment growth in 1996. Based on the difference between outcomes in Salt Lake County and synthetic Salt Lake County, the Olympic Games accounted for an increase of about 17,000 jobs.

Results show that Olympic-generated increases in employment growth dissipated quickly in Fulton and Salt Lake County, consistent with results in previous research Baumann et al. (2012b) and Feddersen and Maennig (2013b). The transitory increase in Fulton and Salt Lake County can be attributed to increased construction activity following selection as the host city, as well as anticipation of increased future tourism as a result of hosting the Games. The absence of sustained increases in employment growth suggests that anticipated persistent increases in tourism do not occur. While Fulton County and Salt Lake County experienced transitory increases in employment growth, Los Angeles County experienced a decrease.

Overall, results presented in this paper call into question the use of the Olympic Games as a tool for local economic development. While hosting the Games may generate transitory increases in employment growth, the decreases in employment growth in Los Angeles provides evidence that potential hosts should proceed with caution when considering a bid to host the Olympic Games.

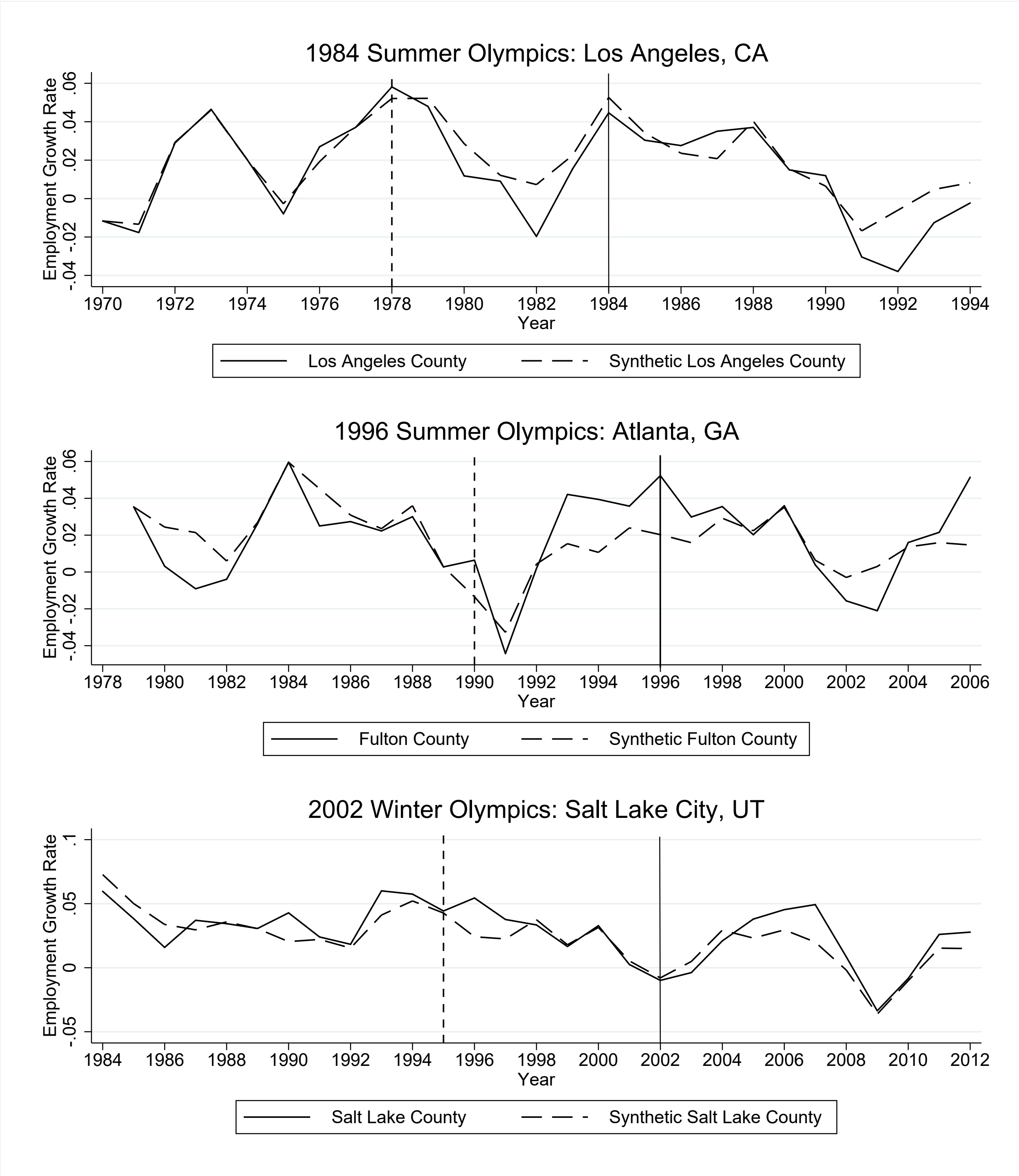

Figure 1.1: Impact of the Olympic Games on Employment Growth: Actual vs Synthetic

Table 1.1: Synthetic Control Weights: Los Angeles County

County	Weights
Middlesex County, MA	0.266
Bronx County, NY	0.239
Santa Clara County, CA	0.212
King County, WA	0.163
San Diego County, CA	0.121

Table 1.2: Predictor Balance and RMSPE: Los Angeles County

Predictor Variables	Actual	Synthetic
Population Growth	0.0046	0.0046
Income Growth	0.0757	0.0765
Employment Growth(1970)	-0.0117	-0.0116
Employment Growth(1974)	0.0202	0.0203
Employment Growth(1977)	0.0371	0.0371
Model Fit Pre-treatment		
Pre-treatment RMSPE		0.0040
Post-treatment/Pre-treatment RMSPE		3.4614
p-value		0.15

RMSPE=Root Mean Squared Prediction Error

Table 1.3: Synthetic Control Weights: Fulton County

County	Weights
Essex County, MA	0.42
Montgomery County, MD	0.355
Fairfield County, CT	0.169
Duval County, FL	0.038
DuPage County, IL	0.017

Table 1.4: Predictor Balance and RMSPE: Fulton County

Predictor Variables	Actual	Synthetic
Population Growth	0.0120	0.0118
Income Growth	0.0883	0.0882
Employment Growth (1979)	0.0354	0.0353
Employment Growth (1984)	0.0596	0.0595
Employment Growth (1989)	0.0028	0.0027
Model Fit Pre-treatment		
Pre-treatment RMSPE		0.0139
Post-treatment/Pre-treatment RMSPE		1.2955
p-value		0.5384

RMSPE=Root Mean Square Prediction Error

Table 1.5: Synthetic Control Weights: Salt Lake County

County	Weights
Pima County, AZ	0.485
DuPage County, IL	0.364
Oklahoma County, OK	0.146
Macomb County, MI	0.004

Table 1.6: Predictor Balance and RMSPE: Salt Lake County

Predictor Variables	Actual	Synthetic
Population Growth	0.0181	0.0181
Income Growth	0.0521	0.0524
Employment Growth (1984)	0.0596	0.0727
Employment Growth (1989)	0.0307	0.0305
Employment Growth (1994)	0.0574	0.0521
Model Fit Pre-treatment		
Pre-treatment RMSPE		0.0115
Post-treatment/Pre-treatment RMSPE		1.1625
p-value: RMSPE		0.4737

RMSPE=Root Mean Square Prediction Error

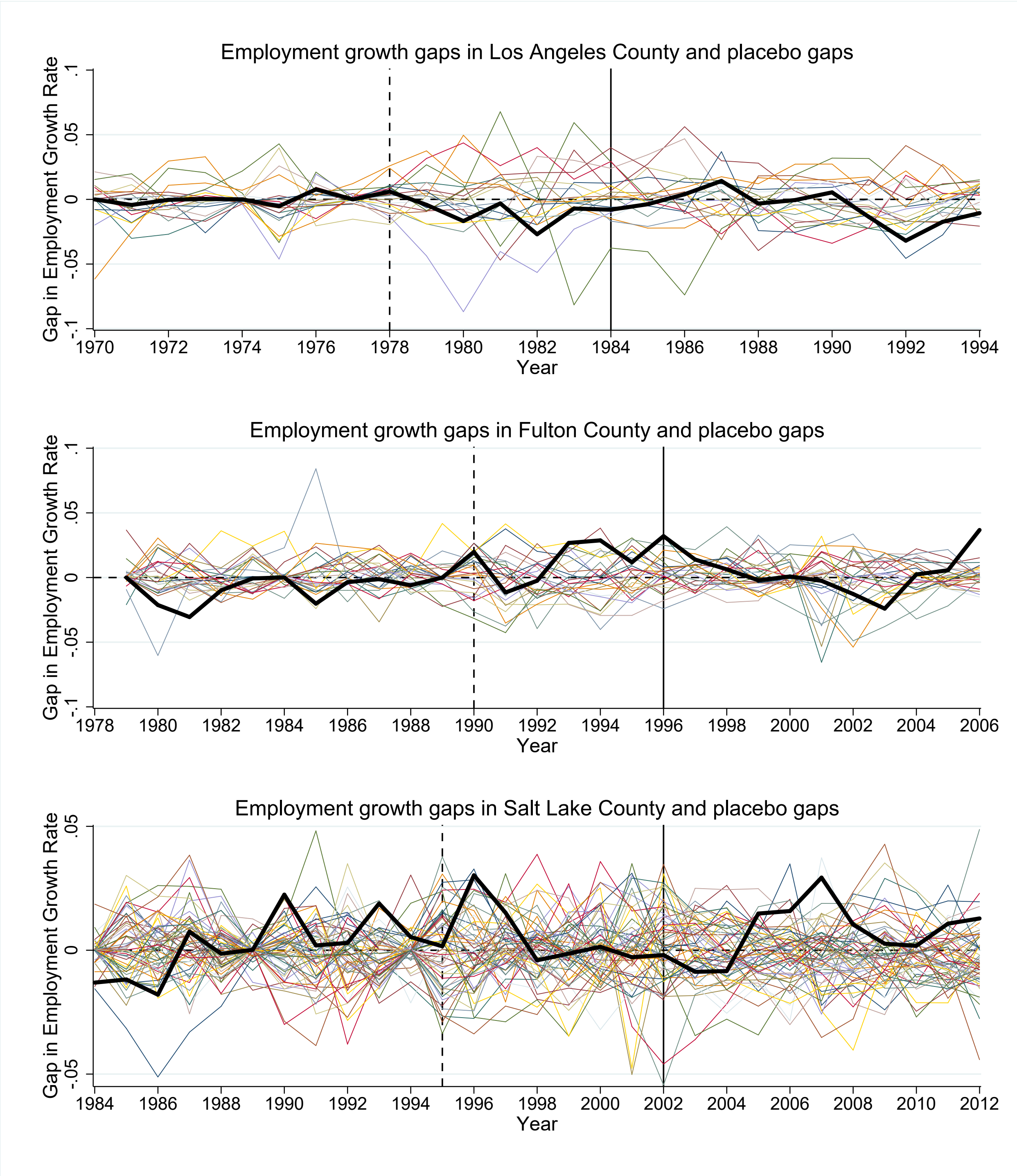

Figure 1.2: Employment Growth Rate Gaps in Host Counties and Placebo Gaps

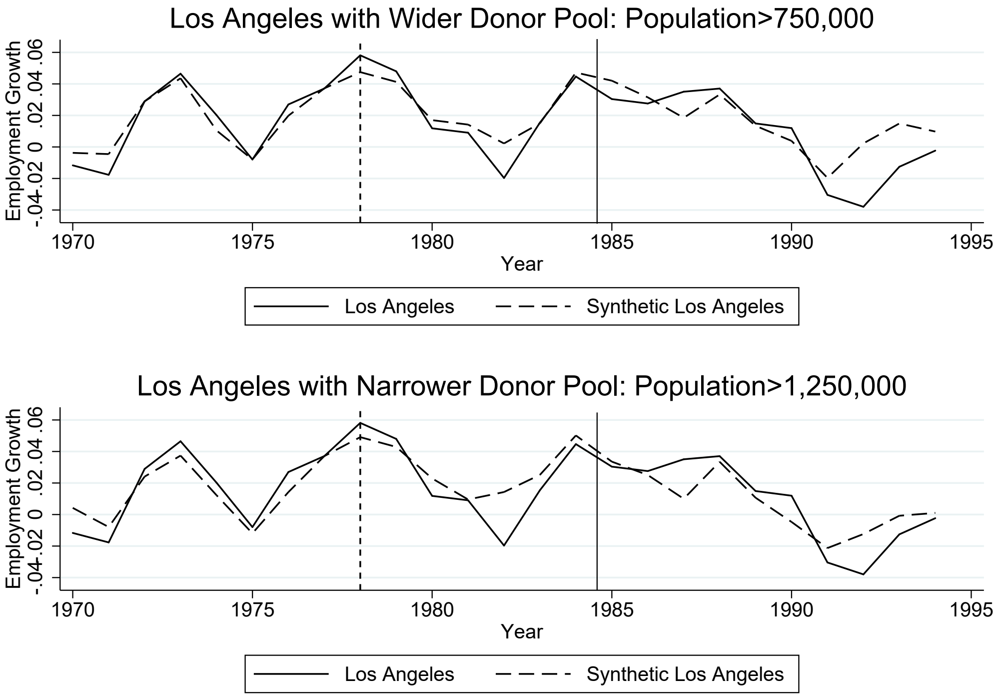

Figure 1.3: Synthetic Control Results with Alternative Donor Pools: Los Angeles County

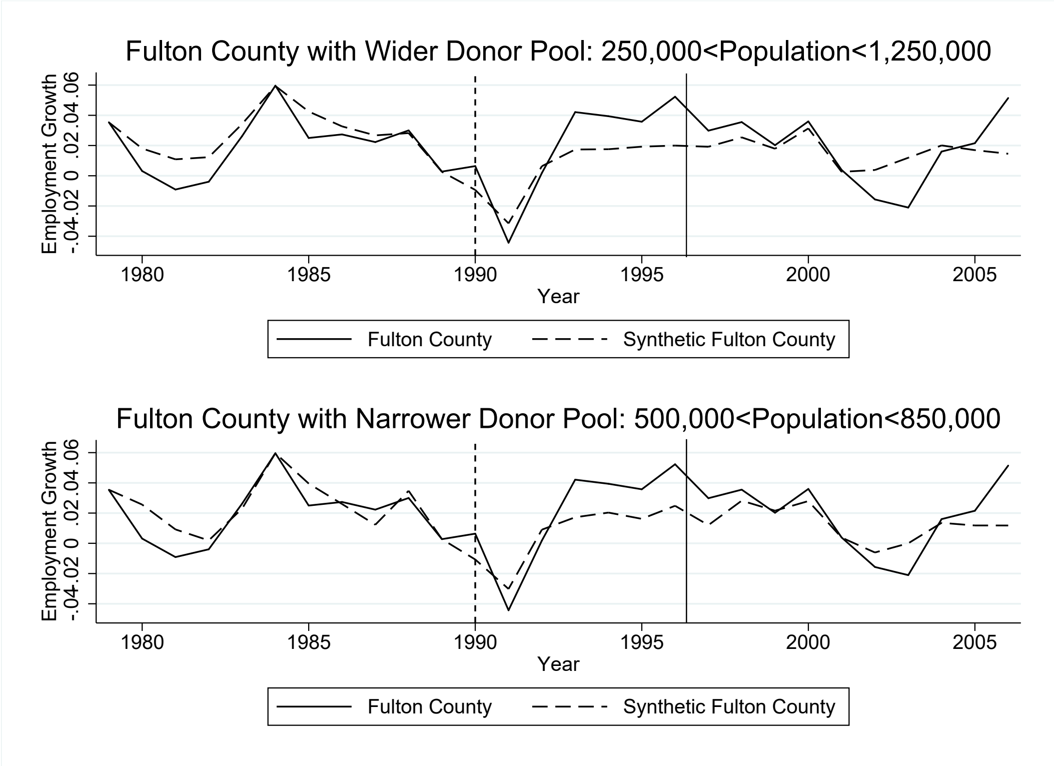

Figure 1.4: Synthetic Control Results with Alternative Donor Pools: Fulton County

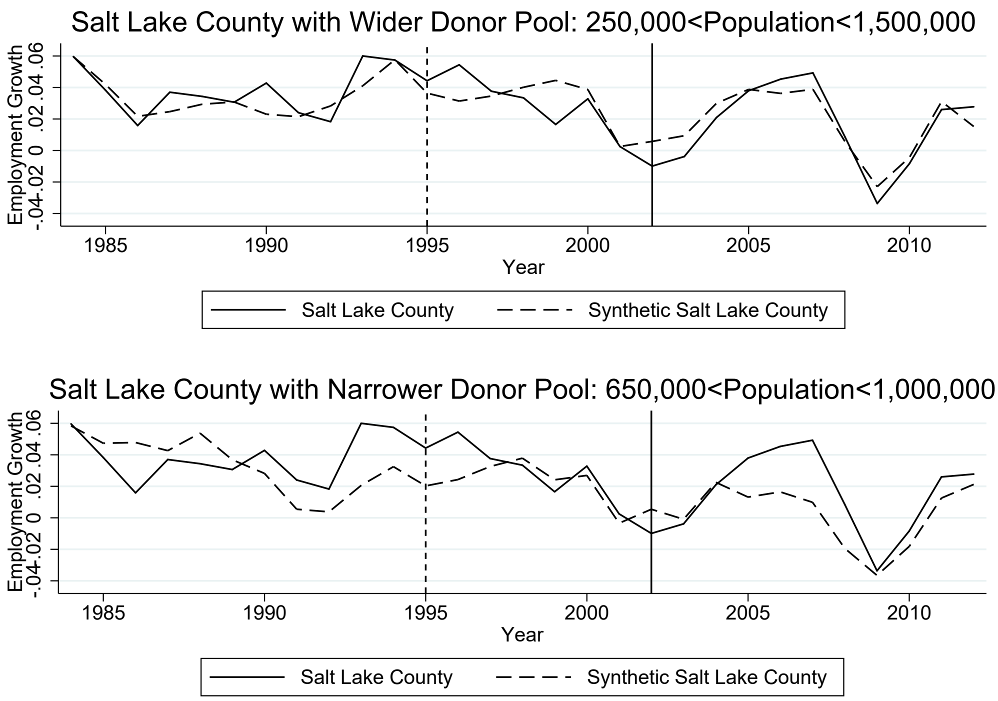

Figure 1.5: Synthetic Control Results with Alternative Donor Pools: Salt Lake County

Table 1.7: Synthetic Control Weights: Hennepin County

County	Weights
Hartford County, CT	0.365
Prince George's County, MD	0.263
Oakland County, MI	0.203
Contra Costa County, CA	0.133
Fairfax County, VA	0.036

Table 1.8: Predictor Balance and RMSPE: Hennepin County

Predictor Variables	Actual	Synthetic
Population Growth	0.0071	0.0071
Income Growth	0.0880	0.0880
Employment Growth (1977)	0.0441	0.0441
Employment Growth (1982)	-0.0177	-0.0097
Employment Growth (1987)	0.0422	.0422
Model Fit Pre-treatment		
Pre-treatment RMSPE		0.0155
Post-treatment/Pre-treatment RMSPE		0.7342
p-value		0.8478

RMSPE=Root Mean Square Prediction Error

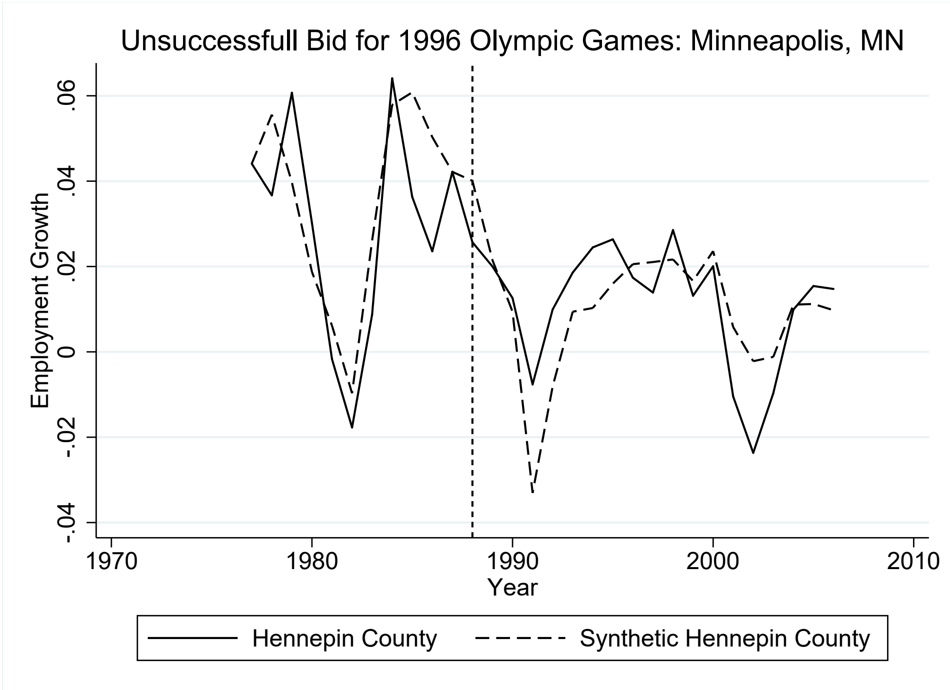

Figure 1.6: Synthetic Control Results: Hennepin County 1996 Olympic Bid

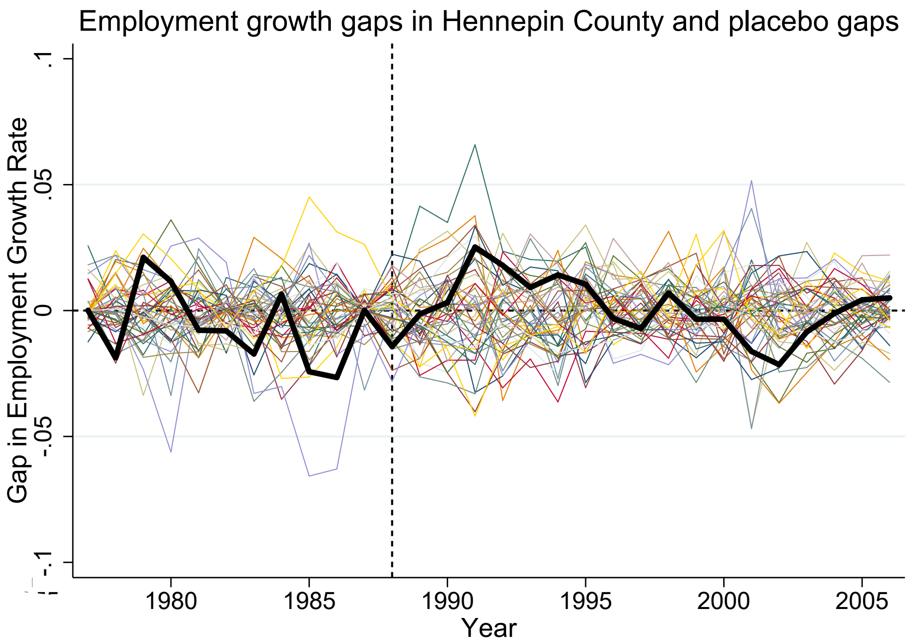

Figure 1.7: Placebo Tests: Hennepin County

Table 1.9: Synthetic Control Weights: Cook County

County	Weights
Allegheny County, PA	0.566
Cuyahoga County, OH	0.223
Palm Beach County, FL	0.085
Orange County, CA	0.077
Santa Clara County, CA	0.05

