Faculty of Economics and Business
Department of Economic Analysis

eman ta zabal zazu

Universidad del País Vasco · Euskal Herriko Unibertsitatea

Three Essays in Economics and Politics

Serhat Hasancebi

PhD Thesis

Supervisors: Cruz Ángel Echevarría Olave and Javier García-Enríquez

Bilbao, October 2021

"The supreme guide in life is knowledge."

Mustafa Kemal Atatürk

For my beloved mother and father ...

Declaration

All sentences or passages quoted in this project dissertation from other people's work have been specifically acknowledged by clear cross referencing to author, work and page(s). I understand that failure to do this amounts to plagiarism and will be considered grounds for failure in this module and the degree examination as a whole.

Name: Serhat Hasancebi

Signed: 

Date: 01/09/2021

Acknowledgment

Throughout the writing of this dissertation I have received a great deal of support and assistance.

Above all, I consider myself privileged enough to have Cruz Ángel Echevarría Olave and Javier García-Enríquez as my thesis supervisors. I am thankful for the continuous advice and assistance they have given me. I will always appreciate the way they structured my thinking. They did not hesitate to give their attention to me whenever I approached them. During my tough moments, they constantly supported me and showed me the route towards progress.

Besides my advisors, I would like to thank to members of the department of economic analysis: Prof. Jesús Vázquez, Prof. Javier Gardeazabal, Prof. Aitor Ciarreta, and Prof. María José Gutiérrez, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I would like to acknowledge my friends Hakan Bay, Luis Herrera, Asif Khan, Unai Tellitu, and Merve Tuncer. They were always present in sharing happiness and sadness. Thank you for not leaving me alone on this long journey.

My sincere thanks also goes to Prof. Veysel Ulusoy and Assistant Prof. Caner Ozdurak, who provided me an opportunity to join their team as a visiting scholar at the Yeditepe University in Turkey.

Last but not the least, I would like to thank my family: my parents and to my sisters for supporting me spiritually throughout writing this thesis and my life in general, without them this achievement would not have been possible.

Serhat Hasancebi, Bilbao

Contents

List of Figures

List of Tables

Introduction of the Thesis

This doctoral thesis consists of three chapters that conduct deep research in the fields of Economics and Political Sciences. In relation to the first line of research, the economic impacts of integration and social violence (wars, terrorism, and revolutions) are studied in accordance with causal inference literature in the first two chapters of this thesis. Quantitative methods rely on constructing counterfactual [or potential evolution of a unit (country, region, state)] in the absence of intervention that can be compared to the evolution actually observed. The difference or comparison between the two provides an accurate estimate of the impact of the intervention. More precisely, the first and second chapters are comparative case studies that examine the treatment effects in Iran and Macao's national economies using causal inference tools. Therefore, it is worth defining briefly what causal inference is.

Causal inference is the use of theory and institutional knowledge to quantify the effect of events and decisions on a particular outcome of interest and dates back to 18. century where the ideas of *cause* and *effect* have started to develop Cunningham (2021). Fisher (1935), Rubin (1974, 1977) among others. It was, however, until Rubin (1974) and Holland (1986) that causal inference was projected as a statistical tool. One of the fundamental problems of causality analysis, as indicated by Holland (1986), is that the unit under some intervention or exposure can not be obtained without mentioned treatment. That is, it is impossible to observe the unit with or without treatment simultaneously. Thus, the challenge of causal inference lies in the best estimation of the counterfactual unit that resembles the treated unit well. As a result of this exigency, the literature on causal inference has rapidly expanded to the present, allowing researchers to apply diverse problems under various methodologies. In order to study the effects of treatments on national economies, two major approaches were used in this thesis. To begin, the *synthetic control method* (SCM) is employed in the first two chapters, Iran's Revolution and Macao's Integration Process, respectively. The SCM was first introduced by Abadie and Gardeazabal (2003) for the purpose of quantifying the economic consequences of terrorism in Basque Country on the economic growth of the region, and it was further developed in Abadie et al. (2010, 2015), where they examine the effects of California's tobacco legislation on cigarette consumption, and the reunification of Germany on the gross domestic product, respectively. The SCM uses the weighted average of control units (known as "*donor pool*") in order to reproduce a counterfactual unit that best resembles the exposed unit by treatment effect. Once the counterfactual unit is estimated, the difference between the actual trajectory and the synthetic unit gives the treatment effect on the outcome of interest. The importance of the synthetic control estimator can be described along the lines of Athey and Imbens (2017), where they demonstrate the SCM as "*arguably the most important innovation in the policy evaluation literature in the last 15 years*". Second, another important methodology, the *Panel Data Approach* proposed by Hsiao et al. (2012) for accessing the impact of a policy intervention by demonstrating the dependence among cross-sectional units, thereby constructing the counterfactual unit. The PDA focuses on classical regression framework to reproduce counterfactual unit, unlike the SCM. Consequently, the second chapter invokes both the SCM and the PDA as an empirical study on assessing the impacts of Macao's transition into mainland China.

Continuing along the second line of research, which contains the third chapter of the thesis, it focuses on the connection between the design of electoral constituencies (provincial or unique at the national level) and the number of representatives elected by each political party. This is a matter of the greatest interest to both political scientists and public opinion: given the distribution of votes by political parties, different designs of the electoral constituencies may give rise to a different number of seats or parliamentarians for each political party. This phenomenon (also known as gerrymandering in the literature) has been the subject of several studies, particularly in the United States. To mention just a few, it is worth mentioning Chen and Rodden (2013), Cox and Katz (2002), or Gelman and King (1994a,b). Nevertheless, the standard asymmetry or advantage of one party over another is only a general statement of principle. For example, and this is the point, the Supreme Court of the United States has never deemed any analytical approach for evaluating asymmetry to be practical. Sam Wang, on the other hand, has devised three statistical contrasts in two recent publications to accurately measure the asymmetry in statewide constituencies: *i)* a non-representative distortion in the number of seats obtained based on expectations of the characteristics of the national districts; *ii)* a discrepancy between the two major parties in the average vote margins across electoral districts; and *iii)* undue and reliable gains for the party in charge of the redistricting, measured by the difference between the average and the median in the percentage of votes, or by an unusually uniform distribution of votes in the districts [see

Wang (2016a,b)]. The formal approach employed by Wang is absolutely novel in the field of political science, so it has not yet been applied to other case studies such as for example, Malta. It is therefore a reasonable and feasible contribution.

Chapter 1: **The Economic Cost of Revolution: The Iranian Case. A Synthetic Control Analysis**

Chapter 1 examines the consequences of the Islamic Revolution on the Iranian economy. The 1979 Revolution affected Iran's economic growth with various factors that caused this economic recession. After the Islamic leader Khomeini's return from exile in France, leading a large crowd, the Islamic Revolution began. The Iranian revolution has affected many factors politically and economically, and has been an important factor that changed Iran's future. The Islamic Revolution of Iran has been the subject of several papers and has been frequently studied by scientists. Both the origin of the revolution, its causes and sociological results are discussed. However, the economic consequences of the Iranian Revolution have not been the subject of any literature with the quantitative-statistical modeling approach. Furthermore, analyzing the impact of the revolution on Iran's economy is essential to understanding the role that political instability plays in economic growth. Consequently, this chapter uses the SCM in order to assess the causal impact of the Iranian Revolution. In addition, and as a *by-product*, several optimization/algorithms techniques have been implemented under the SCM to evaluate the alternative computational procedures.

Chapter 2: **The economic effects of Macao's integration to mainland China: a causal inference study**

Chapter 2, as a case study, examines Macao's integration into Mainland China and enlightens the macroeconomic impacts of such integration on the economy of Macao. Macao was under Portuguese colony for centuries, and in 1999 its administration was transferred to the People's Republic of China. Macao is now the subject of various scholars due to its tremendous economic growth after the mentioned integration process and ranks seventh highest economy in the world in terms of per capita GDP. Thus, it is worth to evaluate such economic development caused by policy evaluation. Furthermore, it represents the first study in which the SCM and the PDA are used to evaluate the impact of the mentioned integration process on the Macao's economy. Accordingly, this chapter seeks to answer the question of what would have happened if the handover of Macao had not taken place?

Chapter 3**: The Maltese Single Transferable Vote Experience: A Case Study of Gerrymandering?**

Regarding the Chapter 3, the objective is to study with quantitative-approach statistics how the design of constituencies affects the composition of parliamentary majorities, taken as a case study of Malta. One of the most discussed problem of political sciences, redistricting the borders of the electoral constituencies in order to favor a political party or incumbent, has been the core issue of modern democracies. Political scientists have been advocating several approaches for detecting such anomalies in order to protect fairness of the elections. Thus, one of the key hallmark of the democracy is that we need to ensure for a given political party's popular vote share should compensate its distribution of the parliamentary seats in the legislature. In the rapidly changing technological environment, governments' collecting data in order to favor their activities during the elections, say, redrawing the electoral districts according to their concentrated vote shares, so that giving rise to possibility of winning in the elections. Determining the existence of such political maneuvering, however, has been made possible by modern software and statistical tools. Consequently, considering Malta's 2013 and 2017 general elections, this last section calls into question the possibility of partisan gerrymandering.

Chapter 1

The Economic Cost of Revolution: The Iranian Case. A Synthetic Control Analysis*

Abstract

In 1978, a revolution in Iran succeeded in toppling Shah Mohammad Reza Pahlavi. After the Shah was forced to leave the country, Ayatollah Ruhollah Khomeini, one of the leaders of the revolution, returned from his exile in France to become the Supreme Leader of Iran. In this paper, we investigate the economic cost of the revolution using the synthetic control method. According to our estimates, we conclude that after the emergence of the revolution, the annual real gross domestic product (GDP) per capita in Iran declined by about 20.15% on average relative to its synthetic counterpart without the revolution in the period 1978-1980. If Iran had not faced such a revolution, the accumulated per capita GDP would have been 6,479 dollars higher, which amounts to an average annual loss of about $2,159 over that period.

Keywords: Comparative Case Study; Synthetic Control Method; Islamic Revolution; Iran.

JEL Classification Number: C13, C21, C23

*Hasancebi, Serhat. 2020. "The Economic Cost of Revolution: The Iranian Case. A Synthetic Control Analysis." *Singapore Economic Review*, [Webpage].

1.1 Introduction

The Islamic Revolution in Iran began on January 1978, overthrowing the Shah's regime. On 1 February 1979, Ayatollah Ruhollah Khomeini returned to Tehran from exile in Paris to be welcomed by several million Iranians. This occurred after series of popular protests had pushed the Shah, Mohammad Reza Pahlavi, to abandon the country [(Moin, 1999)]. On 1 April 1979, following a national referendum (98.2 per cent voted in favour), Iran was declared an "Islamic Republic". Iran's "Islamic Revolution" has had profound implications, domestically, across the Middle East and for wider Islamic–western relations. Khomeini's Islamic regime focused on a jihadhi approach to reorganize and reshape Iran's domestic and foreign policy priorities [(Demirci, 2013)].

Persia was ruled as a monarchy under a Shah starting in the 16th Century. The Qajar dynasty stayed in power until 1925 when the Shah was forced out in a military coup led by a Cossack officer, Reza Khan. He adopted the title Reza Shah Pahlavi, and in 1935 the country's name was changed to Iran [(Arjomand, 1986)]. Iran's economy under the Shah regime exhibited steady growth. Industrialization, infrastructure investment, government aid for the private sector helped to set up a free market system which yielded sustainable growth for Iran's economy [(Amuzegar, 1992)]. In 1963, Mohammad Reza Pahlavi promoted a so-called White Revolution which consisted of land reform, the sale of state factories to private entrepreneurs, and an extension of the vote to women. Thus, as reported by Demirci (2013), the White Revolution aimed to modernize society in the same mold as western countries, although the Shah administration's affairs with Unites States and European States were met with criticism from Khomeini's supporters.

With time, insufficient land reforms led the people to move to the cities from villages, and caused high unemployment rates in big cities. Increasing unemployment and widening inequality between the classes caused the people to start protests against the Shah's regime. Iran's high oil revenues were being transferred to the army budget in order to create a large and powerful army [(Amuzegar, 1992)]. Meanwhile, public discontent with the regime's policies grew, and provoked a revolutionary movement that soon overthrew the monarchy, and resulted in the declaration of an Islamic Republic. The Islamic Revolution was the consequence of demands for change coming from different social groups within Iranian society [(Esfahani and Pesaran, 2008)].

By 1978, Iran had become the second-largest OPEC producer and exporter of crude oil, and the fourth-largest producer in the world, churning out 5,242 billion barrels/day. However, due to the revolution, Iran endured a significant loss of around 39.56% in crude oil production, and reduced production to 3,168 billion barrels/day in 1979.[1] Economic growth and political stability are interconnected and, in particular, the uncertainty associated with an unstable political environment may have reduced investment and economic development [(Alesina et al., 1996)]. Therefore, analysing the impact of the revolution on Iran's economy is crucial to understanding the role that political instability plays in economic growth.

There have been several papers that examine the economic consequences of the Iranian Revolution. Amuzegar (1992) investigates the Iranian economy before and after the revolution, and discusses socio-economic characteristics of the Shah regime, as well as examining the economic development of Iran. Amuzegar, op. cit., highlights that after the revolution, the Iranian economy experienced a deep recession, and absolute poverty increased by 43% during the period from 1979 to 1985. Our article, however, only studies the period 1978-1980. The reason for keeping the post-treatment period only until 1980 was due to the emergence of another treatment: otherwise, the results would have been biased as the effect of the Iran-Iraq war (1981-1988) would have coincided with the influence of the Iranian Revolution. Arjomand (1986) investigates the theoretical significance of the Islamic Revolution in Iran by focusing on the political dynamics of the radical change in Iran's societal structure of domination, along with the moral dynamics of reintegration and collective action. Maloney (2015) examines the complex process relating to the adoption of economic policies that has taken place since 1979. However, the studies carried out so far have not gone beyond in-depth analyses of the revolution, but have followed mainly descriptive approaches.

In this paper, we investigate the economic costs of the Islamic Revolution by using the synthetic control method (SCM) first introduced in Abadie and Gardeazabal (2003) as the analytical tool. The SCM has been implemented for comparative cases in order to measure the consequences of economic shocks, events or policy interventions. Abadie and Gardeazabal (2003) analyse the economic cost of terrorism in the Basque Country in terms of loss of gross domestic product (GDP). Horiuchi and Mayerson (2015) examine the influence of the 2000 Palestinian Intifada upon Israel's economy. Pinotti (2015) estimates the economic cost of organized crime in Southern Italy. Grier and Maynard (2016) study the impact of Hugo Chavez's regime on the Venezuelan economy. Gardeazabal and Vega-Bayo (2017) study the effect of the political and economic integration of Hong Kong with China. Bilgel and Karahasan (2017) analyse the cost of separatist conflict in Turkey in the case of terrorism. Echevarría and García-Enríquez (2019) examine the economic consequences of the Libyan spring. Echevarría and García-Enríquez (2020) analyse the economic cost of the Egyptian episode of the Arab Spring.

[1]Statistics have been obtained from OPEC data set for oil production, https://www.opec.org

The main contribution of this article is that, as far as we know, it represents the first study in which the synthetic control method is used to evaluate the impact of the revolution on the Iranian economy. In addition, and as a by-product, we compare the performance of alternative computational procedures (Matlab©, R, Stata©), and select the one providing the most accurate results. The evidence found in this study implies that after the Islamic Revolution the annual per capita GDP in Iran declined by about 20.15% on average, and by 47.70% in cumulative terms relative to its synthetic counterpart without the revolution in the period 1978-1980. This means that the cumulative per capita GDP loss in Iran was about 6,479 US dollars after the revolution, which amounts to an average annual loss of about $2,159 in the latter period.

The structure of the paper is as follows. Section 1.2 describes the SCM. Data and variables are described in Section 1.3. Section 1.4 shows the main results. Section 1.5 is devoted to the discussion of robustness checks on the results. Section 1.6 concludes. Two formal appendices are included at the end of the paper: Appendix 1.A compares different computational methods to implement the SCM, and Appendix 1.B describes technical details corresponding to Section 1.5.

1.2 Synthetic Control Method

Case studies usually purpose to observe the effect of a treatment in order to examine whether the effect is large or small according to the outcome of interest. Thus, case studies are feasible when some units are under the effect and others are not.

The SCM illustrates a hypothetical counterfactual unit by taking the weighted average of pre-intervention outcomes from selected donor units. The donor units that are combined to form the synthetic control are selected from a pool of potential candidates. Predictor variables that affect the outcome, and the outcome variable itself before the intervention is enacted, determine the selection of donor units and weights.

The following describes the SCM in comparative case studies.[2] Suppose that we observe $J+1$ units. Without loss of generality, suppose also that only the first unit is exposed to the intervention of interest, so that we have J remaining units as potential controls. Borrowing from the statistical matching literature, we refer to the set of potential controls as the "donor pool".

[2]This heavily draws on Abadie et al. (2010), Abadie et al. (2015) and Echevarría and García-Enríquez (2019), Echevarría and García-Enríquez (2020).

Let Y_{it}^0 be the outcome that would be observed for the unit i at time t in the absence of the intervention, for units $i = 1, ..., J+1$, and time periods $t = 1, ..., T$. Let T_0 be the number of the pre-intervention periods, with $t = 1, ..., T_0$. Let Y_{it}^1 be the outcome that would be observed for unit i at time t if unit i were exposed to the intervention periods T_0+1 to T. We assume that the intervention has no effect on the outcome before implementation period so, for $t = 1, ..., T_0$ and all $i = 1, ..., J+1$, we have that $Y_{it}^1 = Y_{it}^0$. In practice, interventions may have an impact prior to their implementation (e.g., via anticipation effects). In these cases, T_0 could be interpreted as the first period in which the outcome may possibly react to the intervention. Implicit in our notation is the usual assumption of no interference between units. That is, we assume that outcomes of the untreated units are not affected by the intervention implemented in the treated unit.

The treatment effect for the treated unit in period t is given by

$$\alpha_{1t} \equiv Y_{1t}^1 - Y_{1t}^0, \tag{1.1}$$

for $t = T_0+1, T_0+2, ..., T$, and where Y_{1t}^1 denotes the (observed) potential outcome under treatment, and Y_{1t}^0 denotes the (unobserved) potential outcome under the hypothesis of no treatment. The way to circumvent this difficulty was first suggested in Abadie and Gardeazabal (2003), which consisted of building up a synthetic treated unit: a weighted average conveniently obtained among the untreated units in such a manner that it mimics the pre-treatment periods as closely as possible. The payoff is an estimate of Y_{1t}^0 which allows one to obtain an estimate for α_{1t}.

Optimal weights are found by solving the following problem:

$$min_{\{w\}_{j=2}^{J+1}} \, (\boldsymbol{X}_1 - \boldsymbol{X}_1^s)'\boldsymbol{V}(\boldsymbol{X}_1 - \boldsymbol{X}_1^s) \tag{1.2}$$

where $\boldsymbol{X}_1 - \boldsymbol{X}_1^s$ is the difference between the P-dimension pre-treatment characteristic vector of the treated unit, $\boldsymbol{X}_1$ denoting the vector of predictors of the outcome variable for the treated unit and $\boldsymbol{X}_1^s \equiv \sum_{j=2}^{J+1} \boldsymbol{X}_j w_j$ standing for the vector of predictors for the synthetic control. Optimal weights, $\{\boldsymbol{w}_j\}_{j=2}^{J+1}$, are restricted to being non-negative, and to add up to one. $\boldsymbol{V}$ is a diagonal, positive semidefinite matrix whose p-th element, $V_p \geq 0$, represents a weight that reflects the relative importance assigned to the p-th variable in vector $\boldsymbol{X}$ as a predictor of the outcome variable, $\boldsymbol{Y}$, and where $\sum_{p=1}^{P} V_p = 1$. The vector of optimal weights, $\boldsymbol{w} = \{w_2, w_3, ..., w_{J+1}\}$, will depend, of course, on the values of $\boldsymbol{V}$. This issue is thoroughly discussed in Appendix 1.A.[3]

Once optimal weights, w_j^*, have been obtained, the effect of the intervention on the treated unit for period t is estimated as

$$\hat{\alpha}_{1t} \equiv Y_{1t}^1 - \sum_{j=2}^{J+1} w_j^* Y_{jt}^0, \tag{1.3}$$

for $t = T_0 + 1, T_0 + 2, ..., T$, and where (by construction and if the donor pool has been correctly specified) no Y_{jt}^0 is affected by the treatment or intervention experienced by unit 1.

A measure of the goodness of fit of the synthetic unit to the observed treated unit, and the one that we follow in this paper as a criterion to rank the alternative computational methods implemented, is the pre-treatment root mean squared prediction error ($Pre - RMSPE$), which is defined as

$$Pre - RMSPE \equiv \left(\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2 \right)^{1/2}. \tag{1.4}$$

Conversely, the post-treatment root mean squared prediction error ($Post - RMSPE$), defined as

$$Post - RMSPE \equiv \left(\frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2 \right)^{1/2}, \tag{1.5}$$

provides us with an approximate measure of the effect (in absolute terms) of the treatment effect, so that the ratio of the latter to the former is interpreted as a natural assessment of the quantitative effect of the treatment.

1.3 Data and Variables

We use annual country-level data which results in a panel data set consisting of 60 countries and 21 yearly observations for each country for the period 1960-1980. Due

[3]See Appendix 1.A for a technical discussion regarding the different computational procedures implemented in this paper to obtain the optimal weights, $\{w_j\}_{j=2}^{J+1}$, in Eq. (1.2).

to the lack of data for periods prior to 1960, our sample starts in that year. The Islamic Revolution took place in 1978 which gives a 19 year pre-intervention period. As mentioned in the introduction, if there had been no other overlapping treatment such as the Iran-Iraq war (1981-1988), longer term analysis would have been feasible. It is not meaningful to mix the two treatments in the sample period, so that we finally set 1980 as the last post-treatment year.

First, as a tentative starting point, we used the Maddison Project Database as our data source which contains per capita GDP for 161 countries. As a next step, we excluded the countries which faced any armed conflict or civilian war during our sample period. If the potential control unit in the donor pool had experienced a sizeable strike, revolution, civilian war or any armed conflict during the same period, it would not have been possible to isolate the effect of the Islamic Revolution in the variable of interest. Nevertheless, structural breaks such as nationwide economic crises that affect both the treated country and the donor pool countries do not invalidate the synthetic control estimates. In addition, we also removed some countries from the sample due to the lack of data for the outcome variable and/or for the predictors which we discuss below.[4]

The donor pool finally consists of 59 control countries after omitting those that are significantly and sustainably affected by some sort of armed conflict. Remaining countries in the donor pool are: Albania, Angola, Australia, Austria, Bahrain, Belgium, Botswana, Cameroon, Canada, Cape Verde, China, Colombia, Costa Rica, Denmark, Djibouti, Egypt, Finland, France, Gabon, Germany, Hungary, India, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Kuwait, Lebanon, Madagascar, Malawi, Malaysia, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Namibia, Netherlands, New Zealand, Norway, Philippines, Puerto Rico, Romania, São Tomé and Principe, Saudi Arabia, Senegal, Singapore, Spain, Sri Lanka, Sweden, Switzerland, Trinidad and Tobago, United Arab Emirates, United Kingdom, United States and Venezuela.

The outcome variable in our case, Y_{jt}, is the per capita GDP in country j at time t [available at Maddison Project]. The per capita GDP is measured in 1990 international dollars which allows us to make international comparisons. Regarding the predictors, first, we use the average fertility rate (births per woman) for the period 1970-1977, and the average of gross capital formation to GDP ratio for the period 1973-1977. Data have been obtained from the World Bank and the United Nations Statistics Division respectively [World Bank].

[4]The whole list of the 48 excluded countries and the circumstance leading to exclusion in each case is shown in a table not included in this paper for the sake of space saving, but which can be obtained from the author upon request.

