Does Hosting The Olympics Help a Country Economically?

Minerva University
NS125 - Research Methods
Prof. Zoogman
April 19, 2024

Does Hosting The Olympics Help a Country Economically?

Abstract

This proposal asks for an examination of the economic impact of hosting the Olympics on a country. The focused treatment unit would be Australia and the outcome variable, "GDP-per-capita." We propose the use of a synthetic control method to compare the performance of Australia against "synthetic Australia," its counterfactual. The study aims to add to the growing body of research on the economic impact of the Olympics.

Introduction & Literature Review

Hosting the Olympics is always a competition within countries (McBride & Manno, 2022). China spent billions of dollars installing a steel hillside and adding fake snow to host the ski event for the Winter Olympics (Bradsher, 2022). So far, literature remains inconclusive on the question of whether hosting the Olympics actually brings economic benefits to the host in the long term. Thus, this study aims to use the robust method of Synthetic controls to test the question, "What is the effect of hosting the Olympics on Australia's GDP-Per-Capita?"

This study has profound implications in terms of policies as well as perhaps rethinking the idea of competing for the Olympics hosting rights. At the same time, for the researcher, this study holds immense importance as this topic lies at the intersection of their passions, data science, economics, and sports. At the same time, the proposed methodology is new and ground-breaking with this study providing a chance to perhaps extend the importance of Synthetic Controls as well as showcase their versatility within the field of Causal Inference.[1]

Currently, two clear groups exist with one claiming positive effects of hosting the Olympics, while the other denies these claims. Olympics supporters cite the development of non-sports-related infrastructure (Baumann & Matheson, 2013; Fowler, 2008), increased revenue from tourism (Taylor, 2012), higher consumer & investor confidence (Ponomarenko & Plekhanov, 2014), and increased employment (Prynn, 2012) as reasons for positive effects.

Meanwhile, these claims are refuted by the use of empirical studies that conclude no positive long-term effects of hosting the Olympics (Billings & Halladay, 2011). Similarly, Baumann et al., 2010 found that an increase in employment is also temporary. The average cost of hosting the Olympics was also overrun by at least 252% for each Olympics post-1976 (inflation-adjusted) (Flyvbjerg & Stewart, 2012). This necessitated bailouts from the Olympic Committee (McBride & Manno, 2021). Matheson & Baade (2004) used the "substitution effect" to criticize the economic effects of the 2000 Olympics in Sydney. They explained that an increase in tourism in Sydney (the host city) was actually accompanied by lower tourism revenue from other regional cities like Adelaide leading to almost no net gain.

---

[1] **#workincontext:** This proposal defines how the study that is being proposed is important to the society at large as well as contributes to the literature. The personal motivation of the researcher is also presented that allows for more context of the work. The motivation is also surrounded by the global importance of the research and how it contributes to policies implemented by many countries as well as the general motivation around hositng the Olympics that allows for the reader to give a certain degree of thought to the topic.

Huang et al., (2022) use a Regression model controlling for development status, forecasted budget costs, and other variables to find the difference between GDP growth rates before and after hosting the Olympics finding negligible positive effects. The analysis also presented a reverse argument for developing countries at the same time due to increased expenditures. Brückner & Pappa (2015) argued that any positive effects are because of the "anticipation effect" that boosts productivity using a panel data regression. Their data spanned from 1950 to 2009 controlling for many variables like government expenditure. Since statistical techniques come with their own assumptions, this study aims to use a more robust method in order to further this particular direction of empirical studies.[2]

Methodology

The proposed methodology for this study is the Synthetic Control method. The choice of method is driven by the idea that we have a small treatment group but a sizable donor pool to draw control group samples. Synthetic controls are preferred over other causal inference methods like Matching as we won't find a perfect match for the treated unit and/or not be able to meet the assumptions of other methods.

This method works by constructing a synthetic control that is a combination of weighted observations from a pool of donors. The weights allocated to each donor unit are chosen to minimize the differences between the treated unit and the synthetic control unit pre-intervention, enabling post-intervention analysis. The biggest advantage of this method is the creation of a "Counterfactual" in a manner that is data efficient and can still be used when the number of predictors is small. A time-series format of data is needed.

Synthetic controls aren't without limitations of their own. The big assumption associated with this study would be that in the absence of the treatment, the treated unit would follow the path of the "Synthetic control." Moreover, it is quite sensitive to the choice of predictors and the pool of donors. As the output post-treatment is also in a time-series format, there is no set treatment effect that the method estimates but rather the focus is on trends and path plots.