Table 1.10: Predictor Balance and RMSPE: Cook County

Predictor Variables	Actual	Synthetic
Population Growth	.042808	.0416506
Income Growth	-.0027651	-.0027719
Employment Growth (1998)	.0200487	.0200708
Employment Growth (2003)	-.0071709	-.0071971
Employment Growth (2008)	-.0051212	-.0051128
Model Fit Pre-treatment		
Pre-treatment RMSPE		0.0722
Post-treatment/Pre-treatment RMSPE		0.9521
p-value		0.6

RMSPE=Root Mean Square Prediction Error

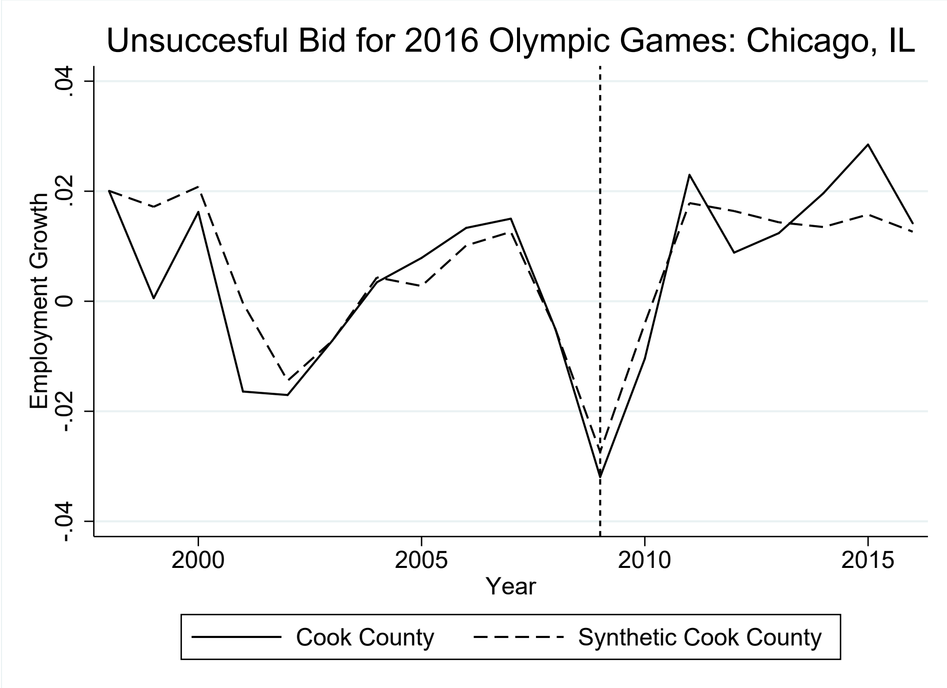

Figure 1.8: Synthetic Control Results: Cook County 2016 Olympic Bid

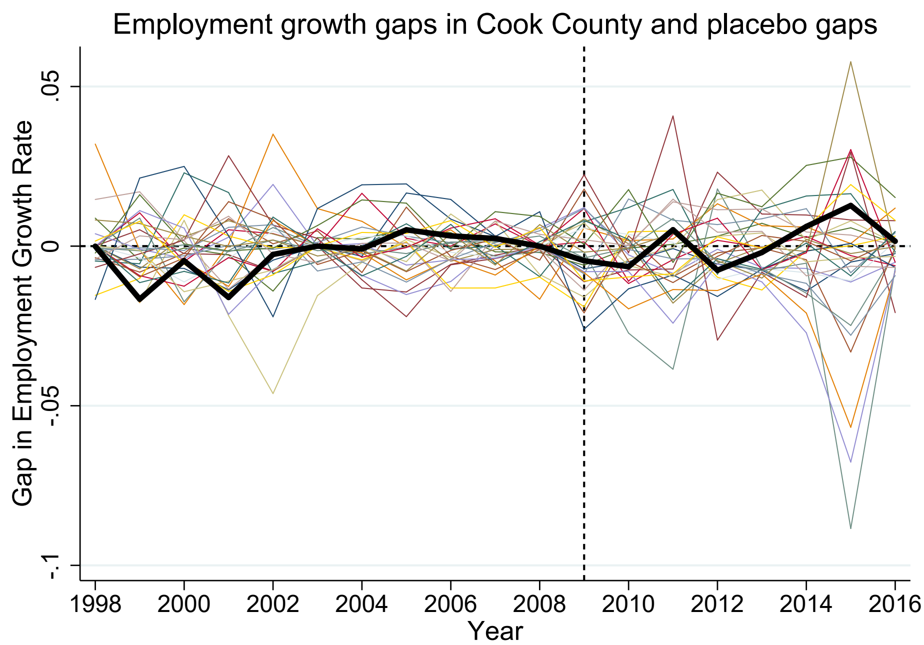

Figure 1.9: Placebo Tests: Cook County

Chapter 2

Wage Discrimination in the NBA: Evidence using Free Agent Signings

2.1 Introduction

The racial structure of host standard metropolitan statistical areas (SMSA) influences the racial structure of National Basketball Association (NBA) teams due to consumer preference to see players of their own race, potentially leading to a large racial wage gap (Burdekin and Idson, 1991). Racial wage gaps, their size, and their existence are essential topics of study in labor economics. Professional sports provides an appropriate setting to examine the potential impact of race on salary. Economists have studied racial wage discrimination in the National Basketball Association (NBA) throughout the 1980s, 1990s, and 2000s, specifically the discrimination against black athletes. Some studies report that black athletes were not paid as highly as their white counterparts; however, the results found across the literature are largely inconsistent. Moreover, this literature has not been examined in recent NBA history. Thus, inconsistent and outdated results motivated this study. We utilize an improved data set and empirical approaches not previously used in the NBA labor market literature to examine the presence and size of the racial wage gap in this labor market.

A portion of our empirical approach most resembles that of Holmes (2011). First, we use the same sample selection process, restricting the sample to include only free agent contract signings. Moreover, we use weighted least squares and quantile regressions to further explore our findings of discrimination in the NBA, the main approaches used by Holmes (2011) to find discrimination within the MLB. Additionally, we use the Oaxaca-Blinder decomposition an approach previously seen in the general labor market and sports literature, but not previously used to analyze the NBA labor market.

Our analysis goes beyond investigating an average racial wage gap. This research, in addition, further investigates discrimination in the NBA by considering three sources of racial discrimination: consumer, employer, and employee discrimination (Becker, 1971). Consumer discrimination is explored using the weighted quantile regressions with an interaction term between black players and the share of the metropolitan statistical area (MSA) population that is white. Employer and employee discrimination are examined by interactions between the race of players with the race of coaches and general managers, which is an approach previously explored by Hamilton (1997).

Our results show that black athletes are paid in the league on average 20.5% less than their counterparts, ceteris paribus. More importantly, 63.9% of this wage gap cannot be explained by observable characteristics and, therefore, is attributed to racial discrimination. Thus, our results indicate the presence of a racial wage gap of 13.1% in the NBA. The wage gap is shown to be robust through various econometric approaches using different specifications including or excluding population characteristics and using alternative statistics for player performance.

We find that consumer discrimination is the primary source for this racial wage gap. This result is derived from our weighted quantile regressions which include an interaction term between the percentage of white population in the employing team's MSA and an indicator variable for whether a player is black. The results indicate that the gap between black and non-black players increases as local share of white population increases. The quantile regressions also show the racial wage gap to only be significant for the upper portions of the salary distribution, which includes role and star players.[1] Role and star players are defined in this research as players with high court visibility relative to bench players, with bench players being located in the lower portion of the salary distribution.

[1]The characterization of players is given based on the distribution of salaries as done by Hamilton (1997). The specific criteria for characterization of players will be further defined in Section 4.

This type of discrimination manifests itself through consumers due to their preference for watching those of the same race on the court (Burdekin and Idson, 1991). The experience of watching a game is the product consumed by customers in this market; hence, the most visible players should be the only ones significantly affected by consumer discrimination if it exists in this market. The conclusions drawn by this paper arise and differ from previous literature due to an important empirical contribution this papers brings to the NBA labor market literature, the use of a data sample which considers only free agents.

The data set we use includes NBA free agency signings from 2011-2017. Data has been a limitation in this literature, as previous papers do not use free agents or usually includes short sample periods. Using free agent signings, previous season performance, and the correct use of other control variables provide an appropriate framework to explore a player's compensation for his expected current level of output.[2] In other words, we are able to more accurately capture his marginal revenue product. Holmes (2011) recognizes this shortcoming of the sports literature regarding the racial wage gap, but investigates the MLB labor market. We are the first to apply this to the NBA setting to examine the racial wage gap.[3] The length of the data set used in this paper must also be highlighted since most NBA labor market papers, with the exception of Hill (2004), Groothuis and Hill (2013), and Hill and Groothuis (2017), investigated wage discrimination against black athletes in the NBA using two or fewer years of data. The data set used here covers free agents from six NBA seasons, which gives us a sample of nearly 800 free agents.

2.2 Literature Review

The amount of papers that study the wage gap between black and non-black men is extensive. Lang and Lehmann (2012) provide a theoretical and empirical review on wage discrimination in the U.S. labor market. The divergence of results in this literature are generally explained by the different control variables and data range used by different authors

[2]For instance, an analysis of a player's pay compared to his current performance that, for example, is in his second year of a three year contract does not yield accurate results.

[3]Johnson and Hall (2018) utilizes free agent signings to examine the impact of variation in state income tax rates on NBA player salaries.

due to theoretical considerations and/or data limitations. After reviewing the literature, Lang and Lehmann conclude that a wage gap of approximately 10% exists between white and black male workers, which is similar to our results for the NBA labor market. Moreover, the authors point out an important result of Lang and Manove (2011), who state that the wages converge for workers with very high and very low levels of education, or human capital, which highlights the importance of analyzing different quantiles of the wage distribution.

A vital difference to be pointed out between typical goods and services market and the NBA is their final goods. Goods and services which can be consumed by individuals generally represent the U.S. labor market's final goods. On the other hand, the NBA labor market offers a final good which sells the experience of watching a basketball game. This is important because the NBA final goods are dependent on the exposure of its workers (players), which is not necessarily true for the U.S. labor market, since buyers frequently do not know which worker specifically produced their good or service they are consuming.

In regards to research on discrimination more specific to sports, the topics covered is broad. For instance, it covers the impact that race has on playing time and salaries in the NFL (Burnett and Van Scyoc (2015); Keefer (2013); and Keefer (2016)[4]), on the probability of an umpire calling a strike in the MLB (Parsons et al., 2011), and on the wages of English soccer players (Szymanski, 2000). Other studies, Hoang and Rascher (1999) and Groothuis and Hill (2004), have focused on exit discrimination finding contrasting results. The literature has also explored the connection between productivity and wage inequality (Berri and Jewell, 2004) and population racial structure and capital investments (stadium reforms) (Bodvarsson and Humphreys, 2013). Results put forward by Price and Wolfers (2010) suggest discrimination among NBA referees. Kahn (1991) provides a review of early studies on this topic related to all sports. Even though discrimination can be studied through several channels, the focus on this paper is to dig deeper on the empirical findings of wage discrimination against black players in the National Basketball Association (NBA).

There exists a substantial literature regarding NBA wage discrimination; however, the inconsistency of their results piqued our interest in this topic. Kahn and Sherer (1988) examine salaries in the 1985-1986 season to find that black players are underpaid by 20%.

[4]Keefer (2016) finds black players start and play more.

Moreover, they also finds that replacing a black player with a white player increases attendance, which indicates the presence of consumer discrimination. Burdekin and Idson (1991) studies consumer based discrimination testing the hypothesis that “whites prefer to see white players." Interestingly, they find that the percentage of white population in the host SMSA is strongly correlated with the percentage of white athletes on the respective NBA team.[5] Gius and Johnson (1998) claim that the racial wage gap was gone by 1996-1997. Hamilton (1997) shows no premium on average received by whites; however, using a quantile regression he highlights a preference from the audience for white players. Groothuis and Hill (2013) study exit discrimination, pay discrimination, and career earnings of NBA athletes using data from 1990-2008, finding conflicting results. Both reverse discrimination and discrimination are found to be potentially present, however the results found are not robust.

Hill (2004) finds that black players are underpaid by 14% to 20% after analyzing a period from 1990 to 2000, but that such wage gap drops out when controlling for height. Hill points out that not controlling for height caused the white indicator variable coefficient to capture the premium that taller players received, due to the fact that white players are on average taller than black players in the NBA. In addition, Kahn and Shah (2005) shows, with a monopsony model, that nonwhite players that were not free agents nor on rookie contracts were underpaid, but the difference was small under rookie contracts and small and insignificant for free agents in the 2001-2002 NBA season. Lastly, Ajilore (2014) focuses on whether white players suffer statistical discrimination finding no statistical differences between black and white.

Recent literature has focused more on the influx of foreign players. Eschker et al. (2004) shows a wage premium for foreign players for the 1996-1997 and 1997-1998 season, and Hoffer and Freidel (2014) finds that foreign players receive an average wage premium of approximately $900,000. Moreover, Hill and Groothuis (2017) find that foreign born who did not attend to college in the U.S. earn a premium in the 1990s, but that such premia disappears in earlier years. Foreign athletes changed how NBA teams scouted, drafted, and

[5]The hypothesis posited in Burdekin and Idson (1991) implies that “blacks prefer to see black players". Murray (2015) replicates the results found in Burdekin and Idson (1991) using data from the NBA for the 2009-10 through 2013-14 seasons finding a similar result.

acquired talent. Our paper addresses this concern by including indicators for both foreign-born players who played US college basketball and for those who developed their talents abroad.

The NBA discrimination literature experiences shortcomings. Most papers focus only one or two seasons[6] and do not use free agents data to determine wage discrimination. Using players that are in the middle of their contract to test for discrimination by using their past season or current season performance as control variables will not accurately estimate the determinants of a player's contract. Player performance in the prior season should not have any power in determining the player's contract value if it is not a newly signed deal during free agency.

A player can regress or improve drastically throughout the duration of his contract making him far outperform or underperform the expectations of his predetermined salary. Player injury is also a concern when considering players in the midst of a current contract.[7] The same intuition is valid for other control variables such as coach's and GM's race, signing team and original team win percentages, age, etc. For instance, this gives a possible explanation for the inconsistency of the presence and size of a racial wage gap presented by the NBA labor market literature. These inconsistencies in data sets that do not capture wage determinants provide us with an opportunity to add to this strand of literature.

2.3 NBA Labor Market

The data utilized in this paper contains free agency signings over a period of 2011 to 2017. The NBA labor market contains many intricacies. New incoming players generally enter the NBA through an entry draft. The structure of a player's contract is determined by his draft position, or a player can be undrafted in which he becomes a free agent. Salaries for first round draft picks follow a rookie salary scale. The value of the contract of a first round pick decreases as the number of the slot they are selected later in the draft, and can be negotiated between 80-120% of the scale value. Contracts for first round selections contain

[6] With the exception of Hill (2004), Groothuis and Hill (2013), and Hill and Groothuis (2017)

[7] For example, Derrick Rose played only 10 games in the 2013-2014 NBA season, while being paid over $17 million as part of a contract extension he and the Chicago Bulls agreed to in 2011.

two guaranteed years followed by team options for each the third and fourth season. Going into the fifth season of the contract a player can sign an extension, sign a qualifying offer, or become a restricted free agent. Second round picks and undrafted players do not receive guaranteed contracts and are able to negotiate their contracts. Rookie contracts are not considered in this paper as they are largely fixed and negotiated without regarding prior NBA performance.

Restricted free agency differs from unrestricted free agency in that players are not able to sign and play for any team. Unrestricted free agents are free to sign with any team they choose, conditional on that team desiring their services. In contrast, a restricted free agent is subject to his team's right of first refusal. Restricted free agents can sign an offer sheet from another team, of which his current team has the ability to match the offer and retain the rights to the player. Restricted free agents are generally paid a higher salary than unrestricted free agents, which is highlighted in our results.

Restricted free agency can impact the free agency period of the player, as well as teams that are interested in pursuing their services. Free agency during our sample period begins on July 1st followed by a short moratorium period. After the moratorium period, players can sign a contract or an offer sheet. For restricted free agents, after signing an offer sheet their current team has a three day period to match the offer. This three day period can affect a teams pursuit of other free agents and potential trade offers.

During free agency periods, teams are constrained by the amount they compensate players. Contracts have a minimum and maximum value that vary based on a player's accolades, NBA experience, whether or not a player is re-signing with their current team, among other characteristics. A maximum contract can be generally 20-35% of the total salary cap space of the team. A player can be incentivized to re-sign with their current team when he is a player that can draw a max contract. For instance, during the 2018 NBA free agency period LeBron James was eligible for a 5 year $205 million contract if he had chosen to re-sign with the Cleveland Cavaliers. He ultimately chose to receive “only" a maximum contract of 4 years for $152 million when he decided to sign with the Los Angeles Lakers.

Players can also receive performance bonuses written into their contracts such as playing a certain amount of games, and keeping a certain level of performance, which is evaluated through their statistics. Ideally, the minimum and maximum values should be censored for in the empirical analysis used in this paper. This required us to individually investigate which player received a max contract every time they appear in our sample. This investigation was done by reading news articles about new contracts signed. Unfortunately, we cannot say with certainty that the media indeed reported all players which received a max contract; hence, this variable may be misrepresenting the sample of players which signed a max contract. Due to this data limitation, our baseline results including an indicator for max contracts are presented in the appendix. The results with the inclusion of max contracts are similar in significance and magnitude.

2.4 Materials and Methods

2.4.1 Data Description

The data on NBA player race was retrieved similarly to Price and Wolfers (2010) and Van Scyoc and Burnett (2013). At least three different observers analyzed the pictures from NBA player bios on the NBA's official website and basketball-reference.com to determine whether a player appears to be black or non-black. This approach is appropriate as players will be discriminated against based on their appearance, and not their genetic race or ethnic background. An indicator is used to identify black players that takes a value of 1 or 0. The same approach is used to determine the race of coaches and general managers. Coaches' and GMs' information were gathered from a combination of basketball-reference.com, basketball.realgm.com, and news articles.

Information on NBA free agency signings from 2011-2017 will be used to find if discrimination exists in the NBA currently. Data on 797 NBA free agent signings was taken from spotrac.com, a website that aggregates data from various reliable sources of NBA information including transactions, signings, and contracts. Player salaries are taken as an average salary value by dividing the total value of their contract by the length of the contract in years. The natural log of a player's annual salary is the dependent variable of all the empirical specifications explored in this paper. The empirical specifications used in this paper also include proper control variables to allow us to better identify the impact of race on player's annual salary. Various statistics and characteristics are used for both team and player. These include: team winning percentage, player characteristics, performance statistics, and information on the MSA that contains the team.

Player characteristics include age, height, race, foreign-born indicators, draft position, and position played. Performance statistics used will be points, rebounds, assists, blocks, steals per game, and field goal percentage. Also included is the amount of games played by the player, their minutes played per game, and in what percentage of games played did the player start. The percentage of games started is used to help control for a starter versus a bench player. Games played is included to help control for players that are signed but do not play whether for skill deficiency or injury. All of the performance and games played statistics are from the season before the players signed their new contracts as their output in the previous season is assumed to be the main driver in their salary following their free agency. Most player statistics, player characteristics, and team winning percentage were all obtained from basketball-reference.com. Height and draft position were obtained from basketball.realgm.com.

An additional statistic is used in various specifications throughout this paper referred to as Value Over Replacement Player (VORP). VORP gives an aggregate measure of a players on court performance and their overall value to their team. This measurement comes from basketball-reference.com and was constructed by Daniel Myers. VORP compares the impact of players to a theoretical replacement player based on their Box Score Plus/Minus (BPM) and the actual percentage of their team's minutes played. Box Score Plus/Minus estimates how well a player performs compared to an average player per 100 possessions, which is defined as 0.0. For example, the highest BPM in the sample is LeBron James in 2015-2016 when he posted a 9.1 BPM, which means James was 9.1 points better per 100 possessions than the average player in the league. For the purposes of VORP, -2 is considered the value of a replacement player. The formula for VORP is $[BPM - \breve{}(-2.0)] * (percentage\ of\ minutes\ played)$*$(team\ games/82)$.

While any box score based metric is not perfect as they can not account for the importance of basketball IQ, fundamental skills, or how effective of a team defender a player may be, VORP is an appropriate measure to be used. In the specifications using VORP, the number of minutes, points, rebounds, assists, steals, and blocks per game and the number of games played in the previous season are all dropped. This is because VORP includes proxies for these performance measures among others in its calculations.

We also ran the regressions using Win Shares (WS), Player Efficiency Ratings (PER), and Wins Produced (WP) as alternative advancement measurements for performance.[8] These advanced performance statistics are able to capture the efficiency and productivity of a player better than the alternative specifications using per game statistics. Nonetheless, our results are robust across the use of any of these performance variables. The results are nearly identical in size and significance when using either VORP or WP. PER and WS yield similar significance, but slightly lower coefficients compared to VORP and WP. VORP, WS, and PER are each retrieved from basketball-reference.com, while WP comes from boxscoregeeks.com.[9]

To investigate the effect of coach and general manager characteristics in the contracting process, this study considers the coach and general manager race as well. Race information for coaches and general managers were retrieved from a combination of basketball-reference.com, basketball.realgm.com, and news articles.

Metropolitan Statistical Area (MSA) data to control for demographic characteristics of the city hosting the team includes total population and percentage of the total population that is white. This approach to control for population characteristics is motivated by Bodvarsson and Humphreys (2013). The population data came from the American Community Survey website. Statistics Canada from the government of Canada was used to gather population characteristics for the Toronto Raptors.