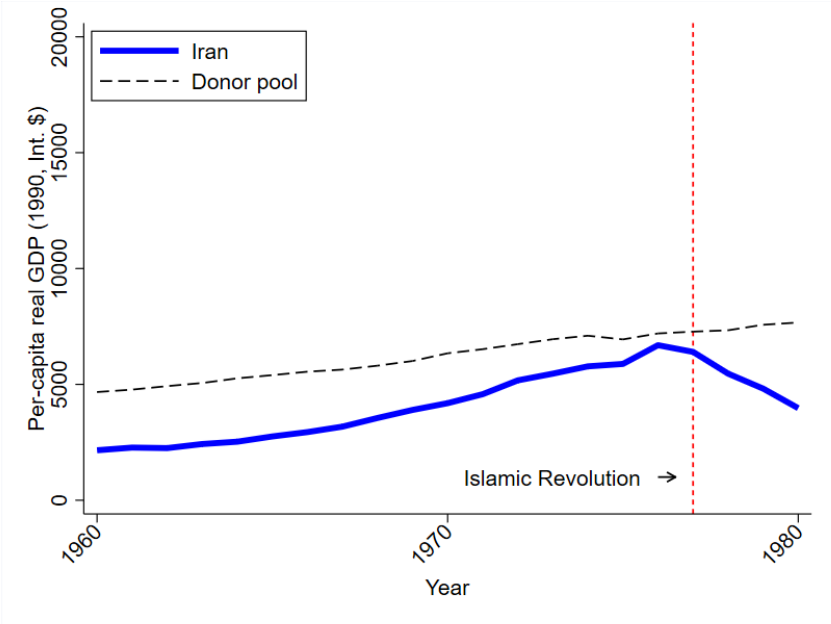
Second, we also include the average of final consumption expenditure to GDP ratio for the period 1973-1977, and the average of retail trade to GDP ratio for the period 1972-1977. In addition, we include the average GDP share of exports of goods and services for the period 1972-1977. All the data have been obtained from the United Nations Statistics Division [available at UNSD]

Finally, we added some (but not all) lagged values of per capita GDP corresponding to 1960, 1963, 1965, 1971, 1974, 1976 and 1977. Otherwise, using all the lagged values of per capita GDP would make all other predictors irrelevant, as pointed out by Kaul et al. (2021). Initially we also experimented with a wider set of additional predictors: population density (people per square km of land area), levels of educational attainment, changes in inventories (relative to GDP), population, total factor productivity, household consumption expenditure (relative to GDP), imports of goods and services to GDP ratio, exchange rate, human capital index based on years of schooling, price level of capital stock and, finally, manufacturing as a proportion of GDP. But, in all cases, their inclusion did not reduce the pre-treatment root mean square prediction error in Eq. (1.4) which we have adopted as a measure of goodness of fit.

1.4 Results

The time paths of per capita GDP for Iran and the average of the countries in the donor pool between 1960-1980 are shown in Figure 1.1. The average per capita GDP among the donor countries approximately mimics Iran's until the Islamic Revolution. After the revolution, however, the difference between the two paths increases. Nevertheless, we will show that a synthetic control can more accurately resemble the pre-revolution per capita GDP path for the observed Iran than the simple average among control units.

Figure 1.1: Trends in per capita GDP: Iran and donors



Key: Time paths for per capita GDP for Iran and the average of the 59 countries in the donor pool.

The values of $Pre - RMSPE$, $Post - RMSPE$ and the ratio of the latter to the former as a natural assessment of the quantitative effect of the treatment were obtained under the alternative computational procedures mentioned in Footnote 3 (see Table 1.3 in Appendix 1.A). Note that, first, different procedures give rise to different control countries and, second, even to different weights for the same control countries.

To sum up the discussion in Appendix 1.A, we conclude that the MSCMT package in R outperforms the other computational procedures in terms of goodness of fit and also in terms of computational speed (not reported for the sake of space saving). Therefore, all the results obtained in the sequel have been obtained using this package to compute synthetic countries. In particular, the synthetic Iran is obtained as a weighted average of five countries, namely: Botswana, Gabon, Japan, Saudi Arabia, and Singapore. The corresponding weights, w_j^*'s, are 30.19%, 9.81%, 15.12%, 0.40% and 44.48% respectively.

The pre-revolution characteristics for predictors and the outcome of interest for Iran, the synthetic Iran and the means for the 59 countries in the donor pool are shown in Table 1.1. As can be seen, the synthetic control resembles the treated country closely in terms of lagged per capita GDP values, gross capital formation and retail trade quite remarkably. Regarding final consumption expenditure relative to GDP, exports of goods and services and fertility rate, the differences between Iran and its synthetic counterpart are higher. Despite such discrepancies, it can be concluded that, in general, both the observed Iran and its synthetic counterpart are very similar in terms of outcome predictors during the pre-revolutionary period. That is, the synthetic Iran provides a better fit for the pre-revolution period than the simple weighted average of the countries in the donor pool. As claimed in Botosaru and Ferman (2019), although the use of all covariates provides an advantage when balancing for constructing the synthetic control estimator, an accurate balance on covariates may not be required for the synthetic control method, "as long as we have a good match on outcomes over an extended period of time prior to the treatment". Finally, the last column in Table 1.1 reports the values of the optimal matrix $\boldsymbol{V}$ in Eq. (1.2). As expected, optimal predictors are higher for lagged values of per capita GDP than for the rest of predictors.

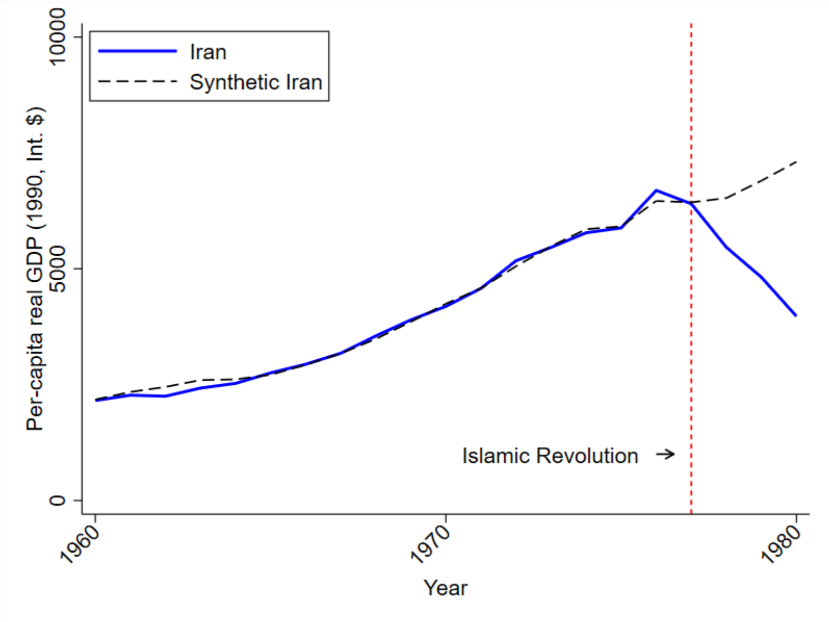
Table 1.1: Pre-revolution characteristics

Predictors	***Iran***	***Synthetic Iran***	***Donor Pool***	$\boldsymbol{V}$
y_{1960}	2,156.48	2,176.94	4,693.88	28.49
y_{1963}	2,426.09	2,596.96	5,414.11	28.49
y_{1965}	2,753.15	2,713.95	6,350.74	28.49
y_{1971}	4,576.78	4,579.79	6,757.12	–
y_{1974}	5,777.71	5,854.13	6,971.01	–
y_{1976}	6,691.35	6,463.41	7,131.94	12.53
y_{1977}	6,402.08	6,433.52	6,969.99	–
$Fertility_I$	3.63	5.13	4.3	–
Gcf_{II}	29.65	28.69	28.3	0.25
Fce_{II}	89.48	83.73	76.1	0.44
$Retail_{III}$	9.9	10.03	14.0	1.16
$Export_{III}$	25.8	29.02	31.4	–

Key: Economic predictors for the pre-revolution period for Iran, the synthetic Iran and donor countries respectively. Predictors are as follows; y_t: per capita GDP at time t; Fertility: average fertility rate (total births per woman); Gcf: average gross capital formation (% GDP); Fce: average final consumption expenditure (% GDP); Retail: retail trade (% GDP); Export: average exports of goods and services (% GDP). Periods are denoted as follows I: 1970-1977; II: 1973-1977: III: 1972-1977. The last column shows the optimal predictor weights in per cent terms; weights below 0.1% are omitted.

The trends in per capita GDP for Iran and its synthetic counterpart over the whole sample period 1960-1980 are displayed in Figure 1.2. The synthetic per capita GDP trajectory matches the actual GDP trajectory quite well in the pre-revolution period (1960-1977), which is the first sign of the success of the synthetic control method. However, following the Islamic Revolution, the two paths start to diverge, the per capita GDP level of synthetic Iran following an upwards trend, while Iran's observed per capita GDP starts out on a heavily downward trajectory.

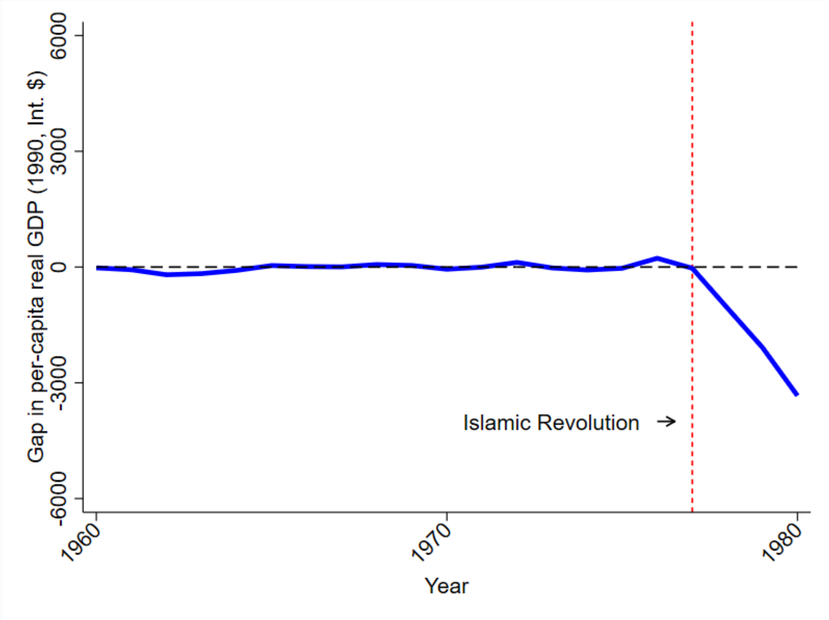
Figure 1.2: Trends in per capita GDP: Iran vs. the synthetic Iran



Key: Time paths of per capita real GDP for Iran and the synthetic Iran for both the pre- and post-treatment periods.

The estimate of the impact of the Islamic Revolution on economic development for Iran is given by the difference between the actual and the synthetic per capita GDP, as shown in Figure 1.3 which compliments Figure 1.2. The per capita GDP gap necessarily mimics the zero-gap line in the pre-revolution period (indicating good fit) and begins to depart from it after the exposure to the Islamic Revolution.

Figure 1.3: Trends in per capita GDP gap: Iran vs. the synthetic Iran



Key: Time path for the gap of per capita real GDP.

All in all, when looking at the numerical solutions, the results become more meaningful. Thus, Table 1.2 displays GDP and economic effects for Iran and its synthetic counterpart for each of the three years in the post-treatment period, along with the accumulated effects until 1980. The first column displays Iran's actual growth rate, while the second column indicates the growth rate of the synthetic Iran. For instance, the observed growth rate of per capita GDP in 1978 was -15.85%, while that of the synthetic Iran in the same year was 1.42%. Likewise, the observed growth rate of per capita GDP for Iran was -47.70% in the three year period, while that of the synthetic counterpart was 12.76%. The third column shows the difference in the growth rates between actual Iran and its counterfactual, thereby giving rise to a loss of 17.27% in the first year of the revolution, and 60.46% in the cumulative value of three years. The difference in the growth rates between actual Iran and its counterpart is indicative of the loss in per capita GDP. In the first year after the revolution, Iran's per capita GDP loss was $1,061. In the case of the cumulative effect the forgone per capita GDP was 6,479 dollars (as measured in 1990 international dollars), that is, in the absence of the revolution Iran would have enjoyed $2,159 higher annual per capita GDP on average over a period of 3 years. Ultimately, in aggregate terms, the real GDP loss amounted to 4,336.519 billion dollars [see columns 4 and 5]. The reduction in Iranian oil production due to the Islamic Revolution caused a drop in national revenues. As mentioned before, Iran suffered a significant loss of crude oil production to the tune of roughly 39.56%, and fell to 3,168 billion barrels/day in 1979.

Table 1.2: Growth and GDP effects

Period	g	g^s	Δg	$\Delta pcGDP$	ΔGDP $(billions)$
1978	-15.849	1.420	-17.270	-1,061	-688.045
1979	-12.598	5.592	-18.190	-2,083	-1,383.656
1980	-19.249	5.748	-24.997	-3,335	-2,264.818
1978-1980	-47.696	12.760	-60.457	-6,479	-4,336.519
Annual average	-15.899	4.253	-20.152	-2,159	-1,445.506

Key: g: growth rate (%) of per capita GDP for Iran; g^s: growth rate (%) of per capita GDP for the synthetic Iran; $\Delta g \equiv g - g^s$: difference in growth rates between Iran and its synthetic counterpart; $\Delta pcGDP$: loss in per capita GDP; ΔGDP: loss in real GDP.

Summarising, we conclude that after the outbreak of the Islamic Revolution the annual per capita GDP in Iran declined by about 20.15% on average relative to its synthetic counterpart without the revolution in the period 1978-1980.

1.5 Robustness Tests

In order to check the validity of our estimates, we run some robustness tests to check whether our results might have been driven by chance alone, or whether the synthetic control method has detected a causality relationship between the Islamic Revolution in Iran and its economy. A summary of the results follows, the reader being referred to Appendix 1.B for further details.

First, following Abadie and Gardeazabal (2003) and Abadie et al. (2010), we run the *in-space placebo test* by applying the synthetic control method to each and every country in the set of control units (which, therefore, were not exposed to a revolution during the sample period of our study), shifting Iran to the donor pool in all cases. That is, we proceed *as if* each one of the countries in the donor pool had been affected by the revolution in 1978. If the placebo test had detected similar results for countries other than Iran, then the observed fall in growth and per capita GDP in Iran would have been driven by factors other than the Islamic Revolution.

We find that the probability of obtaining an average post-treatment gap less than or equal to Iran's would be $1/48 = 0.0208$. In other words, the null hypothesis of no treatment effect of the Islamic Revolution on Iran's per capita GDP would be rejected at a significance level of 5%. Complementary to this, we also computed the distribution of R, which is defined as $R \equiv (Post - RMSPE/Pre - RMSPE)^2$ (see Eq. (1.5) and Eq. (1.4)), for all the countries in the sample: intuition, confirmed by the results, says R should be the highest for Iran. We find that the probability of obtaining a value of R higher than or equal to that of Iran equals $1/60 = 0.0166$. That is to say, the null hypothesis of no treatment effect for Iran would be rejected at a significance level of 5%.

Second, we also run one *in-time placebo test* by shifting the intervention year as if Iran had hypothetically been treated in some generic period prior to 1978, for instance in 1968, which is the midpoint of our true pre-treatment period. Had the synthetic control method properly implemented, no treatment effect should be detected before the revolution (Abadie et al., 2015). We find that the time paths for the observed and the synthetic per capita GDP series are very close to each other both before and after 1968 (see Figure 1.7 in Appendix 1.B). In other words, the pre-treatment fit is more than acceptable and, additionally, no treatment effect is detected. The conclusion is that the SCM has correctly detected the impact of the revolution on Iran's economy.

And, third, we implement a *leave-one-out* exercise in which we iteratively re-estimate the baseline model to construct different counterfactuals, omitting in each iteration one of the five countries that received a positive weight in the synthetic Iran that we obtained in Section 4 (namely, Botswana, Gabon, Japan, Saudi Arabia and Singapore). By excluding countries that received a positive weight some goodness of fit is sacrificed, but this sensitivity check allows us to evaluate to what extent our results are driven by some particular control country. A similar exercise is conducted in Abadie et al. (2015) or Echevarría and García-Enríquez (2019), Echevarría and García-Enríquez (2020) among others. Our results show that the smallest and the largest estimates of the effects of the revolution (in terms of accumulated loss of per capita GDP for the 1978-1980 period) correspond to cases where Singapore and Gabon are removed from the sample respectively. In other words, leaving Gabon out of the donor pool leads to the highest estimate of the gap, while leaving out Singapore leads to the lowest (see Table 1.4 in Appendix 1.B).

Note that there is no monotonic relationship between i) the size of the weight of the left-out unit (Singapore's weight equals 44.5%, the maximum) and ii) the change in the estimated treatment effect after excluding that particular unit from the set of control units (omitting Singapore from the donor pool leads to a minimum effect). The intuition is as follows. For a given weight set obtained in the pre-treatment period, the consequence on the estimate of the treatment effect of leaving one specific unit out of the controls will trivially depend on the post-treatment time evolution of the outcome variable for that unit relative to the average among the rest of control units. Thus, removing a country with a high weight in the unconstrained synthetic country might have no significant effect on the estimated treatment effect provided that the post-treatment period evolution of the outcome variable happened to be close enough to the average for the unconstrained donor pool. However, and conversely, eliminating a country with a low (although not negligible) weight, but with a performance totally different from that of the rest of controls, might lead to a synthetic country completely distinct from the one obtained in the first instance. And, therefore, to a completely different estimate of the treatment effect. The conclusion drawn from the leave-one-out check is that no single control country seems to be driving the results or the estimates found in Section 4.

1.6 Conclusions

The Islamic Revolution, one of the most important conflicts in the history of Iran, began in 1978, overthrowing the Shah's administration. The following year, with the referendum held, Iran was declared an "Islamic Republic". This study was carried out to reveal the economic consequences of the Islamic Revolution in Iran by employing the synthetic control method. The synthetic control method creates a counterfactual unit with the weighted average of the best representative countries in the set of control units. Consequently, the comparison of the established counterfactual unit and the actual Iran allows us to assess the effect of the Islamic Revolution on Iran's economy. In addition, we attempted to achieve the best results by comparing alternative computational methods to obtain the synthetic or counterfactual Iran. Given that the root mean squared prediction error for the pre-treatment period attained with the MSCMT package in R was the lowest, thereby ensuring the best fit, such package turned out to be our choice.

According to our estimates, the numerical results are as follows. The per capita GDP growth rate achieved in the first year following the revolution was -15.85% for Iran, whereas for synthetic Iran it was 1.42%. The cumulative per capita GDP growth rate for the three year period was -47.70%, while for the synthetic counterpart this rate was 12.7%. In other words, the loss in terms of the growth rate of per capita GDP in the three year period was 60.46%. Furthermore, the average per capita GDP loss was $2,159 per year, and the cumulative loss of the three year period was $6,479. That is, if Iran had not been faced with such a revolution, it would have benefited from an annual average of $2,159 higher per capita GDP. Likewise, in aggregate terms, Iran would have seen an annual average of 1,445.5 billion dollars higher real GDP. This corresponds to a total of 4,336.5 billion dollars higher real GDP in the three year period.

In order to check the statistical significance, we run some exercises. First, we tried an in-space placebo analysis by applying the synthetic control method to all the countries in the donor pool as an attempt to assess the hypothetical effects that could emerge had all the countries been exposed to the revolution in 1978. Iran turned out to be the country most affected by the revolution, thereby rejecting the hypothesis of there being no treatment effect of the Islamic Revolution on Iran's economy. Second, we also performed an in-time placebo test by hypothetically shifting the intervention period to 1968, to conclude that the synthetic control method did not find any treatment effect. Finally, we extend the sensitivity tests with a leave-one-out analysis: the results did not differ much from the results obtained in the benchmark case. In summary, all the robustness checks suggest that our conclusion is fairly robust in detecting the negative impact of the Islamic Revolution on Iran's economic development.

The evidence from this study suggests that the Islamic Revolution has a negative impact on the per capita GDP of Iran, and thereby on economic development. Most importantly, according to the results of our research, we have reinforced the importance of political stability in achieving economic development and prosperity in a country. We hope that our analysis will provide a contribution for future researchers.

APPENDICES

1.A Appendix: Alternative Computational Methods

As mentioned in the Introduction, we compare the performance of alternative computational procedures to obtain the synthetic Iran based on two criteria. First, how matrix $\boldsymbol{V}$ is computed: **exogenous** [(i) researcher's choice or (ii) regression-based]; or **endogenous** [(i) nested minimization or (ii) Cross-Validation]. And second, the computer package implemented (R in its different alternatives synth-ipop, synth-LRQP and MSCMT), Matlab© and Stata©.

The optimal weights obtained upon solving the problem in Eq. (1.2), $\boldsymbol{w}^*(\boldsymbol{V}) = \left\{w_j^*(\boldsymbol{V})\right\}_{j=2}^{J+1}$, will depend on the choice for matrix $\boldsymbol{V}$. Several alternatives have been considered in the literature (Echevarría and García-Enríquez 2019, Echevarría and García-Enríquez 2020; Ferman and Pinto, 2016; Firpo and Possebom, 2018). The first alternative is to choose the predictors weights according to the author's preference and knowledge from previous researches. As a second option, a data-driven method can be used to set matrix $\boldsymbol{V}$. After regressing the outcome $\boldsymbol{Y}$ on the set of predictors $\boldsymbol{X}$, the elements of matrix $\boldsymbol{V}$ are obtained by comparing the corresponding OLS coefficients (in modulus or squared) over the sum of all coefficients. We will refer to this alternative as the regression-based method.

The third option amounts to a *nested* (double minimization), where matrix $\boldsymbol{V}$ and control weights $\boldsymbol{w}^*$ are jointly obtained in such a way that $\boldsymbol{w}^*(\boldsymbol{V})$ solves the problem in Eq. (2) and $\boldsymbol{V}$ minimizes the square distance of the outcome between the treated unit and the synthetic counterpart,

$$min_{\{Vp\}_{p=1}^{P}} \quad \sum_{t=1}^{T_0}(Y_{1t} - Y_{1t}^{s})^2 \tag{1.6}$$

where $Y_{1t}^s = \sum_{j=2}^{J+1} Y_{jt} w_j(\boldsymbol{V})$ (see Eq. (1.2)).

As a last alternative, the *cross-validation* option refers to the division of all pre-revolutionary years into two sub-periods training and validation. In the training period, matrix $\boldsymbol{V}$ is minimized and used to obtain optimal $\boldsymbol{w}$ in the validation period (Abadie et al., 2015, Becker and Klößner, 2017).

The results are shown in Table 1.3. Regarding the choice of an appropriate computer package, we first tried the **synth** package in R, the *nested* minimization being the default option.[5] This package allows, in turn, the use of two different algorithms to minimize the $Pre-RMSPE$: Ipop and LowRankQP functions. The obtained $Pre-RMSPE$s were 96.918 and 96.914 respectively. Due to the long pre-treatment period, we also tried the *cross validation* option (once again using the **synth** package in R) by dividing all pre-treatment period into two periods, training and validation. According to our choice, the training period occurs between 1960-1968 and the validation period between 1969-1977. The fit turned out to be worse in both cases (Ipop and LowRankQP) as the resulting $Pre-RMSPE$s were 120.723 and 118.224 respectively. Next, we also tried another package, MSCMT, implemented with R as well (Becker and Klößner, 2017, 2018). In this case, $Pre-RMSPE$ turned out to equal 96.368.