[2] **#litreview:** The proposal clearly articulates the various arguments and findings that exist in the current literature. These are used to build a clear argument for the proposed study. The work clearly demonstrates a skill in summarizing literature, identifying relevant details from them, laying a foundation for the proposed use of new methodologies to explore the impact of Olympics on economic indicators. The different sides of the debate are presented with clear evidence to back up all the claims that allow the reader to form their own opinion while also giving a certain weight to the other side of the argument.

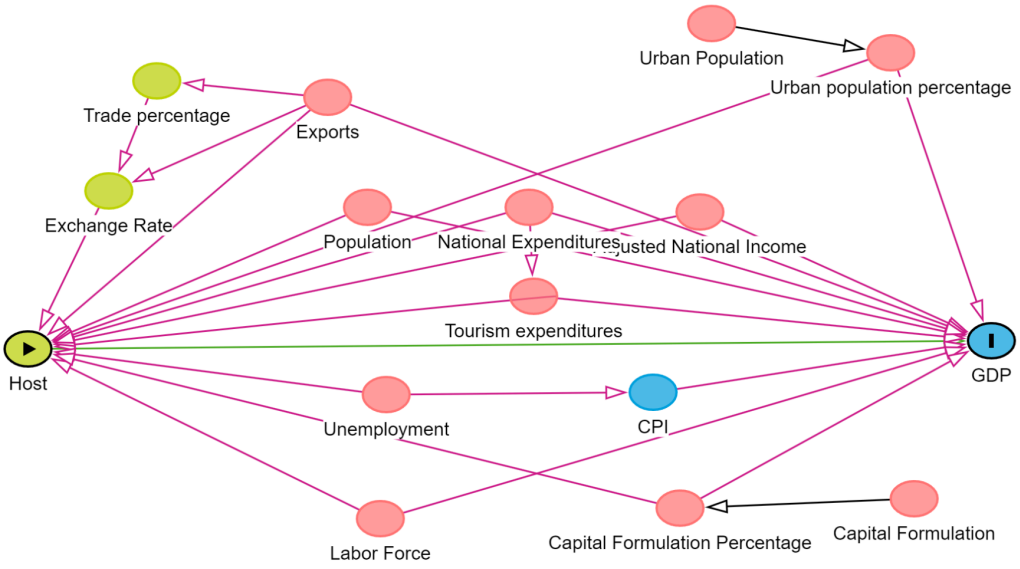

Figure 1: Directed Acyclic Graph showing how the many variables are linked together. We have the GDP (outcome) in blue, the Host (treatment) in green, and various other variables that are seen in the literature on the topic.

DAGs are a good way to show the relationships that are presented between many variables. Economies are widely linked systems with many components and trying to analyze them all without careful consideration can lead to potential confounding biases as shown in the DAG. This diagram also showcases another benefit of using Synthetic controls as they are independent of functional forms and the specific relationship between variables is not important in this method. The main concern is that of parallel trends (treatment and synthetic control follow the same path) in the absence of an intervention. The other assumptions include the similarity of observed & unobserved variables between the two units, no spillover effects on other units in the donor pool, and no anticipation effect (expected effects of treatment shown prior to the treatment).

Data

Most of the indicators for the study can be found in the World Bank DataBank and OECD datasets that are publicly available. However, prior years don't have as much data available which will have to be procured from either governments or other data agencies.

The treatment occurred in 2000 so in order to build a suitable synthetic control, we need as much data as we can procure. Using the satisficing heuristic (The Decision Lab, 2021), one could use the 3 to 1 ratio as required by the method (Abadie, 2021) but ideally, our dataset would span at least pre-treatment data from 1970 onwards and extend to 2010 for a suitable analysis. This allows for a suitable observation of trends and analysis of whether the synthetic control actually resembles the treatment unit.

Several covariates are considered to generate the synthetic control. The proposed variables, their descriptions, classifications, and proposed measurements are shown in Table 1. These are based on the DAG shown in Figure 1.