Table 2.1 shows summary statistics for the entire sample of free agents, black players, and non-black players. The average annual salary received by NBA free agents was over $5 million for each sample, with non-black players having a higher average salary than black players. This initial comparison served as a motivation to further investigate this finding.

[8]The results including these alternatives can be provided upon request.

[9]https://www.basketball-reference.com/about/glossary.html provides more information on the calculation of these VORP, WS, and PER. Wins Produced is discussed in Berri (2010).

Figure 2.1 shows the kernel density of black and non-black athletes. Overall, the differences in density across the salary distribution of non-black players appear to lie slightly above black players in the both the middle and upper sections of the salary distribution.

To produce meaningful results, we must first deter whether our data consists of representative sample of the NBA athletes. To do so we compare the distribution of black players in our sample, 78%, to the distribution of black players in the NBA as a whole, around 75-80%, which is fairly representative (Spears, 2016). As highlighted in the literature, the role of foreign players is important in the NBA. In our sample, almost half of non-black players, 46%, are foreign. Nonetheless, foreigners only constitute 16% of all the players in our sample. Overall, information on 797 free agent signings was obtained. Other noticeable differences between black and non-black players are the frequency which players re-signed, average draft position, and average performance (according to VORP). In general, non-black players perform better, re-sign more, and are drafted later in the draft.

2.4.2 Methodology

Three main econometric specifications are explored to investigate whether a racial wage gap is present in the NBA: a weighted Oaxaca-Blinder decomposition, a weighted least squares (WLS) wage model, and weighted quantile regressions. The weighted twofold Oaxaca-Blinder decomposition follows an approach previously used in the general labor market (e.g.: Neal and Johnson (1996), Neumark (1988), and Boudarbat and Connolly (2013)) and sports literature (e.g.: Van Scyoc and Burnett (2013), Keefer (2013), Burnett and Van Scyoc (2015), and Leeds and Leeds (2017)), but that from the best of our knowledge is for the first time being explored to analyze the NBA labor market. This decomposition was first introduced by Oaxaca (1973) and Blinder (1973). This approach explores how much of the gap between the regressions results of two different samples is explained by observable characteristics (Elder et al., 2010). It allows us to evaluate how much of the wage gap between two groups is not explained by the vector of predictors; in other words, we are able to determine how much of the wage gap is due to discrimination given that we have an appropriate set of control variables. In this section, we will present the standard twofold specification.[10]

First, consider two groups: non-black (1) vs. black (2) players. Let g be an indicator variable for 1 or 2. Let Y_g be the log of the average salary for a member of group g and for it to be defined by $Y_g = X_g'\delta_g + \epsilon_g$. We take the log of the average salary to transform the data to handle skewness. X_g represents a vector with predictors for group g. δ_g is the vector of coefficients for each predictor and the intercept and ϵ_g the residual for group g. Assuming that $E(\delta_g) = \delta_g$ and $E(\epsilon_g) = 0$, after estimating the regression coefficients for both groups, it is straightforward to define the racial wage gap as

$$\overline{G} = \overline{Y}_1 - \overline{Y}_2 = (\overline{X}_1 - \overline{X}_2)\bar{\delta}_1 + \overline{X}_2(\bar{\delta}_1 - \bar{\delta}_2), \tag{2.1}$$

Redefining the two terms on the right-hand side of the equation above as E and U, respectively, allows us to rewrite the equation as

$$\overline{G} = E + U. \tag{2.2}$$

E represents the racial wage gap that is "explained" by systematic differences in the predictors of both groups. In other words, the "endowment effect" (Jann, 2008). U, on the other hand, indicates the log salary differential that is "unexplained" by our predictors, which is defined in the literature which used this decomposition as the "discrimination effect".

In our model, we control for population, team, coach, general manager and player characteristics; race; performance; and season, team, and position fixed effects. Moreover, the model is weighted by the inverse of the amount of contracts signed by a player. It is weighted in this manner to control for players that sign multiple contracts throughout the period so the results are not driven by few players signing multiple short contracts. In the sample 250 players signed one contract, 143 signed two, and one player (Ronnie Price) signed six. The rest of the sample signed between three and five contracts. In the regressions, a player that signs one contract will have a weight of 1, a player signing two contracts will have a weight of 0.5, and so on.

The standard twofold decomposition creates a counterfactual base concern as noted by

[10] For a more detailed explanation of the model please refer to Jann (2008).

Boudarbat and Connolly (2013). Since the result generated from the model specified above is based on the perspective of group 2 (black players), altering the definition of group 1 and 2 can theoretically generate different results to some extent. As an answer to this concern, we instead run a pooled twofold Oaxaca-Blinder decomposition as suggested by Neumark (1988). This specification uses coefficients from a pooled regression (black and non-black together), where an indicator variable for whether a player is black is included (Jann, 2008). In addition, robust standard errors are also applied in the derivation of our results.

Another potential concern is the pooled twofold Oaxaca-Blinder decompositon may potentially understate the discrimination effect compared to OLS. To answer this concern, we use another approach to validate our racial wage gap empirical results, a weighted least squares (WLS) wage model. This empirical framework is very comparible with previously used econometric models seen in the NBA racial discrimination literature. The closest specification to ours, however, was used in the analysis of the MLB labor market by Holmes (2011). To determine the NBA racial wage gap, once again we take the log of the average salary. The log-linear model to test for discrimination based on race is then defined by

$$\begin{aligned} ln(Salary_{ijps}) = \gamma_j + \alpha_i + \tau_s + \beta_1 Performance_{ij(s-1)} + \beta_2 Population_j + \beta_3 Race_i \\ + \beta_4 WinningPct_{j(s-1)} + \beta_5 H_{is} + \beta_6 Coach/GMRace_{js} + e_{ijs}. \end{aligned} \tag{2.3}$$

The model includes all the predictors used in the Oaxaca-Blinder model. More specifically, $Performance_{ij(s-1)}$ contains a vector of performance measures for player i and team j in season $s-1$. $Population_j$ controls for the population level and proportion of the population that is white in the area surrounding team j. $Race_i$ is an indicator variable for player i taking a value of 1 if a player is identified as black. $WinningPct_{j(s-1)}$ controls for the performance of the team a player was under contract with in the previous year, and the team they signs with during free agency. The vector H_{is} contains player characteristics variables. $Coach/GMRace_{js}$ controls for race of coach and general manager in charge of team j at the beginning of season s. γ_j represents a fixed effect for the team a free agent signs with, α_i is a fixed effect for the position player i plays, and season fixed effects is defined by τ_s. This model is weighted by the same approach described in the weighted Oaxaca-Blinder model.

Running an unweighted ordinary least squares regression yields similar results, which can be provided upon request.

To explore whether the discrimination is concentrated in certain types of players, we also run weighted quantile regressions. These regressions are extremely useful in this scenario since their results are not based on the sample mean as the WLS regression; rather, they estimate the function for the natural log of salary quantiles conditional on the control variables specified by the model. According to Holmes (2011), such model diminishes the effect of outliers since they are based on the median of determined quantiles of the distribution of the dependent variable, which in this case is the log of a player's salary. The quantile regressions in this research solve the following minimization problem:

$$min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \rho_\theta(ln(Salary_{ijps}) - A_{ijps}\beta) \tag{2.4}$$

where using an indicator function, I(.), allows us to define ρ_θ as:

$$\rho_\theta(A) = (\theta - I(A \leq 0))A \tag{2.5}$$

For Equation (2.4) and Equation (2.5), A represents a vector including all the control variables specified on the right-hand side of Equation (2.3), and β represents the vector of coefficients of A_{ijps}. Moreover, changing θ defines which quantile we are getting our results based upon. Players once again are weighted by the inverse of the number of contracts which they signed during the sample period we analyze. The following section will analyze empirically such models and further investigate the results they generate.

2.5 Results

To examine the size of the racial wage gap we begin our discussion by interpreting the results from the Oaxaca-Blinder decomposition. We then re-evaluate our results using a weighted least squares (WLS) model. We follow the examination of the size of the racial wage gap by testing for the three different sources of discrimination previously highlighted by Becker (1971): employer, employee, and consumer.

Employer discrimination refers to an employer paying his/her employee less due to racial characteristics. In the NBA, we can think of this channel being translated to the relationship between players and general managers (GM) as well as their coaches, since a GM generally has control of the team, but with input from coaches. This relationship can also serve as the channel for employee discrimination. Consumer discrimination, on the other hand, refers to a decrease in salary explained by consumer preferences. In this league, since around 75-80% of players are black, a premium for non-black players could be seen as a preference of the audience for watching non-black players on the court. This paper follows Becker (1971), in which each of the channels of discrimination defines discrimination as being correlated to individuals' tastes for characteristics similar to theirs.

To investigate the relationship between the race of players with the races of coaches and GMs to test for employer and employee based discrimination, the WLS specification is used. In addition, we use quantile regressions to test for consumer discrimination.

2.5.1 Oaxaca-Blinder Decomposition Model

The results of our pooled twofold Oaxaca-Blinder decomposition are reported in Table 2.2. Columns 1-4 show different specifications, adjusting for alternative performance measures and adding population controls. Each specification includes player, team, GM, and coach characteristics. Control variables for columns (1) and (3) for performance include points, assists, rebounds, blocks, and steals per game as well as field goal percentage. Columns (2) and (4) use VORP instead of per game statistics to control for performance. In specifications using VORP, games played and field goal percentage are also dropped, as they are included in VORP. We also run regressions with different specifications using alternative advanced metrics such as WP, WS, and PER, which yield similar results and can be provided upon request. In columns (3) and (4), population controls are added, including population size and the percentage of the population that is white from the MSA hosting the player's signing team.

We select Column 4 as our preferred specification, proposing that it includes the most appropriate control variables based on the current data availability. In this paper, we select

VORP as our preferred performance measurement for two reasons: (1) its ability to be interpreted as a measure of a player's value and (2) its aggregation characteristic, which allows us to have a more concise measure of performance of a player.

Table 2.2 indicates the existence of a 20.5% gap in the mean average salary between black and non-black players, with the latter receiving this premium. 36.2% of this gap is attributed to systematic differences in characteristics of black and non-black athletes; however, this is not significant. The remaining portion of this gap is unexplained by the predictors used in our decomposition model. Following the papers in the literature which explore the Oaxaca-Blinder decomposition, we interpret this unexplained portion of the wage gap as the discrimination that black players suffer in the NBA labor market. Thus, this model indicates that black NBA players on average receive 13.1% less than non-black NBA players, all else equal.

2.5.2 Weighted Least Squares Regression

We estimate a weighted least squares model to further explore the existence of the racial wage gap in the NBA using Equation (2.3). This allows us to compare our results with the previous literature, which used similar econometric specifications, but data sets containing certain limitations as discussed. Moreover, Elder et al. (2010) shows that the twofold Oaxaca-Blinder may overstate the contribution of observed characteristics, thus understating the discrimination effect; hence, this specification also function as a robustness check for the Oaxaca-Blinder decomposition results. The results are shown in Table 2.3.

The results for player and team characteristics are as expected. Age is found to have a quadratic relationship with salary, also seen in Johnson and Hall (2018). A player is compensated more when re-signing with their current team, when they are restricted free agents, or when signing multi-year contracts. An explanation for players that re-sign receiving a higher salary comes from these players being eligible for a higher maximum salary from their respective team. Also, if a team values a player they can pay more to retain his services. This result shows that players do not on average give teams a “hometown discount". Restricted free agents are generally young players completing their rookie contracts that have higher potential than other free agents. Their current team has the ability to match an offer sheet and the player will be forced to stay with this team. This, as well as the NBA salary cap structure, leads to these players receiving large contract offers from other teams to make it more difficult for their current team to match.

We posit that multi-year contracts have a positive effect on wages because a player on a one year contract may be signed roster filler or otherwise not included in the long term plan for his team, while players with multi-year contracts will be on the team long-term. Contract length and salary are shown to be positively related as in Krautmann and Oppenheimer (2002). Points, rebounds, assists, and minutes per game are found to be positive as expected. Due to the amount of variables, per game statistics are not reported. In specifications using VORP, a higher VORP increases salaries. We also find that players signing after playing for a team with a high winning percentage receives a higher salary, and players that sign with a team that has a high winning percentage will receive a lesser salary. The lesser salary from signing with a better team could be a partial result of "ring chasing" behavior in the NBA. [11]

In each specification, discrimination is shown to be present in the NBA, ranging from black players being underpaid between 11.6 - 13.1% with our preferred specification being column 4. Column 4 shows a wage gap of 13.1%. The wage gap is identical when using both the Oaxaca-Blinder decomposition and WLS regression. Our general findings of wage discrimination goes against the qualitative results found by Ajilore (2014) and Hill and Groothuis (2017), but it agrees qualitatively with Kahn and Sherer (1988). Groothuis and Hill (2013) finds similar quantitative results, but their results are not robust. We believe our results diverge from previous literature due to the use of a data set including free agents only, which is something not explored by previous authors. As previously highlighted, our sample of free agents can more properly control for the marginal revenue productivity of each player when determining wages. Moreover, interestingly our results are comparable with the average wage gap found in the U.S. labor market (Lang and Lehmann, 2012). We however do not find any significant results on the effect of GM and coach race on player salary.

[11] For instance, veteran player David West opted out of a $12.6 million dollar contract, then signed with a San Antonio Spurs for the veterans minimum of approximately $1.5 million.

2.5.3 Employer and Employee Discrimination

We further explore the impact of coach's and GM's race to examine different sources of discrimination, such as through employer and employee preferences for working with an individual of the same race. In the WLS results shown in Table 2.3, there are no results found to show a significant effect of coach's or GM's race on player salaries. However, we are motivated to explore this relationship by Hamilton (1997). Hamilton finds no evidence of these variables being significant in determining a player's wage. We explore this relationship further because using data with free agents only could yield different results. Thus, we run WLS regression models as specified in column (4) of Table 2.3, but including an interaction term between player's and GM's race or between player's and coach's race. These results are shown Table 2.4 columns (1) and (3). Additionally, we use a logit regression to determine if black players are more likely to re-sign with a black coach or GM relative to a white coach or GM. As in the WLS regressions, players here are weighted based on the inverse of the number of contracts each player signed during the sample period which the data was collected. These results are shown Table 2.4 columns (2) and (4). The interaction terms between player with coach's and GM's race yields no significant results on player salaries or their likelihood to re-sign. When adding an interaction term between player and coach the wage gap is found to increase, but it lowers when interacting player's and GM's race.

2.5.4 Consumer Discrimination

In this subsection we investigate whether the results found above are due to consumer discrimination. Since around 75% of players in the NBA are black, we hypothesize that the discrimination results from a preference of the audience to observe a group of players on the court similar to themselves, as in Burdekin and Idson (1991). We do not assume that audiences have disdain for the opposite race (Becker, 1971). We posit the idea that consumer discrimination connects with players' visibility, in other words, it connects with the amount of time a player is on the court. Hence, we run quantile regressions, an approach previously used in the sports wage discrimination literature, as specified by Equation (2.4) and (2.5) to investigate if the discrimination result is concentrated on more visible players, players who spend more time on the court.

We assume here that the lower quantiles (10th and 25th) capture bench players, players who do not consistently play a large amount of minutes, as previously assumed by Hamilton (1997). We allow the 90th quantile to capture star players. This label is due to results from Humphreys and Johnson (2020) and Hausman and Leonard (1997), who show that star players are drivers of attendance in the NBA. Star players are the most visible players on the team, thus we can expect them to be subject to the highest amount of discrimination. Figure 2.2 plots minutes per game and salary, showing that players who receive a higher salary generally play more minutes per game. Additionally, the correlation between minutes per game and salary is 0.6268. Initially, utilizing quantiles to capture star players raises concern regarding the proportional of players that are black and non-black across the salary distribution, particularly star players. The racial breakdown in the top 10% of salaries and the entire sample are 77% and 78%, respectively. We refer to the 50th and 75th quantiles as role players, who can be seen as players who stay on the court the longest after the teams' star players (90th quantile). If consumer discrimination is present, we expect such result to be found on the 50th, 75th, and 90th quantiles due to the higher visibility of those players. Our quantile regressions results can be seen in Table 2.5.

Given our classification of different players based on distribution of salaries this table provides support for our hypothesis of presence of consumer discrimination in the NBA. This result can be observed in columns 3 through 5, which reports that black players who are role or star players (50th, 75th, and 90th quantiles, respectively) receive a significantly lower pay due to their race relative to their counterparts, all else equal. More importantly, the empirical results show that as a black player becomes more visible to the audience, higher is the racial wage gap he faces. In addition, the result of no discrimination found for bench players seems plausible due to their low average of minutes played per game.

To state with confidence that the NBA faces consumer discrimination warrants further analysis. In response, we ran additional quantile regressions including an interaction term between black players and the share of white population in the MSA which the player's team is located. The result of a negative and significant coefficient for the interaction term would indicate that the gap between black and non-black athletes' salaries increases in MSAs with a higher share of white population. The results of the quantile regressions including this interaction term are reported in Table 2.6.

In general, a higher share of white population in the MSA where the team is located is associated with higher salaries for both black and non-black players; nonetheless, this effect is not symmetric. Salaries for black athletes increase at a slower rate than non-black athletes as white population increases when interacting the two variables, increasing the racial wage gap. While the discrimination on the 50th percentile from Table 2.5 loses significance, the interaction term is found to be negative and significant for both the 75th and 90th quantiles, with the 90th percentile experiencing a larger effect. Burdekin and Idson (1991) supports this result as the authors find consumer preferences to watch their own race play shapes NBA team structure. This consumer preference leads to a higher value placed on non-black athletes as white population increases, due to the NBA being approximately 75% black. Since the final good consumed during an NBA basketball game is watching players play in the game, consumers are concerned with the race of players who are on the court. The preference of consumers to interact with those of their own race is apparent in the results shown in Table 2.6 and provides empirical evidence of the existence of consumer discrimination in the NBA.

Tables 2.5 and 2.6 indicate a possible premium being given to foreign-born players. At the bottom of the salary distribution foreign-born players with no U.S. college experience receive a premium, while foreign-born players with U.S. college experience who we label as role players receive a premium. A premium for foreign players is not found in other specifications. We do not make any claims to a foreign premium existing, as the results are not robust.

2.6 Conclusion

This study investigates empirically wage discrimination against black players in the NBA using an empirical method not previously used to study the NBA, as well as methods commonly used to strengthen the results found. We also use a more suitable data set considering only free agency signings. This allows us to more properly capture how a player is compensated for his marginal revenue product. Using both the Oaxaca-Blinder decomposition and a weighted linear wage model controlling for player performance, player, team, and employer characteristics, it is found that black NBA athletes are on average underpaid by 13.1% compared to their non-black counterparts. Moreover, our results suggest the presence of consumer discrimination in the NBA, finding an increase in the racial wage gap as the share of white population in the player's team MSA increases.

Figure 2.1: Kernel Density Function by Race

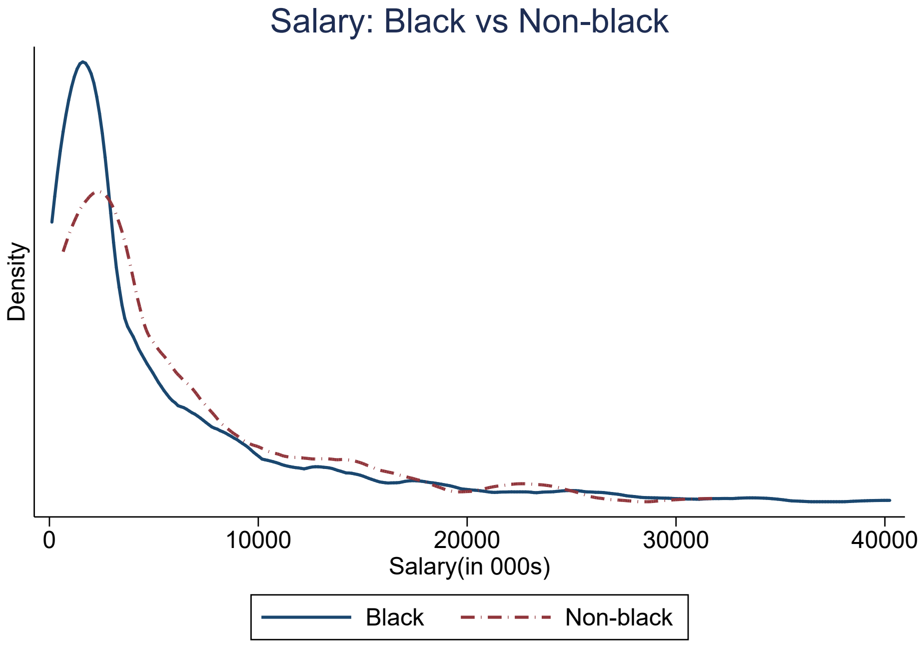

Table 2.1: Mean of variables used in the regressions, by race

	Non-black	Black	Total
Avg Salary (in 000s)	5590	5144	5242
Player is Black	0.000	1.000	0.780
Foreign-Born – No College	0.38	0.04	0.11
Foreign-Born – College	0.08	0.05	0.05
Age	28.66	27.69	27.90
Games Played	58.25	56.64	57.00
% of Games Started	0.377	0.374	0.375
Minutes Played Per Game	19.96	20.75	20.58
Points Per Game	7.741	8.244	8.134
Rebounds Per Game	3.842	3.437	3.526
Assists Per Game	1.766	1.826	1.813
Blocks Per Game	0.421	0.387	0.394
Steals Per Game	0.547	0.671	0.644
VORP	0.605	0.571	0.579
Previous Team Win %	0.518	0.508	0.511
Signing Team Win %	0.529	0.517	0.520
Re-sign	0.469	0.350	0.376
Height in Inches	80.62	78.45	78.93
Draft Position	33.69	28.59	29.71
Head Coach is Black	0.251	0.278	0.272
GM is Black	0.143	0.204	0.191
Population (in 000s)	6414	6157	6213
White Population	66.66	65.86	66.04
Restricted Free Agent	0.194	0.130	0.144
Multi-year Contract	0.606	0.592	0.595
Observations	175	622	797