Stata© has its own version of the synth package, which offers two options to compute matrix $\boldsymbol{V}$. First, $\boldsymbol{V}$ can be exogenously obtained by means of a regression based procedure, the default option, leading to a $Pre-RMSPE$ equal to 98.588. Second, matrix $\boldsymbol{V}$ also can be endogenously obtained by implementing a *nested* minimization procedure (with respect to both $\boldsymbol{V}$ and $\boldsymbol{w}$) which requires providing (up to three) initial or guessed values for $\boldsymbol{V}$, returning the one which results in the lowest $Pre-RMSPE$. In this case, we obtained a value of $Pre-RMSPE$ equal to 96.375. Performing this nested minimization, however, we found that Stata© was not robust to alternative ordering of predictors. That is, when the predictors' order is different then the optimal weights change (McClelland and Gault, 2017).

[5]The synth packages for Matlab©, Stata© and R are available at `http://www.mit.edu/~jhainm/synthpage.html`.

Table 1.3: The synthetic Iran: Alternative computational methods

1- Nested optimization: R (Ipop-synth)					
Botswana: 0.318	Gabon: 0.103	Japan: 0.156	Saudi Arabia: 0.010	Singapore: 0.413	
Pre − RMSPE: 96.918; *Post − RMSPE*: 2282.487 Ratio: 23.551					
2- Nested optimization : R (LowRankQP-synth)					
Botswana: 0.315	Gabon: 0.105	Japan: 0.155	Saudi Arabia: 0.008	Singapore: 0.417	
Pre − RMSPE: 96.914; *Post − RMSPE*: 2284.461; Ratio: 23.572					
3- Cross Validation: R (Ipop-synth)					
Botswana: 0.288	Cameroon: 0.014	China: 0.020	Gabon: 0.156	Japan: 0.151	Singapore: 0.371
Pre − RMSPE: 120.723; *Post − RMSPE*: 2081.480; Ratio: 17.242					
4- Cross Validation: R (LowRankQP-synth)					
Botswana: 0.321	Gabon: 0.152	Japan: 0.169	Mongolia: 0.018	Singapore: 0.339	
Pre − RMSPE: 118.224; *Post − RMSPE*: 2065.311; Ratio: 17.469					
5- Nested optimization : R (MSCMT)					
Botswana: 0.302	Gabon: 0.098	Japan: 0.151	Saudi Arabia: 0.004	Singapore: 0.445	
Pre − RMSPE: 96.368; *Post − RMSPE*: 2351.978; Ratio: 24.406					
6- Nested optimization: Matlab					
Botswana: 0.276	Gabon: 0.137	Japan: 0.100	Singapore: 0.487		
Pre − RMSPE: 121.197; *Post − RMSPE*: 2223.187; Ratio: 18.344					
7- Regression-based: Stata					
Botswana: 0.328	Gabon: 0.116	Japan: 0.155	São Tomé and P. : 0.015	Senegal: 0.386	
Pre − RMSPE: 98.588; *Post − RMSPE*: 2209.888; Ratio: 24.415					
8- Nested optimization (Allopt): Stata					
Botswana: 0.305	Gabon: 0.097	Japan: 0.151	São Tomé and P. : 0.007	Senegal: 0.440	
Pre − RMSPE: 96.375; *Post − RMSPE*: 2351.207; Ratio: 24.396;					

Key: Synthetic Iran (countries and weights) under alternative computational methods. See main text.

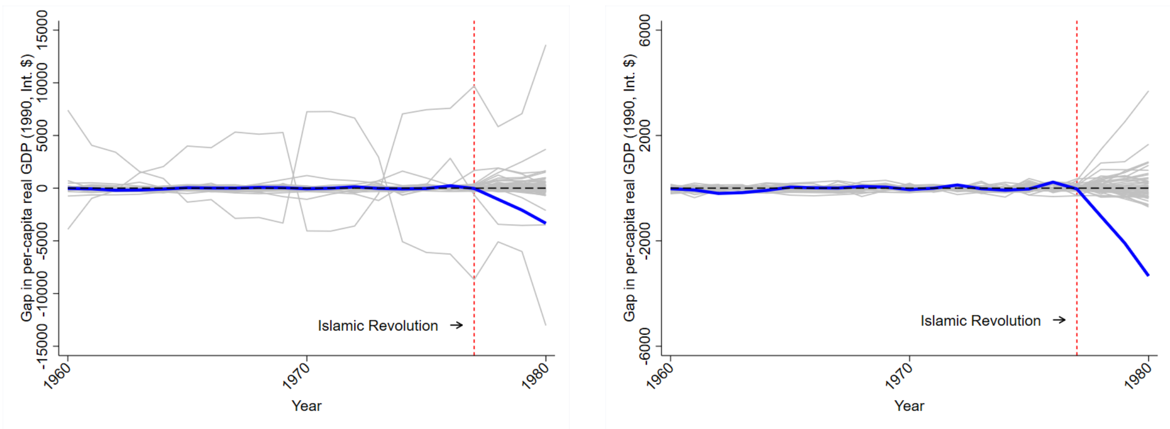
As for the Matlab© package, it also implements a joint minimization with respect to both $\boldsymbol{V}$ and $\boldsymbol{w}$, which needs the corresponding guessed solution as the starting values for $\boldsymbol{V}$. Using a regression-based method to find such starting values, we obtained $Pre-RMSPE$ equal to 121.197. Unfortunately, different starting values (which must be set by the researcher) led to very different solutions. Moreover, using alternative versions of Matlab© (and even different machines) again resulted in different optimal weights. All this makes Matlab© the least reliable computational method among those considered here. In summary, MSCMT, the R package, has proven to be the most effective tool to compute the synthetic control in this paper.

1.B Appendix: Robustness tests

In-space placebo test. The graphical results are shown in Panel (a) in Figure 1.4, where each grey line illustrates the estimated gap in per capita GDP for each of the 59 countries in the donor pool, while the blue line represents the estimated gap for Iran.

Figure 1.4: In-space placebo tests

(a) (b)



Key: Per capita *GDP* gap in Iran and placebo gaps. *Panel (a)*: Iran plus 59 control countries. *Panel (b)*: Iran plus 47 control countries after excluding 12 countries whose $Pre-RMSPE$ is at least twice Iran's.

The estimated per capita GDP gap of Iran closely follows the zero gap line before the intervention which is one of the requirements for proper use of the SCM as an estimation method [(Abadie, 2021)]. Notice also that the fit for some of the countries in the donor pool is not particularly precise. Therefore, confidence in the method is compromised in some instances. Thus, we discard those placebo runs for which the $Pre-RMSPE$ (see Eq. (1.4)) is more than twice that of Iran. As a result, the sample is reduced to Iran plus 47 control countries, so that 12 countries are excluded. The idea is that countries with a too high $Pre-RMSPE$ (or, equivalently, with a poorer fit than Iran's) are not reliable, in the same way that a poor fit for Iran would also reduce our confidence in the validity of our results. This exercise can be found, amongst others, in Abadie et al. (2010) or Acemoglu et al. (2016).[6] The result is shown in Panel (b) in Figure 1.4. As expected, now all the estimated gaps for the pre-treatment period are lower and closer

[6] Acemoglu et al. (2016) follow a similar criterion as only control units with a $Pre-RMSPE$ less than or equal to $3^{1/4}$ times that of the treated unit are considered to test the statistical significance of the treatment effect.

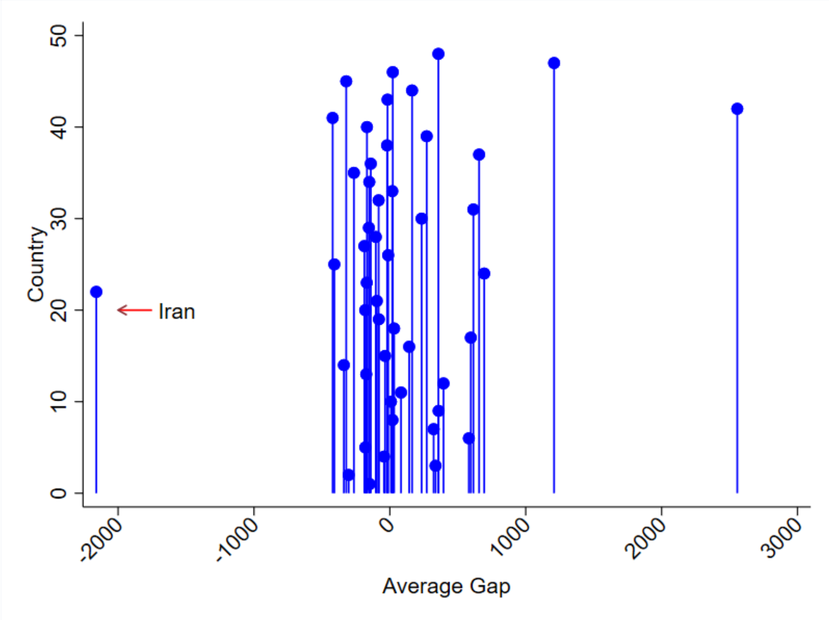
to Iran's. And, additionally, after inspecting the plots of the estimated gaps for the post-treatment period for all countries, it becomes apparent that the Islamic Revolution had a negative impact on Iran that the 47 control countries did not experience.

In order to formally test the statistical significance of the effect and go beyond a simple graphical analysis, we computed the distribution of the average post-treatment gap (AG) for the 48 countries represented in Panel (b) in Figure 1.4. Since the difference between the observed outcome and its synthetic counterpart in the Iranian case will be (in absolute value) the largest amongst that set of countries, the average post-treatment gap in Iran is expected to be the lowest. More precisely, the AG of unit k is defined as

$$AG_k \equiv \frac{\sum_{t=T_0+1}^{T} \left(Y_{kt} - \sum_{q=1,q\neq k}^{K+1} Y_{qt} w_{kq}^* \right)}{T - T_0}, \tag{1.7}$$

where w_{kq}^* denotes the optimal weight for control unit q and for synthetic unit k, and where $\left\{1, 2, ..., k, ..., K+1\right\}$ denotes the set of countries in the sample left after removing those with a poor pre-treatment fit. Iran's AG was the lowest amongst the 48 countries, and the probability of obtaining an AG less than or equal to Iran's would be $1/48 = 0.0208$ [see Figure 1.5]. In other words, conditional on control countries resulting in $Pre-RMSPE$ values less than twice that of Iran, the null hypothesis of no treatment effect of the Islamic Revolution on Iran's per capita GDP would be rejected at a significance level of 5%.

Figure 1.5: Average post-treatment gap per country



Key: The average post-treatment gap of the 48 countries (including Iran) obtained in the last stage of the in-space placebo test. Numbers along the y-axis denote the country IDs.

A complementary way to formally test the statistical significance of the impact of the Islamic Revolution on Iran's economy is to show the ratio of post-treatment mean square prediction error to pre-treatment mean square prediction error for all countries: intuitively, a sizeable impact should result in a high ratio. For this purpose, we obtained the entire distribution of ratios for all countries. More precisely, from Eq. (1.4) and Eq. (1.5), the ratio for country j, R_j, is defined as

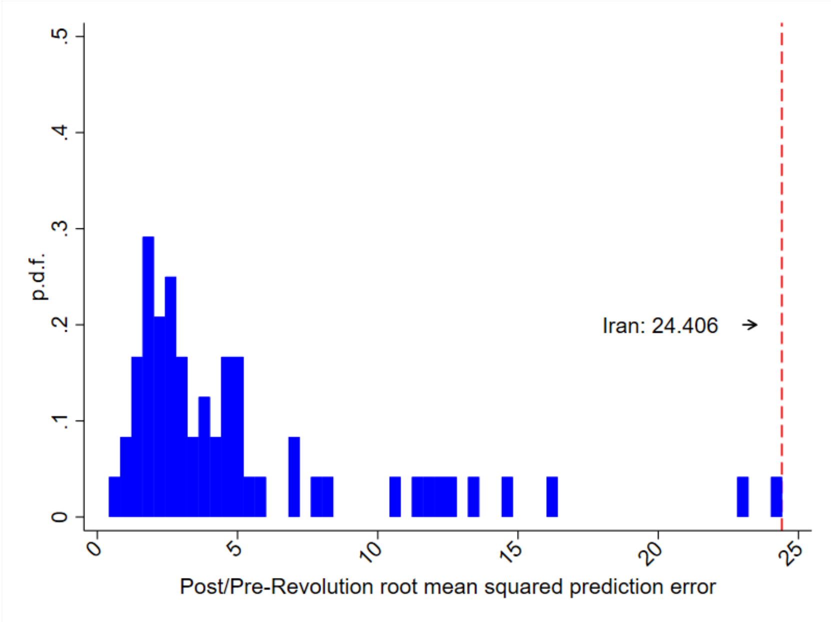
$$R_j \equiv \frac{\sum_{t=T_0+1}^{T}\left(Y_{jt} - \sum_{q=1,q\neq j}^{J+1} Y_{qt}w_{jq}^{*}\right)^2/(T-T_0)}{\sum_{t=1}^{T_0}\left(Y_{jt} - \sum_{q=1,q\neq j}^{J+1} Y_{qt}w_{jq}^{*}\right)^2/T_0} = \left[\frac{Post - RMSPE_j}{Pre - RMSPE_j}\right]^2. \tag{1.8}$$

Figure 1.6 displays the distribution of the ratio for the 60 countries in our sample. Formally, the probability of finding a ratio higher than or equal to Iran's is given by

$$p(R_1) \equiv \frac{\sum_{k=1}^{J+1}[R_k \geq R_1]}{J+1}. \tag{1.9}$$

In our case, $p(R_1)$ equals $1/60 = 0.0166$. That is to say, the null hypothesis of no treatment effect for Iran would be rejected at a significance level of 5%.

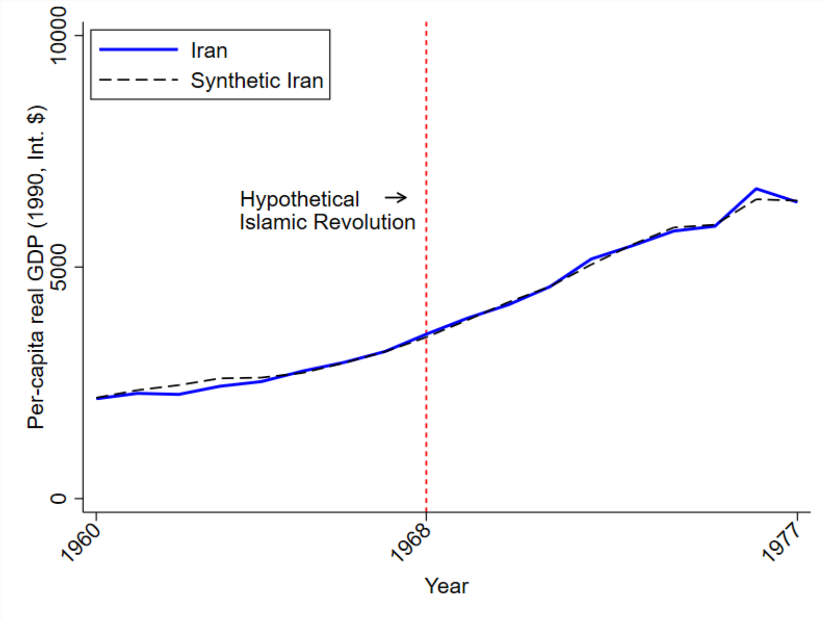
Figure 1.6: Distribution of R



Key: Ratio of the post-Islamic Revolution to the pre-Islamic Revolution $RMSPE$: Iran with 59 countries in the donor pool.

In-time placebo. In the case, we pose the treatment year in 1968 in order to see whether the synthetic control method captures a treatment which did not take place. If that were the case, the SCM would have failed to find the exact effect of the Islamic Revolution. A similar exercise of out-of-sample validation is carried out in Abadie et al. (2015) where they set the hypothetical date of the German reunification in 1975, i.e. 15 years before reunification actually took place. The graphical result is shown in Figure 1.7. We observe that the time paths for the observed per capita GDP and for the counterfactual follow the same pattern both before 1968 (so that the fit is remarkably good and the synthetic Iran is reliable as a counterfactual), and for the period 1968-1977 (so that no treatment effect is detected). We interpret this result as showing that the SCM has correctly detected the impact of the revolution on Iran's economy.

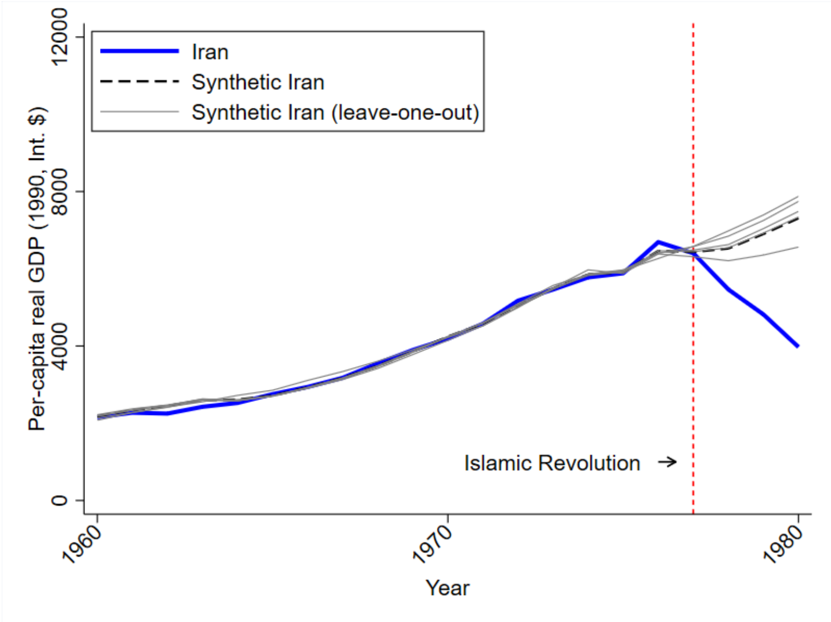
Figure 1.7: In-time placebo test



Key: In-time placebo test under the counterfactual that the Islamic Revolution had occurred in 1968.

Leave-one-out. Here we recalculate the synthetic Iran five times excluding alternatively Botswana, Gabon, Japan, Saudi Arabia and Singapore. Figure 1.8 displays the results, where the blue solid line represents the observed Iranian per capita GDP in terms of evolution over time, the black dashed line represents the synthetic Iran with the above five controls (the same as in Figure 1.2), and the grey lines represent the five restricted counterfactuals.

Figure 1.8: Leave-one-out



Key: Time paths of per capita GDP for Iran, the synthetic Iran with all donor pool countries with positive weights and the synthetic Iran dropping one country at a time.

The fit in the pre-treatment period (remarkably good) is quite similar in all cases. As for the post-treatment period, however, some small differences in the estimated counterfactuals (and, consequently, in the estimated gaps) appear. Thus, the exercise provides a sort of confidence band for the estimated post-treatment gap which contains, by construction, the estimate obtained in Section 4.

The numerical results of the exercise (the quantitative counterpart to Figure 1.8) are shown in Table 1.4. More precisely, the Table presents the smallest and the largest estimates of the effects of the revolution (in terms of accumulated loss of per capita GDP for the period 1978-1980). These are obtained when Gabon and Singapore respectively are left out of the sample: i.e. leaving Gabon out of the donor pool leads to the highest estimate, while leaving out Singapore leads to the lowest. For instance, the (post-treatment) average annual growth rate of per capita GDP of the synthetic unit formed when Gabon is removed from the donor pool is 17.98%, while that of the synthetic Iran formed when Singapore is taken out of the donor pool is 3.88% [see columns 2 and 6]. For completeness, columns 4 and 8 display the estimated losses of per capita GDP when Gabon and Singapore are left out of the donor pool respectively. Thus, the removal of Gabon would result in an annual average loss of per capita GDP of $2,665, while leaving out Singapore would imply a loss of $1,624.

Table 1.4: Leave-one-out analysis

		Leaving out Gabon (*maximum effect*)				***Leaving out Singapore*** (*minimum effect*)			
Period	g	g^s	Δg	$\Delta pcGDP$	ΔGDP ($billions$)	g^s	Δg	$\Delta pcGDP$	ΔGDP ($billions$)
1978	-15.85	5.93	-21.78	-1,517.2	-983.2	-1.66	-14.19	-744.7	-482.6
1979	-12.60	5.75	-18.35	-2,576.9	-1,711.1	2.34	-14.44	-1,538.4	-1,021.5
1980	-19.25	6.31	-25.56	-3,901.7	-2,649.3	3.20	-22.45	-2,588.6	-1,757.7
1978-1980	-47.70	17.98	-65.68	-7,995.9	-5,343.6	3.88	-51.58	-4,871.8	-3,261.9
Annual ave.	-15.90	5.99	-21.89	-2,665.3	-1,782.2	1.29	-17.19	-1,623.9	-1,087.3

Key: See Key to Table 1.2.

Chapter 2

The economic effects of Macao's integration in mainland China: a causal inference study*

Abstract

Macao was a Portuguese colony until 1999, when its sovereignty was ceded to China. In July 2000, the "Macao Gaming Committee" was formed with the final aim of liberalizing the gaming industry, starting a bidding process with the result that three gaming concessions were granted in 2002. Later, in June 2003, Macao signed the Closer Economic Partnership Arrangement (CEPA), and the Individual Travel Scheme and Removal of Preferential Tariff, a milestone in the integration process of the former colony into mainland China. In this article, we try to estimate the consequences of this socio-economic process in terms of per capita *GDP*. We build up a panel data set, spanning the period 1970 and 2012, with 25 countries, setting 2000 as the initial treatment year for the integration process. The analysis is carried out by means of two alternative methodologies: the Synthetic Control Method and the Panel Data Approach. We find that the integration treatment had a significant, positive effect on Macao's per capita *GDP*. Finally, we also analyze the effects of integration on the per capita net inflow of foreign direct investment, the unemployment rate and the per capita exports and imports of goods and services as additional outcome variables.

Keywords: Comparative Case Studies; Synthetic Control Method; Panel Data Approach; Causal Inference; Integration; Macao.

JEL codes: C14, C15, C20, C34

*This chapter was written with the collaboration of my advisors, Cruz Ángel Echevarría Olave and Javier García-Enríquez.

2.1 Introduction

Macao (often referred to in statistical data bases as Macao SAR, China) currently exhibits the status of Special Administrative Region (SAR) under the central government of the People's Republic of China (PRC). Historically, Macao became a Portuguese colony when the first Portuguese traders settled in 1557 and rented the territorial dominion from the Ming officials, becoming a transshipment center for East-West trade. But it was not until 1849 that Portugal increased its influence at the expense of the local Portuguese inhabitants, becoming a fully-blown colony in its own right. From a strict point of view, nevertheless, Portugal never had the sovereignty of Macao and, unlike Hong Kong, Macao was never ceded to the Portuguese [Yee (2001)].

In April 1987, the PRC and Portugal signed the Joint Declaration by the Government of the Portuguese Republic and the Government of the People's Republic of China on the Question of Macau which stipulated that the sovereignty over the territory would be ceded to the PRC in December 1999. Such a Joint Declaration initiated a "smooth transition" towards a future status of Macao as a SAR (as opposed to Hong Kong's rough transition), thereby enjoying a high degree of autonomy, with the exception of foreign and defense affairs, which remained the exclusive reserve of the national government in Beijing, and (as in the case of Hong-Kong) becoming another instance of "One Country, Two Systems" [Edmonds and Yee (1999)].