Variable	**Name**	**Measurement (Principle/Practice)**	**Label**
Country	country	-	Treatment group indicator
Year	year	years (discrete/discrete)	Time variable (treatment indicator)
GDP per capita	gdp_per_capita	current US$ (continuous/discrete)	Outcome variable
Country population	million_population	million persons (discrete/discrete)	Covariate
Consumer Price Index (Inflation indicator)	cpi	annual growth rate (%) (continuous/discrete)	Covariate
Currency exchange rate	exchange_rate	national currency per US dollar (Continuous/continuous)	Covariate
Total area of land	land_area	sq. km (discrete/discrete)	Covariate
International tourism expenditures	tourism_expenditure	% of total imports (continuous/discrete)	Covariate
Exports of goods and services	exports_percentage	% of GDP (continuous/discrete)	Covariate
Total labor force	labor_force	Persons (discrete/discrete)	Covariate
Gross national expenditures	national_expenditure	% of GDP (continuous/discrete)	Covariate
Gross capital formation	capital_formulation	current US$ (continuous/discrete)	Covariate
Gross capital formation	Cfp	% of GDP (continuous/discrete)	Covariate

percentage			
Adjusted net national income	adj_national_income	current US$ (continuous/discrete)	Covariate
Total trade	trade_percentage	% of GDP (continuous/discrete)	Covariate
Total country unemployment	unemployment	% of total labor force (continuous/discrete)	Covariate
Urban population	urban_pop_percentage	% of total population (continuous/discrete)	Covariate
Urban population	urban_pop	Persons (discrete/discrete)	Covariate

Table 1: Table of all the variables that are being used in the study. This list can be expanded if more data can be accessed as the method is insensitive to the function.

The donor pool also requires careful consideration of units to include. They have to be fairly similar to the treatment unit but some analyses do benefit from including units that are not as similar (Abadie, 2021). The proposed set of donors for this study is: NewZealand, United Kingdom, Denmark, Brazil, Sweden, Mexico, Japan, South Africa, Singapore, Malaysia, and Argentina. Countries like Canada are similar in terms of covariates but hosted the Olympics in 1988 so can't be considered. Older data might be limited but we propose imputation of missing values as advancements in statistical methods allow for considerable accuracy with such procedures.

Data Analysis

In order to explore the efficacy of this study, we have collected some data from the World Bank website that was used to see whether the predicted trends actually show in the data. This data only spans from 1986 to 2005, a twenty year period around the Olympics.

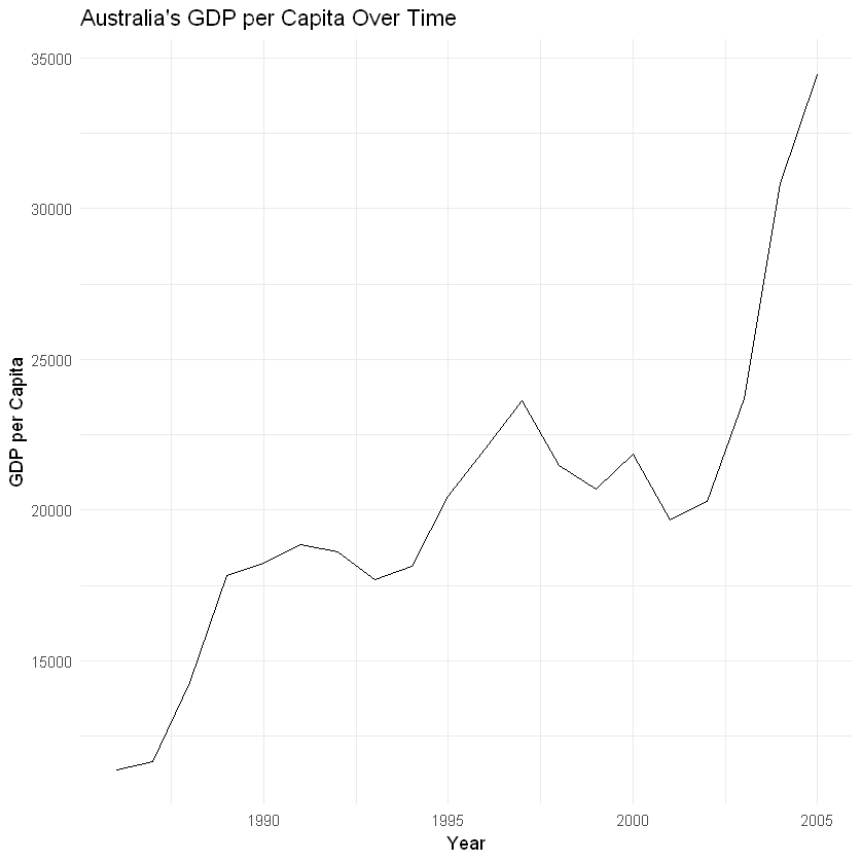

Figure 2: Australia's GDP per capita over time. While the country has followed a generally positive trend, we do see a bigger positive rate of rise in the years following 2000 when Australia hosted the Olympics.

Figure 2 shows that Australia does indeed see its GDP per capita rise in the years following the Olympics and the rate of growth also increases post the Olympics. However, there is also a slight decrease following the year 2000 but small periods of negative trends can often be associated with business cycles (Cerra et al., 2020).