Table 2.2: Pooled Twofold Oaxaca-Blinder Decomposition

	(1)	(2)	(3)	(4)
Non-black	15.07***	15.07***	15.07***	15.07***
	(205.34)	(202.87)	(205.45)	(203.44)
Black	14.86***	14.86***	14.86***	14.86***
	(361.58)	(359.25)	(361.49)	(359.23)
Difference	0.205**	0.205**	0.205**	0.205**
	(2.44)	(2.41)	(2.44)	(2.42)
Explained	0.0889	0.0780	0.0878	0.0743
	(1.12)	(1.02)	(1.11)	(0.97)
Unexplained	0.116**	0.127**	0.117**	0.131**
	(2.39)	(2.08)	(2.41)	(2.14)
Observations	797	797	797	797

t statistics in parentheses

Position, team, and year fixed effects included

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: Weighted Least Squares Regression Results

	(1) Salary	(2) Salary	(3) Salary	(4) Salary
Age	1.573**	3.210***	1.573**	3.208***
	(2.23)	(4.13)	(2.24)	(4.15)
Age^2	-0.261**	-0.542***	-0.262**	-0.543***
	(-2.16)	(-4.05)	(-2.18)	(-4.09)
Multi-year Contract	0.397***	0.541***	0.399***	0.546***
	(7.93)	(9.97)	(7.92)	(10.05)
Foreign-Born – No College	-0.00949	0.0468	-0.00386	0.0560
	(-0.14)	(0.59)	(-0.06)	(0.70)
Foreign-Born – College	0.123	0.0380	0.121	0.0305
	(1.50)	(0.35)	(1.49)	(0.28)
Restricted Free Agent	0.296***	0.392***	0.297***	0.392***
	(4.97)	(5.75)	(4.99)	(5.77)
Games Played	0.00128		0.00125	
	(0.94)		(0.92)	
% of Games Started	-0.0252	0.594***	-0.0194	0.598***
	(-0.28)	(7.75)	(-0.22)	(7.80)
Field Goal Percentage	0.103		0.0968	
	(0.34)		(0.32)	
Previous Team Win %	0.846***	0.319*	0.855***	0.341**
	(5.67)	(1.93)	(5.74)	(2.09)
Signing Team Win %	-0.773***	-0.720***	-0.777***	-0.710***
	(-4.40)	(-3.64)	(-4.39)	(-3.55)
Re-sign	0.156***	0.153***	0.156***	0.152***
	(3.35)	(2.79)	(3.34)	(2.76)
Draft Position	-0.000579	-0.00690***	-0.000593	-0.00689***
	(-0.54)	(-6.22)	(-0.55)	(-6.23)
Height in Inches	0.0198	0.00182	0.0191	-0.000596
	(1.49)	(0.13)	(1.44)	(-0.04)
Head Coach is Black	-0.0413	-0.0233	-0.0527	-0.0549
	(-0.65)	(-0.32)	(-0.78)	(-0.72)
GM is Black	-0.0690	-0.109	-0.0555	-0.0903
	(-0.87)	(-1.15)	(-0.67)	(-0.91)
Black	-0.116**	-0.127**	-0.117**	-0.131**
	(-2.29)	(-2.00)	(-2.31)	(-2.05)
VORP		0.287***		0.284***
		(11.19)		(11.13)
Population (000,000s)			-0.134	-0.0463
			(-0.56)	(-0.15)
White Population			0.0231	0.0539
			(0.82)	(1.57)
Per Game Performance	Y	N	Y	N
Position, Team, & Year Fixed effects	Y	Y	Y	Y
Observations	797	797	797	797
R^2	0.769	0.699	0.769	0.701

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: General Manager and Coach Relationship

	(1)	(2)	(3)	(4)
	Salary	Re-sign	Salary	Re-sign
Player is Black	-0.138**	-0.336	-0.124*	-0.301
	(-2.01)	(-1.16)	(-1.92)	(-1.08)
Head Coach is Black	-0.0600	-0.318	-0.0535	0.107
	(-0.48)	(-0.59)	(-0.70)	(0.32)
GM is Black	-0.0791	0.463	-0.0152	0.0822
	(-0.80)	(0.93)	(-0.09)	(0.12)
Black Player and Coach Interaction	0.00866	0.509		
	(0.07)	(0.98)		
Black Player and GM Interaction			-0.0737	0.461
			(-0.47)	(0.76)
Player & Team Characteristics	Y	Y	Y	Y
Position, Team, & Year Fixed effects	Y	Y	Y	Y
Observations	797	797	797	797

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.2: Salary and Minutes Played per Game Scatterplot

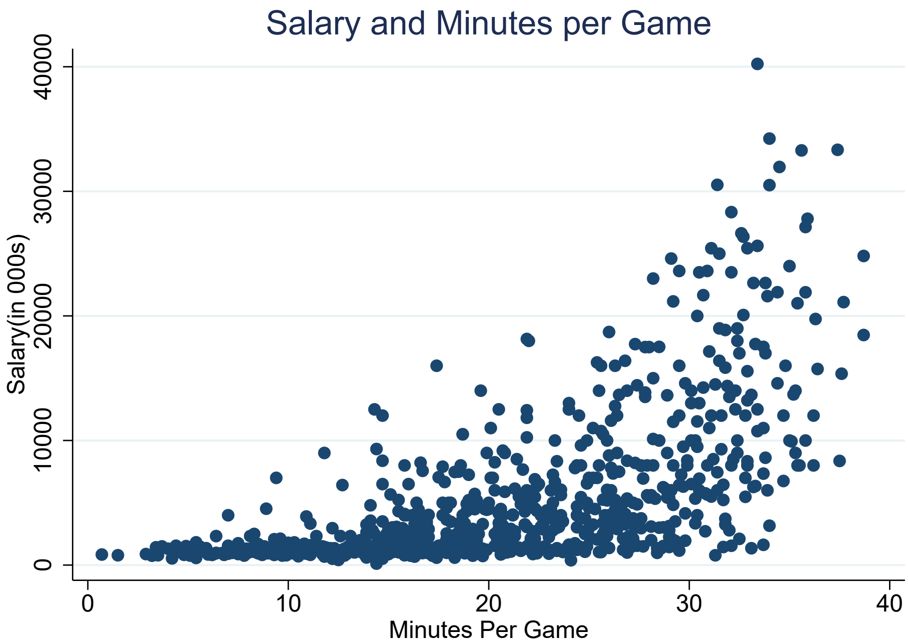

Table 2.5: Quantile Regression Results

	(10%) Salary	(25%) Salary	(50%) Salary	(75%) Salary	(90%) Salary
Age	3.667***	2.926***	4.288***	2.811***	1.916*
	(4.69)	(4.53)	(9.12)	(3.41)	(1.77)
Age^2	-0.601***	-0.479***	-0.727***	-0.466***	-0.322*
	(-4.62)	(-4.32)	(-9.31)	(-3.25)	(-1.75)
Population (000,000s)	-0.162	-0.180	0.0922	-0.00956	0.219
	(-0.42)	(-0.55)	(0.34)	(-0.03)	(0.55)
White Population	0.0188	0.0561**	0.0468	0.0478**	0.0623
	(0.40)	(2.16)	(1.24)	(2.01)	(1.40)
Multi-year Contract	0.374***	0.477***	0.643***	0.664***	0.584***
	(6.05)	(11.15)	(13.94)	(14.43)	(7.97)
Foreign-Born – No College	0.188*	0.0967	0.0590	0.0497	-0.00705
	(1.65)	(1.45)	(0.65)	(0.68)	(-0.08)
Foreign-Born – College	-0.0193	-0.0584	0.0458	0.191**	-0.0679
	(-0.15)	(-0.45)	(0.54)	(2.48)	(-0.51)
Restricted Free Agent	0.465***	0.460***	0.416***	0.340***	0.298***
	(4.46)	(7.99)	(6.64)	(5.28)	(2.76)
% of Games Started	0.510***	0.606***	0.612***	0.621***	0.604***
	(3.88)	(9.26)	(9.97)	(8.13)	(6.71)
VORP	0.273***	0.285***	0.235***	0.271***	0.270***
	(7.23)	(11.31)	(9.59)	(8.79)	(7.19)
Previous Team Win %	0.296	0.114	0.242	0.474**	0.404*
	(1.22)	(0.64)	(1.50)	(2.52)	(1.83)
Signing Team Win %	-0.623**	-0.645***	-0.560**	-0.760***	-0.703**
	(-2.13)	(-3.30)	(-2.48)	(-3.62)	(-2.01)
Re-sign	0.146**	0.130***	0.154***	0.147***	0.133*
	(2.29)	(3.13)	(2.85)	(3.00)	(1.89)
Draft Position	-0.00372**	-0.00454***	-0.00694***	-0.00732***	-0.0108***
	(-2.30)	(-4.53)	(-6.88)	(-6.17)	(-7.53)
Height in Inches	-0.0231	-0.0259*	-0.0141	0.0182	0.0247
	(-1.23)	(-1.93)	(-1.05)	(1.32)	(1.20)
Head Coach is Black	-0.0911	-0.0421	-0.0570	0.0148	0.0511
	(-0.75)	(-0.75)	(-0.82)	(0.18)	(0.51)
GM is Black	0.0765	-0.0434	-0.141	-0.126	-0.202
	(0.55)	(-0.42)	(-1.38)	(-1.46)	(-1.52)
Black	0.0416	-0.0136	-0.105*	-0.163**	-0.218***
	(0.49)	(-0.20)	(-1.76)	(-2.31)	(-2.63)
Position, Team, & Year Fixed effects	Y	Y	Y	Y	Y
Observations	797	797	797	797	797

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Quantile Regression Results: Black Player and White Population Interaction

	(10%) Salary	(25%) Salary	(50%) Salary	(75%) Salary	(90%) Salary
Age	3.673***	3.084***	4.313***	2.315***	1.885**
	(3.24)	(6.28)	(7.65)	(2.70)	(2.45)
Age^2	-0.601***	-0.507***	-0.732***	-0.391***	-0.311**
	(-3.16)	(-6.09)	(-7.71)	(-2.64)	(-2.24)
Population (000,000s)	-0.149	-0.171	0.0969	0.0893	0.0784
	(-0.39)	(-0.59)	(0.31)	(0.26)	(0.25)
White Population	0.0230	0.0519**	0.0500	0.0604**	0.102***
	(0.66)	(2.27)	(1.29)	(1.99)	(3.00)
Multi-year Contract	0.373***	0.475***	0.636***	0.674***	0.621***
	(5.97)	(11.92)	(13.73)	(12.47)	(10.78)
Foreign-Born – No College	0.188	0.114*	0.0552	0.0321	0.0374
	(1.36)	(1.78)	(0.56)	(0.38)	(0.41)
Foreign-Born – College	-0.0266	-0.0803	0.0454	0.161*	-0.0259
	(-0.27)	(-0.62)	(0.47)	(1.72)	(-0.22)
Restricted Free Agent	0.445***	0.460***	0.422***	0.282***	0.292***
	(3.71)	(9.01)	(6.25)	(3.30)	(3.48)
% of Games Started	0.510***	0.603***	0.612***	0.650***	0.574***
	(4.67)	(11.20)	(8.34)	(9.21)	(7.43)
VORP	0.270***	0.283***	0.240***	0.256***	0.264***
	(6.91)	(16.17)	(8.98)	(8.72)	(8.87)
Previous Team Win %	0.298	0.0824	0.249	0.493***	0.492***
	(1.20)	(0.51)	(1.49)	(2.98)	(2.62)
Signing Team Win %	-0.614	-0.627***	-0.562**	-0.733***	-0.755**
	(-1.63)	(-3.86)	(-2.38)	(-3.22)	(-2.53)
Re-sign	0.148**	0.141***	0.145**	0.156***	0.125**
	(2.55)	(3.64)	(2.57)	(3.26)	(2.19)
Draft Position	-0.00357**	-0.00464***	-0.00703***	-0.00803***	-0.0111***
	(-2.13)	(-4.82)	(-6.16)	(-6.56)	(-8.16)
Height in Inches	-0.0213	-0.0265**	-0.0148	0.0162	0.0132
	(-0.96)	(-2.07)	(-1.10)	(1.10)	(0.72)
Head Coach is Black	-0.0891	-0.0413	-0.0531	-0.00397	0.0647
	(-0.85)	(-0.93)	(-0.71)	(-0.04)	(0.74)
GM is Black	0.109	-0.0587	-0.145	-0.156	-0.232**
	(0.70)	(-0.86)	(-1.41)	(-1.57)	(-1.98)
Black	-0.00562	0.109	-0.0161	0.484	0.546
	(-0.01)	(0.36)	(-0.04)	(1.53)	(1.48)
Black Player and White Pop Int	0.000724	-0.00176	-0.00152	-0.00965**	-0.0114**
	(0.10)	(-0.42)	(-0.28)	(-2.05)	(-2.18)
Position, Team, & Year Fixed effects	Y	Y	Y	Y	Y
Observations	797	797	797	797	797

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Can Legalization “Weed" Out Risky Behaviors? Determining Whether Marijuana Acts as a Substitute or Complement

3.1 Introduction

Potential economic, political, and social outcomes associated with legalization of recreational marijuana appears frequently in debates. Support for marijuana stems from multiple sources including medicinal properties, allowing law enforcement to focus on more serious crimes, advancing freedom of choice, and generating large tax revenues. Opponents to marijuana legalization claim that marijuana would increase traffic accidents, harms users, and acts as a complement to other drugs and risky behavior. The introduction of legal recreational marijuana potentially impacts the consumption of alcohol and tobacco for individuals. Thus, legalization provides an important topic since tobacco and alcohol represent the first and third most common causes of preventable deaths in the U.S.[1]

Legalization began in November 2012 during the Presidential election. Colorado and Washington (state) each held votes that successfully legalized recreational marijuana. Le-

[1]https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/alcohol-facts-and-statistics

gal sales of marijuana eventually became implemented in January 2014 in Colorado, and July 2014 in Washington. Following the legalization of Washington and Colorado, Oregon, Washington D.C., and Nevada followed among others. These states are the subject of this study as they represent the only states that have both legalized and implemented marijuana into the legal market. These states will be included in the treatment group in this paper, with the control group consisting of states that have liberalized marijuana through legalizing medical marijuana or decriminalizing marijuana. This control group provides a sample of states that are more comparable to states that legalize marijuana than states that have made no progress toward legalization.

The behaviors examined include moderate and risky consuming alcohol and tobacco. Examining these behaviors determines whether marijuana can act as a substitute or complement for these other legal goods. Previous literature found that marijuana can be a complement or a substitute to various risky behaviors (Anderson et al., 2013; Chu, 2015; Chan et al., 2020). This paper uses a differences-in-differences approach to examine whether marijuana acts as a substitute or complement. In order to achieve covariate balance between the treated and control samples, entropy balancing is used (Hainmueller, 2012). Entropy balancing creates a better matched sample to satisfy selection on observables.

This paper contributes to the literature by exploring a wide array of risky behaviors across multiple legal goods following legalization of recreational marijuana. Previous literature largely focuses on medical marijuana laws and generally studies the impact of marijuana on one particular good. In addition, diff-in-diff with entropy balancing determines the impact of legalization of marijuana on residents of legal states compared to liberalized states, allowing for well defined treated and control groups. This allows to test for possible casual evidence of the impact of legalization on risky behaviors. Results show a decrease in overall alcohol consumption, the use of smokeless tobacco, and less drunk driving indicating that marijuana acts as a substitute.

3.2 Literature Review

Extensive research exists on the relationship between marijuana and other good, yielding mixed results. Subbaraman (2016) provides an interdisciplinary review of literature studying substitution and complementarity of alcohol and cannabis. Of the 39 papers reviewed, substitution is supported by 16 while 10 support complementarity.[2]

Early studies show a complementary relationship between marijuana and other substances. Pacula (1998) and Williams et al. (2004) find that the demand for both alcohol and marijuana decrease as price of alcohol increases indicating the two substances act as complements. Saffer and Chaloupka (1999) look at the cross price effect of marijuana, alcohol, cocaine, and heroin generally finding a complementary relationship with marijuana and the other substances.

Results put forth by more recent literature contradicts previous studies. Anderson et al. (2013) find that following the introduction of marijuana laws, traffic fatalities decrease, with this effect being larger for alcohol related fatalities. This result suggests that alcohol and marijuana are substitutes. Using a regression discontinuity, Crost and Guerrero (2012) find marijuana consumption drops after the age of 21 while alcohol consumption increases, indicating alcohol substituting for marijuana. Choi et al. (2019) examines the substitution of marijuana and cigarette by studying medical marijuana laws. Medical marijuana laws cause a decrease in adult cigarette consumption by 1 to 1.5 percentage points. The reductions leads to substantial cost savings of $4.6-6.9 billion.

Dragone et al. (2019) examines the impacts of recreational marijuana in counties of states that legalize. Dragone et al. (2019) focuses their study on crime, finding a decrease in rape, property crime, and thefts.[3] To explore the mechanisms, they consider marijuana a substitute for alcohol and other drugs.[4] While marijuana is associated with relaxation, alcohol is associated with aggression. Dragone et al. (2019) potentially understates due to interstate trafficking of recreational marijuana, particularly in a study using border counties.

[2] 12 papers support neither and one paper supports both.

[3] Further studies find the marijuana legalzation is not associated with an increase in crime (Morris et al., 2014; Kepple and Freisthler, 2012; Freisthler et al., 2013).

[4] Dragone et al. (2019) does not specify what drugs are included when discussing “other drugs"

The intensity of interstate trafficking appears in Hansen et al. (2017), finding that following legalization in Oregon, marijuana retailers along the border of Washington and Oregon experienced an immediate decrease in sales by 41%. Hansen et al. (2020) continues research on recreational marijuana looking at recreational marijuana legalization and traffic fatalities using the synthetic control method. Constructing both a synthetic Washington and Colorado, no difference in traffic fatalities involving alcohol or marijuana appear.

Chu (2015) examines the relationship between the passing of medical marijuana laws and usage of marijuana and hard drugs. Results show a decrease in arrests for possessing heroin or cocaine. The decrease in arrests for heroin and cocaine, and the increase in marijuana usage suggest that marijuana is a substitute for these hard drugs. Further evidence for marijuana being a substitute is drawn from a decrease in treatment for heroin. Chu (2015) also finds that marijuana usage increases following the passage of medical marijuana laws. Chu (2014) results show an increase in marijuana use as well.

Results presented in Powell et al. (2018) show that broad access to medical marijuana facilitates marijuana acting as a substitute for opioids. Marijuana's ability to substitute for strong and addictive opioids lies in the liberal allowances for dispensaries to provide marijuana. Powell et al. (2018) find that medical marijuana laws reduce both daily opioid doses filled and opioid overdose deaths. Chan et al. (2020) further exhibits marijuana's ability to act as a substitute for opioids. Results show large decreases in opioid mortality following the introduction of medical marijuana dispensaries and further decreases after introducing recreational marijuana dispensaries. Marijuana again appears as a substitute for opioids in Livingston et al. (2017).

This paper extends upon this literature as it examines the impact of legalization of recreational marijuana on the consumption of other legal goods at the individual level. Previous studies have not focused on the variety of consumption behaviors studied here, especially smokeless tobacco. Also, the focus on this paper is not the date of legalization, but the date in which marijuana became available in the market. Looking at the introduction of marijuana into the legal market, combined with using individual level data, allows the relationship between marijuana and other legal good to be thoroughly explored. Using the date marijuana enters the legal market follows Chan et al. (2020). Chan et al. (2020) finds no impacts on opioids following the passing of medical marijuana or recreational marijuana laws. However, the significance appears following dispensaries opening.

3.3 Empirical Analysis

3.3.1 Data Description

The data used in this paper comes from the 2010-2017 waves of the Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS is conducted annually to a randomly selected sample of the population, providing a representative sample of the U.S. population. The Center for Disease Control and Prevention, with assistance from state governments, conducts the BRFSS by phone, including individuals over the age of 18 in each state in the U.S. The BRFSS is rich with data for the individuals selected including personal characteristics such as age, income, education, health status, family structure, and many other economic and demographics characteristics. Table 3.1 reports the means of personal characteristics in legal, liberalized, and all states.

The data includes risky behaviors in which individuals partake. The risky behaviors include use of alcohol, cigarettes, and smokeless tobacco. An indicator variable for alcohol use equals 1 if an individual consumed any alcohol in the last 30 days. Furthermore, data includes how many days out of 30 that they consumed alcohol. In addition to use and frequency of alcohol consumption, additional variables indicate binge drinking, the amount of binge drinking, an indicator variable for drunk driving, and incidences of drunk driving. Variables for tobacco include whether an individual is a smoker or user of smokeless tobacco, and an indicator variable if individuals use it daily or not. An additional indicator variable shows if an individual had made a serious attempt to quit smoking in the last 12 months.

It is worth noting that not all of the variables for risky behaviors are available in each wave of the BRFSS. Data on drinking and driving is only available in the 2010, 2012, 2014, and 2016 waves. Also, individuals taking this survey do not always answer the questions regarding these risky behaviors, leading to incomplete data. However, the large amount of observations included yields a representative sample. Ideally, the BRFSS would include information on marijuana use. While the survey includes a question regarding marijuana use, the question appears sporadically. The data does not include enough information to be useful in this study. Many states do not include any information nor does information appear in each year of the BRFSS.

Descriptive Statistics are reported in Table 3.2, containing means for each outcome variable in legal, liberalized, and all states. There are noticeable difference between the means of outcome variables in the treated and control groups. For instance, 58.6% of the treated sample consumed alcohol, while the control group consumed at a rate of 49.8%. Also, the amount of days consuming alcohol in the last 30 days was 4.518 days in legal states, and 3.423 in liberalized states. Table 3.3 reports summary statistics for legalizing states before and after recreational marijuana was introduced into the market.

3.3.2 Methodology

A differences-in-differences approach examines the impact of the legalization of recreational marijuana on individuals partaking in the risky behavior of other legal goods such as alcohol and tobacco. The model estimated is defined by

$$Y_{ismt} = \alpha_0 + \beta_1 Legalized_{st} + \beta_2 X_{it} + \tau_s + \lambda_m + \gamma_t + \epsilon_{ist} \tag{3.1}$$

in which Y_{ismt} represents outcome variables including a variety of uses of alcohol and tobacco. $Legalized_{st}$ captures the period following the introduction of recreational marijuana into the market indicating treatment. The coefficient shows whether marijuana acts as a substitute or complement for the good considered. Colorado, Washington (state), Oregon, Nevada, and Washington D.C represent the treated group. States that pass laws to legalize marijuana may not be considered treated. This is due to the time lag between passing a legalization legislation and introducing recreational marijuana into the market. These states include California, Maine, Massachusetts, Michigan, and Vermont.