Macao's economy experienced a remarkable transformation as a consequence of this process. The average annual growth rate of *GDP* between 1999 and 2012 was about 12.73%, thereby allowing Macao to post a per capita *GDP* at current US$74,061 in 2016, the seventh highest in the world according to The World Bank. On top of this, the unemployment rate fell from 6.3% in 1999 to 1.97% in 2012.[1] Along the same lines, the annual average net inflow of foreign direct investment in per capita terms between 1982 and 1999 was $8.7, and $7,032.5 between 2000 and 2012 (measured at chained PPPs in 2017US$).[2] The process also affected the inflow of foreign visitors: thus, the yearly average of foreign visitors for the period 1995-1999 was 7.5 million, while it reached 19.5 million for the period 2000-2012.[3] The main reason for such huge economic development can be found in the gaming industry, and we will elaborate on this in more detail in Section 2.2. At the time of the start of the integration process into the PRC, Macao's economy rested on four sectors: manufacturing, construction and real estate, financial services, and gaming. By 2013, however, the gaming industry represented more than 60%

[1] See The World Bank. The effect on unemployment will be specifically discussed in Section 2.6.
[2] See Penn World Table 10.0 and The World Bank.
[3] See The World Bank.

of *GDP*. The year 2013, however, was the beginning of a turning point, as annual growth rates of casino gaming revenues dropped sharply: -2.56%, -34.33% and -3.31% in 2014, 2015 and 2016 respectively [see Center for Gaming Research at University of Nevada, Las Vegas]. Several factors help explain this extraordinary change of trend. First, the anti-corruption campaign which started in late 2012, which dissuaded mainland high rollers from conspicuous consumption. Second, the tighter transit visa policy and the removal of China UnionPay devices in casinos and to fight money laundering as from mid 2014. Third, the increasing competition in the gaming sector in neighboring countries such as Indonesia, Malaysia, Thailand, and Singapore. Finally, the smoking ban on casino main floors introduced in October 2014 [see Liu et al. (2015) and Sheng and Gu (2018)].

In this article we aim to go one step further than the descriptive-style works found in the literature that have dealt with the case of Macao by providing a quantitative assessment along the lines of causal inference literature, which is more and more present in current research. More precisely, we estimate the consequence in terms of i) per capita *GDP*, ii) per capita net inflow of foreign direct investment, and iii) the unemployment rate, and iv) per capita exports and imports of goods and services of the intervention outlined above i.e. the integration process into mainland China. The closest precedents to our work are Hsiao et al. (2012) and Gardeazabal and Vega-Bayo (2017), both of them analyzing the, to some extent, parallel case of Hong Kong.

Hsiao et al. (2012) study the consequences in terms of the per capita *GDP* growth rate of two major interventions or treatments with a sample period which spans from 1993:Q1 to 2008:Q1. The first is the *political integration* (change of sovereignty) into mainland China in July 1997. The second (the *economic integration*) is related to the signing of the Closer Economic Partnership Arrangement (CEPA) with mainland China and the Individual Travel Scheme and Removal of Preferential Tariff signed in June 2003 (but whose implementation would start in January 2004). In both cases, the method used to estimate the counterfactual, or the potential outcome variable in the absence of intervention, is what they term the *Panel Data Approach* (*PDA*). In order to be more precise, in the case of political integration, the pre-treatment period spans from 1993:Q1 to 1997:Q2, setting the post-treatment period from 1997:Q3 to 2003:Q4, i.e. until the beginning of the economic integration treatment. Hsiao et al. (2012) find that political integration had a statistically non-significant impact on Hong Kong's economic growth because the predicted Hong Kong saw the third largest effect out of eight controls considered. This negative result allows them to extend the pre-treatment period for economic integration from the very beginning of the sample period (1993:Q1) right through to 2003:Q4. Regarding economic integration, instead, they find that the real *GDP* growth rate of Hong Kong increased by more than 4% compared to the hypothetical growth rate for the period 2004:Q1-2008:Q1 in the absence of the CEPA agreement with mainland China.

Gardeazabal and Vega-Bayo (2017) revisit the same issue (and the same data set), but using an alternative methodology to estimate the counterfactual, namely the *Synthetic Control Method* (*SCM*). They show that the results depend on which method is used to obtain the unobserved per capita *GDP* growth rate in the absence of treatment. Thus, the authors find that, under the *SCM*, the effect of political integration is statistically significant at the 10% level: the synthetic Hong Kong has the largest gap out of the other 13 countries in the sample. The immediate consequence is that, if the *SCM* is implemented instead of the *Panel Data Approach* (*PDA*), the pre-treatment period to compute the effect of economic integration must be reduced (in particular to 2000:Q2–2003:Q4). The effect of economic integration seems to be smaller when estimating it using the *SCM* than when using the *PDA*, and the statistical significance results are inconclusive. As a byproduct, an additional finding is that the results obtained under the *PDA* are less robust to changes in the set of controls than the *SCM*.

Macao's integration in PRC differs, however, from the Hong-Kong case in one key aspect: the timing. As described in the previous paragraphs, two landmarks can be clearly distinguished in the experience as regards Hong Kong: the political integration and the economic integration, separated by 6 years and 2 quarters. Macao's integration, by contrast, is a much faster process. As will be discussed in Section 2.2, the time lapse between the cession of Macao to the PRC and the signature of the CEPA are separated by 2 years and a half. Additionally, and this is the key point, deregulation of the gaming industry Macao, giving rise to a huge increase in economic activity, as manifested by the gaming tax revenues, took place in between. This leads us to refer to a unique process of integration.

In this paper we build up a panel data set which spans the time period from 1970 to 2012, and contains 25 countries: Macao and 24 control units (namely, Indonesia, Japan, Malaysia, Thailand, United States, Taiwan, Philippines, Germany, United Kingdom, France, Australia, Austria, Brazil, Canada, Denmark, Finland, Italy, Mexico, Netherlands, New Zealand, Spain, Sweden, Switzerland and Norway). According to the above discussion, we set 2000 as the initial treatment year for the integration. We first set the per capita *GDP* as the outcome variable to focus on. We next reinforce our analysis by considering the effects on the net inflows of foreign investment, unemployment, and exports and imports as additional outcome variables. Another variable which might be worth considering is that of travel and tourism, a potential driver of economic activity. Unfortunately, the series available for the total travel and tourism contribution to *GDP* starts in 1996 [see The World Bank], which prevents us from estimating the counterfactual. The analysis is carried out following a similar computational strategy as the one in Gardeazabal and Vega-Bayo (2017), that is, estimating the effects of the intervention by means of two alternative methodologies: *SCM* and *PDA*.

Our main results follow:

1. The integration process of Macao into the PRC did have a statistically significant, positive treatment effect on both the level of per capita *GDP* and its long-run growth rate for the period 2000-2012. The qualitative aspect of the result is fairly robust to the method followed to compute the counterfactual (whether *SCM* or *PDA*). The estimates of average gaps in the level of per capita *GDP* and its growth rate vary according to the method chosen to compute the counterfactual. Following the best pre-treatment fit criterion (which in our particular exercise will always be the *PDA-AICc*), we find an approximate yearly average per capita *GDP* gap of $42,321 for the period. Similarly, we find a statistically significant difference of 11.4% between the observed growth rate of per capita *GDP* and that of the counterfactual series.

2. We have also analyzed some additional outcome variables. Thus, we have found a statistically significant, positive treatment effect on foreign investment. According to our *PDA-AICc* results, the integration of Macao into China caused an approximate yearly average increment of $6,358 in the per capita net inflow of foreign investment during the post-treatment period.

3. As for unemployment, we have found a seemingly negative annual average effect (-1.392%) or fall in the unemployment rate. The assessment, however, largely varies with the method implemented to compute the counterfactual and, even after focusing on the *PDA-AICc* approach, the apparent effect is not statistically significant.

4. In order to back up our results, we have also considered two more additional outcome variables: exports and imports of goods and services. Regarding the former, we have found a statistically positive yearly effect on the per capita level of approximately $42,115. As for the latter, the (at first sight) positive effect on per capita imports strongly depends on the method implemented to compute the counterfactual, so its statistical significance is discarded.

5. As an overall conclusion, we have learned that the methodological approach implemented in causal inference studies of this type does matter, and it is strongly recommended that researchers conduct deep, rigorous sensitivity analyses to test

the internal validity of their results.

The rest of the paper is organized as follows. Section 2.2 describes the main steps in the integration process. Section 2.3 describes the twofold empirical strategy implemented to estimate the counterfactual. Section 2.4 describes the data. Results of the integration are discussed in Section 2.5. Section 2.6 discusses the effects on foreign direct investment, unemployment, exports and imports. Section 2.7 concludes.

2.2 Macao's integration process

As pointed out in the Introduction, Macao's integration process formally started on 20 December 1999 when the sovereignty of the territory was ceded to mainland China, thus honoring the Joint Declaration signed by the two parties involved twelve years previously, on 13 April 1987 and which explicitly followed the principle of "one country, two systems". Such a joint agreement established, among other provisions, Macao's pataca as the legal currency and its perfect convertibility, the exclusive right of the autonomous government to set its own tax system. Additionally, the executive power would be vested in the Government of the Macao SAR, which would be formed by the local population, with the person in the highest position of responsibility appointed by the Central People's Government on the basis of the results of elections or consultations held in Macao. This marked the birth of the Special Administrative Region of Macao of the People's Republic of China.

Gambling was first legalized in Macao in 1847 where for decades it operated as a monopoly concession. But right after the cession of sovereignty (actually, the very next day), the Chief Executive announced his plan of liberalizing Macao's gaming industry, and the Macao Gaming Committee was formed in July 2000, its main function being "to conduct studies on the development, legal issues, administrative regulations and policies related to gaming".[4] As a direct consequence, a bidding process was opened in late 2001, with two concessions being granted by mid 2002, followed by successive sub-concessions which by the end of 2012 allowed the existence of 35 casinos [see Gaming Inspection and Coordination Bureau, Macao SAR and Hsieh and Wang (2014).] This put an end to the monopoly regime that had existed up to that point, and (as can be seen in Section 2.6), made it possible to attract a substantial amount of foreign direct investment and foreign visitors [Simpson (2018)]. Despite some decline in the following years, in 2016 the casino

[4]In the words of the Gaming Inspection and Coordination Bureau itself, the aim was to reinforce the policy direction set by the Macao SAR: "tourism, gaming, conventions and exhibitions as the "head", and the service industry as the "body", driving the overall development of other industries."

gaming participation was still about 47% of that economy [Sheng and Gu (2018)]. In the words of Lo (2009), Macao became "[a] casino-driven capitalist economy". Some further figures to emphasize the magnitude of the sector follow. *i*) In 2013 the gross gaming revenue in Macao relative to Las Vegas was 6.4 times higher. *ii*) In spite of the aforementioned decline, the gross gaming revenue was still 4.0 times higher in 2016 [see `https://www.casinonewsdaily.com/`]. *iii*) The result was that the average annual growth rate of casino gaming revenues in Macao between 2002 and 2013 was 28.86%. *iv*) The revenue from gambling tax in real terms increased by 20.38% in 2000, 13.71% in 2001, 26.75% in 2002, 38.39% in 2003, 42.62% in 2004 and 30.44% in 2005.[5] This last series of figures reinforces the concept of an integration process or a smooth transition as opposed to the Hong Kong case.

Another landmark within this process is represented by the Closer Economic Partnership Arrangement that, after a period of consultations, the two sides signed on October 2003, and which became effective in January 2004. The CEPA, a sort of free trade agreement arrangement, and with the purpose of "promoting stable economic development as well as the improvement of living standard for the two sides", focused on three main lines: i) reduction or progressive elimination of tariff and non-tariff barriers in all trade in goods between the two parties; ii) progressive liberalization of trade in services by reducing or eliminating all discriminatory measures between the two parties; and iii) promotion of trade and investment facilitation [see Government Printing Bureau and CEPA].

Two immediate consequences of the CEPA followed. The first one was that the liberalization allowed mainland Chinese citizens (those residents in Beijing, Shanghai and some cities in the province of Guangdong) to visit Macao individually, that is, they did not need to be part of an organized trip. This is the so-called Individual Visit Scheme, which is explicitly covered in Article 14 of the CEPA [see Liu et al. (2015)]. Another remarkable consequence was an arrangement for the zero tariff for trade in goods, developed in Annex 1 of the CEPA. As a result, Macao would continue to exempt from customs duty all goods imported from the Mainland, and the Mainland would progressively exempt from customs duties goods imported from Macao.

[5]See Macao's Year Book of Statistics for several years. In order to obtain the figures in real terms, we used the consumer price index (base 2010) available at the The World Bank.

2.3 Empirical strategy

As set out in the Introduction, the results in this paper have been obtained by carrying out two alternative empirical strategies largely used in the literature to assess the impact of an intervention on one unit by previously estimating the corresponding counterfactual or potential outcome of the treated agent in the absence of treatment: namely, the *SCM* and the *PDA*.[6] The former was introduced in a seminal paper in Abadie and Gardeazabal (2003), and further developed in Abadie et al. (2010) and Abadie et al. (2015) amongst others in a rapidly increasing stream of literature.[7] The latter was introduced in Hsiao et al. (2012), and later compared to the performance of the former in Gardeazabal and Vega-Bayo (2017), Wan et al. (2018) and Hsiao and Zhou (2019).

The starting point is the same in both cases. Assume that there exist $J+1$ units (countries, regions,...) where, without loss of generality, unit 1 (and only unit 1) is exposed to some sort of intervention or treatment at some time $t = T_0$. We denote the outcome variable of interest for unit j and at time t upon which we want to measure the effect of intervention by $Y_{j,t}$. Specifically, $Y^1_{j,t}$ and $Y^0_{j,t}$ will stand for the potential outcome of unit j and at time t *with* and *without* treatment respectively. Assume further that we can observe $Y^0_{1,t}$ for $t = 1, 2, ..., T_0$ (i.e. all pre-intervention values of Y for the treated unit), $Y^1_{1,t}$ for $t = T_0+1, T_0+2, ..., T$ (i.e. all post-intervention values of Y for the treated unit), and $Y^0_{j,t}$ for $t = 1, 2, ..., T_0, T_0+1, ..., T$ and for $j = 2, 3, ..., J+1$ (i.e. both pre- and post-intervention values of Y for the J control units). The question posed is simple: how to predict $Y^0_{1,t}$ for $t = T_0+1, T_0+2, ..., T$ (i.e. the unobserved outcome Y for the treated unit in the absence of intervention)? We assume that there are no covariates other than lagged values of the outcome variable which either cause or are correlated with it.[8]

The *SCM* assumes that there exist weights $\boldsymbol{w} = \{w_j\}_{j=2}^{J+1}$ such that

$$\widehat{Y}^0_{1,t} = \sum_{j=2}^{J+1} w_j Y^0_{j,t}. \tag{2.1}$$

[6]This heavily draws on Section 2 in Gardeazabal and Vega-Bayo (2017) and Section 2 in Wan et al. (2018).

[7]In the expression of a quote that has become a classic: "*The [SCM] ... is arguably the most important innovation in the policy evaluation literature in the last 15 years*" [Athey and Imbens (2017), p. 9]

[8]We are all aware of the potential bias that using all lagged values of the outcome variable as the only predictors might generate under the *SCM*, despite the fact that the goodness of fit might be greatly improved in general. However, we wanted to use the same information with the two, *SCM* and *PDA*, methodologies, so that the difference in the results might be explained exclusively by the difference in the methods to estimate the counterfactual, as in Gardeazabal and Vega-Bayo (2017) and Wan et al. (2018). See the discussion in Kaul et al. (2021).

Denoting the pre-treatment vector of the outcome variable of the treated unit and the j-th control unit by $\boldsymbol{Y}_1^0$ and $\boldsymbol{Y}_j^0$, respectively, the weight vector, $\boldsymbol{w}$, is obtained as the solution to the program

$$\min_{\{w_j\}_{j=2}^{J+1}} \left(\boldsymbol{Y}_1^0 - \sum_{j=2}^{J+1} w_j \boldsymbol{Y}_{j,t}^0 \right)' \boldsymbol{V} \left(\boldsymbol{Y}_1^0 - \sum_{j=2}^{J+1} w_j \boldsymbol{Y}_{j,t}^0 \right), \tag{2.2}$$

subject to $w_j \geq 0$ and $\sum_{j=2}^{J+1} w_j = 1$. $\boldsymbol{V}$ (the predictors weight matrix) is a diagonal positive-definite matrix representing the relative importance of (in this case) each lagged value in the prediction of $\boldsymbol{Y}_1^0$.[9]

The *PDA*, like the *SCM*, assumes that the intervention in unit 1 affects neither pre-intervention values of Y_1 nor pre- and post-intervention values of Y_j for any of the control units. The *PDA* additionally *i*) allows for the existence of a constant term, α, (to control for differences in individual fixed-effects between the treated unit and the control units), and *ii*) imposes no restrictions on the slope regression coefficients, $\boldsymbol{\beta}$, in the following regression model

$$Y_{1,t} = \alpha + \beta_2 Y_{2,t} + \beta_3 Y_{3,t} + ... + \beta_{J+1} Y_{J+1,t} + u_{1,t}, \tag{2.3}$$

for $t = 1, 2, ..., T_0$ and where $u_{1,t}$ stands for the i.i.d. random idiosyncratic component of unit 1 and $E(u_{1,t}) = 0$. Once the *OLS* coefficients $(\widehat{\alpha}, \widehat{\beta}_2, ..., \widehat{\beta}_{J+1})$ in Eq. (2.3) are estimated, the prediction of $Y_{1,t}^0$ provided by the *PDA* is simply given by

$$\widehat{Y}_{1,t}^0 = \widehat{\alpha} + \widehat{\beta}_2 Y_{2,t} + \widehat{\beta}_3 Y_{3,t} + ... + \widehat{\beta}_{J+1} Y_{J+1,t}, \tag{2.4}$$

for $t = T_0 + 1, T_0 + 2, ..., T$.

Note the key difference between the estimated counterfactuals under the two procedures. In Eq. (2.1) the control weights are restricted (no intercept, non-negativity of weights and controls weights add up to 1). In Eq. (2.4), instead, the regression coefficients are unconstrained. If such constraints are not binding, the *SCM* will be more efficient.[10] If,

[9] We have used version 4.0.5 of R throughout all the paper and the **MSCMT** package, Version 1.3.4, to compute the weight vector, $\boldsymbol{w}$, [see Becker and Klößner (2017) and Becker and Klößner (2018)]. For alternative methods to compute $\boldsymbol{V}$ see Echevarría and García-Enríquez (2020). The program in Eq. (2.2) is usually referred to as the "inner" optimization part in the *SCM* literature. In our case, the optimal $\boldsymbol{V}$ is endogenously obtained as the solution to the "outer" minimization part within a nested minimization procedure [see Becker and Klößner (2018)].

[10] If this were the case (i.e. the constraints on the regression coefficients were not effective and a perfect fit were obtained), the *SCM* and the *OLS* regression would necessarily lead to the same control coefficients. This is so because the predictor weights of matrix $\boldsymbol{V}$ in the outer optimization problem referred to in n. 9 would become meaningless, and only the control weights, $\boldsymbol{w}$, would be relevant

on the contrary, constraints turn out to be binding, the *SCM* will give rise to a biased prediction of the counterfactual. Note also that, first, not restricting the coefficients in Eq. (2.4) leads to extrapolation bias: the estimated coefficients may (and surely will) fall out of the $[0, 1]$ interval, which allows for extrapolation out of the support of the data. And, second, a mechanical use of the *SCM* can lead to an interpolation bias if the control set contains units with characteristics very different from those of the treated unit. Thus, limiting the donor pool to similar units is highly recommended or, alternatively, penalizing the choice of controls dissimilar to the treated unit [see Abadie (2021)].[11] Unfortunately, there seems to be no clear-cut conclusion regarding which approach is preferable. In the words of Hsiao and Zhou (2019), p. 480, *"Statistical inference could be fragile and sensitive to inferential procedures [...] We are still only in the process of groping toward the truth, not discovering the truth."*

One additional problem posed by the *PDA* is that of the appropriate choice of the controls or untreated units to estimate the counterfactual, and where the researcher must take into account the limitations raised along three dimensions. First, the number of pre-treatment observations may not be large: 30 in our case. Second, the number of potential controls and, therefore, regression coefficients to estimate, J, might be very large depending on the case under study (up to 24 in our exercise). And, finally, sparsity or parsimony (i.e. a reduced number of non-zero coefficients) as a desirable property already embedded by construction in the synthetic control method [see the discussion on this issue in Abadie (2021)].

The method implemented here is the same as in Hsiao et al. (2012), Gardeazabal and Vega-Bayo (2017) and Wan et al. (2018). Briefly exposed, it can be described as follows. Assuming a set of J potential control units, in a first stage, all possible combinations of $m = 1, 2, 3, ..., M \leq J$ slope regressors (i.e. in addition to the constant term) are considered up to a maximum number of controls which is given by $M \equiv \min\{J, T_0 - g - 1\}$, where T_0 denotes the number of pre-treatment observations, and the desired minimum of degrees of freedom is denoted by g. And then, for each set of such $m-$tuples, the model specification attaining the highest R^2 is selected. Finally, in a second stage, the resulting models are ranked according to some criterion, the minimum $AICc$ in our case, although other alternatives such as the AIC and the BIC could also be considered [see Konishi and Kitagawa (2008)].[12]

[see Eq. (2.2)].

[11] It would be desirable to build up the set of control units only with countries belonging to the same "club", or as similar as possible, to minimize the interpolation issue. Assuming no external effects of Macao's integration process on the rest of the country, Chinese provinces would be the optimal choice. Once again, data limitation prevents us from ideal solutions: provincial GDP data is available starting 2001 [see National Bureau of Statistics of China].

[12] Note that the number of model specifications to be executed may be considerable: $\sum_{j=1}^{M} \frac{J!}{j!(J-j)!}$.

An increasingly popular alternative method of model selection with the aforementioned property of sparsity is provided by the *LASSO* regression originally introduced in Tibshirani (1996). By adding a penalty term in the minimization of the sum of squared residuals to the standard *OLS* case, $\lambda \sum_{j=1}^{J} |\beta_j|$, where $\lambda \geq 0$ represents a tuning parameter and $|\beta_j|$ stands for the absolute value of the slope regression coefficient of the $j-$th control unit in Eq. (2.3), non-zero coefficients (regardless of their signs) are penalized. The method requires first standardizing the predictors, in such a way that each of them (i.e. for $i = 2, 3, ..., J + 1$) is both centered ($T_0^{-1} \sum_{t=1}^{T_0} Y_{i,t} = 0$) and has unit variance ($T_0^{-1} \sum_{t=1}^{T_0} Y_{i,t}^2 = 1$). Otherwise, the *LASSO* coefficients would depend on the units in which the predictors are measured. If, additionally, the outcome variable is centered ($T_0^{-1} \sum_{t=1}^{T_0} Y_{1,t} = 0$), the constant term α in Eq. (2.3) can be suppressed. Thus, for the optimal solution on centered data ($\widehat{\boldsymbol{\beta}} \equiv \widehat{\beta}_2, \widehat{\beta}_3, ...\widehat{\beta}_{J+1}$), the optimal solution for uncentered data could be easily retrieved. Note that $\widehat{\boldsymbol{\beta}}$ would be the same, and the constant term $\widehat{\alpha}$ would be given by $\overline{Y}_1 - \sum_{i=2}^{J+1} \overline{Y}_i \widehat{\beta}_i$, where $\left\{\overline{Y}_j\right\}_{j=1}^{J+1}$ denote the *original* means.[13]

The result is that increasing values of λ lead to a lower number of control units in the regression. In the particular case that $\lambda = 0$, one will obtain the *OLS* estimator. If, instead, λ were large enough, all slope coefficients would equal zero, therefore only the constant term, α, would be left. Needless to say, the final choice will depend on the particular value at which the tuning parameter is set, which is usually accomplished by following a cross-validation procedure [see Li and Bell (2017) for formal definitions and details concerning such a procedure]. Regarding the distribution of the *LASSO* estimator, this is known only if the number of controls is relatively smaller than sample size [Knight and Fu (2000)], and even then, only asymptotically. Moreover, although there have been a lot of studies on its asymptotic distributional properties, as Jagannath and Upadhye (2018) points out, the asymptotic results can give an incorrect picture of the *LASSO* estimator's actual finite sample behavior, so inference based on this estimator is still an open question. *LASSO* estimators are in particular recommended when the number of control units is large enough. Given that this does not seem to be the case in our exercise, we claim that the selection of controls can be conveniently implemented by following the above described information criterion.