In this case, it might be better to analyze whether these growth trends were global or local to Australia. Using data from our donor pool, we visualize this in Figure 3.

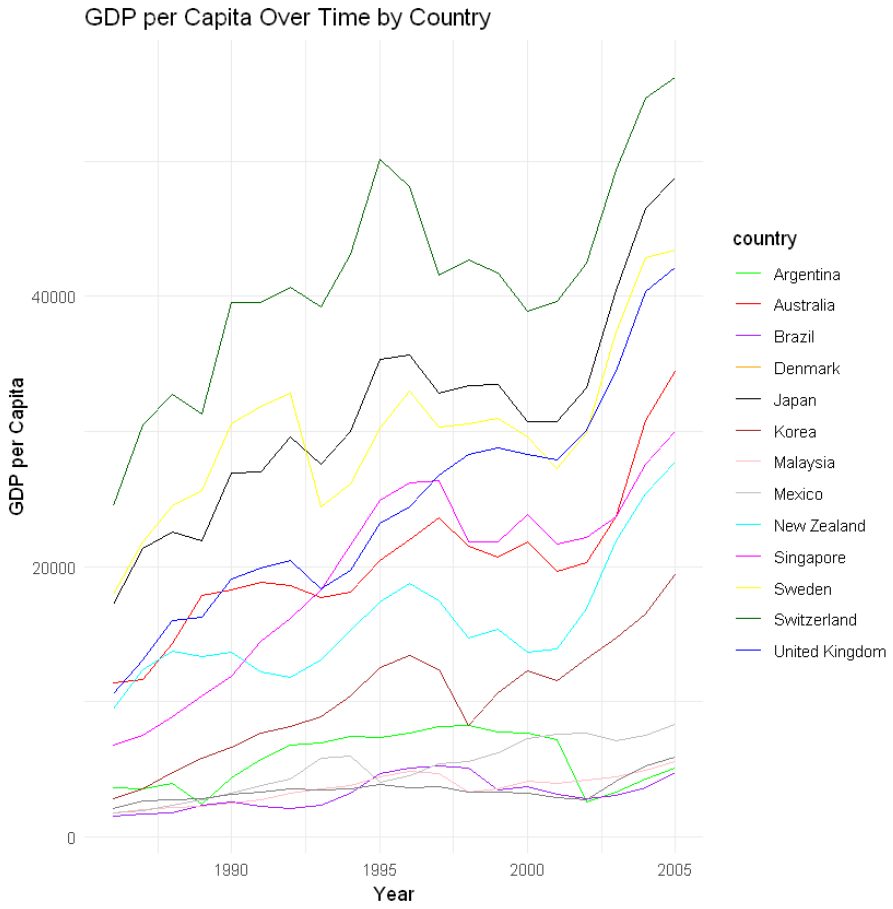

Figure 3: GDP per capita for all the countries in our dataset. The increased rate of GDP growth seems to be a global trend with countries like Switzerland, Sweden, Japan, and Brazil showing nearly identical trends,

Figure 3 shows us that the increased rate of GDP growth doesn't seem to be limited to Australia. Many countries including Japan, Sweden, and Switzerland demonstrated similar trends. Thus, we do observe that this increase can't be associated with only the Olympics and there might be other global events like the dot-com bubble that led to this rise.

In order to explore this data further, it might be beneficial to look at variables that are commonly associated with increases in GDP per capita. National income is one of the variables often associated with this as it makes up a big chunk of the country's GDP (Maverick, 2023). Similarly, trade percentages are highly correlated with GDP trends (Ross, 2019). Thus, a logical next step would be to observe these trends around the same time period to see whether we can isolate the effect or find a common occurring theme that can be associated with this period of growth.