Equation (1) includes a vector of individual characteristics (X_{it}). Variables include information regarding individuals age, education, employment, household structure, and reported health status. Household structure included whether the individual is married or not, as well as children in the household. Health status includes reported overall good health, as well as the number of days in the past 30 days that an individual reported poor physical and mental health.

To control for variation over time γ_t represents fixed effects for year of the phone interview conducted. Month fixed effects, λ_m, control for variation of consumption of the goods considered throughout the year. With different states having different laws regarding alcohol, along with social and political factors associated with marijuana legalization, state fixed effects are included, τ_s (Spetz et al., 2019).

State fixed effects will not capture all of the social and political difference between states that legalize marijuana and those in which marijuana remains illegal. The control group must be constructed to handle for the large potential large differences between these states. To do so, the control group is constructed of states that have liberalized marijuana through legalizing medical use, or decriminalizing possession of marijuana. Overall, this excludes 13 contiguous states from the sample. The sample of legal, liberalized, and illegal states is shown in Figure 3.1.[5]

A potential issue with the diff-in-diff strategy shown above is the possibility of an unbalanced treatment and control group. To correct for this, entropy balancing is utilized to balance the samples. Proposed by Hainmueller (2012), entropy balancing creates a nearly identical control group to be compared to the treated group. Balancing is done by reweighting scheme that calibrates unit weights. This technique is utilized by Grossman et al. (2019), studying the impact of investments made by the Appalachian Regional Commission (ARC) on ARC counties compared to non-ARC counties.

Individuals are balanced on multiple personal characteristics. These characteristics include: age, sex, marital status, employment, income, education, and the number of children in the household. Table 3.4 shows entropy balancing using the full sample. Before entropy balancing there exists a significant difference between the sample of legal and liberalized states. The significance disappears following entropy balancing. Due to inconsistency in reporting of the behaviors studied, entropy balancing is conducted to balance legal and liberalized states prior to analyzing each outcome considered. The tables showing the balancing

[5]The sample does not consider Alaska or Hawaii.

process are not included here due to the large amount of extra tables that would need to be included.

3.3.3 Results

The regression results for Equation (1) are shown in Table 3.5-3.10. Table 3.5 shows the impact of legalization on overall use of alcohol. Columns 1-3 show the results considering any consumption of alcohol, while columns 4-6 displays results on the number of days drinking in the last 30 days. When controlling for reported health in column 3, a decrease in any consumption of alcohol appears significant. In each regression using the number of days drinking, a decrease in the number of day drinking is found to be significant. These results indicate that marijuana acts as a subsitute for alcohol. Introducing recreational marijuana into the market leads to a decrease in alcohol consumption. Considering Colorado after legalization, the Colorado Department of Revenue reported fewer gallons of alcohol sold in 2015 than in 2014. A decrease occurs again in 2016 compared to 2015.

Further considering the amount of alcohol consumed, Table 3.6 reports results on the maximum number of alcoholic beverages consumed in a sitting. When controlling for education, employment status, and measures of health the maximum number of drinks in a sitting decreases significantly. Table 3.7 shows the results of the impact of legalization on binge drinking, with no significant impact being found.

Table 3.8 shows the result of possibly the riskiest behavior studied, drinking and driving. Columns 1-3 show any occurrence of drinking and driving in the last 30 days, and columns 4-6 shows the number of incidence of drinking and driving reported. Across each specification, it is found that drinking and driving decreases due to legalization. Using Colorado as an example, according to the Colorado Task Force on Drunk & Impaired Driving, the number of fatal crashes increased from 407 in 2011 to 558 in 2016. However, while fatal crashes have increased, the number of fatal crashes in which a BAC of 0.08 or higher was reported only increased from 160 in 2011 to 161 in 2016. A limitation of this paper is the inability to directly examine the consumption of marijuana. Thus, marijuana-related traffic accidents substituting for alcohol-related traffic accidents represents a potential concern. Hansen et al.

(2020) partially mitigates this concern, as findings indicate no difference between traffic fatalities involving marijuana following legalization.

Results show that marijuana acts as a substitute for alcohol, but what about tobacco products? Legalization and its impact on tobacco products are reported in Tables 3.9 and 3.10. Table 3.9 examines cigarette smoking (columns 1-3) and the use of smokeless tobacco (columns 4-6). Legalization appears to not have an impact on smoking cigarettes, but across all specifications a significant decrease in the use of smokeless tobacco appears. Table 3.10 further shows a decrease in the use of smokeless tobacco, finding that individuals in legalized states are less likely to be daily users of smokeless tobacco. While marijuana does not appear to substitute nor complement cigarettes, results show marijuana and smokeless tobacco to be substitutes.

3.3.4 Falsification Test

Utilizing the same approach considering failed recreational marijuana votes serves as a falsification test. State with failed votes will represent the "treated group". In November 2015, voting in Ohio included a measure to legalize recreational marijuana. The vote to legalize recreational marijuana in Ohio failed with 63.65% voting no. In Arizona, 51.32% voted no causing a vote to legalize marijuana to fail in 2016. These failed votes provide a setting to use as a falsification test.[6]

Ohio and Arizona are compared to other states that have liberalized marijuana. Legal marijuana states and states in which no marijuana liberalization has occurred are excluded from the sample. Table 3.11 presents the results of the falsification test. Results clearly illustrate that no significant changes occurred following a failed vote to legalize marijuana. This strengthens the results reported above showing that marijuana acts as a substitute, lessening risky behaviors involving alcohol and tobacco.

While Arizona passed medical marijuana laws in the 1990s, Ohio recently passed a medical marijuana law in 2016. However, the first license to sale medical marijuana was not issued until 2019. Considering the results found by Chan et al. (2020), the results of this

[6]Information on voting outcomes are obtained from Ballotpedia.

falsification are not surprising. Chan et al. (2020) finds that effects do not appear until marijuana laws are implemented, not when the law itself passes.

3.3.5 Mechanisms

Two primary mechanisms drive the results presented in this paper. First, the increased availability of marijuana causes marijuana to more easily be substituted for other legal goods. While marijuana use occurs in states without legalization, being able to easily access marijuana makes substitution much easier. When illegal, consuming marijuana involves explicitly breaking the law. Without this legal barrier, individuals wanting to consume marijuana are able to do so without fear of legal consequences.

There also exists a negative stigma from consuming marijuana when illegal (Brown, 2015). Marijuana becoming legal potentially lessens the stigma associated with consumption, increasing demand. With the stigma lessened and availability increased, marijuana appears to become a competitor for alcohol and smokeless tobacco. These mechanisms explain why marijuana appears to act as a substitute when introduced into the legal market.

3.4 Policy Implications

The legalization of marijuana is a much debated topic. Overall, the results support legalizing recreational marijuana, showing that the consumption of other legal, potentially risky goods decrease as a result of legalization. To further consider the potential benefits of legalizing, tax revenues generated from legalizing can be incredibly large.

Marijuana is a heavily taxed good with a large market. In 2017, the state of Colorado received nearly $225 million in tax revenues. The tax revenues have grown every year since legalization occurred, indicating a growing market for legal, recreational marijuana. Marijuana became subjected to a 15% recreational marijuana sales tax and a 15% excise tax starting in 2017. Before 2017, marijuana was subject to 2.9% sales tax, a 10% recreational marijuana sales tax and a 15% excise tax.[7] These high tax rates are not exclusive to Col-

[7]Marijuana tax information for Colorado was retreived from https://www.colorado.gov/pacific/revenue/colora marijuana-tax-data

orado. As of January 2018, marijuana had a 37% sales tax in Washington, 17% in Oregon, and 15% excise and 10% sales tax in Nevada for example.[8]

A potential issue with marijuana taxes is the cannibalization of tax revenues from tobacco and alcohol. However, looking at tax revenues in Colorado, tobacco has decreased, but only by $7,000,000, and taxed 2 million fewer packs of cigarettes from 2014 to 2017. This is much less than the tax revenues gained from marijuana.[9] Alcohol tax revenue in Colorado increase by around $1,500,000 in the same timeframe.[10] Cannibalization of tax revenues does not appear to be a concern, at least in Colorado. While tobacco experiences a small decrease and alcohol tax revenue increases, in the same time period marijuana tax revenue increased from $35 million to nearly $225 million.

3.5 Conclusion

This paper explores the impact of the legalization of marijuana on other risky behaviors that are legal for individuals to engage in. To test the impact, data from the BRFSS is used for diff-in-diff using entropy balancing. Overall, it is found that the legalization of recreational marijuana can significantly reduce other risky behaviors involving the use of other legal substances, suggesting the substituting relationship between marijuana and other legal goods potentially risky to consume.

Individuals subject to legalization are less likely to consume alcohol and smokeless tobacco. Further analyzing abusive alcohol behaviors, it is found that the maximum number of alcoholic beverages consumed in one sitting decreases, as well as drinking and driving. Overall, results indicate that the legalization of recreational marijuana can weed out risky behaviors involving the consumption of alcohol and smokeless tobacco, supporting marijuana legalization. Using failed recreational marijuana votes held in Ohio and Arizona, results are shown to hold up to a falsification test.

---

[8] General marijuana tax information comes from https://taxfoundation.org/state-marijuana-taxes-2018/

[9] https://www.colorado.gov/pacific/revenue/annual-report

[10] https://www.colorado.gov/pacific/revenue/colorado-liquor-excise-taxes

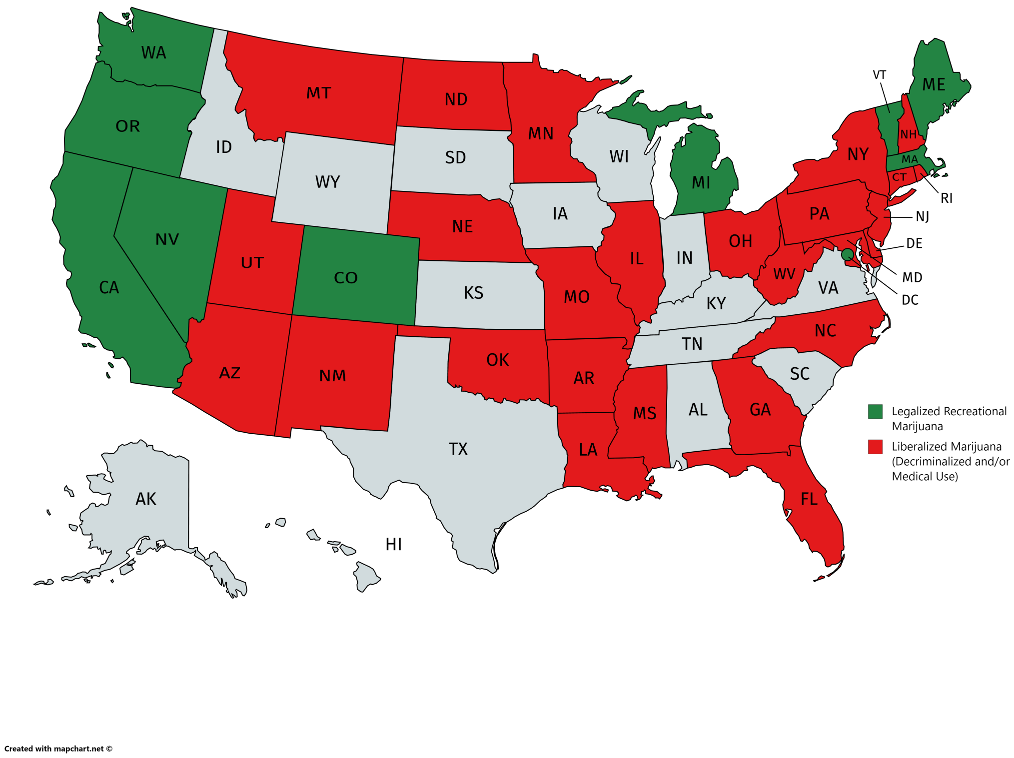

Figure 3.1: Legalization of Recreational Marijuana By State

Table 3.1: Summary Statistics: Means of Characteristics in Legal and Liberalized States

	Liberalized	Legal	Total
Children in Household	0.516	0.497	0.511
Age	55.66	54.99	55.47
Male	0.410	0.423	0.414
Married	0.532	0.517	0.527
Good health	0.806	0.833	0.814
Days with Bad Physical Health	4.384	4.173	4.324
Days with Bad Mental Health	3.440	3.466	3.447
Employed	0.411	0.413	0.411
Self-employed	0.082	0.093	0.085
Out of work $>$ 1 year	0.028	0.032	0.029
Out of work $<$ 1 year	0.024	0.027	0.025
Homemaker	0.063	0.060	0.062
Student	0.023	0.025	0.023
Retired	0.297	0.285	0.293
Unable to Work	0.073	0.065	0.071
Did not graduate HS	0.083	0.070	0.080
Graduated HS	0.297	0.247	0.283
Attended college or technical school	0.272	0.270	0.271
Graduated college or technical school	0.348	0.413	0.366
Less than $10,000	0.052	0.051	0.052
$10,000 to $14,999	0.059	0.055	0.058
$15,000 to $19,999	0.081	0.069	0.077
$20,000 to $24,999	0.099	0.087	0.095
$25,000 to $34,999	0.114	0.107	0.112
$35,000 to $49,999	0.146	0.142	0.145
$50,000 to $74,999	0.157	0.161	0.158
$75,000 or more	0.292	0.329	0.303
Observations	2,697,117		

Means reported

Table 3.2: Summary Statistics: Means of Behaviors in Legal and Liberalized States

	Liberalized	Legal	Total
Any drinking in the last 30 days	0.498	0.586	0.523
Number of days drinking in the last 30 days	3.423	4.518	3.731
Maximum number of drinks in a sitting in the last 30 days	3.194	3.056	3.150
Any binge drinking in the last 30 days	0.122	0.135	0.126
Number of days binge drinking in the last 30 days	1.053	0.959	1.024
Any drinking and driving in the past 30 days	0.028	0.032	0.029
Incidence of drinking and driving in the past 30 days	0.070	0.105	0.081
Smoker	0.156	0.142	0.152
Daily Smoker	0.113	0.100	0.109
Serious attempt to quit smoking in the last 12 months	0.569	0.572	0.569
User of Smokeless Tobacco	0.032	0.021	0.029
Daily user of smokeless tobacco	0.018	0.011	0.016
Observations	2,633,735		

Means reported

Table 3.3: Summary Statistics: Means of Behaviors in Legal States Before and After Implementation of Recreational Marijuana

	Pre-legalization	Post-legalization	Total
Any drinking in the last 30 days	0.584	0.595	0.586
Number of days drinking in the last 30 days	4.489	4.657	4.518
Maximum number of drinks in a sitting in the last 30 days	3.071	2.981	3.056
Any binge drinking in the last 30 days	0.135	0.134	0.135
Number of days binge drinking in the last 30 days	0.964	0.932	0.959
Any drinking and driving in the past 30 days	0.033	0.029	0.032
Incidence of drinking and driving in the past 30 days	0.113	0.059	0.105
Smoker	0.144	0.127	0.142
Daily Smoker	0.103	0.086	0.100
Serious attempt to quit smoking in the last 12 months	0.572	0.567	0.572
User of Smokeless Tobacco	0.020	0.027	0.021
Daily user of smokeless tobacco	0.010	0.015	0.011
Observations	742,206		

Means reported

Table 3.4: Entropy Balancing: Means of Characteristics in Legal and Liberalized States

	Not Balanced		Entropy Balanced	
	Legal States	Liberalized States	Legal States	Liberalized States
Children in household	0.5252	0.5478*	0.5252	0.5252
Age	54.73	55.15*	54.73	54.73
Male	0.4325	0.4229*	0.4325	0.4325
Married	0.528	0.5406*	0.528	0.528
Self-employed	0.09522	0.08466*	0.09522	0.09522
Out of Work > 1 year	0.03113	0.02721*	0.03113	0.03113
Out of Work < 1 year	0.02655	0.02402*	0.02655	0.02655
Homemaker	0.05582	0.05757*	0.05582	0.05582
Student	0.02304	0.02029*	0.02304	0.02304
Retired	0.2716	0.2797*	0.2716	0.2716
Unable to work	0.06304	0.07202*	0.06304	0.06304
Graduated HS	0.2396	0.2906*	0.2396	0.2396
Attended college or technical school	0.2714	0.274*	0.2714	0.2714
Graduated college or technical school	0.4224	0.3577*	0.4224	0.4224
$10,000 to $14,999	0.05555	0.05902*	0.05555	0.05555
$15,000 to $19,999	0.06868	0.08118*	0.06868	0.06868
$12,000 to $24,999	0.08625	0.09915*	0.08625	0.08626
$25,000 to $34,999	0.1073	0.1153*	0.1073	0.1073
$35,000 to $49,999	0.1435	0.1468*	0.1435	0.1435
$50,000 to $74,999	0.1612	0.1577*	0.1612	0.1612
$75,000 or more	0.3265	0.289*	0.3265	0.3265
Observations	418,645	1,022,427		

Means reported

* indicates significant differences between the means in legal states and liberalized states.

Table 3.5: Regression Results: Overall Alcohol Use

	Any drinking in the last 30 days			Days drinking in the last 30 days		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized	-0.00484	-0.00479	-0.00519*	-0.145**	-0.155***	-0.166***
	(-1.80)	(-1.85)	(-1.99)	(-3.18)	(-3.45)	(-3.64)
Children in Household	-0.0312***	-0.0246***	-0.0245***	-0.381***	-0.344***	-0.344***
	(-62.04)	(-49.97)	(-49.48)	(-54.23)	(-48.35)	(-47.80)
Age	-0.00449***	-0.00343***	-0.00310***	0.0115***	0.0115***	0.0159***
	(-149.34)	(-90.57)	(-80.01)	(24.36)	(18.61)	(25.26)
Male	0.0985***	0.0786***	0.0802***	1.942***	1.801***	1.845***
	(107.55)	(86.77)	(87.59)	(126.19)	(115.48)	(116.62)
Married	0.0883***	-0.0219***	-0.0212***	0.728***	-0.242***	-0.219***
	(94.70)	(-21.71)	(-20.82)	(47.85)	(-14.29)	(-12.77)
Good health			0.0823***			0.765***
			(54.68)			(32.87)
Days with Bad Mental Health			0.00180***			0.0419***
			(28.32)			(39.26)
Days with Bad Physical Health			-0.00241***			-0.0178***
			(-37.11)			(-16.83)
Constant	0.720***	0.482***	0.418***	4.766***	2.824***	1.949***
	(187.55)	(106.22)	(86.92)	(74.86)	(38.09)	(24.88)
Year, month fixed effects	Y	Y	Y	Y	Y	Y
State fixed effect	Y	Y	Y	Y	Y	Y
Employment and education controls	N	Y	Y	N	Y	Y
Observations	1394197	1394197	1355916	1255443	1255443	1222930
R^2	0.063	0.139	0.143	0.047	0.072	0.075

T-stats in parentheses

Robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Outcome mean in full sample: Any drinking=0.523, Number of days drinking=3.731

Table 3.6: Regression Results: Maximum Number of Drinks

	Maximum number of drinks in a sitting		
	(1)	(2)	(3)
Legalized	-0.0295	-0.0385*	-0.0400*
	(-1.61)	(-2.11)	(-2.18)
Children in Household	-0.117***	-0.131***	-0.132***
	(-23.71)	(-26.47)	(-26.62)
Age	-0.0553***	-0.0580***	-0.0571***
	(-200.62)	(-161.76)	(-157.70)
Male	1.505***	1.456***	1.477***
	(213.81)	(207.14)	(206.39)
Married	-0.509***	-0.486***	-0.471***
	(-70.71)	(-59.97)	(-57.96)
Good health			-0.0517**
			(-3.20)
Days with Bad Mental Health			0.0216***
			(30.10)
Days with Bad Physical Health			-0.00178**
			(-2.73)
Constant	5.678***	6.732***	6.608***
	(186.74)	(128.64)	(118.99)
Year, month fixed effects	Y	Y	Y
State fixed effect	Y	Y	Y
Employment and education controls	N	Y	Y
Observations	731091	731091	717757
R^2	0.158	0.167	0.170

T-stats in parentheses

Robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Outcome mean in full sample: Maximum amount drank in a sitting=3.150

Table 3.7: Regression Results: Binge Drinking

	Any binge drinking in the last 30 days			Days binge drinking in the last 30 days		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized	0.00185	0.00125	0.000538	-0.00825	-0.0143	-0.0221
	(0.97)	(0.66)	(0.28)	(-0.35)	(-0.60)	(-0.93)
Children in Household	-0.0185***	-0.0183***	-0.0185***	-0.105***	-0.120***	-0.121***
	(-46.92)	(-45.83)	(-45.64)	(-18.22)	(-20.69)	(-20.76)
Age	-0.00542***	-0.00539***	-0.00531***	-0.0265***	-0.0282***	-0.0270***
	(-234.27)	(-180.08)	(-173.62)	(-90.81)	(-73.70)	(-70.17)
Male	0.0854***	0.0797***	0.0815***	0.802***	0.763***	0.783***
	(127.69)	(116.33)	(117.02)	(93.01)	(87.94)	(89.14)
Married	-0.0230***	-0.0428***	-0.0421***	-0.497***	-0.383***	-0.360***
	(-35.68)	(-58.83)	(-56.94)	(-54.38)	(-38.77)	(-36.49)
Good health			0.0132***			-0.205***
			(13.57)			(-9.77)
Days with Bad Mental Health			0.00162***			0.0305***
			(34.14)			(32.18)
Days with Bad Physical Health			-0.000560***			-0.000356
			(-13.52)			(-0.41)
Constant	0.405***	0.403***	0.384***	2.312***	3.376***	3.255***
	(149.94)	(123.58)	(111.51)	(63.94)	(55.41)	(50.32)
Year, month fixed effects	Y	Y	Y	Y	Y	Y
State fixed effect	Y	Y	Y	Y	Y	Y
Employment and education controls	N	Y	Y	N	Y	Y
Observations	1383396	1383396	1346067	743565	743565	729864
R^2	0.082	0.088	0.089	0.035	0.045	0.049

T-stats in parentheses

Robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Outcome mean in full sample: Any binge drinking=0.126, Number of days binge drinking=1.024