For instance, if $J = 24$, $T_0 = 30$, assuming that $g = 3$, and keeping the maximum number of controls reasonably low, $M = 10 < \min\{24, 26\}$, the procedure implies that a total of 4,540,385 regressions must be run. Computations have been made with the R **pampe** package, Version 1.1.2 [see Vega-Bayo (2015)].

[13] See Hastie et al. (2016) for details.

2.4 Data

We build up a balanced panel data with 43 yearly observations (between 1970 and 2012) and 25 countries: Macao and 24 control units. The beginning of the sample period is given by the first year for which we have found available data for per capita *GDP*, our main outcome variable in this paper, in Penn World Table 10.0. The common measure unit of this variable is chained PPPs 2017 US$, which allows us to make comparisons both across units and over time. Year 2012 gives us the end of the sample period because, as explained in the Introduction, 2013 marks the start of a turning point in Macao's economy as a result of some major reforms. Otherwise, extending the post-treatment period beyond that date would contaminate any estimate of Macao's counterfactual in the absence of integration.[14]

In order to choose the country members of the control set we have followed the criterion of geographical and/or economic proximity. Thus, we have focused on countries which on average for the period 1991-1999 exhibit a minimum of trade relationships with Macao (in terms of exports or/and imports), i.e. Macao's trade partners and which, preferably, are located in the Pacific basin in the same spirit of Hsiao et al. (2012).[15] Data on export and import partner shares have been obtained from the WITS (World Integrated Trade Solution) database available at The World Bank. Some countries were removed from the initial choice of controls for different reasons. Vietnam had its own treatment as a consequence of the war that it underwent between 1955 and 1975. Portugal, as the former occupying state, was also exposed to the same treatment as Macao. Singapore and South Korea, two of the four Asian tigers, experienced abnormally high GDP growth rates during the sample period. As regards the other two, Hong Kong was a relatively small economy which also went through its own treatment as already discussed in the introduction; as concerns Taiwan, the time evolution of its per capita *GDP* series during the pre-treatment period is very similar to that of Macao, following a stable trend during the post-treatment period [see panel (*b*) in Figure 2.1]. As a result, the donor pool consists of Indonesia, Japan, Malaysia, Thailand, United States, Taiwan, Philippines, Germany, United Kingdom, France, Australia, Austria, Brazil, Canada, Denmark, Finland, Italy, Mexico, Netherlands, New Zealand, Spain, Sweden, Switzerland and Norway. Given that the size of the countries in the control set is large enough relative to the treated country, we believe that none of them are affected by the

[14] An increased number of observations might be desirable, but quarterly data for Macao's GDP are only available starting in 2001 [see Macao Statistics and Census Service].

[15] Given the reduced number of pre-intervention observations when studying the political integration of Hong Kong in PRC, they choose countries that are either in the region or economically closely associated with Hong Kong [see Hsiao et al. (2012), p. 720,]. In our case, however, observations are not the issue, but rather interpolation bias.

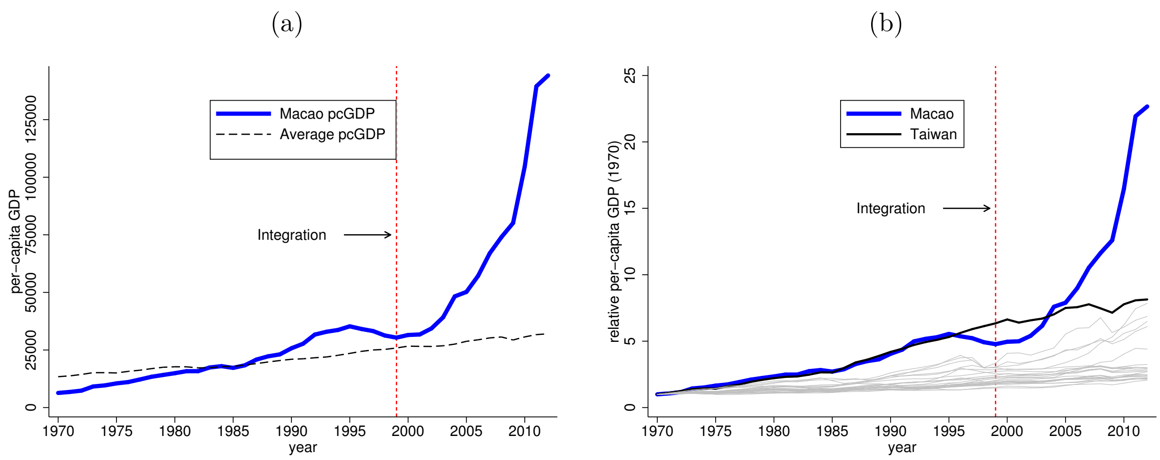
interventions in Macao, so that the estimated counterfactual is free of the potential bias externalities or spillover effects that might be induced.

As a means of lessening the interpolation bias as a result of putting together very dissimilar countries in the donor pool, we have reduced the number of potential controls to a reasonable number, at the expense of a poorer fit when estimating the counterfactual which, otherwise, might result in a case of overfitting. Clearly, enlarging the set of controls is always too easy a way to obtain a good match between the observed and the counterfactual during the pre-treatment period [see Abadie (2021)]. An additional consequence of having a relatively small set of controls is that the chances of running statistical significance tests and placebo robustness checks which are standard in the causal inference literature are, if not eliminated, at least diminished [see Billmeier and Nannicini (2013), p. 987].

2.5 Results

According to the timing of events described in the Introduction, the pre-treatment period for integration spans from 1970 to 1999, while the post-treatment period covers 2000 to 2012. The time paths for the outcome variable for both the observed values and the population-weighted average for Macao and all countries in the sample for both the pre- and post-intervention period are represented in Figure 2.1, panel (a). The difference between the two is patent, a clear indication that there is room for improvement in the search for a counterfactual. Note that during the pre-treatment period, Macao's per capita GDP and the average of all the countries in the data set follow similar trends. In the post-treatment period, however, both trends clearly diverge: while the rest of the countries continue on a steady growth path, Macao experiences growth rates substantially higher. For completeness, panel (b) in Figure 2.1 shows the series of per capita GDP normalized to 1970 levels. As discussed in the previous Section, Macao and Taiwan follow similar patterns during the pre-treatment period; during the post-treatment period, however, Taiwan does not show a change in trend, whereas Macao displays a clear increase in trend.

Figure 2.1: Per capita GDP for Macao and control countries



Key: Panel (a). The blue line represents Macao's observed path of per capita *GDP*. The black dashed line depicts the population-weighted average per capita *GDP* of the control countries in the sample. Panel (b) Per capita *GDP* normalized to 1970 levels. The blue line corresponds to Macao's observed path. The gray lines represent the corresponding series for the control countries other than Taiwan. The black line represents Taiwan's.

The results are shown in Tables 2.1 and 2.2. The first column in Table 2.1 represents Macao's actual per capita *GDP*, whereas the second column represents the (population-weighted) average for the 24 countries in the donor pool. The third, fourth and fifth columns display the estimated values under the three methods above discussed: synthetic control method (*SCM*) in column three, and panel data approach (*PDA*) in columns four and five. *PDA-AICc* (resp. *PDA-LASSO*) is implemented using *AICc* (resp. *LASSO*) as the model selection criterion in column four (resp. five). As can be seen, both methodologies' predictions were closer to the Macao's observed per capita *GDP* than the simple average of the donor pool countries. The fit, however, is better under the *PDA*, in particular *PDA-AICc*: the root of the mean square prediction error (*RMSE*) is substantially lower than that obtained under the *SCM* and *PDA-LASSO* (405.084 compared to 2,007.300 and 841.747, respectively).[16] This is the expected result, of course: by construction, the *PDA* admits extrapolation and a non-zero constant term, while the *SCM* does not. Moreover, and this will be the general case in *all* the computations (i.e. all the outcome variables considered in this paper), the fit obtained under the *PDA-AICc* will be slightly better than under the *PDA-LASSO* and, moreover, with a lower number of non-zero coefficients (except for the foreign direct investment case in which only the constant term turns out to be non-zero). Needless to say, the worst fit corresponds to

[16]By definition, $RMSE \equiv \left[\sum_{t=1}^{T_0}(Y_{1,t} - \widehat{Y}_{1,t}^{0})^2/T_0\right]^{1/2}$.

the trivial counterfactual obtained as the mean among the 24 control units ($RMSE =$ 5,287.969) [see third row from the bottom in Table 2.1].

As an alternative to the *RMSE* just introduced in footnote 16 (and which is the standard measure of pre-treatment fit reported in all software packages computing the *SCM*), we can compute the goodness-of-fit measure suggested by Ferman et al. (2020):

$$\widetilde{R}^2 \equiv 1 - \frac{\sum_{t=1}^{T_0}(Y_{1,t} - \widehat{Y}_{1,t}^0)^2}{\sum_{t=1}^{T_0}(Y_{1,t} - \overline{Y}_1)^2}, \tag{2.5}$$

where $\overline{Y}_1 \equiv T_0^{-1} \times \sum_{t=1}^{T_0} Y_{1,t}$. Note that, compared to the *RMSE*, *i*) $\widetilde{R}^2$ does not depend on the unit of measurement of the outcome variable. *ii*) $\widetilde{R}^2 \leq 1$, and possibly negative if, as might be the case, the fit obtained with the *SCM* is poor enough.[17] *iii*) $\widetilde{R}^2 = 1$ would denote a perfect fit. *iv*) $\widetilde{R}^2 = 0$ if, for instance, the fit obtained under the *PDA* were such that the estimated coefficients of all *slope* regressors happened to be identically equal to zero (i.e. $\widehat{Y}_{1,t}^0 = \widehat{\alpha} = \overline{Y}_1$ for all t).[18] And *v*) $\widetilde{R}^2$ would allow one to set a minimum threshold level for the computation of the counterfactual to be considered as acceptable. For instance, Ferman et al. (2020) consider two lower bounds in their numerical exercise, 0.80 and 0.95.

The $\widetilde{R}^2$ obtained under the *PDA-AICc* is the highest (0.998, vs 0.992 and 0.953 under *PDA-LASSO* and *SCM* respectively) [see second row from the bottom in Table 2.1]. The table also shows the observed and the estimated annual average growth rate of per capita GDP for the pre-treatment period. As expected *PDA-AICc*, provides the closest estimated value to the observed one, 5.947% vs 5.967%, [see last row in Table 2.1].

[17] This will be the case when estimating the effect of intervention upon exports and imports by applying the *SCM* as discussed in Section 2.6.

[18] And this will be precisely the case when estimating the effect of intervention upon foreign domestic investment via *PDA-LASSO* as discussed in Section 2.6.

Table 2.1: Per capita GDP pre-treatment values

Year	***Actual***	***Sample Mean***	***SCM***	***PDA-AICc***	***PDA-LASSO***
1970	6,361.435	14,018.900	8,517.113	6,835.791	6,829.343
1971	6,798.270	14,499.050	9,110.073	6,568.518	7,354.069
1972	7,375.458	15,142.250	9,941.479	7,754.016	8,382.555
1973	9,172.819	15,944.160	10,810.471	8,866.897	9,262.238
1974	9,636.785	16,202.840	10,858.118	9,430.161	9,778.670
1975	10,514.421	16,087.190	11,000.354	9,641.516	10,213.390
1976	11,082.733	16,764.680	11,956.051	11,778.672	11,095.786
1977	12,199.522	17,161.850	12,696.756	12,566.868	12,124.612
1978	13,375.349	17,811.450	13,849.628	13,127.837	13,119.991
1979	14,178.809	18,530.350	14,777.174	14,421.575	13,857.085
1980	14,934.289	19,050.370	15,577.507	14,862.754	15,264.305
1981	15,809.837	19,067.180	15,812.486	15,733.996	16,414.205
1982	15,814.070	18,835.380	15,729.487	16,170.814	15,940.571
1983	17,433.713	19,051.620	15,959.522	17,282.128	16,765.910
1984	17,986.902	19,502.430	16,496.573	17,625.962	17,762.220
1985	17,246.930	19,580.230	16,681.898	16,932.304	17,735.223
1986	18,351.347	20,327.160	18,433.900	18,537.076	18,366.513
1987	20,857.386	21,137.320	20,338.497	20,520.760	19,664.421
1988	22,260.359	21,890.290	21,925.685	22,502.991	21,062.133
1989	23,101.346	22,630.650	23,524.268	23,328.670	23,450.777
1990	25,753.531	23,417.250	25,322.526	26,008.516	26,579.385
1991	27,740.359	23,501.380	26,777.029	28,258.109	28,539.853
1992	31,699.718	23,900.250	28,102.428	30,852.974	30,324.766
1993	32,924.308	24,213.880	29,144.130	32,713.030	31,466.516
1994	33,717.316	25,102.120	30,084.245	34,252.831	32,792.392
1995	35,288.605	26,259.530	31,489.878	34,866.867	33,865.492
1996	34,133.416	27,260.690	33,201.314	34,176.263	34,984.164
1997	33,226.595	28,323.610	34,131.127	33,727.461	35,458.787
1998	31,305.890	28,778.360	34,613.898	30,711.852	30,806.037
1999	30,374.731	29,981.500	35,269.474	30,599.039	31,394.838
RMSE		5,287.969	2,007.300	405.084	841.747
$\widetilde{R}^2$		0.675	0.953	0.998	0.992
Ave. growth	5.967	2.428	5.125	5.947	5.801

Key: Estimation results of integration by the *SCM*, the *PDA-AICc* and the *PDA-LASSO*, as well as Macao's actual per capita *GDP* and the (population- weighted) average of the 24 countries in the donor pool. *RMSE*: see footnote No. 16. $\widetilde{R}^2$: see Eq. (2.5). Annual growth: average growth rate for the pre-treatment period (in per cent terms).

Table 2.2 shows the *SCM* weights and the *PDA* coefficients. As can be observed, the synthetic Macao is the result of a weighted average of only two countries, Japan and Taiwan, with weights equal to 0.410 and 0.590, respectively. Under the *PDA-AICc*, however, nine countries display (unconstrained) non-zero coefficients: four with a positive sign (Indonesia, Germany, Italy and New Zealand) and five with a negative one (Australia, Brazil, Netherlands, Switzerland and Sweden), a clear sign of the presence of extrapolation. Results under the *PDA-LASSO* differ: seven countries obtain positive coefficient estimates (Indonesia, Japan, USA, Taiwan, Canada, Italy and Mexico), and three negative (UK, Brazil and Netherlands). Regarding the effect of integration, the two methods provide the same sign effect, although quantitatively disparate. Defining the post treatment average gap (AG) as

$$AG_j \equiv \frac{\sum_{t=T_0+1}^{T}(Y_{j,t} - \widehat{Y}_{j,t}^0)}{T - T_0}, \tag{2.6}$$

where $Y_{j,t}$ and $\widehat{Y}_{j,t}^0$ denote the observed outcome and counterfactual for country j, respectively, for $j = 1, 2, ..., J + 1$, we find that the *SCM* estimates a positive effect of the intervention (on average, per capita *GDP* would be $30,333.553 higher for each of the 14 post-intervention years). The *PDA* estimates also imply positive effects, $42,321.878 under the *PDA-AICc* and $28,634.32 under the *PDA-LASSO*) [see last row in Table 2.2].[19] The conclusion we see from these results is neat: all the methods considered (*SCM*, *PDA-AICc* and *PDA-LASSO*) predict a positive effect of the intervention on per capita *GDP*, although they provide quantitatively different figures. Were the goodness-of-fit the right criterion to choose our best estimate of the effect (either one of the two discussed above), *PDA-AICc* would give us the answer: an approximate yearly average of $42,321 between 2000 and 2012.

[19] For completeness, Table 2.2 also shows λ_{min}, the value obtained by λ (the tuning parameter of the *PDA-LASSO* estimator) which minimizes the average squared error over all the pre-treatment periods [see Li and Bell (2017), Section 4.2].

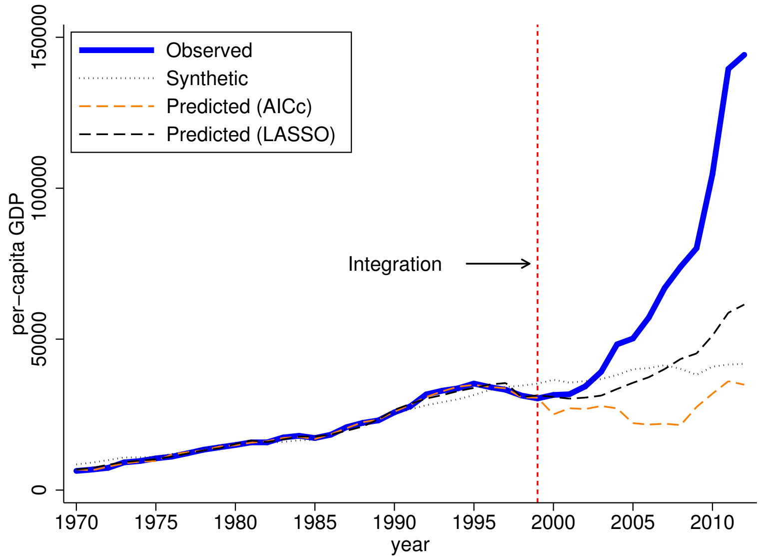
Table 2.2: Per capita GDP: weights and coefficients

Country	***SCM weights***	***PDA-AICc***		***PDA-LASSO***
(Intercept)		7,430.252	(**)	-2,990.999
Indonesia		4.149	(***)	5.125
Japan	0.410			0.372
Malaysia				
Thailand				
United States				0.104
Taiwan	0.590			0.254
Philippines				
Germany		1.561	(***)	
United Kingdom				-0.402
France				
Australia		-0.431	(*)	
Austria				
Brazil		-1.056	(***)	-0.710
Canada				0.160
Denmark				
Finland				
Italy		0.958	(***)	0.032
Mexico				0.031
Netherlands		-1.324	(***)	-0.029
New Zealand		1.516	(***)	
Spain				
Sweden		-0.683	(***)	
Switzerland		-0.529	(***)	
Norway				
AG	$30,333.553	$42,321.878		$28,634.32

Key: Only non-zero values are shown. *SCM* weights are the solution to the program in Eq. (2.2), and *PDA-AICc* and *PDA-LASSO* coefficients are the *OLS* estimates of the model in Eq. (2.3). *AG* refers to average post-treatment gap [see Eq. (2.6)]. Significance levels of *PDA-AICc* coefficient estimates: (***) 0.1%, (**) 1%, and (*) 5%. $\lambda_{min} = 41.34$ for the *LASSO* regression.

The graphical counterpart of Tables 2.1 and 2.2 is shown in Figure 2.2. As mentioned above, it is patent that the matching for the pre-treatment period with the *SCM* is worse than those with the *PDA*, whilst *PDA-AICc* is slightly better than *PDA-LASSO*.

Figure 2.2: Per capita GDP: observed and counterfactuals

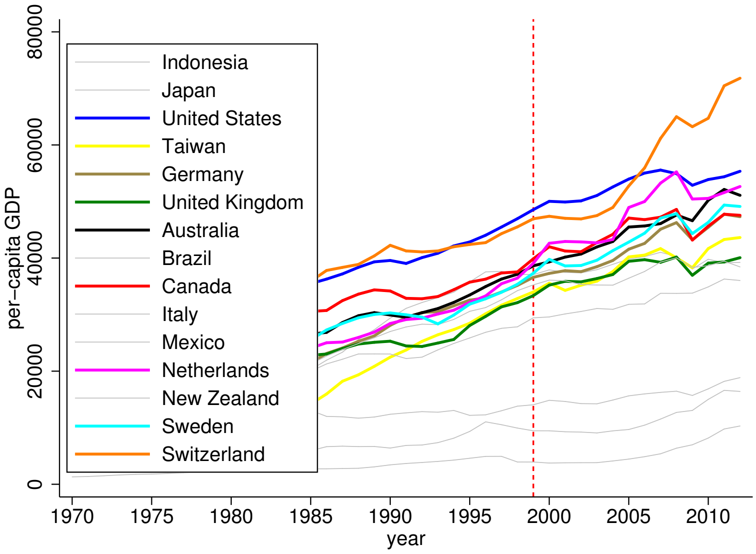


Key: The solid blue line stands for Macao's observed per capita *GDP*. The black dotted line represents the *SCM* counterfactual; the orange dashed line depicts the *PDA-AICc* counterfactual; and the *PDA-LASSO* counterfactual is represented by the black dashed line.

The difference in the results obtained under the two implementations of the *PDA*, or the fact that the model selection criterion plays a role, requires some deeper thinking. First, one should note that both methods provide different controls with non-zero coefficients (even though whenever a control country appears with a non-zero coefficient under both criteria, the sign is the same). Why such a disparity and, in particular, why is the counterfactual estimate generated by the *LASSO* regression substantially higher, and smoother, than the one generated by the *AICc* regression? Second, this would lead us to study in detail the differential role that some control countries might be playing. Figure 2.3 displays the series of per capita *GDP* for the 15 countries which appear under at least one of the two model selection criteria considered, and in which some specific countries are highlighted.

Switzerland, the US and Australia, for instance, display some of the highest levels of *pcGDP* during the post-treatment period within this subset. It turns out that the *LASSO* coefficient is positive for the US (and zero for Switzerland and Australia) and that the *AICc* coefficient is negative for Switzerland and Australia (and zero for the USA). The case is similar with Taiwan: high levels of *pcGDP* with a positive coefficient in the *LASSO* regression, and a zero coefficient in the *AICc* counterfactual. Consider now the Netherlands: this control displays a high level of per capita *GDP* and a (lower) negative coefficient in the *AICc* regression than in the *LASSO* regression. All this partly explains why *LASSO* counterfactual is higher than the *AICc* counterfactual. Note that were it not for the negative coefficient of the UK in the *LASSO* regression, the difference between both counterfactuals would have been even larger. Finally, the time evolution of per capita *GDP* for Netherlands, Sweden and Australia and their *AICc* coefficients might also explain why the *AICc* counterfactual evolves less smoothly than the *LASSO* coefficient.

Figure 2.3: PDA-AICc vs PDA-LASSO controls

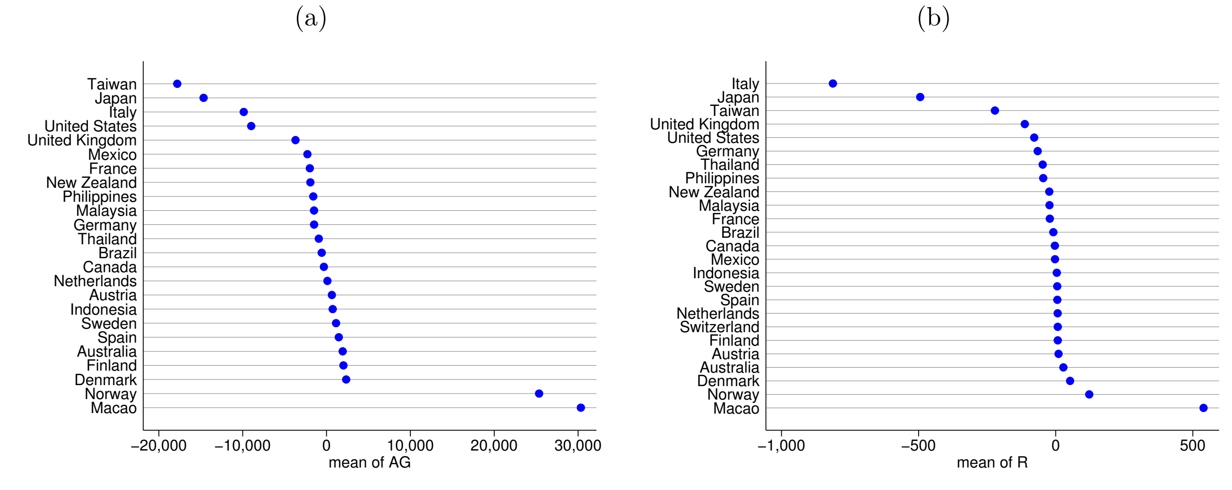


Key: Series of per capita GDP for the 15 countries which appear under at least one of the two selection criteria implemented. Specific countries are shown in color.