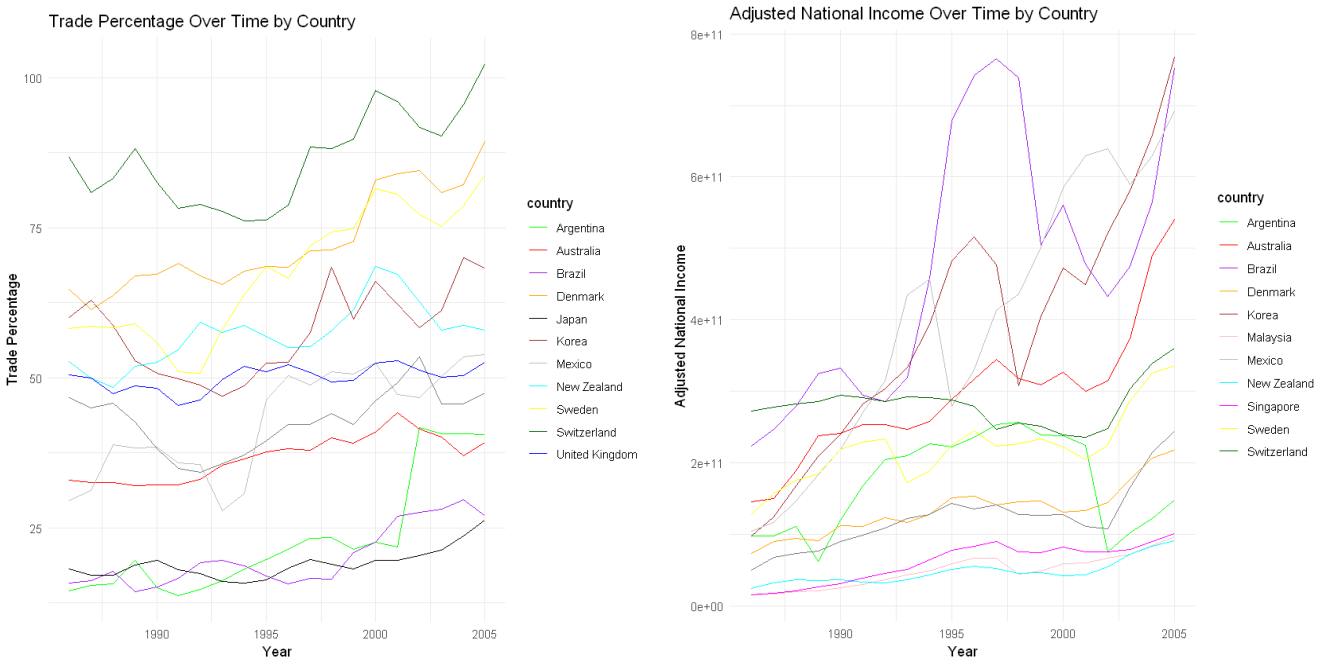

Figure 4: Left: Trade percentages for the countries over time. These are percentages relative to GDP. Right: Adjusted National Income for the countries over time. These are the two variables most commonly associated with growth in GDP. Note that Malaysia & Singapore are not included in the trade percentages graph and the same is the case for Japan and the UK for Adjusted National Income. This is to increase interpretability of the graphs.

Figure 4 gives us two more sets of observations in our investigation. Australia did not see a considerable change in trade percentages (left) around the Olympics in 2000 thus, we can't associate the change in GDP as being because of that. However, Australia did see a sharp increase in National income (right) during this time which could be one of the causes behind the increase in GDP. Similar trends are also observed in Brazil, Mexico, South Korea, and Malaysia.

Pilot Study[3]

Despite the analysis above, one cannot isolate the cause behind the rise in GDP. Thus, there is still a possibility that the effect in Australia was because of the Olympics hosting. This is where Synthetic controls are very useful, as they allow us to explore the effect of Olympics without worrying about functional forms[4]. Thus, we have carried out a pilot study that showcases the same steps that we aim to follow in our research but on a smaller dataset taken from the World Bank.

The main Question that the pilot study tackles is the same as the main proposed study, "What is the economic impact of hosting the Olympics on Australia?" There are three main Alternative hypotheses associated with this study. The primary hypothesis posits that hosting the Olympics has a positive economic impact, increasing GDP per capita due to heightened investment and international attention. The alternative hypotheses suggest that the effects are negligible, or even negative, considering the significant costs associated with hosting.

The Logic comes from the Literature Review section at the start of the paper in the form of existing literature as well as from the preliminary data analysis that we have conducted from Figures 2 to 4. These build on the proposed mechanisms through research and that is what the pilot study aims to add to.

The Method that we are proposing is a quantitative method in the form of Synthetic Controls. This method will construct a counterfactual scenario representing what Australia's economic performance might have been without the Olympics. This synthetic control was constructed using a donor pool of countries that did not host the Olympics during the study period but were similar in other economic and demographic characteristics. Key predictors included population size, CPI, exchange rate, land area, and several other economic indicators, weighted to match the pre-treatment characteristics of Australia as closely as possible. The weights associated with each of the variables and the specific donor pool countries are shown below.

---

[3]**#observationalstudy:** The proposal is the design of an observational study done through a quantitative lens. The Literature review allows the reader to contextualize the observed phenomena and then follow the procedure designed to investigate the factors that could be causing it. The data plots also showcase this clear design. A statistical procedure is identified and explained along with its applicability. Finally, to showcase the practicality and proof of concept, a pilot study based on the same principles is conducted and presented in the proposal.