Table 3.8: Regression Results: Drinking and Driving

	Any drinking and driving in the last 30 days			Incidence of drinking and driving in the last 30 days		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized	-0.00702***	-0.00713***	-0.00732***	-0.0765***	-0.0772***	-0.0777***
	(-4.15)	(-4.22)	(-4.28)	(-10.44)	(-10.52)	(-10.44)
Children in Household	-0.00315***	-0.00340***	-0.00340***	0.00203	-0.000138	-0.0000830
	(-8.94)	(-9.49)	(-9.43)	(0.75)	(-0.05)	(-0.03)
Age	-0.000810***	-0.000771***	-0.000723***	-0.00160***	-0.00172***	-0.00153***
	(-39.07)	(-28.19)	(-26.09)	(-13.17)	(-10.41)	(-9.14)
Male	0.0267***	0.0258***	0.0267***	0.0782***	0.0748***	0.0785***
	(46.18)	(43.77)	(44.38)	(26.83)	(25.34)	(25.35)
Married	-0.0151***	-0.0184***	-0.0180***	-0.0528***	-0.0492***	-0.0469***
	(-25.62)	(-27.05)	(-26.28)	(-16.73)	(-13.96)	(-13.37)
Good health			-0.000312			-0.0197**
			(-0.28)			(-2.71)
Days with Bad Mental Health			0.000921***			0.00384***
			(18.20)			(10.25)
Days with Bad Physical Health			-0.000148**			-0.000243
			(-3.23)			(-0.77)
Constant	0.0599***	0.0585***	0.0513***	0.0753***	0.185***	0.167***
	(25.37)	(17.67)	(14.45)	(7.02)	(9.14)	(7.86)
Year, month fixed effects	Y	Y	Y	Y	Y	Y
State fixed effect	Y	Y	Y	Y	Y	Y
Employment and education controls	N	Y	Y	N	Y	Y
Observations	452834	452834	444182	452834	452834	444182
R^2	0.016	0.018	0.019	0.011	0.012	0.013

T-stats in parentheses

Robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Outcome mean in full sample: Any drinking and driving=0.028, Incidence of drinking and driving=0.070

Table 3.9: Regression Results: Tobacco Use

	Smoke cigarettes			Use smokeless tobacco		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized	-0.00101	-0.00160	-0.00120	-0.00436***	-0.00457***	-0.00473***
	(-0.54)	(-0.88)	(-0.65)	(-4.88)	(-5.13)	(-5.25)
Children in Household	-0.00357***	-0.00676***	-0.00636***	0.000929***	0.000481**	0.000495**
	(-9.52)	(-18.22)	(-17.13)	(5.12)	(2.58)	(2.63)
Age	-0.00262***	-0.00288***	-0.00264***	-0.000666***	-0.000738***	-0.000743***
	(-123.74)	(-102.39)	(-92.98)	(-63.57)	(-54.15)	(-53.51)
Male	0.0260***	0.0286***	0.0315***	0.0460***	0.0452***	0.0454***
	(39.20)	(43.18)	(47.23)	(146.12)	(143.40)	(141.80)
Married	-0.0987***	-0.0515***	-0.0497***	-0.00854***	-0.00679***	-0.00676***
	(-145.54)	(-69.61)	(-66.50)	(-29.66)	(-20.33)	(-19.99)
Good health			-0.0203***			-0.00114*
			(-16.83)			(-2.28)
Days with Bad Mental Health			0.00413***			0.000118***
			(73.33)			(5.30)
Days with Bad Physical Health			0.000510***			0.0000352
			(9.80)			(1.65)
Constant	0.343***	0.458***	0.436***	0.0448***	0.0640***	0.0639***
	(126.99)	(129.15)	(116.49)	(36.84)	(41.24)	(39.10)
Year, month fixed effects	Y	Y	Y	Y	Y	Y
State fixed effect	Y	Y	Y	Y	Y	Y
Employment and education controls	N	Y	Y	N	Y	Y
Observations	1419241	1419241	1379606	1423058	1423058	1383113
R^2	0.039	0.091	0.100	0.032	0.036	0.036

T-stats in parentheses

Robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Outcome mean in full sample: Smoke=0.156, Use smokeless tobacco=0.032

Table 3.10: Regression Results: Daily Tobacco Use and Attempts to Quit

	Daily smoker			Made an attempt to quit			Daily use of smokeless tobacco		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Legalized	0.000200	0.00141	0.00131	-0.000351	-0.000867	-0.000463	-0.00271***	-0.00284***	-0.00284***
	(0.03)	(0.19)	(0.17)	(-0.22)	(-0.55)	(-0.29)	(-3.96)	(-4.16)	(-4.09)
Children in Household	0.0130***	0.0139***	0.0142***	-0.00210***	-0.00479***	-0.00450***	0.000862***	0.000614***	0.000613***
	(10.43)	(11.01)	(11.15)	(-6.50)	(-14.83)	(-13.90)	(6.46)	(4.47)	(4.42)
Age	-0.00295***	-0.00321***	-0.00333***	-0.00165***	-0.00180***	-0.00160***	-0.000298***	-0.000308***	-0.000311***
	(-33.98)	(-30.20)	(-30.54)	(-92.68)	(-74.86)	(-65.80)	(-41.14)	(-32.37)	(-32.05)
Male	-0.0403***	-0.0354***	-0.0340***	0.0195***	0.0215***	0.0239***	0.0280***	0.0274***	0.0275***
	(-16.87)	(-14.40)	(-13.56)	(33.84)	(37.04)	(40.81)	(118.86)	(116.83)	(115.45)
Married	-0.0136***	0.0103***	0.00985***	-0.0706***	-0.0353***	-0.0337***	-0.00331***	-0.00291***	-0.00288***
	(-5.48)	(3.78)	(3.56)	(-119.02)	(-54.55)	(-51.67)	(-15.53)	(-11.73)	(-11.44)
Good health			-0.0412***			-0.0154***			-0.000740*
			(-12.15)			(-14.09)			(-1.99)
Days with Bad Mental Health			0.000417**			0.00343***			0.0000508**
			(3.21)			(66.49)			(3.08)
Days with Bad Physical Health			0.00119***			0.000292***			0.0000376*
			(7.96)			(6.19)			(2.34)
Constant	0.701***	0.707***	0.729***	0.232***	0.321***	0.301***	0.0177***	0.0276***	0.0277***
	(69.52)	(61.74)	(60.22)	(100.24)	(102.32)	(91.14)	(20.20)	(24.91)	(23.76)
Observations	224506	224506	217540	1419241	1419241	1379606	1423058	1423058	1383113
R^2	0.015	0.021	0.024	0.026	0.068	0.076	0.021	0.024	0.024
Year, month fixed effects	Y	Y	Y	Y	Y	Y	N	Y	Y
State fixed effect	Y	Y	Y	Y	Y	Y	N	Y	Y
Employment and education controls	N	Y	Y	N	Y	Y	N	Y	Y

T-stats in parentheses

Robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Outcome mean in full sample: Smoke daily=0.113, Use smokeless tobacco daily=0.016, Attempt to quit smoking=0.569

Table 3.11: Falsification Check: Failed Recreational Marijuana Votes

Outcome	“Treated" Coefficient	Observations
Any drinking in the last 30 days	-0.00835	963,310
	(-1.34)	
Number of days drinking in the last 30 days	-0.0702	866,014
	(-0.76)	
Maximum number of drinks in a sitting in the last 30 days	0.0170	486,145
	(0.26)	
Any binge drinking in the last 30 days	-0.00503	956,195
	(-1.22)	
Number of days binge drinking in the last 30 days	-0.0483	495,156
	(-0.70)	
Any drinking and driving in the past 30 days	0.00199	301,427
	(0.44)	
Incidence of drinking and driving in the past 30 days	0.0329	301,427
	(1.17)	
Smoker	0.00338	978,636
	(0.73)	
Daily Smoker	0.00153	980,974
	(0.37)	
Serious attempt to quit smoking in the last 12 months	-0.0243	159,843
	(-1.50)	
User of Smokeless Tobacco	-0.000298	978,636
	(-0.13)	
Daily user of smokeless tobacco	0.00263	980,974
	(1.48)	

T-stats in parentheses

Robust Standard Errors

Each regression includes full controls from Equation (1).

Chapter 4

Appendices

4.1 Appendices to Chapter 1

4.1.1 Synthetic Control Results with One Employment Growth Lag

Kaul et al. (2015) warn against using all past values of the outcome variable as this results in all other predictors having no contributing weight. Kaul et al. (2015) recommends using one lag for the outcome variable, selecting the year prior to treatment. Results shown in Section 4 select three years of employment growth before treatment occurs. This approach follows Islam (2019).

Table 4.1 reports pre-treatment RMSPE for each Olympic hosting county using one lag of employment growth the year prior to hosting and three lags of employment growth as seen in Islam (2019). In each county using three years of employment growth generated a lower RMSPE in each Olympics, indicating a better fit. Figure 4.1 show synthetic control results using one year of employment growth, average population growth, and average income growth to create the synthetic county. For each Olympic Games study, results remain consistent.

Examining Figure 4.2 Los Angeles County appears to experience a reduction in employment growth in 1982. While this impact appears less significant than seen in Figure 1.2, the lower pre-treatment RMSPE makes results presented in Section 4.2 preferable. Results seen in Section 4.2 for Fulton County and Salt Lak County remain in Figure 4.2. Fulton County experiences an increase in employment growth in 1993, 1994, and 1996. Salt Lake County experiences an increase in 1996.

Table 4.1: Pre-treatment RMSPE: One Lag of Employment Growth vs Three Lags of Employment Growth

County	One Lag of Employment Growth	Three Lags of Employment Growth
Los Angeles County	0.0076	0.004
Fulton County	0.0142	0.0139
Salt Lake County	0.0186	0.0115

Figure 4.1: Synthetic Control Results: One Lag of Employment Growth

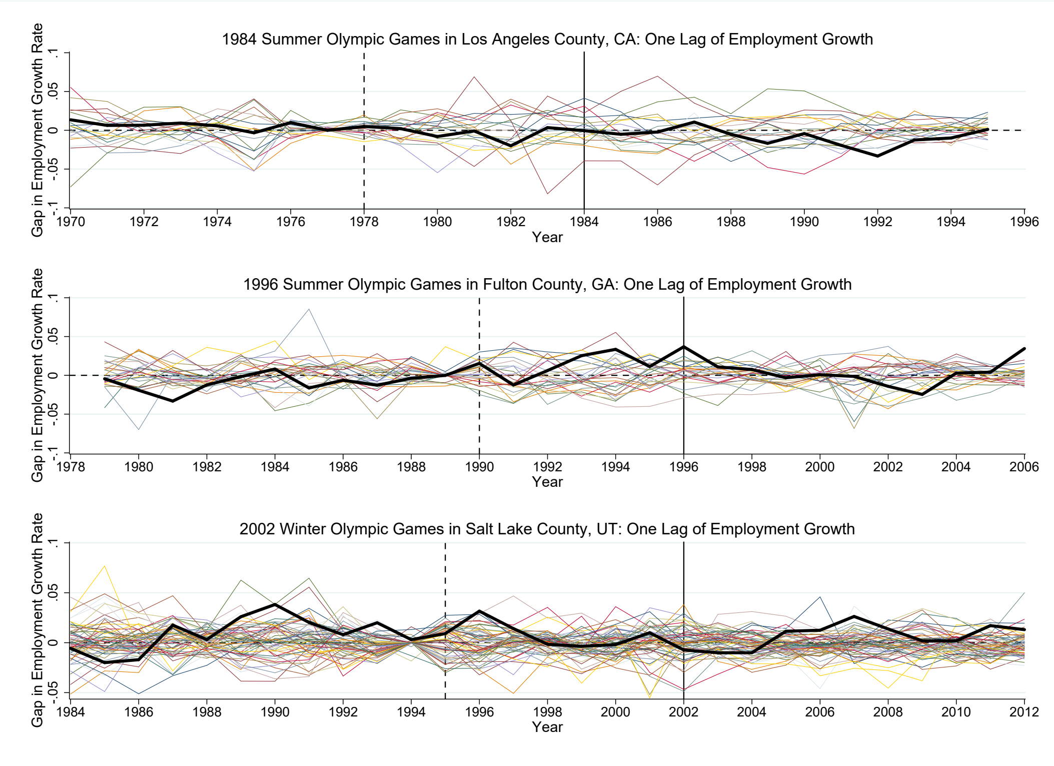

Figure 4.2: Placebo Tests: One Lag of Employment Growth

4.1.2 Placebo Tests Dropping Donor Counties with High MSPE

The following Figures present placebo tests dropping placebo counties that were poor pre-treatment fits, measured by MSPE, as in Abadie et al. (2010). Excluding poor fitting placebo counties highlights years in which an increase in employment growth is experienced. Placebo tests are presented dropping placebos with MSPE two times, five times, and twenty times higher for Los Angeles County. 1982 represents a significant decrease in employment growth from the Olympics in Los Angeles County. When dropping counties with MSPE five times higher and two times higher, Los Angeles County appears to experience a decrease in employment growth in 1980 as well. For Fulton County placebos with MSPE two times and five times higher are dropped, while placebos with MSPE two times higher are dropped for Salt Lake County. Results remain the same for Fulton County and Salt Lake County when dropping poorly fit counties. Fulton County experiences an increase in growth in 1993, 1994, and 1996. An increase in growth in Salt Lake County occurs in 1996, the year following being awarded the Olympic Games.

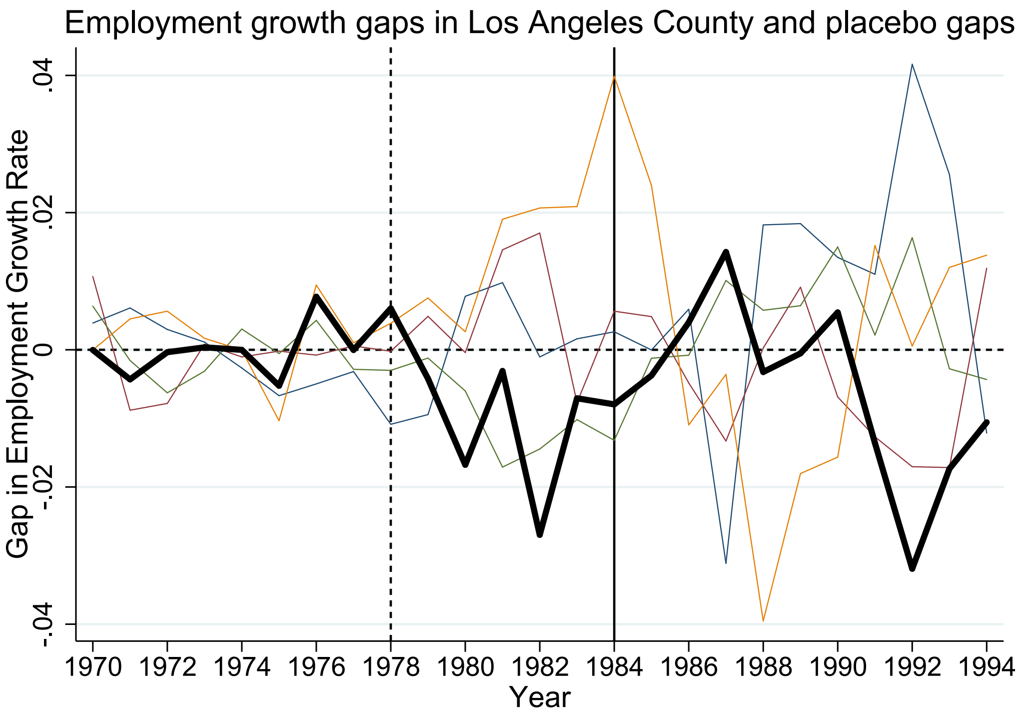

Figure 4.3: Placebo Tests: Los Angeles County Dropping Counties with MSPE Two Times as High

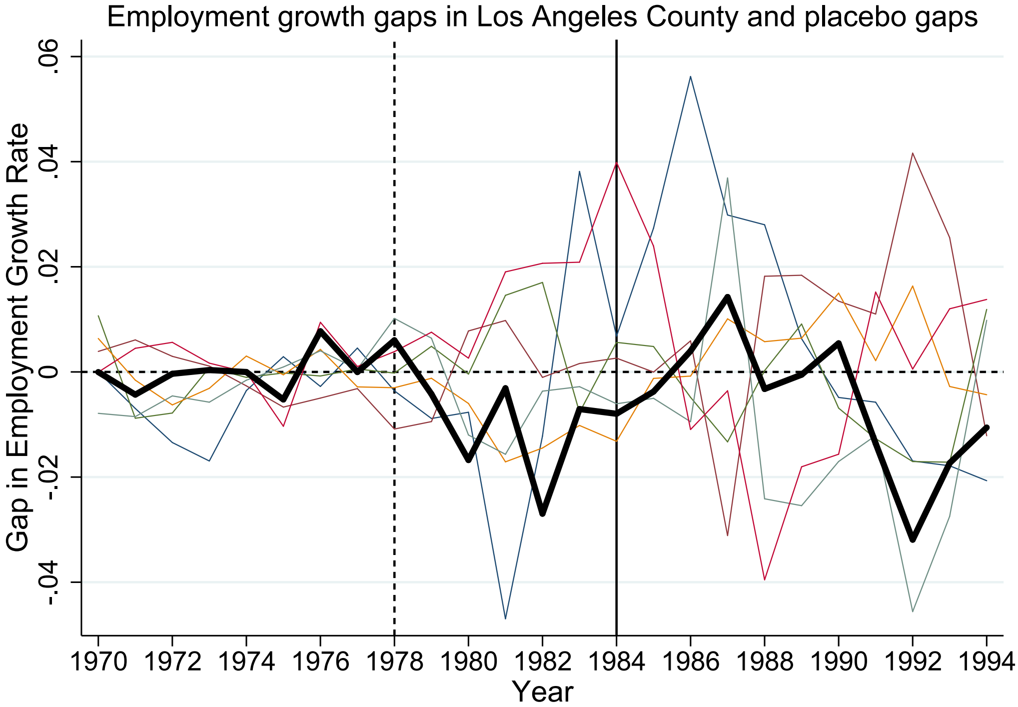

Figure 4.4: Placebo Tests: Los Angeles County Dropping Counties with MSPE Five Times as High

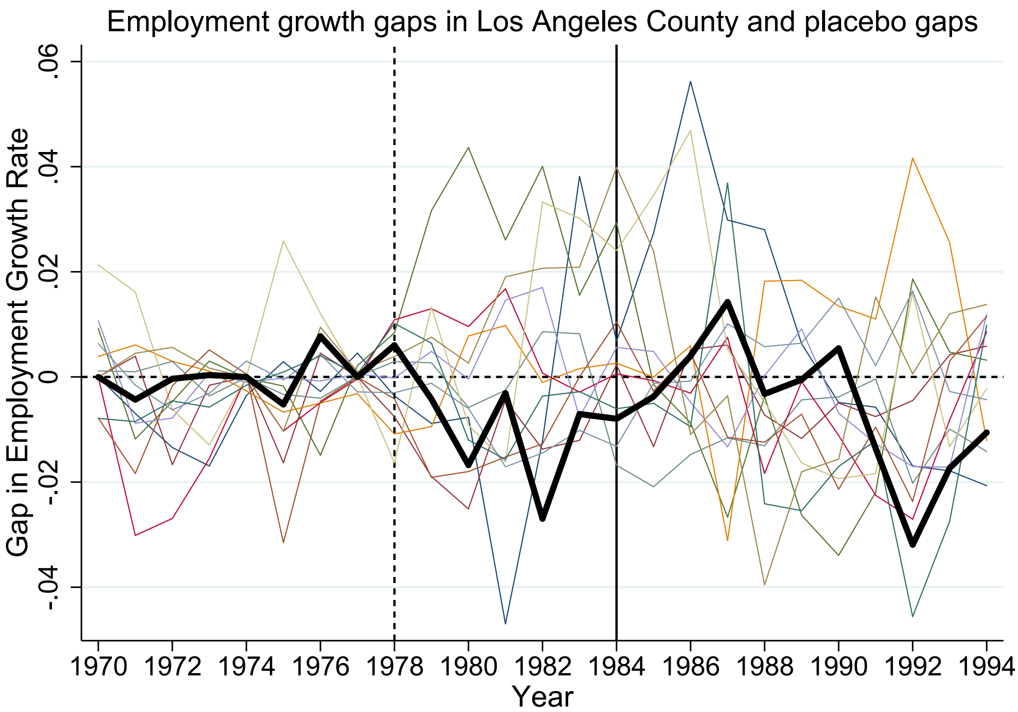

Figure 4.5: Placebo Tests: Los Angeles County Dropping Counties with MSPE Twenty Times as High

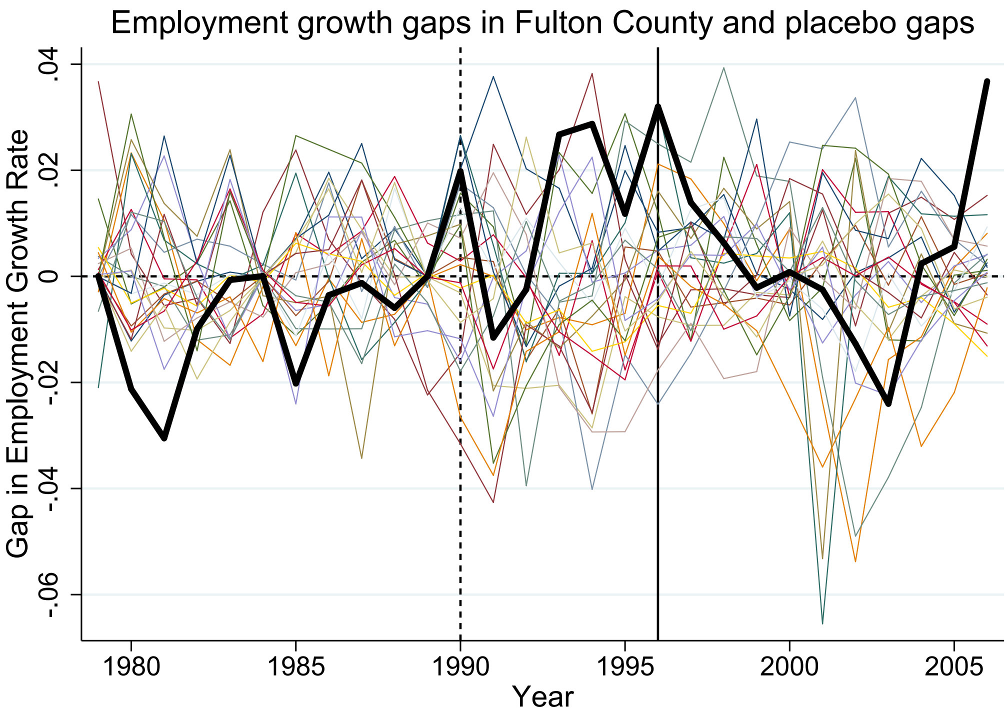

Figure 4.6: Placebo Tests: Fulton County Dropping Counties with MSPE Two Times as High