We next perform an in-space placebo test, the standard procedure in the related literature to check the statistical significance of our previous estimates. Thus, Macao is moved to the donor pool and then we estimate, sequentially, the effect of the intervention in every other country in the pool. We run this exercise both for the *SCM* and *PDA-AICc*, leaving aside *PDA-LASSO*, in order not to overload the paper. Hence, the exercise allows us to obtain a distribution of effects for the countries which did not experience the same treatment, in such a way that our results would be validated if the probability of finding other significant treatment effects in non-treated countries were low enough. It is worth noting that when performing this exercise, Macao has not been excluded from the donor successive pools, as in Abadie et al. (2010), although the opposite might be a valid alternative.

We first start with the *SCM*. Two types of tests are usually implemented in the literature. The first one focuses on the distribution of the previously defined post treatment average gap (AG). Prior to that, we remove those placebo cases for which the fit in the pre-treatment period is poor (relative to Macao's) otherwise confidence in the results of post-treatment analysis and placebo tests might be reduced [see Abadie et al. (2010) and Acemoglu et al. (2016) among others]. Thus, we are willing to exclude those control countries for which the *RMSE* is at least as high as $\sqrt{2}$ times that of Macao.[20] The results are shown in Figure 2.4, panel (a). Formally, the probability of finding an AG_j higher than or equal to Macao's equals $p = 1/24 = 0.04$. Thus, the null hypothesis of no positive effect of the political treatment upon Macao's per capita *GDP* would be rejected at a 5% significance level.

Figure 2.4: *SCM* and in-space placebo test



Key: Panel (a) represents the distribution of the average post-treatment gap, AG_j, among the 24 countries in the sample whose *RMSE* is less than $\sqrt{2}$ times that of Macao [Eq. (2.6)]: Switzerland is left out of the sample. Panel (b) represents the distribution of ratio R_j across all 25 countries in the sample [Eq. (2.7)].

Complementary to this test, we can consider the entire distribution of the *signed* ratios of post- to pre-treatment mean squared prediction error, R_j, which is defined as

$$R_j \equiv \frac{\sum_{t=T_0+1}^{T}(Y_{j,t} - \widehat{Y}^0_{j,t})^2/(T-T_0)}{\sum_{t=1}^{T_0}(Y_{j,t} - \widehat{Y}^0_{j,t})^2/T_0} \times \frac{AG_j}{|AG_j|}. \tag{2.7}$$

Note, first, that this time no control country has been left out, as a strong enough post-treatment effect can more than offset a poor fit during the pre-treatment period. Second,

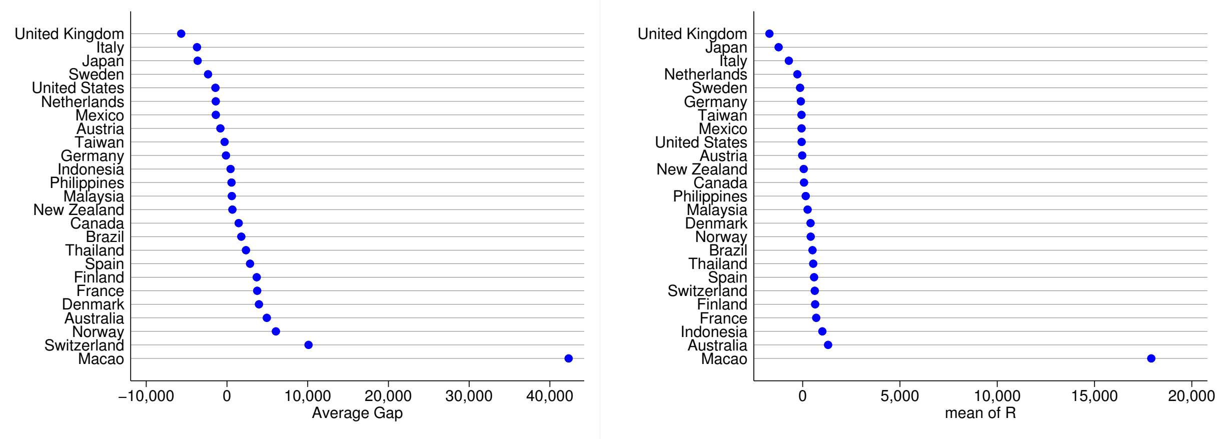
[20] Only one control is lost (namely, Switzerland) with the *SCM*, but none with the *PDA-AICc*.

and most important, the sign of R_j depends on the sign of AG_j, thereby allowing for a one-sided test, which may result in a substantial gain of power [see Abadie (2021)]. The result is shown in Figure 2.4, panel (b). On this occasion Macao ranks first as well, so that the probability of finding an effect larger than or equal to Macao's equals $p = 1/25 = 0.04$.

We next proceed to carry out the same exercise, this time implementing the *PDA-AICc* method and the result is shown in Figure 2.5. Regarding the distribution of AG_j, all placebo runs obtain a good fit, so that none are disposed of as a result of applying the above mentioned $\sqrt{2}$ rule. Macao exhibits the highest value: in other words, the probability of finding an effect larger than or equal to Macao's equals $p = 1/25 = 0.04$ once again [see panel (a)]. The same result is obtained after computing the empirical distribution of R_j [see panel (b)]. The consequence after this placebo test run is clear: regardless of the method implemented to compute the counterfactual of per capita *GDP* in the absence of integration, the null hypothesis of no effect of the intervention on Macao's per capita *GDP* is rejected.

Figure 2.5: *PDA-AICc* and in-space placebo test

(a) (b)



Key: Panel (a) represents the distribution of the average post-treatment gap, AG_j, among the 24 countries in the sample whose *RMSE* is less than $\sqrt{2}$ times that of Macao [Eq. (2.6)]. Panel (b) represents the distribution of ratio R_j across all 24 countries in the sample [Eq. (2.7)].

For completeness, Table 2.3 details the predicted effects of Macao's integration i) on the per capita *GDP* (in terms of both levels and growth rates), ii) with the *SCM* and the *PDA* methods implemented here to compute the counterfactual (and, in the latter case, with the two model selection criteria described above, *AICc* and *LASSO*), and iii) for each year in the post-treatment period.

Thus, the *accumulated* difference in the per capita *GDP* growth rates between the actual or observed Macao and its counterfactual ranges between 360.9% (*PDA-AICc*) and 278.9% (*PDA*-LASSO). The figure for the *SCM* is closer to the former, 356.1%. In terms of average *annual* growth rates for each of the 10 years such numbers would mean 11.4%, 7.2% and 12.4% respectively. The other side of the coin is the effect in terms of increment in per capita *GDP,* where (necessarily) results differ depending on the particular procedure followed to obtain the counterfactual. Had it not been for the integration process, the *accumulated* per capita real *GDP* in the period 2000–2012 would have been between 550,184.4 dollars lower (42,321.9 dollars per year on average) in the *PDA-AICc* case, and 372,246.1 dollars lower (28,634.3 dollars per year on average) in the *PDA-LASSO* case.

Table 2.3: Effect on per capita *GDP* levels and growth rates.

	SCM		***PDA-AICc***		***PDA-LASSO***	
Year	Δg^S	$\Delta\, pcGDP^S$	Δg^P_{AICc}	$\Delta\, pcGDP^P_{AICc}$	Δg^P_{LASSO}	$\Delta\, pcGDP^P_{LASSO}$
2000	0.2	-5,005.3	21.8	6,455.2	5.5	669.1
2001	3.2	-3,848.2	-7.4	4,643.6	2.4	1,442.0
2002	6.7	-1,745.7	9.1	7,507.0	7.1	3,717.0
2003	12.2	2,426.1	10.2	11,316.7	11.8	7,881.8
2004	19.4	10,139.0	26.0	21,217.1	16.2	14,795.8
2005	-0.9	10,170.4	22.0	28,046.5	-2.1	14,670.7
2006	12.9	16,761.9	15.9	35,457.8	8.6	19,781.4
2007	14.7	25,594.0	15.8	44,998.4	9.8	26,873.4
2008	13.4	33,850.4	12.6	52,476.1	2.1	30,564.5
2009	13.1	41,932.1	-19.5	52,618.7	4.0	34,864.5
2010	23.5	63,760.0	14.5	72,719.2	17.5	53,486.4
2011	31.6	97,933.8	20.4	103,456.4	18.5	80,771.7
2012	2.7	102,367.4	6.5	109,271.1	-1.2	82,727.2
Accumulated	356.1	394,336.2	360.9	550,184.4	278.9	372,246.1
AG		30,333.6		42,321.9		28,634.3
$\widehat{\gamma}_1$	12.4		11.4		7.2	
t-statistic	16.0		14.8		22.8	

Key: Δg^S: difference in growth rates between actual and synthetic Macao; $\Delta\, pcGDP^S$: gap in per capita *GDP* according to *SCM*; Δg^P_{AICc}: difference in growth rates between actual Macao and predicted Macao according to *PDA-AICc*; $\Delta\, pcGDP^P_{AICc}$: gap in per capita *GDP* according to *PDA-AICc*; Δg^P_{LASSO}: difference in growth rates between actual Macao and predicted Macao according to *PDA-LASSO*; $\Delta\, pcGDP^P_{LASSO}$: gap in per capita *GDP* according to *PDA-LASSO*. *Accumulated*: accumulated $\Delta pcGDP$ for all the post-treatment period. *AG*: yearly average of $\Delta pcGDP$. $\hat{\gamma}_1$: annual average difference of per capita *GDP* growth rates [see see Eq. (2.9)], and t-statistic (computed with robust standard errors). All growth rates are in per cent terms. Per capita *GDP*, as in the rest of the paper, measured in PPP 2017 US dollars.

An issue that is open to question here is that of the extent of the long-run effects of integration and its statistical significance along the lines of Hsiao et al. (2012). Thus, defining $\widehat{\Delta}Y_{1,t} \equiv Y_{1,t} - \widehat{Y}^0_{1,t}$, for $t = T_0 + 1, T_0 + 2, ..., T$, we could fit the following $AR(p)$ model for the estimated treatment effects,

$$\widehat{\Delta}Y_{1,t} = a + \sum_{i=1}^{p} b_i \widehat{\Delta}Y_{1,t-i} + \eta_{1,t}, \tag{2.8}$$

where $\eta_{1,t}$ denotes an i.i.d. random term with $E(\eta_{1,t}) = 0$, and where the maximum order of such autoregressive process would depend on the number of observations in the post-treatment period, $T - T_0$. Note that, by the very definition of long run, one would have that $\widehat{\Delta Y}_{1,t} = \widehat{\Delta Y}_{1,t-1} = \ldots = \widehat{\Delta Y}_{1,t-p} \equiv \widehat{\Delta Y}_1^{LR}$. This would allow us in turn to easily compute the long-run effect of the intervention just by applying the delta method in a straightforward manner to estimate $\widehat{\Delta Y}_1^{LR} \equiv \widehat{a} / \left(1 - \sum_{i=1}^{p} \widehat{b}_i\right)$.[21] For instance, in the case of the *PDA-AICc* (where the pre-treatment fit attained is the highest), and assuming an $AR(2)$, we obtain

$$\widehat{\Delta Y}_{1,t} = \underset{(2413.4062)}{4997.6422} + \underset{(0.4564)}{0.9122} \ \widehat{\Delta Y}_{1,t-1} + \underset{(0.5372)}{0.2601} \ \widehat{\Delta Y}_{1,t-2} + \widehat{\eta}_{1,t},$$

where the high estimated robust standard errors shown in parentheses imply that neither of the coefficients is statistically significant. The corresponding long-run effect would be -28994.547, with a t-statistic equal to -0.9465. Not surprisingly, this would not be statistically significant either.

The intuition is simple: the gap between the observed series of per capita *GDP* and any of the counterfactuals considered is far from being constant, but increasing [see Figure 2.2]. Therefore, it would make sense to look for the long-run effect not in the level, but in the *growth rate* of per capita *GDP*. Thus, assuming constant yearly growth rates in the post-treatment period for the observed and the counterfactual series (respectively g and $\widehat{g}^0$), so that $Y_{1,t}^1 = Y_{1,0}^1(1 + g^1)^t$ and $\widehat{Y}_{1,t}^0 = Y_{1,0}^0(1 + \widehat{g}^0)^t$ for $t = T_0 + 1, T_0 + 2, ..., T$, we estimate the following model

$$\widehat{\Delta} y_{1,t} = \gamma_0 + \gamma_1 t + \kappa_{1,t}, \tag{2.9}$$

where $\widehat{\Delta} y_{1,t} \equiv \ln Y_{1,t}^1 - \ln \widehat{Y}_{1,t}^0$, $\gamma_0 \equiv \ln Y_{1,0}^1 - \ln Y_{1,0}^0$, and $\gamma_1 \equiv \ln(1+g^1) - \ln(1+\widehat{g}^0) \approx g^1 - \widehat{g}^0$, and where $\kappa_{1,t}$ is an i.i.d. random term with $E(\kappa_{1,t}) = 0$. The results are shown in the last 2 rows of Table 2.3. The statistically significant estimates of γ_1 range between 12.0% under the *SCM*, 7.2% under the *PDA-LASSO* method, and 11.4% under the *PDA-AICc* method.

Some conclusions can be derived from the discussion in this Section. First, the integration process of Macao into the PRC did have a statistically significant, positive treatment effect on both the level of per capita *GDP* and its long-run growth rate for the period 2000-2012. The qualitative aspect of the result is robust to the method followed to compute the counterfactual (whether *SCM* or *PDA*). The estimates of average gap (in the

[21] Computations for the delta method have been carried out with the R **car** package, Version 3.0-10 [see CRAN].

level of per capita *GDP*) and the growth gaps depend on the method chosen to compute the counterfactual, however. Adopting the *PDA-AICc* as the method implemented to assess the counterfactual (on the grounds that its pre-treatment fit outperforms that of the *SCM* and *PDA-LASSO*) we find an approximate yearly average per capita *GDP* gap of $42,322 for the period. Following the same criterion, we find that the annual growth rate of the per capita *GDP* was 11.4% higher than that theoretically observed in the absence of treatment.

2.6 Other outcome variables

One might wonder what the driving forces behind the observed effect are on per capita *GDP*. As mentioned in the Introduction of this paper, foreign direct investment skyrocketed in the years following the start of the integration: the average annual per capita foreign direct investment for the period 2000-2012 was 80,594% higher than that of the period 1982-1999. Similarly, the unemployment rate fell from 6.3% in 1999 to 1.97% in 2012, and per capita exports and imports of goods and services increased by 462.7% and 177.6%, respectively, in the same period. Thus, to support the results in Section 2.5, next we consider the effects of Macao's integration process on the three aforementioned economic indicators as additional variable outcomes i.e. 1) net inflows of foreign investment, 2) the unemployment rate, and 3) exports and imports of goods and services. A variable which, in principle, might be worth considering is that of travel and tourism, another potential engine of economic activity. Unfortunately, the series available for travel and tourism total contribution to *GDP* starts in 1996 [see The World Bank], which rules out the possibility of having a reasonable pre-treatment period and, therefore, a reliable estimate of the counterfactual.

Per capita Foreign Direct Investment. As regards per capita net inflows of foreign investment, Taiwan drops out of the sample due to the lack of data, thereby reducing the number of untreated countries to 23.[22] The main results are shown in Table 2.4. Regarding the fit, the $\widetilde{R}^2$ for the *SCM* and the *PDA-LASSO* procedures are extremely poor (-0.162 and 0.000 respectively).[23] Following the *PDA-AICc* method, instead, the fit seems quite acceptable as $\widetilde{R}^2 = 0.978$, which leads us to a predicted annual average gap for per capita net inflow of foreign direct investment for the post-treatment period of $6,358. Graphically, the results are shown in Figure 2.6

[22]Data for the variable have been obtained from Penn World Table 10.0 and The World Bank; the latter does not provide this information for Taiwan.

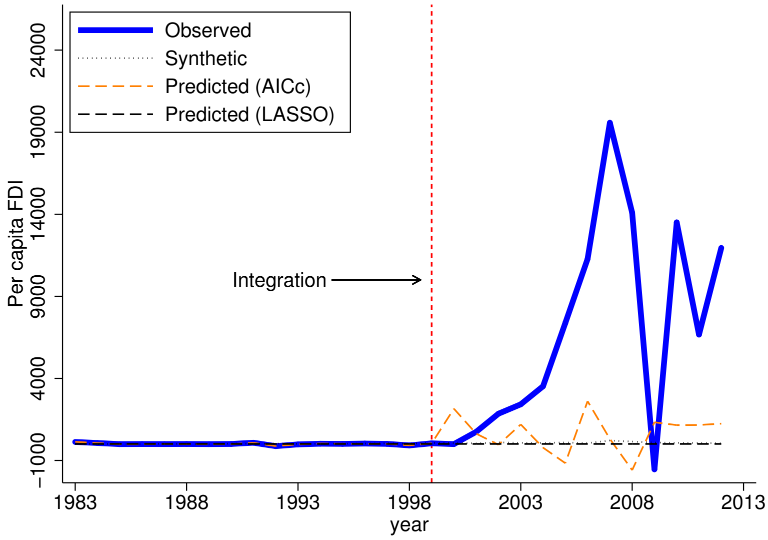
[23]In other words, the *SCM* fit is worse than the one the simple pre-treatment average gives, and the *PDA-LASSO* exactly equals the latter.

Table 2.4: Per capita FDI: weights and coefficients

Country	SCM weights	PDA-AICc		PDA-LASSO
(Intercept)		110.349	(***)	8.715
Indonesia	0.174			
Japan	0.700	-4.417	(***)	
Malaysia				
Thailand		0.217	(***)	
United States				
Philippines				
Germany				
United Kingdom		0.420	(***)	
France		-0.235	(***)	
Australia				
Austria				
Brazil		-0.422	(***)	
Canada				
Denmark				
Finland				
Italy	0.126			
Mexico		1.523	(***)	
Netherlands				
New Zealand				
Spain		-0.236	(***)	
Sweden				
Switzerland				
Norway		0.206	(***)	
RMSE	59.545	8.153		55.231
$\widetilde{R}^2$	-0.162	0.978		0.000
AG	$6,945.000	$6,358.200		$7,023.800

Key: Only non-zero values are shown. *SCM* weights are the solution to the program in Eq. (2.2), and *PDA-AICc* and *PDA-LASSO* coefficients are the *OLS* estimates of the model in Eq. (2.3). *RMSE*: see footnote No. 16. $\widetilde{R}^2$: see Eq. (2.5). *AG* refers to average post-treatment gap [see Eq. (2.6)]. Significance levels of *PDA-AICc* coefficient estimates: (***) 0.1%, (**) 1%, and (*) 5%. λ_{min} = 24.95 for the *LASSO* regression.

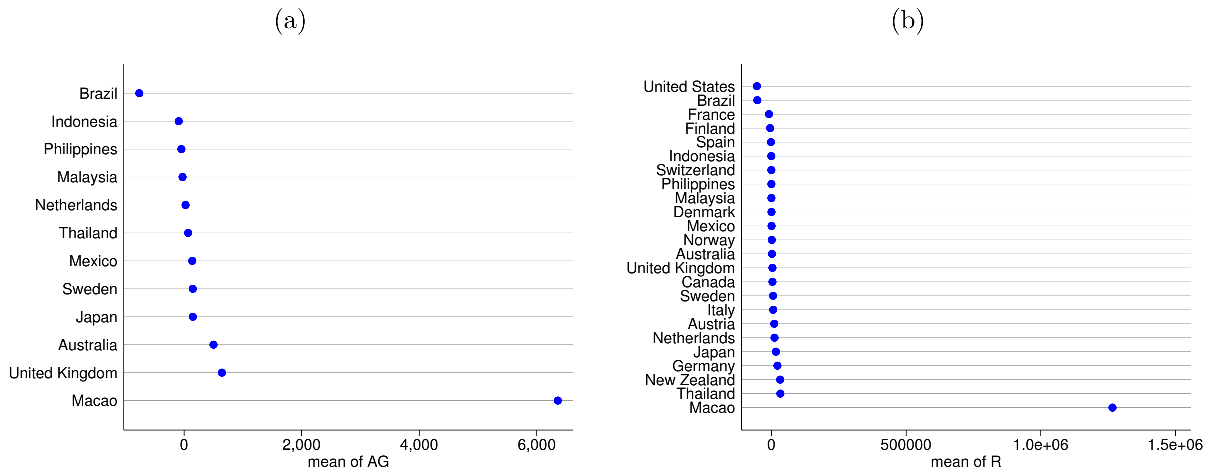
Figure 2.6: Per capita FDI: observed and counterfactuals



Key: The solid blue line depicts Macao's observed per capita FDI. The black dotted line represents the *SCM* counterfactual; the orange dashed line refers to the *PDA-AICc* counterfactual; and the *PDA-LASSO* counterfactual is shown by the black dashed line.

Regarding the in-space placebo robustness check, the procedure followed is the same as in the case of per capita *GDP*, although on this occasion (given the previous discussion) we will only focus on the *PDA-AICc* results [see Figure 2.7]. Based on the distribution of the average post-treatment gap, AG, the probability of finding an effect larger than or equal to Macao's equals $p = 1/12 = 0.08$ (this time, 12 countries drop out as a result of a poor fit). Looking at the distribution of the ratio R, instead, the probability of finding an effect larger than or equal to Macao's equals $p = 1/24 = 0.04$.

Figure 2.7: Per capita FDI: PDA (AICc) and in-space placebo test



Key: Panel (a) represents the distribution of the average post-treatment gap, AG_j, among the 12 countries in the sample whose *RMSE* is less than $\sqrt{2}$ times that of Macao [Eq. (2.6)]. The sample reduces by 50%: United States, Spain, Switzerland, France, Finland, Denmark, Norway, Canada, Austria, Germany, New Zealand and Italy have been removed. Panel (b) represents the distribution of ratio R_j across all 24 countries in the sample [Eq. (2.7)].

Unemployment rate. The numerical results for the effect on the unemployment rate are shown in Table 2.5: the sample data refer to the same 25 countries considered when studying the effect on per capita *GDP*, although the sample period has been reduced: the pre-treatment period is 21 years shorter, as it starts in 1991, which seriously compromises the reliability of the counterfactual and, therefore, the assessment of the effect.[24] As in all the previous calculations, the best fit is obtained with the *PDA-AICc* ($\widetilde{R}^2 = 0.968$), implying a yearly average difference between the unemployment rate observed and the counterfactual of -1.392%, in between those obtained with the other methods.

[24]Data for the unemployment rate have been obtained from The World Bank for all the countries in the sample, except for Taiwan which have been obtained from National Statistics (Republic of China) Taiwan.