[4]**#fruitfuldirections:** This study innovatively applies the synthetic control method to analyze the economic impact of hosting the Olympics, showcasing its suitability for the complex data involved. By synthesizing pre-treatment and post-treatment observations, it robustly defends this approach as ideal for isolating the effects of such large-scale events. This methodology not only addresses critical gaps in existing research but also sets a precedent that encourages further explorations into the economic implications of mega-events, thereby promising fruitful directions for future studies. Moreover, by drawing on various studies, the methodology is defended as to why it is ideal for this dataset as well as question that we are tackling.

Country	Weight
Argentina	0.002
Brazil	0.002
Denmark	0.003
Japan	0.002
South Korea	0.006
Malaysia	0.003
Mexico	0.003
New Zealand	0.542
Singapore	0.002
South Africa	0.004
Sweden	0.008
Switzerland	0.012
United Kingdom	0.411

Variable	Weight
Population Size	0.074
CPI	0.068
Exchange rate	0.014
Land Area	0
Tourism Expenditure	0.363
Exports Percentage	0
Labor Force	0.004
National Expenditure	0.333
Capital Formation	0.084
Adjusted National Income	0.001
Trade Percentage	0.001
Unemployment	0
Urban Population	0.058

Table 2: Weights associated with specific variables to predict Synethetic Australia are shown (right). The specific weights associated with each of the donor pool countries are shown (left). The most important variables are Tourism Expenditure & National Expenditure (% of GDP). Meanwhile, the most important countries in the donor pool are the United Kingdom & New Zealand.

The Results actually show a divergence between Australia and Synthetic Australia. Initially, Australia performs worse than its counterfactual but by 2005, the performance is the same. This indicates that the economic costs associated with hosting had been counteracted by this time. Figure 5 below showcases this prediction in the form of a path plot.

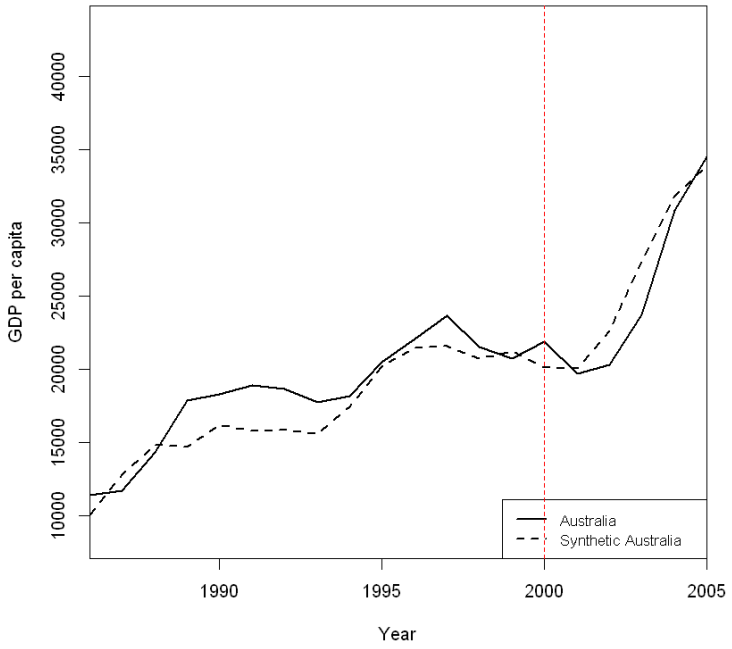

Figure 5: Path plot showing the performance of Ausralia vs its Synethetic counterfactual. The year of treatment is shown by the red line. We see the inital divergence of Australia but by 2005, both the lines seem to have met again.[5]

Inferences are hard to draw from this analysis given its short time period. This pilot study does showcase that hosting the Olympics might have had an initial negative impact on Australia's economy but not anything long-term. However, we can see that the path plots don't necessarily match at all points so we need to expand the dataset to include variables over a longer period and if possible, more donor countries in the control group to make the analysis better.[6]

[5]**#dataviz:** All my data visualization follow best practices with white backgrounds, clearly labeled axes, titles, and clear figure captions. Moreover, each country's color remains constant throughout the figures allowing the reader to interpret them easily. The choice of plots was also done with a lot of consideration by using path plots instead of gap plots to show fit of the algorithm as it was a pilot study rather than a full paper. Similarly, in some figures outliers were taken out of the plots in order to increase their interpretability.