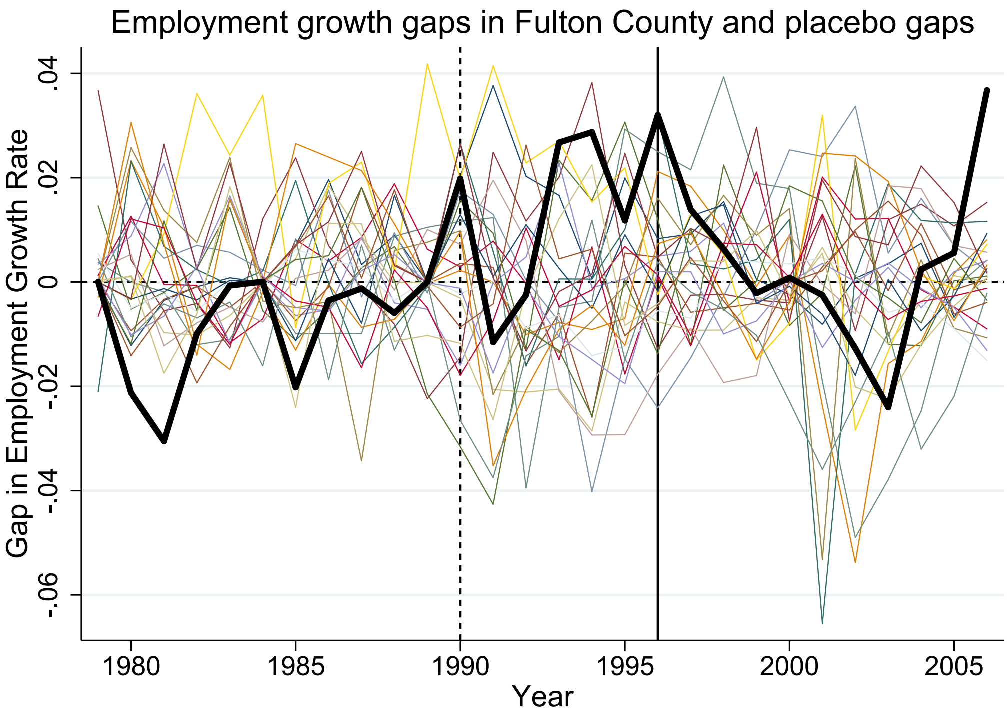

Figure 4.7: Placebo Tests: Fulton County Dropping Counties with MSPE Five Times as High

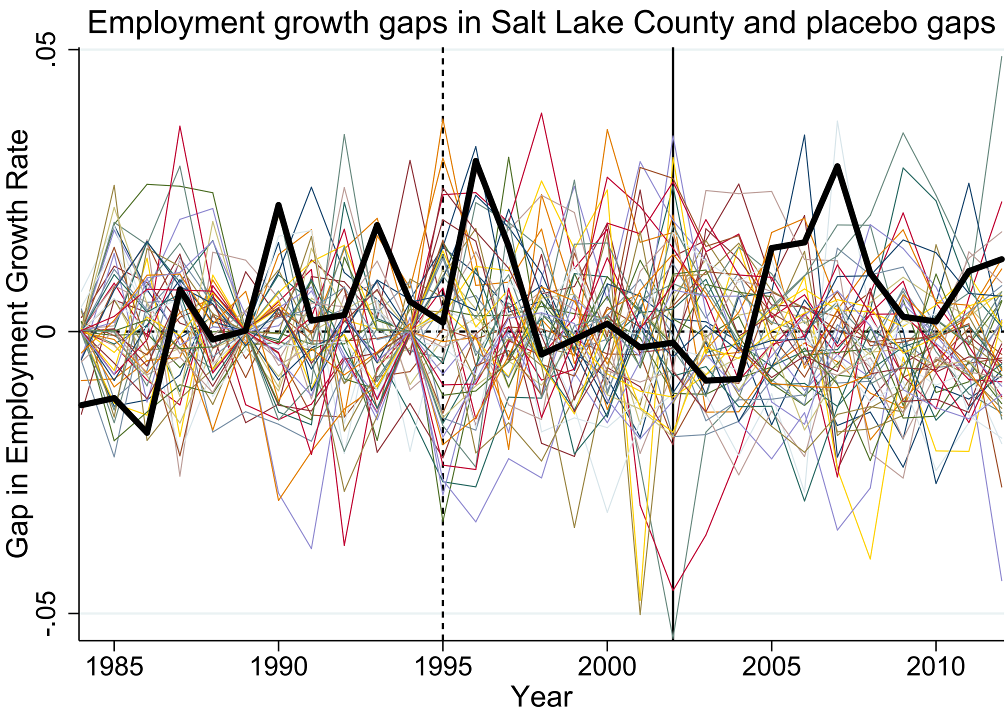

Figure 4.8: Placebo Tests: Salt Lake County Dropping Counties with MSPE Two Times as High

4.1.3 Spillover Effects: State-level Synthetic Control Analysis

While this paper studies of the county which represents the focal point of each Olympic Games studied, various counties in the hosting state held events. For example, the sailing events in the 1996 Olympic Games took place over three hours from Atlanta, GA in Savannah, GA. Considering the size of the event that the Olympics represents and events being held throughout Olympic hosting states, analyzing spillover effects becomes important. The following figures posits two approaches to test for spillover effects. The first approach, seen in Figures 18 and 19, analyzes the host state in its entirety, including the host county. The second approach (Figures 20 and 21) uses the synthetic control approach on host states excluding the host county. Deducting the host county from the state mitigates concern of state-level results being driven by the host county. Tables 4.2 and 4.3 presents synthetic weight for each approach. For each state, a portion of the synthetic counterpart is comprised of similar states regardless of the approach used. Results show no clear, significant spillover effects experienced in either California or Utah. However, the positive impact felt from the lead up to the 1996 Olympic Games appears to have been felt throughout Georgia.

Table 4.2: Synthetic Control Weights: State-Level Analysis Including Host County

Donor State	California Weights	Georgia Weights	Utah Weights
Connecticut	0.448	–	–
Delaware	0.004	–	–
Florida	–	0.332	–
Idaho	0.17	–	–
Michigan	–	0.261	–
Nevada	0.125	–	0.397
New Hampshire	–	0.306	–
New Jersey	–	–	0.259
Rhode Island	–	0.101	–
South Dakota	–	–	0.118
Washington	0.248	–	–
Wyoming	–	–	0.226

Table 4.3: Synthetic Control Weights: State-Level Analysis Excluding Host County

Donor State	California Weights	Georgia Weights	Utah Weights
Arizona	–	0.086	0.189
Connecticut	0.271	–	–
Florida	–	0.256	–
Idaho	0.327	–	–
Michigan	–	0.184	–
Nevada	0.131	–	0.411
New Hampshire	–	0.223	–
South Dakota	–	–	0.4
Virginia	–	0.25	–
Washington	0.158	–	–
Wyoming	0.113	–	–

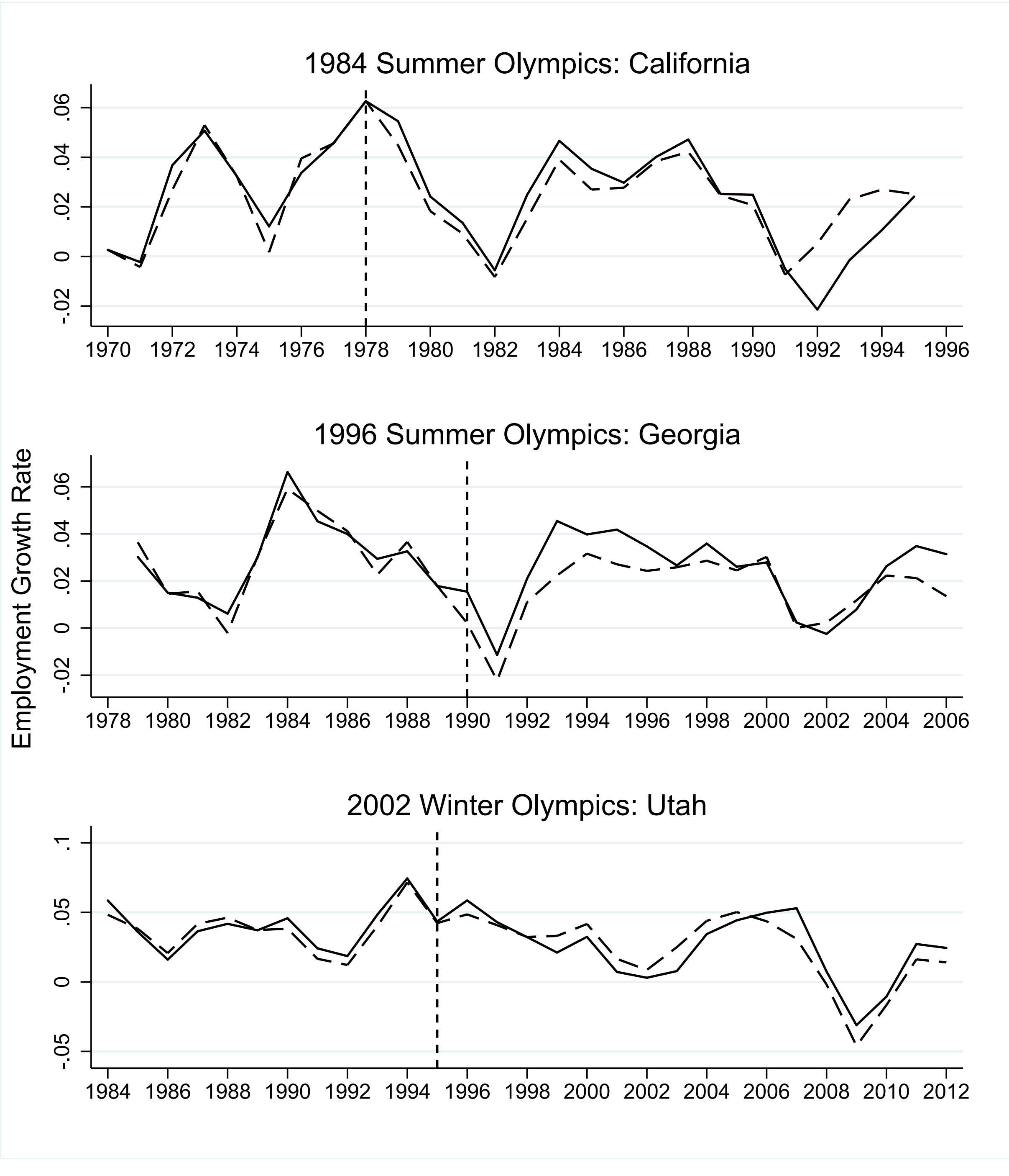

Figure 4.9: State-level Synthetic Control Results

The solid line indicates employment growth rates experienced in host states, while the dashed line represents their synthetic counterpart.

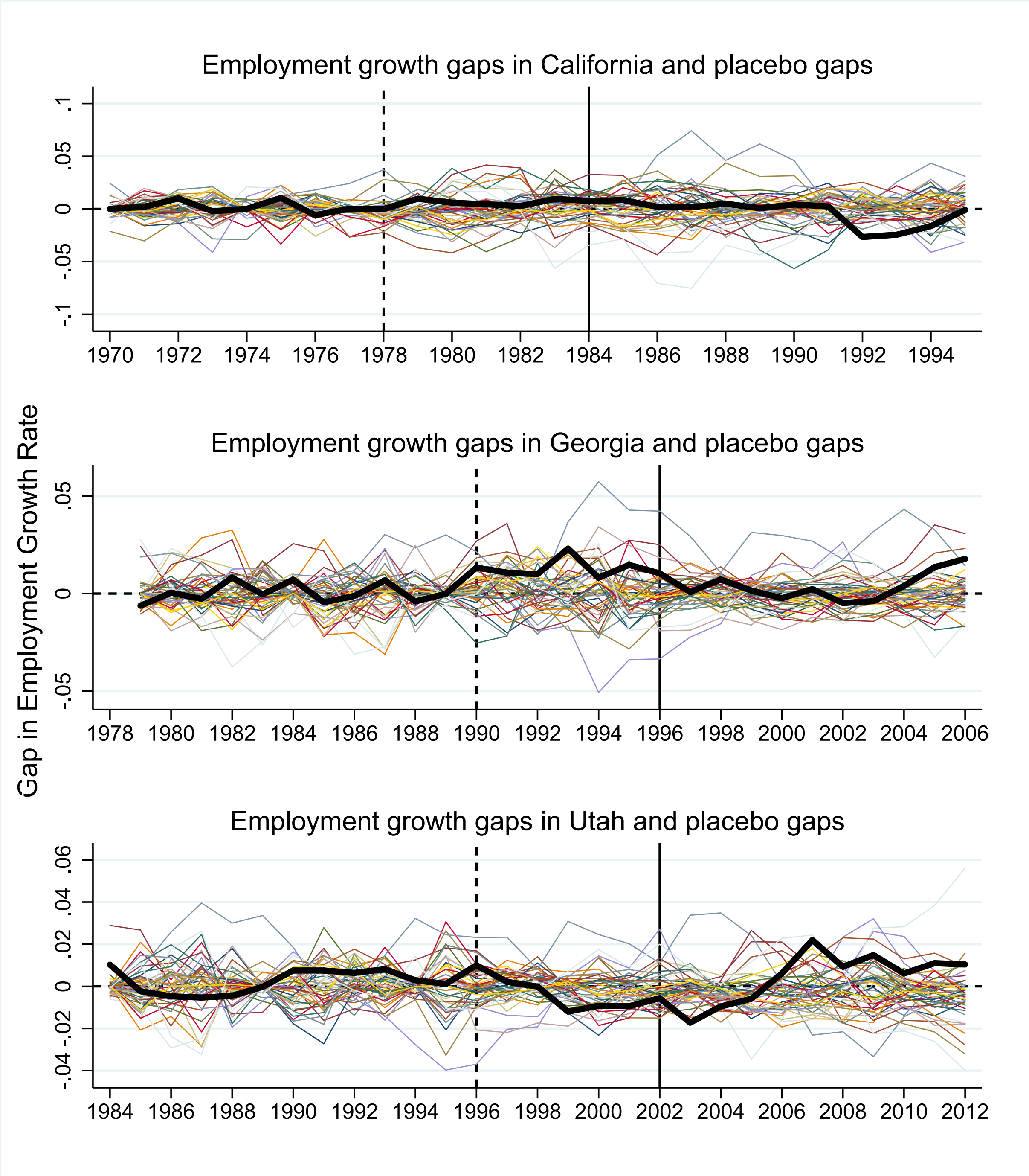

Figure 4.10: State-level Placebo Tests

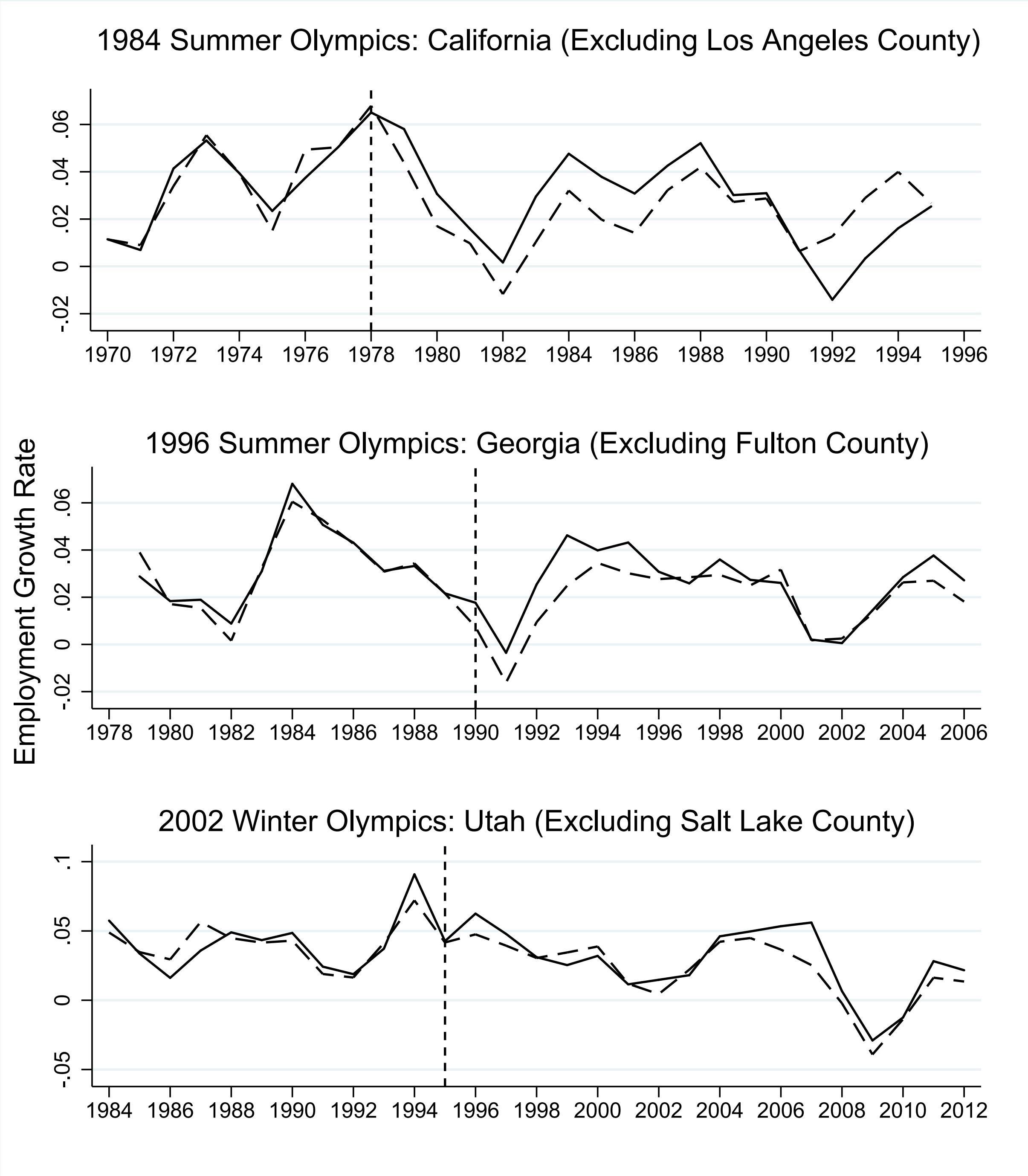

Figure 4.11: State-level Synthetic Control Results Excluding Olympic Host County

The solid line indicates employment growth rates experienced in host states, while the dashed line represents their synthetic counterpart.

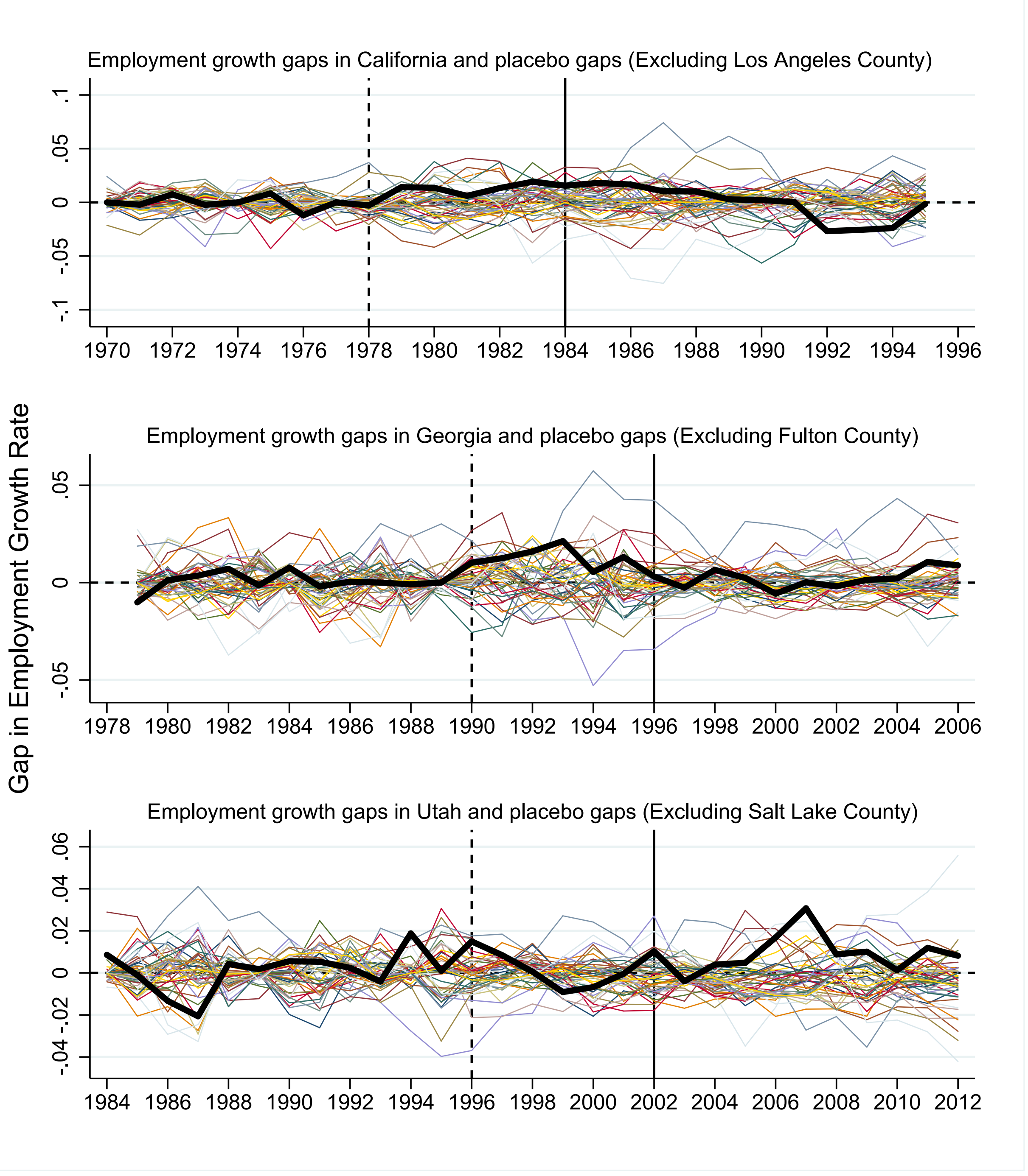

Figure 4.12: State-level Placebo Tests Excluding Olympic Host County

4.2 Appendices to Chapter 2

4.2.1 Results with Max Contracts

Table 4.4: Pooled Twofold Oaxaca-Blinder Decomposition: With Max Contracts

	(1)	(2)	(3)	(4)
Non-black	15.07***	15.07***	15.07***	15.07***
	(205.79)	(204.00)	(205.85)	(204.59)
Black	14.86***	14.86***	14.86***	14.86***
	(363.16)	(361.76)	(363.08)	(361.77)
Difference	0.205**	0.205**	0.205**	0.205**
	(2.44)	(2.42)	(2.44)	(2.43)
Explained	0.0925	0.0806	0.0910	0.0766
	(1.17)	(1.06)	(1.15)	(1.00)
Unexplained	0.112**	0.124**	0.114**	0.128**
	(2.34)	(2.07)	(2.37)	(2.14)
Observations	797	797	797	797

t statistics in parentheses

Position and year fixed effects included

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.5: Weighted Least Squares Regression Results: With Max Contracts