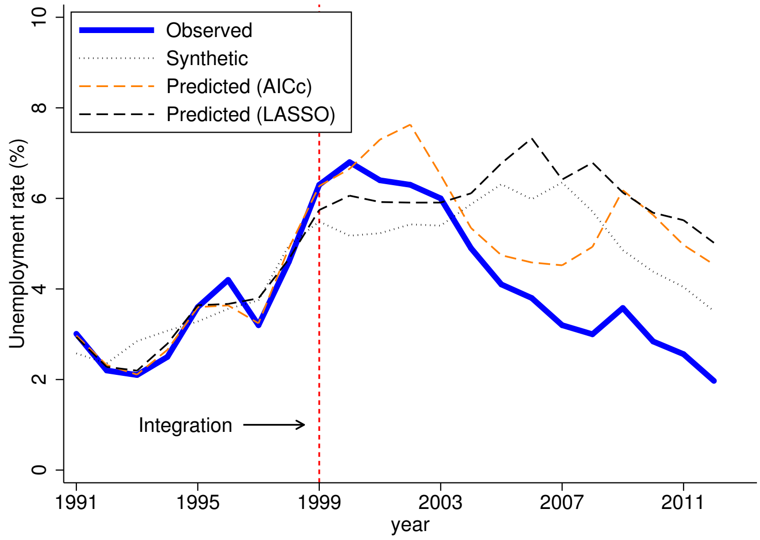
Table 2.5: Unemployment rate: weights and coefficients

Country	**SCM weights**	**PDA-AICc**		**PDA-LASSO**
(Intercept)		0.834		2.3019
Australia				-0.466
Austria				
Brazil	0.004			
Canada				
Switzerland		-0.791	(**)	
Germany				
Denmark				
Spain				-0.069
Finland				
France				
United Kingdom				
Indonesia	0.748			
Italy				
Japan		1.687	(***)	0.169
Mexico				
Malaysia				
Netherlands				
Norway				
New Zealand				
Philippines				1.74
Sweden				
Thailand	0.248			0.074
United States				
Taiwan				
RMSE	0.548	0.224		0.342
$\widetilde{R}^2$	0.814	0.968		0.927
AG (%)	-0.984	-1.392		-1.855

Key: Only non-zero values are shown. *SCM* weights are the solution to the program in Eq. (2.2), and *PDA-AICc* and *PDA-LASSO* coefficients are the *OLS* estimates of the model in Eq. (2.3). *RMSE*: see footnote No. 16. $\widetilde{R}^2$: see Eq. (2.5). *AG* refers to average post-treatment gap [see Eq. (2.6)]. Significance levels of *PDA-AICc* coefficient estimates: (***) 0.1%, (**) 1%, and (*) 5%. λ_{min} = 0.1265 for the *LASSO* regression.

The graphical representation of the observed, synthetic and predicted series is illustrative: the patterns displayed by the three counterfactuals are quite dissimilar, and even the predicted sign of the effect during the first years of the post-treatment periods (until 2003) differs [see Figure 2.8].

Figure 2.8: Unemployment rate: observed and counterfactuals



Key: As in all previous cases, the solid blue line depicts Macao's observed outcome variable (the unemployment rate in this case). The black dotted line refers to the *SCM* counterfactual; the orange dashed line shows the *PDA-AICc* counterfactual; and the *PDA-LASSO* counterfactual is represented by the black dashed line.

Finally, regarding the statistical significance of the seeming reduction in the unemployment rate (and focusing on the *PDA-AICc* results), we have conducted the same in-space placebo analysis as with the previous outcome variables considered. Regarding the distribution of the average gap, AG_j, we find that the probability of finding a reduction in the unemployment rate at least as large as that in Macao equals $p = 7/25 = 0.28$; and regarding the distribution of the ratio R_j, the probability of attaining a value less than or equal to that in Macao equals $p = 15/24 = 0.625$.[25]

Per capita Exports and Imports. Finally, we next address the assessment of the effect of the integration on exports and imports of goods and services. We first tried the same set of countries in the donor pool as in the case of *GDP*, and the original data base, Penn World Table 10.0, which contains data on the shares of imports and exports of merchandises *alone*. That is to say, it provides neither data on services nor on merchandises *plus* services. Additionally, the pre-treatment fit obtained with the three methods under consideration was quite poor. This led us to the WDI DataBank which does contain data on exports and imports of goods *and* services (as shares of *GDP*), but which does not show data for Taiwan, thereby reducing the donor pool to 23 countries

[25]The figure is suppressed for the sake of brevity.

as in the case of foreign direct investment, the sample period starting in 1982.[26] A summary of the findings obtained follows.

Regarding per capita exports of goods and services, our next outcome variable, the results are shown in Table 2.6. The fit provided by the *SCM* is quite poor ($\widetilde{R}^2$= -4.723): the restriction imposed by the method, so that the treated unit must be within the convex hull of the control units, prevents a good fit in cases like this where the treated unit displays higher values of the outcome variable than the controls. More precisely, synthetic Macao is made up of one single control, Switzerland. Once again, the best fit is reached by using the *PDA-AICc* method, with $\widetilde{R}^2$=0.960, and even with a lower number of non-zero coefficients than with the *PDA-LASSO* procedure ($\widetilde{R}^2$=0.937). These two methods provide different estimates for average difference between the observed and the predicted series, *AG*: \$42,115.421 and \$37,025.770 respectively. See panel (a) in Figure 2.9 for a graphical counterpart of Table 2.6.[27]

[26]More precisely, we used the following sources to compute the series of per capita exports and imports of goods and services: export share of *GDP* (The World Bank), import share of *GDP* (The World Bank), *GDP* (Penn World Table 10.0) and population (The World Bank). The sample period begins in 1982, the first year for which we have data on export and import *GDP* shares. In order to lengthen the pre-treatment period as much as possible, the source of data on *GDP* is Penn World Table 10.0, not The World Bank, because the corresponding series in the latter starts eight years later. Recall from Section 2.4 that *GDP* is measured at chained PPPs 2017US\$.

[27]The graphs are omitted, but it could be shown that (were the result obtained with the *PDA-AICc* method reliable) the probabilities of finding a larger effect on exports than that of Macao would be $p = 1/24 = 0.04$ and $p = 9/24 = 0.25$ according to the distributions of AG_j and R_j respectively.

Table 2.6: Per capita exports of goods & services: weights and coefficients

Country	**SCM weights**	**PDA-AICc**		**PDA-LASSO**
(Intercept)		9957.920	(***)	17815.876
Australia				
Austria				
Brazil				3.816
Canada				
Switzerland	1.000			
Germany				
Denmark				
Spain				
Finland		-1.241	(***)	-0.836
France				-0.688
United Kingdom				
Indonesia				0.854
Italy				
Japan				-2.077
Mexico		1.941	(**)	0.238
Malaysia				
Netherlands		-1.076	(***)	
Norway				
New Zealand		2.623	(***)	
Philippines				
Sweden				
Thailand				
United States		4.934	(***)	5.023
RMSE	6,348.056	529.276		668.205
$\widetilde{R}^2$	-4.723	0.960		0.937
AG	\$27,504.031	\$42,115.421		\$37,025.770

Key: Only non-zero values are shown. *SCM* weights are the solution to the program in Eq. (2.2), and *PDA-AICc* and *PDA-LASSO* coefficients are the *OLS* estimates of the model in Eq. (2.3). *RMSE*: see footnote No. 16. $\widetilde{R}^2$: see Eq. (2.5). *AG* refers to average post-treatment gap [see Eq. (2.6)]. Significance levels of *PDA-AICc* coefficient estimates: (***) 0.1%, (**) 1%, and (*) 5%. λ_{min} = 46.2 for the *LASSO* regression.

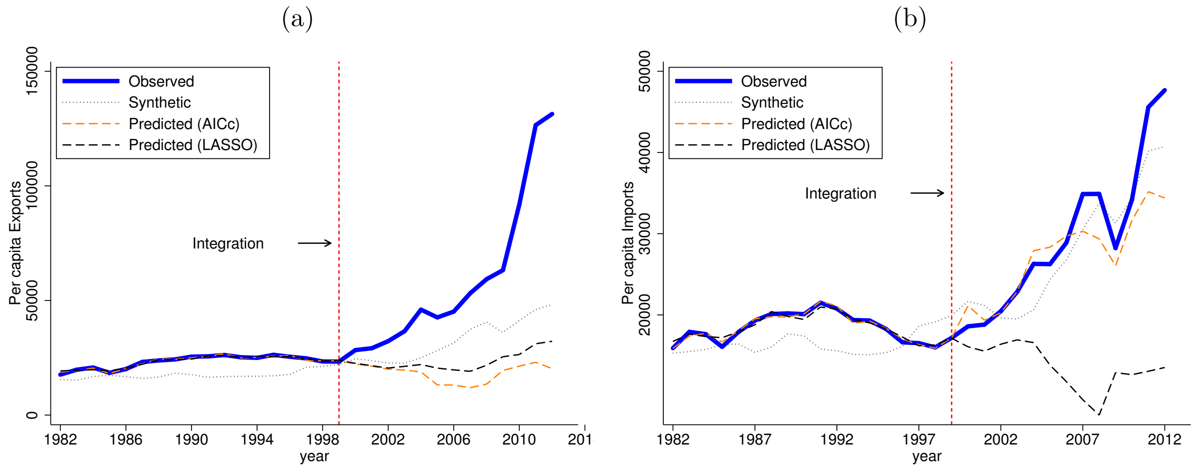
As for per capita imports, the last outcome variable studied in this paper, the results are shown in Table 2.7. As is the case with exports, the goodness of fit that results after applying the *SCM* is extremely low ($\widetilde{R}^2$=-2.428), so that any single attempt to draw any kind of assessment regarding the effect of the treatment would be pointless. Both the *PDA-AICc* and the *PDA-LASSO* methods provide better fits for the pre-treatment period ($\widetilde{R}^2$=0.970 and $\widetilde{R}^2$=0.924 respectively). Contrary to the export case, however, the predicted effects (as measured by *AG* in Eq. (2.6)) are quite different (\$2,389.713 and \$16,237.920 respectively), so that once again it does not seem meaningful to draw any conclusions, even as an approximation, as regards the extent of the effect on imports [see panel (*b*) in Figure 2.9].

Table 2.7: Per capita imports of goods & services: weights and coefficients

Country	**SCM weights**	**PDA-AICc**		**PDA-LASSO**
(Intercept)		19,916.000	(***)	27,563.924
Australia		-1.361	(*)	
Austria		1.337	(***)	
Brazil				-1.449
Canada		-2.542	(***)	
Switzerland	1.000			
Germany				1.091
Denmark				-0.183
Spain				
Finland				-0.164
France				
United Kingdom				
Indonesia				
Italy				-0.809
Japan		-3.627	(***)	-0.712
Mexico				0.878
Malaysia		0.377	(***)	
Netherlands				
Norway				-0.998
New Zealand				-0.106
Philippines				
Sweden				
Thailand				0.425
United States		5.574	(***)	0.567
RMSE	3,189.498	296.783		475.295
$\widetilde{R}^2$	-2.428	0.970		0.924
AG	1,761.588	2,389.713		16,237.920

Key: The same as in Table 2. Only non-zero values are shown. *SCM* weights are the solution to the program in Eq. (2.2), and *PDA-AICc* and *PDA-LASSO* coefficients are the *OLS* estimates of the model in Eq. (2.3). *AG* refers to average post-treatment gap [see Eq. (2.6)]. Significance levels of *PDA-AICc* coefficient estimates: (***) 0.1%, (**) 1%, and (*) 5%. λ_{min} = 40.85 for the *LASSO* regression.

Figure 2.9: Per capita exports and imports: observed and counterfactuals



Key: Panel (a): The solid blue line represents Macao's observed per capita *exports* of goods and services. The black dotted line shows the *SCM* counterfactual; the orange dashed line represents the *PDA-AICc* counterfactual; and the *PDA-LASSO* counterfactual is represented by the black dashed line. Panel (b): As panel (a), but referred to per capita *imports* of goods and services.

Some conclusions can be drawn from the results in this Section. First, there is a statistically significant positive effect on the net inflow of per capita foreign direct investment with an approximate average gap between 2000 and 2012 of $6,358, the *PDA-AICc* being the most reliable method to compute the counterfactual. Second, there seems to be an apparent average reduction in the unemployment rate of about 1.392% in those years. Figures greatly vary depending on the method followed to compute the counterfactual, however. And even focusing on the method for which the best pre-treatment fit is attained (*PDA-AICc*), the effect is not statistically significant. Third, a statistically significant positive effect on the per capita exports of goods and services is obtained, resulting in an approximate average gap in the post-treatment period of $42,115, if one follows the test based on the *AG* distribution. Relying on the *R* distribution, however, the effect would not be statistically significant. As with foreign direct investment, the *PDA-AICc* seems to be the right choice as a method to assess the effect of the intervention, rather than the *SCM*. And, fourth, no clear-cut effect on the per capita imports of goods and services is found: despite both *PDA-AICc* and *PDA-LASSO* attaining reasonably good pre-treatment fits, the positive predicted effects are so distant from each other that even attempting to search for the statistical significance is not recommended.

2.7 Conclusions

In this paper we have assessed the economic effects of the integration process of Macao into China that started on 20 December 1999, until then a Portuguese colony, becoming a Special Administrative Region and the second instance of "one country, two systems" after Hong Kong. The deregulation of gaming in 2002 and the signature of the Closer Economic Partnership Arrangement (CEPA) in 2003 represent two additional milestones in this process of "smooth transition".

As a consequence of this political-economic process, Macao's economy experienced a significant transformation. The average annual growth rate of *GDP* between 1999 and 2016 was about 10.32%, thereby allowing Macao to post the seventh highest per capita *GDP* in the world in 2016. Regarding the net inflow of foreign direct investment in per capita terms, the annual average between 1982 and 1999 was $8.7, and $7,032.5 between 2000 and 2012 (measured at chained PPPs in 2017US$). The process also affected the inflow of foreign visitors: while the annual average of foreign visitors for the period 1995-1999 was 7.5 million, it reached 19.5 million for the period 2000-2012. A reason key to explaining the above economic development was the gaming industry which by 2013 represented more than 60% of Macao's *GDP*.

In this article we have provided a quantitative assessment along the lines of the causal inference literature. We have built up a panel data set, consisting of Macao and 24 control countries, for a sample period starting in 1970 and extending until 2012, and in which 2000 represents the first post-treatment year. In order to compute the counterfactual (or Macao's economic theoretical performance in the absence of intervention), two alternative methodologies have been implemented: the synthetic control method and the panel data approach, the latter displaying the best pre-treatment fit when the corrected Akaike Information Criterion is set as the model selection procedure.

With this data set and these analytical tools, we have assessed the effect on per capita *GDP*, as our main outcome variable of interest, as well as additional outcome variables. A summary of the conclusions follows.

First, the integration process of Macao did have a statistically significant, positive treatment effect on both the level of per capita *GDP* and its long-run growth rate for the period 2000-2012. The qualitative aspect of the result proves robust to the method followed to compute the counterfactual (whether *SCM* or *PDA*). The numerical estimates of average gap in the level of per capita *GDP* and the gap of the growth rate vary with the method chosen to compute the counterfactual, however. In other words, the method

(*SCM* vs *PDA*), and also the model selection criteria (*AICc* vs *LASSO*) matter. Following the best pre-treatment fit criterion (which in our particular exercise has always led us to the *PDA-AICc*), we have found an approximate yearly average per capita *GDP* gap of $42,321 for the referred to period. Similarly, we find a statistically significant difference of 11.4% between the observed growth rate of per capita *GDP* and that of the counterfactual series.

Second, we have also found a statistically significant, positive treatment effect on foreign investment. According to our *PDA-AICc* results, the integration of Macao into China gave rise to an approximate annual average increment of $6,358 in the per capita net inflow of foreign investment during the post-treatment period at issue.

Third, as an additional outcome variable of interest, we have also analyzed the effect of the integration on the unemployment rate. We have found a seemingly negative average effect (-1.392%) or fall in unemployment, but the assessment largely varies with the method implemented to compute the counterfactual. And, even after focusing on the *PDA-AICc* approach, the effect is not statistically significant.

Fourth, in order to underpin our results, we have also considered two more additional outcome variables: exports and imports of goods and services. Regarding the former, we cannot conclude that the effect is statistically significant, as alternative test procedures provide different results. Concerning the latter, the (at first sight) positive effect on per capita imports strongly depends on the method implemented to compute the counterfactual, so that its statistical significance is cast into doubt.

As an overall conclusion, we have learned that the methodological approach implemented in causal inference studies does matter, and that researchers should conduct deep, rigorous sensitivity analyses to test the internal validity of their results. Or, quoting Wan et al. (2018), p. 123, *"We are still only in the process of groping toward the truth, not discovering the truth."*

Chapter 3

The Maltese Single Transferable Vote Experience: A Case Study of Gerrymandering?*

Abstract

Re-drawing the electoral boundaries to provide benefit to one particular political party and thereby damaging the principle of representation in democracy has been a core issue in political science in the recent years. For years social scientists have been advocating the idea of measuring or preventing the potential for damage that may arise from the existence of the above-mentioned redistricting process. Following this discussion, we investigate the possible gerrymandering phenomenon that might have arisen, or whether there are any asymmetries or partisan biases due to boundary delimitation of the electoral constituencies in the case of the Maltese general elections. From the evidence of various statistical tests and simulations, our conclusion is that we find no evidence of gerrymandering in the 2013 and 2017 Maltese general elections.

Keywords: Gerrymandering, redistricting, single transferable vote, Malta.

JEL codes: K16

*The author gratefully appreciates comments from Annick Laruelle, and seminar participants at I Quantitative Finance and Economics Workshop of the University of Basque Country.

3.1 Introduction

Redistricting the electoral constituencies in favor of a certain group or party has always been a critical concept for political scientists, who have sought to minimize the opportunities which could help partisan advantages. The process of altering the borders of electoral districts to give a political party or incumbent a significant advantage is called Gerrymandering (Grofman, 1985). Although gerrymandering can weaken any electoral system implemented, it is also detrimental to the concept of democracy's representativeness. In other words, political maneuvering by gerrymandering limits the integrity of electoral processes in which seat-vote casting is not adequately carried out.

For example, in the United States there are 50 states and several constituencies in each state, so that minor alterations in each electoral district may have a significant chance of influencing the overall outcome. The abundance of electoral districts and the presence of a two-party electoral system make the United States a suitable ground for gerrymandering practices (Wang, 2016a,b). Especially the election of only one candidate from each electoral district, that is, the existence of single-member districts (SMD) further exacerbates this situation.

On the contrary, multi-party systems minimize situations such as gerrymandering, but cannot completely eliminate them (Wong, 2019). Delimiting electoral boundaries in multi-party systems mostly affects the minority group, their concentration in certain regions or their distribution as a minority in different regions making it difficult for them to get representatives elected. Taylor et al. (2003), p. 292, state the importance of the redistribution of parliamentary constituencies: "*Election results are determined not just by the quantity of votes returned to each party but also by the spatial distribution of those votes and the location of constituency boundaries around them.*"

The idea of gerrymandering dates back to 1812, when the governor of Massachusetts, Elbridge Gerry, took advantage of redrawing the electoral boundaries [Cox and Katz (2002)]. Accordingly, the origins of the gerrymandering phenomena derives from the fusion of its name and the imagery of the salamander [see Figure 3.1].[1]

[1]See origins of the word "gerry-mander", BOS. GAZETTE

Figure 3.1: Imaginary graphic of Southern Essex state after redesigning the electoral constituencies.

Key: In reply to the newly constituted Southern Essex Senate district by a Massachusetts legislature to benefit the Democratic-Republican Party, this cartoon was produced in March 1812. The caricature mocked the unusual form of the constituency as a "dragon-monster", which was then attributed to a salamander by federalist journalists.[2]

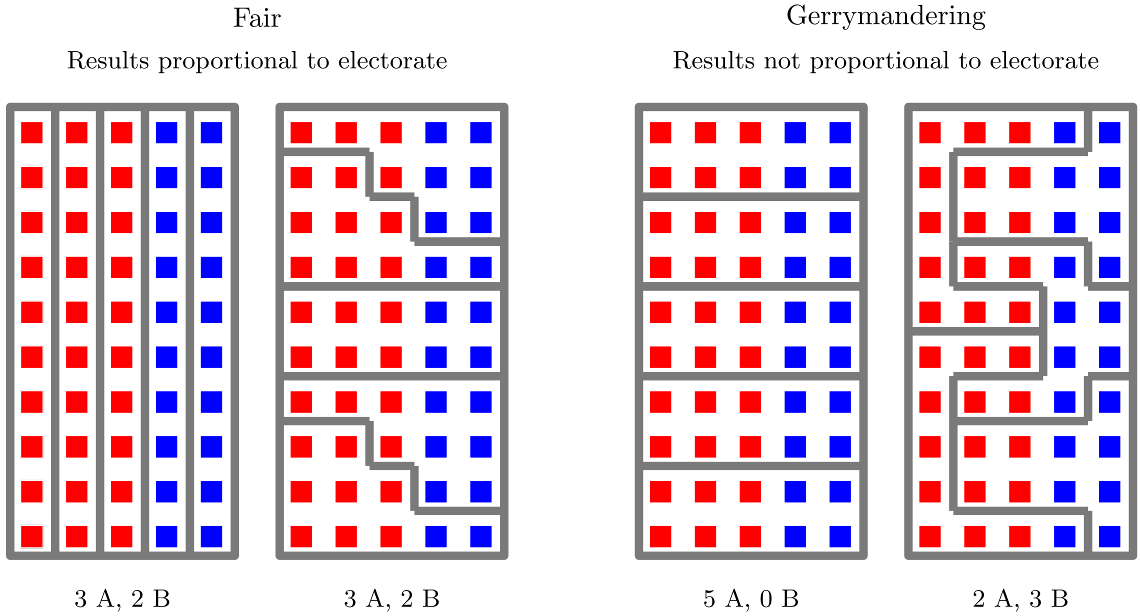
Strategies to change the electoral districts have been commonly used, and the attempts of one party to triumph over the other have continued right up to the present day. For instance, and for the particular case of the US representative democracy, Gelman and King (1994b) conclude that, first, redistricting increases electoral responsiveness. And, second, gerrymandering biases electoral systems in favor of the incumbent party as compared to what would have happened if the other party controlled it, although any type of redistricting reduces partisan bias as compared to an electoral system without redistricting. Nevertheless, such modifications should not be made in a partisan manner to favor a party or an incumbent. The ideal decision should be to strive to ensure that the population scales of the districts match each other.

A simple example in which 50 electors will present their choices between hypothetical political parties A and B, as seen in Figure 3.2, explains why redistricting matters. Suppose that the electoral district scheme will allocate 60% and 40% of the 50 electors

[2] Yet political benefits may still be achieved without the strange or bizarre shape of an election map. In other words, an attractive electoral constituency design can hide malevolent intentions, as in the case of the 1958-1964 Texas congressional redistricting. Consequently, political manipulations may also be conducted on a tidy electoral map [for further information see Bickerstaff (2020)].

in four combinations. In each combination, five electoral districts are designed, one unique candidate being elected in each district by majority. The first and second boxes show that these fifty voters can be evenly allocated on a proportional basis, with the consequence that 3 representatives to party A and 2 representatives to party B can be appointed, respectively. In the third and fourth boxes, however, we observe that the electoral composition is based on a strategy of gerrymandering that favors one party against the other, with five party A representatives and three party B representatives. In order to avoid these disparities and to optimize the fairness of the election outcomes, it is critical to remove the distortions that gerrymanders can cause. Especially in terms of democracy, fairness of the election is critical in the sense of representing the election process and the proportionality of the seats-votes into the parliament, as it allows voters to elect those who rule with their own free will. In addition, concepts such as party favoritism and partisan bias would have been avoided. In the words of Bowler and Grofman (2010), p. 11, *"Elections are, of course, central to conceptions of democracy in the modern era, so they are consequently important in a definitional sense. Some might even argue that elections are the key hallmark of democracy."*

Figure 3.2: Four ways to divide 50 electors into 5 districts



Key: A simple example of how regulation of electoral districts can have distorted outcomes for parties A and B.

In this regard, two of the main tactics for gerrymandering are *i) packing*, which implies that the electoral districts are overloaded with the electors of the opposing party, hence the party will have excessive wasted votes in the electoral districts it wins. Consequently, the party at issue will reduce its chance of winning in other constituencies. Additionally, *ii) cracking* can be described as happening when the rival party's popular vote is distributed over many electoral districts, preventing it from winning the majority in certain electoral districts. This will minimize the chance of the party returning representatives to the congress [Grofman (1985), Wang (2016a,b)].