[6]**#qalmri:** This LO was actually applied in two different ways. First, I used the QALMRI method to organize my paper starting from the Question, moving to the Alternatives presented in Literature as well as the Logic. Then, I moved to the Methodology section and finally Results and Inferences. Moreover, I also used this LO to present the pilot study allowing for a more concise and clear presentation that the reader would be able to easily follow. This allowed for the use of the LO in an in-depth manner by making use of the framework to make the readability of the paper better and easier.

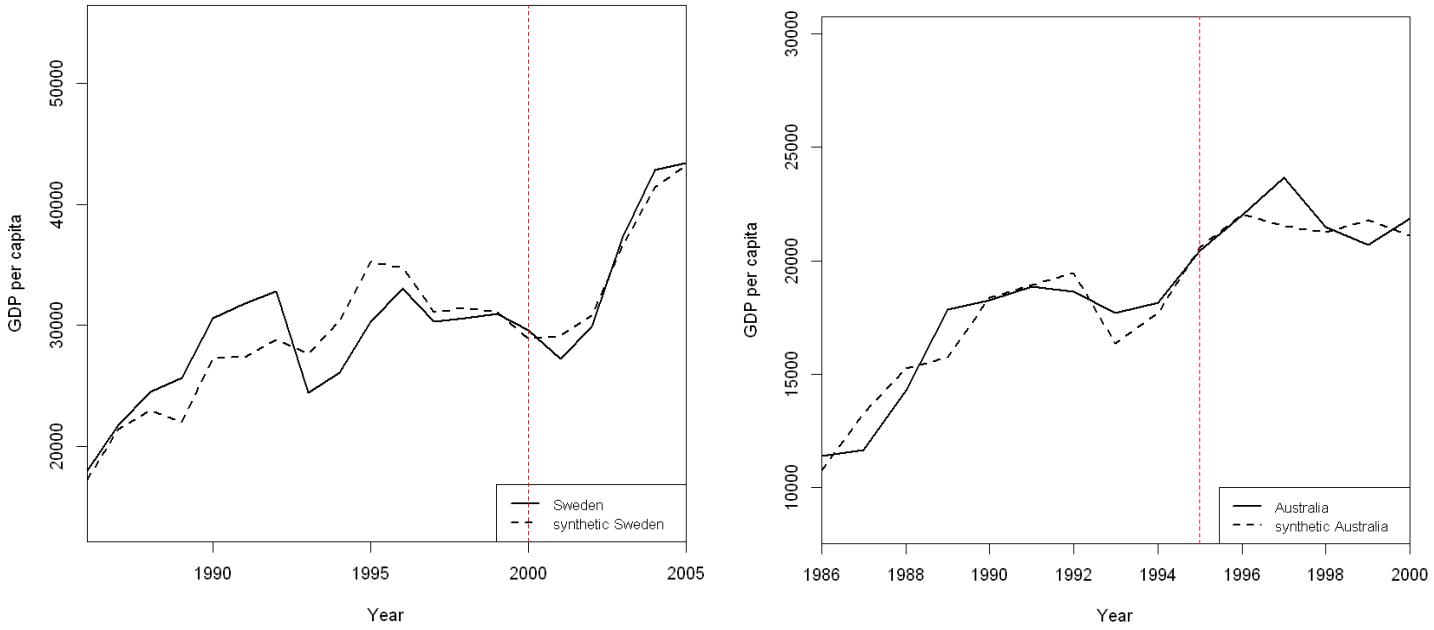

Figure 6: Path plots for the in-place placebo test (left) and the in-time placebo test (right). The in-place placebo test uses Sweden as the treatment and acts as if the treatment occured in Sweden. The in-time placebo test assumes the treatment occured in 1995 instead of 2000 in Australia. The lack of divergences in both plots is indicative that the Synthetic Control method actually works and the trend is not being observed due to errors in the algorithm.

Moreover, we also did two placebo tests in order to test the robustness of this pilot study. This helps to make sure and indicate that the algorithm does indeed work and the gaps are because of missing data that can be expanded about. These are shown in Figure 6. The in-place placebo test shows that the effect of Olympics won't show up in a country that didn't host the Olympics. The in-time placebo test shows that if we try to analyze the effect before the treatment (1995 in our example), the treatment effect doesn't show up. These robustness checks demonstrate that the methodology works and expanding the dataset will be beneficial to the study's purpose of adding to existing literature.

AI Statement: No AI was used in this assignment.