	(1) Salary	(2) Salary	(3) Salary	(4) Salary
Age	1.574**	3.246***	1.579**	3.246***
	(2.24)	(4.18)	(2.26)	(4.21)
Age^2	-0.259**	-0.543***	-0.261**	-0.544***
	(-2.15)	(-4.07)	(-2.17)	(-4.11)
Multi-year Contract	0.400***	0.547***	0.403***	0.552***
	(7.99)	(10.06)	(7.99)	(10.15)
Foreign-Born – No College	0.00988	0.0701	0.0151	0.0791
	(0.15)	(0.88)	(0.22)	(0.98)
Foreign-Born – College	0.120	0.0470	0.117	0.0383
	(1.46)	(0.43)	(1.43)	(0.36)
Restricted Free Agent	0.292***	0.390***	0.293***	0.390***
	(4.91)	(5.75)	(4.92)	(5.75)
Max Contract	0.487***	0.640***	0.487***	0.643***
	(5.90)	(6.46)	(5.95)	(6.59)
Games Played	0.00167		0.00163	
	(1.24)		(1.22)	
% of Games Started	-0.0390	0.564***	-0.0336	0.568***
	(-0.43)	(7.35)	(-0.37)	(7.40)
Field Goal Percentage	0.163		0.160	
	(0.54)		(0.53)	
Previous Team Win %	0.812***	0.323**	0.821***	0.344**
	(5.48)	(1.97)	(5.55)	(2.12)
Signing Team Win %	-0.750***	-0.687***	-0.751***	-0.673***
	(-4.28)	(-3.50)	(-4.25)	(-3.39)
Re-sign	0.148***	0.142***	0.147***	0.141***
	(3.18)	(2.62)	(3.17)	(2.59)
Draft Position	-0.000703	-0.00680***	-0.000728	-0.00681***
	(-0.66)	(-6.23)	(-0.69)	(-6.24)
Height in Inches	0.0178	-0.00186	0.0169	-0.00444
	(1.34)	(-0.13)	(1.28)	(-0.31)
Head Coach is Black	-0.0329	-0.0143	-0.0462	-0.0477
	(-0.51)	(-0.20)	(-0.68)	(-0.63)
GM is Black	-0.0633	-0.0992	-0.0509	-0.0813
	(-0.80)	(-1.05)	(-0.62)	(-0.83)
Black	-0.112**	-0.124**	-0.114**	-0.128**
	(-2.24)	(-1.98)	(-2.26)	(-2.03)
VORP		0.262***		0.258***
		(10.64)		(10.54)
Population (000,000s)			-0.0935	-0.00587
			(-0.39)	(-0.02)
White Population			0.0251	0.0559*
			(0.89)	(1.65)
Position & Year Fixed effects	Y	Y	Y	Y
Observations	797	797	797	797
R^2	0.774	0.708	0.774	0.710

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2.2 Data Sources

The data used in this paper comes from a variety of sources. Salary and contract information was obtained from Spotrac.com. This includes variables used such as average salary value, length of contract, restricted free agency status, re-signing, and the number of contracts signed in our sample period. Realgm.com provides every NBA draft since 1947, offering information on each draft eligible player. Information relevant to this paper from realgm.com includes the player's height, draft position, place of birth, and college experience (or lack thereof). Player's place of birth and college experience are then used to create the indicator variables "Foreign-Born – No College" and "Foreign-Born – College".

Performance statistics for both team and player retrieved from basketball-reference.com includes each per game statistic, games played, games started, and team winning percentages. Basketball-reference.com also provides player age. VORP, the only performance statistic not taken from basketball-reference.com, was obtained from boxscoregeeks.com. The race of players, coaches, and general managers were determined using pictures from various websites. Pictures for players appear on NBA.com (the official NBA website) and basketball-reference.com. Coach and general manager pictures come from a wide variety of websites. In addition to NBA.com and basketball-reference.com, news websites such as USAtoday.com provide pictures.

Final data collected includes population characteristics. These were largely obtained from the American Community Survey through census.gov. This data included population counts, as well as the percentage of the population that is white. The American Community Survey provided population characteristics for all but one NBA team, the Toronto Raptors. Due to the Toronto Raptors playing in Canada, population characteristics for Toronto comes from statcan.gc.ca. The following table summarizes variables used and their sources.

Table 4.6: Sources Used to Retrieve Variables

Variable	Source
Avg Salary	Spotrac.com
Player is Black	NBA.com and Basketball-reference.com
Foreign-Born – No College	Realgm.com
Foreign-Born – College	Realgm.com
Age	Basketball-reference.com
Games Played	Basketball-reference.com
% of Games Started	Basketball-reference.com
Minutes Played Per Game	Basketball-reference.com
Points Per Game	Basketball-reference.com
Rebounds Per Game	Basketball-reference.com
Assists Per Game	Basketball-reference.com
Blocks Per Game	Basketball-reference.com
Steals Per Game	Basketball-reference.com
VORP	Boxscoregeeks.com
Previous Team Win %	Basketball-reference.com
Signing Team Win %	Basketball-reference.com
Re-sign	Realgm.com
Height in Inches	Realgm.com
Draft Position	Realgm.com
Head Coach is Black	Various websites/news articles
GM is Black	Various websites/news articles
Population	Census.gov and Statcan.gc.ca
White Population	Census.gov and Statcan.gc.ca
Restricted Free Agent	Spotrac.com
Multi-year Contract	Spotrac.com

Bibliography

Abadie, A., Diamond, A., and Hainmueller, J. (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., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132.

Ajilore, O. (2014). Do white NBA players suffer from reverse discrimination? *Economics Bulletin*, 34(1):558–566.

Anderson, D. M., Hansen, B., and Rees, D. I. (2013). Medical marijuana laws, traffic fatalities, and alcohol consumption. *The Journal of Law and Economics*, 56(2):333–369.

Associated Press (2017). AP Analysis: Rio de Janeiro Olympics cost $13.1 billion. *USA Today*.

Atkinson, G., Mourato, S., Szymanski, S., and Ozdemiroglu, E. (2008). Are we willing to pay enough to back the bid'?: Valuing the intangible impacts of London's bid to host the 2012 Summer Olympic Games. *Urban Studies*, 45(2):419–444.

Atlanta Committee for the Olympic Games (1997). *The Official Report of the Centennial Olympic Games*. Peachtree.

Baade, R., Baumann, R., Matheson, V., et al. (2010). Slippery slope? Assessing the economic impact of the 2002 Winter Olympic Games in Salt Lake City, Utah. *Région et Développement*, 31:81–91.

Baade, R. A., Matheson, V., et al. (2002). Bidding for the Olympics: Fool's gold. *Transatlantic sport: The comparative economics of North American and European sports*, 127.

Baade, R. A. and Matheson, V. A. (2004). The quest for the cup: Assessing the economic impact of the World Cup. *Regional Studies*, 38(4):343–354.

Baade, R. A. and Matheson, V. A. (2016). Going for the Gold: The economics of the Olympics. *Journal of Economic Perspectives*, 30(2):201–18.

Baade, R. A. and Sanderson, A. R. (2012). An analysis of the political economy for bidding for the Summer Olympic Games: Lessons from the Chicago 2016 bid. *International Handbook on the Economics of Mega Sporting Events*, pages 85–107.

Bauman, A., Murphy, N., and Matsudo, V. (2013). Is a population-level physical activity legacy of the London 2012 Olympics likely? *Journal of Physical Activity and Health*, 10(1):1–3.

Baumann, R., Ciavarra, T., Englehardt, B., and Matheson, V. A. (2012a). Sports franchises, events, and city livability: An examination of spectator sports and crime rates. *The Economics and Labour Relations Review*, 23(2):83–97.

Baumann, R., Engelhardt, B., and Matheson, V. A. (2012b). Employment effects of the 2002 Winter Olympics in Salt Lake City, Utah. *Jahrbücher für Nationalökonomie und Statistik*, 232(3):308–317.

Becker, G. (1971). *The Economics of Discrimination. Second Edition*, University of Chicago Press, Chicago.

Berri, D. J. (2010). Measuring performance in the National Basketball Association. *The Oxford Handbook of Sports Economics.*

Berri, D. J. and Jewell, R. T. (2004). Wage inequality and firm performance: Professional basketball's natural experiment. *Atlantic Economic Journal*, 32(2):130–139.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, pages 436–455.

Bodvarsson, Ö. B. and Humphreys, B. R. (2013). Labor market discrimination and capital: The effects of fan discrimination on stadium and arena construction. *Contemporary Economic Policy*, 31(3):604–617.

Boudarbat, B. and Connolly, M. (2013). The gender wage gap among recent post-secondary graduates in Canada: a distributional approach. *Canadian Journal of Economics/Revue canadienne d'économique*, 46(3):1037–1065.

Brown, S. A. (2015). Stigma towards marijuana users and heroin users. *Journal of Psychoactive Drugs*, 47(3):213–220.

Burdekin, R. C. and Idson, T. L. (1991). Customer preferences, attendance and the racial structure of professional basketball teams. *Applied Economics*, 23(1):179–186.

Burnett, N. J. and Van Scyoc, L. J. (2015). Compensation discrimination for defensive players: Applying quantile regression to the National Football League market for linebackers and offensive linemen. *Journal of Sports Economics*, 16(4):375–389.

Chan, N. W., Burkhardt, J., and Flyr, M. (2020). The effects of recreational marijuana legalization and dispensing on opioid mortality. *Economic Inquiry*, 58(2):589–606.

Choi, A., Dave, D., and Sabia, J. J. (2019). Smoke gets in your eyes: Medical marijuana laws and tobacco cigarette use. *American Journal of Health Economics*, 5(3):303–333.

Chu, Y.-W. L. (2014). The effects of medical marijuana laws on illegal marijuana use. *Journal of Health Economics*, 38:43–61.

Chu, Y.-W. L. (2015). Do medical marijuana laws increase hard-drug use? *The Journal of Law and Economics*, 58(2):481–517.

Crost, B. and Guerrero, S. (2012). The effect of alcohol availability on marijuana use: Evidence from the minimum legal drinking age. *Journal of Health Economics*, 31(1):112–121.

Cunningham, S. (2018). Causal inference: The mixtape (V. 1.7). *Tufte-Latex.GoogleCode.com.*

Dragone, D., Prarolo, G., Vanin, P., and Zanella, G. (2019). Crime and the legalization of recreational marijuana. *Journal of Economic Behavior & Organization*, 159:488–501.

Elder, T. E., Goddeeris, J. H., and Haider, S. J. (2010). Unexplained gaps and Oaxaca–Blinder decompositions. *Labour Economics*, 17(1):284–290.

Eren, O. and Ozbeklik, S. (2016). What do right-to-work laws do? Evidence from a synthetic control method analysis. *Journal of Policy Analysis and Management*, 35(1):173–194.

Eschker, E., Perez, S. J., and Siegler, M. V. (2004). The NBA and the influx of international basketball players. *Applied Economics*, 36(10):1009–1020.

Feddersen, A. and Maennig, W. (2013a). Employment effects of the Olympic Games in Atlanta 1996 Reconsidered. *International Journal of Sport Finance*, 8(2).

Feddersen, A. and Maennig, W. (2013b). Mega-Events and sectoral employment: The case of the 1996 Olympic Games. *Contemporary Economic Policy*, 31(3):580–603.

Flyvbjerg, B., Stewart, A., and Budzier, A. (2016). The Oxford Olympics Study 2016: Cost and cost overrun at the games. *Said Business School WP 2016-20*.

Freisthler, B., Kepple, N. J., Sims, R., and Martin, S. E. (2013). Evaluating medical marijuana dispensary policies: Spatial methods for the study of environmentally-based interventions. *American Journal of Community Psychology*, 51(1-2):278–288.

Giesecke, J. A. and Madden, J. R. (2011). Modelling the economic impacts of the Sydney Olympics in retrospect–Game over for the bonanza story? *Economic Papers: A Journal of Applied Economics and Policy*, 30(2):218–232.

Gius, M. and Johnson, D. (1998). An empirical investigation of wage discrimination in professional basketball. *Applied Economics Letters*, 5(11):703–705.

Grier, K. and Maynard, N. (2016). The economic consequences of Hugo Chavez: A synthetic control analysis. *Journal of Economic Behavior & Organization*, 125:1–21.

Groothuis, P. A. and Hill, J. R. (2004). Exit discrimination in the NBA: A duration analysis of career length. *Economic Inquiry*, 42(2):341–349.

Groothuis, P. A. and Hill, J. R. (2013). Pay discrimination, exit discrimination or both? Another look at an old issue using NBA data. *Journal of Sports Economics*, 14(2):171–185.

Grossman, D. S., Humphreys, B. R., and Ruseski, J. E. (2019). Out of the outhouse: The impact of place-based policies on dwelling characteristics in Appalachia. *Journal of Regional Science*, 59(1):5–28.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46.

Hamilton, B. H. (1997). Racial discrimination and professional basketball salaries in the 1990s. *Applied Economics*, 29(3):287 – 296.

Hansen, B., Miller, K., and Weber, C. (2017). The grass is greener on the other side: How extensive is the interstate trafficking of recreational marijuana? *NBER Working Paper No. 23762*.

Hansen, B., Miller, K., and Weber, C. (2020). Early evidence on recreational marijuana legalization and traffic fatalities. *Economic Inquiry*, 58(2):547–568.

Hausman, J. A. and Leonard, G. K. (1997). Superstars in the National Basketball Association: Economic value and policy. *Journal of Labor Economics*, 15(4):586–624.

Hill, J. R. (2004). Pay discrimination in the NBA revisited. *Quarterly Journal of Business and Economics*, pages 81–92.

Hill, J. R. and Groothuis, P. A. (2017). Is there a wage premium or wage discrimination for foreign-born players in the NBA? *International Journal of Sport Finance*, 12:204–221.

Hoang, H. and Rascher, D. (1999). The NBA, exit discrimination, and career earnings. *Industrial Relations: A Journal of Economy and Society*, 38(1):69–91.

Hoffer, A. J. and Freidel, R. (2014). Does salary discrimination persist for foreign athletes in the NBA? *Applied Economics Letters*, 21(1):1–5.

Holmes, P. (2011). New evidence of salary discrimination in Major League Baseball. *Labour Economics*, 18(3):320–331.

Hotchkiss, J. L., Moore, R. E., and Rios-Avila, F. (2015). Reevaluation of the employment impact of the 1996 Summer Olympic Games. *Southern Economic Journal*, 81(3):619–632.

Hotchkiss, J. L., Moore, R. E., and Zobay, S. M. (2003). Impact of the 1996 Summer Olympic Games on employment and wages in Georgia. *Southern Economic Journal*, pages 691–704.

Humphreys, B. R. and Howard, D. R. (2008). *The Business of Sports: Volume 1, Perspectives on the Sports Industry*. Praeger Perspectives Series.

Humphreys, B. R. and Johnson, C. (2020). The effect of superstars on game attendance: Evidence from the NBA. *Journal of Sports Economics*, 21(2):152–175.

Humphreys, J. M. and Plummer, M. K. (1995). The economic impact on the state of Georgia of hosting the 1996 Summer Olympic Games. *Mimeograph*.

Isidore, C. (2002). Salt Lake City's five-star gamble. *CNNMoney*.

Islam, M. Q. (2019). Local development effect of sports facilities and sports teams: Case studies using synthetic control method. *Journal of Sports Economics*, page 1527002517731874.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4):453–479.

Johnson, C. and Hall, J. (2018). Do national basketball association players need higher salaries to play in high tax states? Evidence from free agents. *Applied Economics Letters*, 25(5):359–361.

Kahn, L. M. (1991). Discrimination in professional sports: A survey of the literature. *Industrial and Labor Relations Review*, 44(3):395–418.

Kahn, L. M. and Shah, M. (2005). Race, compensation and contract length in the NBA: 2001–2002. *Industrial Relations: A Journal of Economy and Society*, 44(3):444–462.

Kahn, L. M. and Sherer, P. D. (1988). Racial differences in professional basketball players' compensation. *Journal of Labor Economics*, 6(1):40–61.

Kang, Y.-S. and Perdue, R. (1994). Long-term impact of a mega-event on international tourism to the host country: A conceptual model and the case of the 1988 Seoul Olympics. *Journal of International Consumer Marketing*, 6(3-4):205–225.

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2015). Synthetic control methods: Never use all pre-intervention outcomes together with covariates. *University of Hohenheim Working Paper*.

Keefer, Q. A. (2016). Race and NFL playing time. *International Journal of Sport Finance*, 11(2):144.

Keefer, Q. A. W. (2013). Compensation discrimination for defensive players: Applying quantile regression to the National Football League market for linebackers. *Journal of Sports Economics*, 14(1):23–44.

Kepple, N. J. and Freisthler, B. (2012). Exploring the ecological association between crime and medical marijuana dispensaries. *Journal of Studies on Alcohol and Drugs*, 73(4):523–530.

Krautmann, A. C. and Oppenheimer, M. (2002). Contract length and the return to performance in Major League Baseball. *Journal of Sports Economics*, 3(1):6–17.

Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., and Sutton, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Economics*, 25(12):1514–1528.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature*, 50(4):959–1006.

Lang, K. and Manove, M. (2011). Education and labor market discrimination. *American Economic Review*, 101(4):1467–96.

Leeds, E. M. and Leeds, M. A. (2017). Monopsony power in the labor market of Nippon Professional Baseball. *Managerial and Decision Economics*, 38(5):689–696.

Livingston, M. D., Barnett, T. E., Delcher, C., and Wagenaar, A. C. (2017). Recreational cannabis legalization and opioid-related deaths in Colorado, 2000–2015. *American Journal of Public Health*, 107(11):1827–1829.

Maennig, W. and Richter, F. (2012). Exports and Olympic Games: Is there a signal effect? *Journal of Sports Economics*, 13(6):635–641.

Matheson, V. (2012). Assessing the infrastructure impact of mega-events in emerging economies. *Economics Department Working Papers*, Paper 8.

Morris, R. G., TenEyck, M., Barnes, J. C., and Kovandzic, T. V. (2014). The effect of medical marijuana laws on crime: Evidence from state panel data, 1990-2006. *PloS One*, 9(3):1–7.

Munasib, A. and Rickman, D. S. (2015). Regional economic impacts of the shale gas and tight oil boom: A synthetic control analysis. *Regional Science and Urban Economics*, 50:1–17.

Murray, M. (2015). Customer preference and race in the NBA. *FanSided*.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–895.

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, 23(3):279 – 295.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, pages 693–709.

Pacula, R. L. (1998). Does increasing the beer tax reduce marijuana consumption? *Journal of Health Economics*, 17(5):557–585.

Parsons, C. A., Sulaeman, J., Yates, M. C., and Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review*, 101(4):1410–35.

Peri, G. and Yasenov, V. (2015). The labor market effects of a refugee wave: Applying the synthetic control method to the Mariel boatlift. Working Paper 21801, National Bureau of Economic Research.

Powell, D., Pacula, R. L., and Jacobson, M. (2018). Do medical marijuana laws reduce addictions and deaths related to pain killers? *Journal of Health Economics*, 58:29–42.

Price, J. and Wolfers, J. (2010). Racial discrimination among NBA referees. *Quarterly Journal of Economics*, 125(4):1859 – 1887.

Pyun, H. (2018). Exploring causal relationship between Major League Baseball games and crime: A synthetic control analysis. *Empirical Economics*, pages 1–19.

Rose, A. K. and Spiegel, M. M. (2011). The Olympic effect. *The Economic Journal*, 121(553):652–677.

Saffer, H. and Chaloupka, F. (1999). The demand for illicit drugs. *Economic Inquiry*, 37(3):401–411.

Scandizzo, P. L. and Pierleoni, M. R. (2018). Assessing the Olympic games: The economic impact and beyond. *Journal of Economic Surveys*, 32(3):649–682.

Schwarz, H. (2015). Taxpayers are cool with hosting the Olympics - until they have to pay for them. *The Washington Post*.

Smith, A. (2009). Theorising the relationship between major sport events and social sustainability. *Journal of Sport & Tourism*, 14(2-3):109–120.

Spears, M. J. (2016). Where are all the white American NBA players? *The Undefeated*.

Spetz, J., Chapman, S. A., Bates, T., Jura, M., and Schmidt, L. A. (2019). Social and political factors associated with state-level legalization of cannabis in the United States. *Contemporary Drug Problems*, pages 165–179.

Subbaraman, M. S. (2016). Substitution and complementarity of alcohol and cannabis: A review of the literature. *Substance Use & Misuse*, 51(11):1399–1414.

Szymanski, S. (2000). A market test for discrimination in the English professional soccer leagues. *Journal of Political Economy*, 108(3):590–603.

Tirunillai, S. and Tellis, G. J. (2017). Does offline TV advertising affect online chatter? Quasi-experimental analysis using synthetic control. *Marketing Science*, 36(6):862–878.

US General Accounting Office (2000). *Olympic Games Federal Government Provides Significant Funding and Support*.

Van Scyoc, L. J. and Burnett, N. J. (2013). How times have changed: Racial discrimination in the market for sports memorabilia (baseball cards). *Applied Economics Letters*, 20(9):875–878.

Williams, J., Liccardo Pacula, R., Chaloupka, F. J., and Wechsler, H. (2004). Alcohol and marijuana use among college students: Economic complements or substitutes? *Health Economics*, 13(9):825–843.

Zimbalist, A. (2016). *Circus maximus: The economic gamble behind hosting the Olympics and the World Cup*. Brookings Institution Press.