Words – 2044 words

References

Abadie, A. (2021). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. Journal of Economic Literature, 59(2), 391–425. https://doi.org/10.1257/jel.20191450

Baumann, R., Engelhardt, B., & Matheson, V. (2010). The Labor Market Effects of the Salt Lake City Winter Olympics. https://web.holycross.edu/RePEc/hcx/HC1002-Matheson-Baumann-Engelhardt_SLCOlympics.pdf

Baumann, R., & Matheson, V. (2013). Infrastructure Investments and Mega-Sports Events: Comparing the Experience of Developing and Industrialized Countries. https://web.holycross.edu/RePEc/hcx/HC1305-Baumann-Matheson_MegaEventsDeveloping.pdf

BILLINGS, S. B., & HOLLADAY, J. S. (2011). SHOULD CITIES GO FOR THE GOLD? THE LONG-TERM IMPACTS OF HOSTING THE OLYMPICS. Economic Inquiry, 50(3), 754–772. https://doi.org/10.1111/j.1465-7295.2011.00373.x

Bradsher , K. (2022, February 14). For China, Hosting the Olympics Is Worth Every Billion (Published 2022). The New York Times. https://www.nytimes.com/2022/02/14/business/economy/olympics-china-economics.html

Brückner, M., & Pappa, E. (2015). News Shocks in the Data: Olympic Games and Their Macroeconomic Effects. Journal of Money, Credit and Banking, 47(7), 1339–1367. https://doi.org/10.1111/jmcb.12247

Cerra, V., Fatas, A., & Saxena, S. (2020). Hysteresis and Business Cycles. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3630157

Flyvbjerg, B., & Stewart, A. (2012). Olympic Proportions: Cost and Cost Overrun at the Olympics 1960-2012. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2238053

Fowler, G. A. (2008, July 16). China Counts the Cost of Hosting the Olympics. WSJ; The Wall Street Journal. https://www.wsj.com/articles/SB121614671139755287

Maverick, J. B. (2023, July 26). Measuring Economic Conditions: GNI or GDP? Investopedia. https://www.investopedia.com/ask/answers/062315/gross-national-income-gni-or-gross-domestic-product-gdp-better-measure-economic-condition-country.asp#:~:text=of%20the%20country.-

OECD (2024), Population (indicator). doi: 10.1787/d434f82b-en

OECD (2024), Inflation (CPI) (indicator). doi: 10.1787/eee82e6e-en

OECD (2024), Exchange rates (indicator). doi: 10.1787/037ed317-en

Prynn, J. (2012, October 17). The Olympics boom created 100,000 jobs in London. Evening Standard; Evening Standard. https://www.standard.co.uk/news/london/the-olympics-boom-created-100-000-jobs-in-london-8214954.html

Ross, S. (2019). What impact does the balance of trade have on GDP calculations? Investopedia. https://www.investopedia.com/ask/answers/061515/what-impact-does-balance-trade-have-gdp-calculations.asp

Taylor, A. (2012, July 26). Business Insider; Insider. https://www.businessinsider.com/how-the-olympic-games-changed-barcelona-forever-2012-7

The Decision Lab. (2021). Satisficing. The Decision Lab. Retrieved April 12, 2024, from https://thedecisionlab.com/reference-guide/psychology/satisficing

The World Bank (2023), World Development Indicators https://databank.worldbank.org/source/world-development-indicators.

Appendix

[7]**GitHub Repository:** https://github.com/AbdullahKhurram30/research-proposal/

[8]**Dataset Link:** https://docs.google.com/spreadsheets/d/11J2mMOC0rFnNx5ZgzlAaKJW-62hcgmqH4fU3Sm6k0Kk/edit?usp=sharing

[7]**#dataanalysis:** My data analysis spans two different directions. Firstly, I applied this LO to do exploratory analysis on the data, making plots of the GDP per capita trend for Australia, comparing it to other countries, and then also doing a similar comparison for two other variables commonly associated with GDP per capita. These plots were also explained and their inferences discussed in the paper. The second part was doing the pilot study that was done using Synthetic control. Here, the algorithm code was implemented and alongside, various plots and tables were generated to present the results. This allowed the reader to easily interpret the study and follow along with what the proposed methodology is doing.

[8]**#dataprep:** My dataprep application consists of the transition from the initial dataset to the final dataset. This is showcased in the attached notebook. The code is clearly commented as to what it is doing. This dataprep is done in Python as it allowed for a more algorithmic approach for the dataset. First, teh dataset is converted to a long format. Then, we extracted the year column, pivot the dataset to a wide dataset, and reset the index. A column mapper is used to rename columns algorithmically and then the dataset is converted to numeric formats.