There have been several papers that investigate possible distortions or gerrymandering scenarios that occurred during redistricting process in various cases. Johnston (2015) investigates how redesigning the electoral districts creates benefits for the Labour Party in the UK, and how conservatives were affected by malapportionment and gerrymandering scenarios in the 2015 general election. Ratto Trabucco (2019) examines the consequences of redrawing electoral constituencies in the Genoa region of Italy in terms of gerrymandering. Giugăl et al. (2017) analyze the malapportionment and gerrymandering for political parties in Romania as a consequence of changing the electoral system with the 2008 electoral reform. Penadés and Santiuste (2013) discuss the advantage rate (the difference between the percentage of votes and the percentage of seats obtained) of political parties in the Spanish electoral system and highlight the advantages or disadvantages of concentrating political parties' votes in certain electoral constituencies. Goderbauer (2016) explores the optimal number of electoral districts in the case of the German Bundestag in order to reduce the average deviation across constituency populations. Wong (2019) studies the gerrymandering scenario in the case of Hong Kong to question the partisan bias at the level of residential buildings. Magar et al. (2017) analyze how the plurality component of Mexico's electoral system gives an advantage to political parties and causes systematic bias in the votes-to-seats translation in congressional elections.

In these debates, which have been going on for years, the need for a quantitative evaluation of gerrymandering applications has been expanded. For instance, and focusing on the US case, Gelman and King (1990) analyze the effects of redistricting as revealed in the votes received by the Democratic and Republican candidates for state legislature, and develop measures of partisan bias and the responsiveness of the composition of the legislature to changes in statewide votes. Their statistical model provides estimates of partisan bias and responsiveness along with measures of their variabilities from only a single year of electoral data. Gelman and King (1994a) emphasize that the reclassification of electoral districts increases responsiveness, whereas gerrymandering can prejudice the electoral system if those who regulate electoral changes re-engineer the electoral system. However, it is also stressed that any improvement in the electoral districts decreases partisan bias relative to areas where no change has been made. Another useful approach is the *efficiency gap (EG)* introduced by Stephanopoulos and McGhee (2015), which analyzes disparity in each party's votes by identifying the difference between the losing and winning parties' wasted votes and dividing them by the total number of votes. If $EG > 0$ for a party, this metric would imply that the party at issue loses more votes than the other, and so the electoral system favors the opposing party. This gap should be close to zero under a fair system and shows that the wasted votes are evenly distributed between the two parties.

The closest precedents to this article are Wang (2016a,b), where they study partisan gerrymandering scenarios in the United States. Wang (2016a) examines possible gerrymandering scenarios with statistical tests in Maryland congressional districts and the Wisconsin State Assembly. The initial point of the study is examining the 1812 Massachusetts State Senate elections, where the original gerrymandering story initiated. With the applied statistical tests, Wang, *op. cit.* reveals how the Democratic-Republican party benefited from redesigning the electoral districts in the 1812 Massachusetts State Senate elections. As a second case, Wang (2016a) investigates post-2010 Maryland congressional districts and whether the 2011 law amendment creates possible partisan gerrymandering scenarios, and concludes that the Democrats increased their seat numbers for the given vote share in comparison with the pre-redistricting. As a final case, Wisconsin State Assembly districts were investigated in a historical context from 1984 to 2014 elections. He reveals that the highest differences in winning vote shares for both parties were in 2012 and 2014 (after 2011 redistricting), indicating that the gerrymandered constituencies were finally in favor of the Republicans.[3]

In this article, we examine the two recent legislative elections of Malta, the last years for which there are available data, in terms of whether the 2017 and 2013 redistricting processes gave any particular political party any advantage.[4] As a starting point, we hypothetically applied different electoral formulas other than the single transferable vote (STV), which are frequently mentioned in the electoral systems literature, both on the condition that the electoral constituencies remain the same, and as one vast constituency (i.e. nationwide). In other words, we questioned whether the election outcome would have changed if other electoral formulas were implemented in the same electoral districts, and if, alternatively, Malta had had a single nationwide constituency. Second, and as a follow-up to Wang (2016a,b), we applied three gerrymandering statistical tests to measure whether there was any partisan gerrymandering scenario or partisan bias in Malta's 2013 and 2017 general elections. After obtaining the necessary information from the Gerrymandering measurements, we carried out simulations in order to detect possible anomalies that might be caused by the redistricting process. We conclude that the 2013 and 2017 general elections of Malta were fair for political parties in terms of reallocation of constituencies and showed no signs of possible gerrymandering or partisan bias.

[3] For detailed information see Princeton gerrymandering project.

[4] As the electoral constituencies of Malta were redesigned before the general elections in 2013, the results of the 2013 general elections are also presented briefly in the following sections.

The rest of the paper is organized as follows. Section 3.2 describes the single transferable vote system for Malta. Section 3.3 describes the data and the detail of the statistical tests carried out in the paper. Results of gerrymandering measurements and statistical tests are discussed in Section 3.4. Section 3.5 concludes. Finally, two formal appendices are included at the end of the paper: Appendix 3.A describes how the single transferable vote system works with a given empirical example, and Appendix 3.B presents the results of the analysis if the 2013 general elections.

3.2 The Single Transferable Vote in Malta

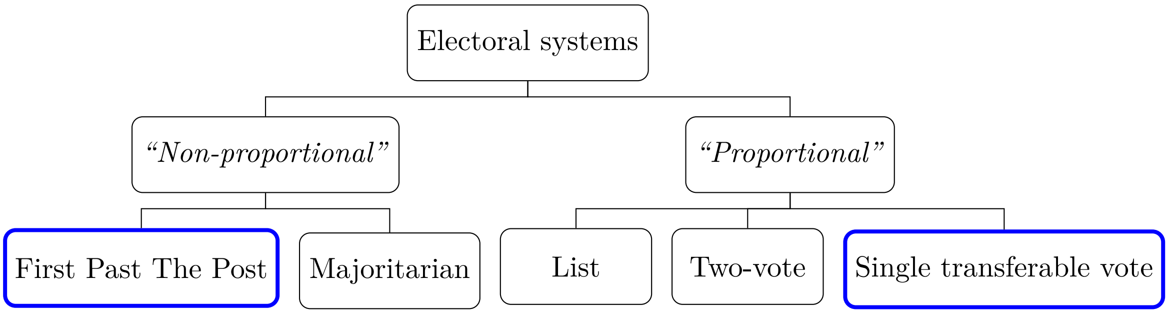
This section is devoted to Malta's electoral system, the single transferable vote, and its place in the world. Why does the Maltese case deserve some attention? First, Malta is good example of *de facto* bipartidism. No parties other than the *Partit Nazzjonalista/Nationalist Party (PN)* and the *Malta Labour Party (MLP)* have obtained representation in parliament since Malta became independent in 1964, so that Malta has proven that it is one of the purest two-party electoral systems in the worldwide [Gallagher (2010)].[5] In this respect, it seems to be a suitable ground for partisan gerrymandering scenarios, and eventually it has proximity to Wang (2016a,b) as referenced in this paper. Second, there is a long held concern in the Maltese political class about the likely existence of gerrymandering in the design of the electoral district map for general elections. See Bickerstaff (2020), De Miño and Lane (1996), Grofman and Lijphart (2003) and De Miño and Lane (2010) among others. Along these lines, and as we will discuss below in this section, two constitutional reforms have been implemented with the purpose of achieving a higher degree of proportionality between the vote and seat distribution in the Maltese House of the Representatives.

Farrell (1997) describes 5 main electoral systems as displayed in Figure 3.3. We will not examine all electoral systems in this section, but it is useful to make some general definitions. Let us first dwell on the reason for the separation of electoral systems as the concepts of *proportionality* and *non-proportionality.* The most common non-proportional electoral systems consist of First Past The Post (FPTP) and majoritarian systems, while proportional systems include list, two-vote and single transferable vote systems. The difference between the two systems is generally based on different objectives. While non-proportional systems often try to consolidate the stability of governments by obtaining a plurality or overall majority from particular electoral districts, proportional systems, on

[5] For convenience, we will refer to the Malta Labour Party and the Nationalist Party as MLP and PN, respectively, throughout the article.

the contrary, aim to obtain the closest number of seats corresponding to the proportion of votes obtained. Generally speaking, the debates on electoral systems are therefore centered on representativeness and stability of the governments formed [Farrell (1997)]. Gerrymandering tests are typically performed for the FPTP system, an example of which is the United States. As we extend these statistical tests to the case of Malta's single transferable vote system, it is also worth describing the electoral systems implemented in both countries.

Figure 3.3: The five main types of electoral systems

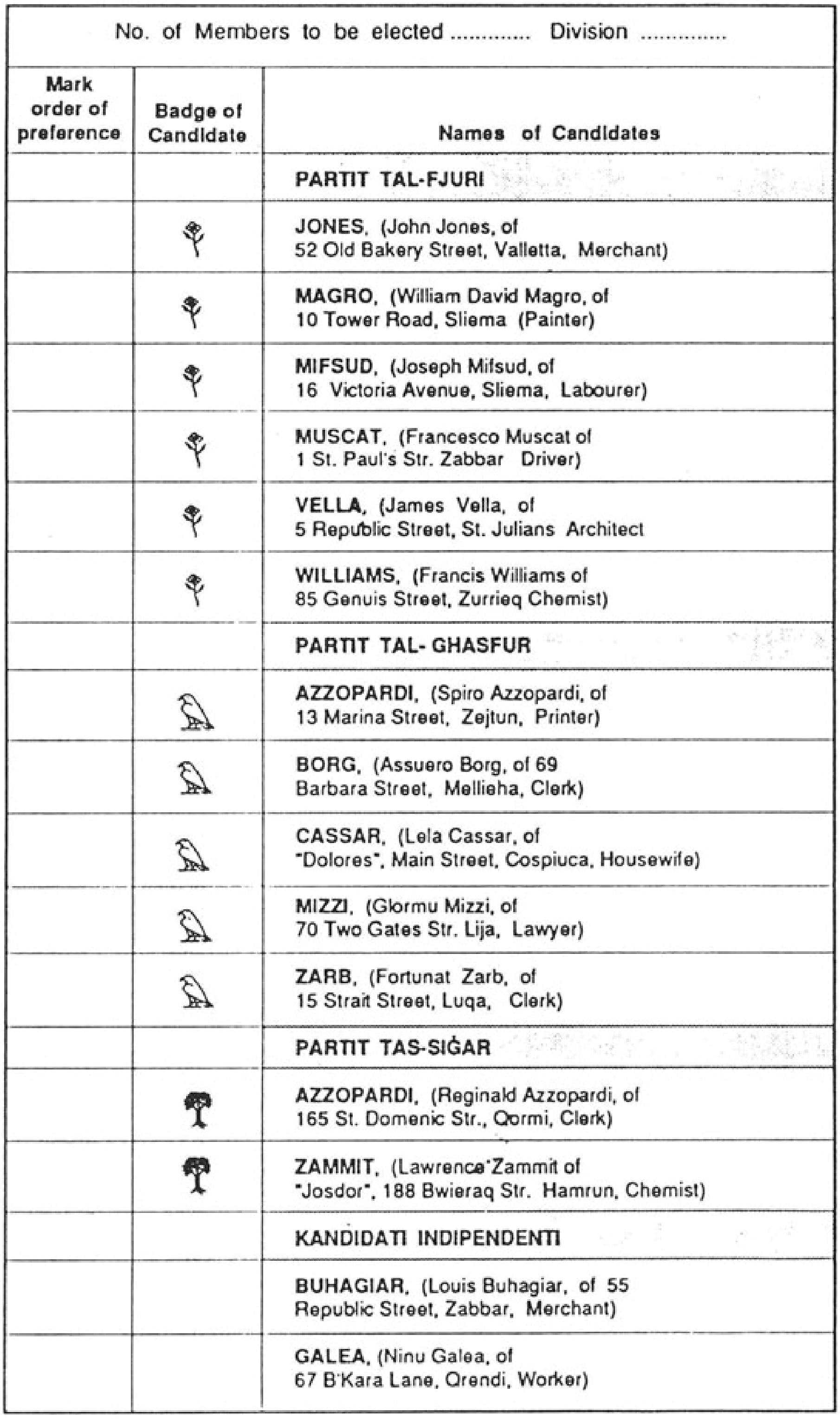


The FPTP system is often used in countries such as the United States, Canada, and Britain. By definition it is based on pluralism, meaning that candidates will be considered elected from a constituency with the highest vote share among their competitors [Diamond and Plattner (2006)]. For example, in the case of the United States, in a two candidate race in a particular constituency, the candidate with superior vote share will be elected as a representative from that electoral district. The rights of representation of minority groups are one of the most critical arguments in this system. More precisely, all the votes of candidates who do not win generate surplus votes while minimizing the opportunity for smaller parties to be represented, resulting in disproportionate election outcomes [Farrell (1997)]. On the other hand, factors such as the single-party government (other than coalition) or the stability of governments are shown as the advantages of the FPTP system.

The single transferable vote (STV) is a proportional representation (PR) system and originates from the 19th century philosophers Thomas Hare and John Stuart Mill. The STV electoral system is generally implemented in countries with small populations such as Ireland and Malta, as well as in the Australian state of Tasmania. It is also used in Cambridge (Massachusetts) to elect school committees in the United States [Tideman (1995)].

The principal reason why the STV is part of proportional electoral systems is that electoral constituencies appoint more than one candidate for a legislative term. In other words, in a given multi-member district (MMD), electors may choose more than one candidate. Thus, the STV differs from other electoral systems such as the FPTP and the majoritarian systems, as it does not feature single-candidate electoral districts. STV's ballot structure and electoral formula are further variances from other electoral systems. STV allows voters to list their preferences on their ballot papers. Among the candidates nominated in each constituency, electors can select their votes as a first, second, or third choice, so ensuring that minority groups are somewhat represented [Bowler and Grofman (2010)]. A simple example of the ballot paper is represented in Figure 3.4, where each candidate is inserted under the name of the party and ordered alphabetically [Farrell (1997), p. 85].

Figure 3.4: Malta single transferable vote ballot paper



No. of Members to be elected .... Division ....

Mark order of preference	Badge of Candidate	Names of Candidates
		PARTIT TAL-FJURI
		JONES, (John Jones, of 52 Old Bakery Street, Valletta, Merchant)
		MAGRO, (William David Magro, of 10 Tower Road, Sliema (Painter)
		MIFSUD, (Joseph Mifsud, of 16 Victoria Avenue, Sliema, Labourer)
		MUSCAT, (Francesco Muscat of 1 St. Paul's Str. Zabbar Driver)
		VELLA, (James Vella, of 5 Republic Street, St. Julians Architect
		WILLIAMS, (Francis Williams of 85 Genuis Street, Zurrieq Chemist)
		PARTIT TAL- GHASFUR
		AZZOPARDI, (Spiro Azzopardi, of 13 Marina Street, Zejtun, Printer)
		BORG, (Assuero Borg, of 69 Barbara Street, Mellieha, Clerk)
		CASSAR, (Lela Cassar, of "Dolores", Main Street, Cospiuca, Housewife)
		MIZZI, (Glormu Mizzi, of 70 Two Gates Str. Lija, Lawyer)
		ZARB, (Fortunat Zarb, of 15 Strait Street, Luqa, Clerk)
		PARTIT TAS-SIĠAR
		AZZOPARDI, (Reginald Azzopardi, of 165 St. Domenic Str., Qormi, Clerk)
		ZAMMIT, (Lawrence Zammit of "Josdor", 188 Bwieraq Str. Hamrun, Chemist)
		KANDIDATI INDIPENDENTI
		BUHAGIAR, (Louis Buhagiar, of 55 Republic Street, Zabbar, Merchant)
		GALEA, (Ninu Galea, of 67 B'Kara Lane, Qrendi, Worker)

As indicated previously, a further distinction between the STV and the other electoral systems is that an electoral formula is needed to determine which candidate is to be elected. A quota (called Droop quota) indicates the minimum amount of first-preference votes needed to select a candidate in a certain constituency. More precisely, the Droop quota, DQ, is defined as

$$DQ \equiv \left[\frac{total\, valid\, votes}{total\, number\, of\, seats + 1} \right] + 1. \tag{3.1}$$

The first candidate from that district to meet the quota is therefore automatically elected when computing the overall number of first-preference votes [Herron et al. (2018)].[6]

Let us consider a basic example below to understand the droop quota better. Suppose that in a constituency with 500 electors, 5 candidates must be elected. Accordingly, the first candidate to receive $(500/(5+1))+1 = 84.3$ votes will be directly elected from that electoral constituency. The process of counting votes continues until the total number of representatives to be elected from that constituency is completed by transferring the surplus votes of the elected candidate to other candidates. That is, according to the candidate's second, third, etc. preferential votes on the ballots, proportional distribution of the votes for candidates will be spread until the last phase. Thus, essentially, the STV system can also minimize the number of wasted votes which is one of the most discussed about issues with regard to electoral systems [Bowler and Grofman (2010), Farrell and Katz (2014)]. Still, the wasted votes concept can also be an issue in all electoral systems, including the STV system. In the STV case, for instance, the wasted vote problem arises in three situations. The first is the total votes of the candidates who did not reach the quota until the last count, but accumulated votes. Second, representatives elected in the last count produce wasted votes by the amount exceeding the quota. Lastly, the sum of the non-transferable votes accumulated until the last count [Mair and Laver (1975)].

In Malta, the PR-STV system has been implemented since 1921 and elects 65 Members of Parliament (MPs) from the 13 multi-member constituencies, five from each, to the unicameral parliament (known as the House of Representatives).[7] The legislative basis for the electoral process is defined by two bodies of legislation: the Constitution of Malta and the General Elections Act of 1991. An Electoral Commission, whose members are appointed by the Prime Minister, establishes the electoral district borders [De Miño and Lane (1996, 2010)]. In these regulation issues, which are frequently mentioned in the electoral systems literature, criteria such as equal population ratios, compactness,

[6] See Appendix 3.A for the explanation of how the STV works in a given example.
[7] We will use the terms MPs and representatives interchangeably.

interests and representation of other minority groups are taken into consideration.

At this point, another important issue is that of the proportionality of the election results. In other words, the popular vote share of a party in the elections and the number of seats obtained should correspond to each other. There is a general consensus amongst scholars on this issue that the district magnitude should consist of at least five representatives in terms of promising proportional results [Mair (2003), Taagepera and Shugart (1989)]. In this context, the STV system implemented by Malta seems to be a fair practice for political parties from a theoretical point of view. Nevertheless, there were some election processes with disproportionate results. The first one was in the 1981 general elections: while the PN had the majority of votes throughout the country (50.9%), it could not achieve the majority to gain office. And the MLP received 49.1% of the first-preference votes, but it secured the majority of seats in the parliament and formed the government. So much so, that this situation suggested that electoral districts were organized according to gerrymandering practices during the MLP government [Grofman and Lijphart (2003)]. After the boycotts of the PN and the political stalemate, the concept of bonus seats was introduced with the 1987 amendment in order to ensure seat-vote proportionality (see Article 52 of the Constitution). Consequently, a decision was made to give additional seats to a party obtaining a majority of first-preference votes in order to ensure the majority of seats in the House of Representatives [De Miño and Lane (2010)]. For instance, 4 bonus seats were given to the PN following the general elections in 1987 and 2008 in order to secure the majority in the Parliament, so that the governments were formed by the PN. Furthermore, 4 bonus seats were granted to the PN in the elections of 2013 and 2 bonus seats in the general elections of 2017, although this time to ensure the proportionality of vote-seat instead of the parliamentary majority.[8]

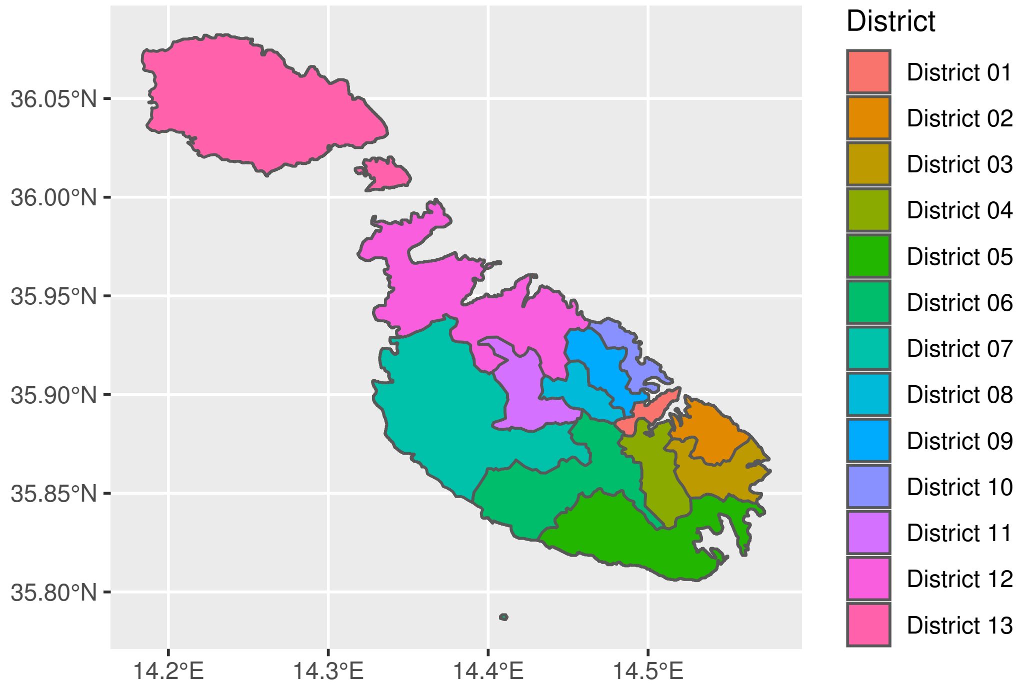
The 13 constituencies of Malta, determined by the 2017 constitutional amendment, are displayed in Figure 3.5.[9] The number of registered voters in the 13 constituencies ranges from 24,884 to 28,680, with a standard deviation of 867.3 and a coefficient of variation of 3.30. The minimal demographic variation in the electoral districts is a prerequisite to first-glance investigation into a gerrymandering situation, thereby preventing potential malapportionment by ensuring that the number of electors does not vary considerably between electoral districts [Wang (2016b), Johnston (2003)].[10]

[8] See data for the bonus seats.

[9] See the details for the 2017 constitutional amendment on redesigning the electoral constituencies.

[10] Malapportionment can be sometimes described as one of the gerrymandering techniques, but this is not the case for the Maltese parliamentary elections as demographics in all electoral districts do not differ much, and are homogeneous.

Figure 3.5: Malta's electoral constituencies



Key: Malta's electoral constituencies as decided for the 2017 legislative elections. There is a total of 13 constituencies, each represented by a different colour.

For example, in the 2017 general elections, the highest quota among the 13 electoral districts was 4,306 votes, and the minimum quota to be reached was 3,847, with a standard deviation of 121.7 and a coefficient of variation of 3.06. While Malta had a total of 8 constituencies in 1921, it was raised to a total of 13 constituencies in 1976 and has continued unchanged until today. For completeness, Table 3.1 provides some descriptive statistics for Malta's 2017 general elections. The second column indicates the total number of registered voters for each constituency, while the third column shows the quota corresponding to the constituencies. The fourth and fifth columns refer to the first-preference votes and the number of seats the political parties obtained, respectively.

Table 3.1: Descriptive statistics for the 2017 Maltese general elections

Constituency	Registered voters	Quota	First-preference votes		Seats obtained	
			MLP	PN	MLP	PN
1	26,598	4,033	13,844	10,094	3	2
2	26,396	4,062	17,353	6,761	4	1
3	25,404	3,894	16,328	6,775	4	1
4	26,095	4,033	16,383	7,528	4	1
5	25,295	3,870	15,259	7,720	3	2
6	24,884	3,847	13,717	9,164	3	2
7	27,106	4,147	14,042	10,509	3	2
8	25,982	3,968	10,830	12,591	2	3
9	25,636	3,853	9,712	13,007	2	3
10	26,460	3,887	8,873	14,058	2	3
11	26,244	3,986	10,282	13,207	2	3
12	27,076	3,899	11,059	11,982	2	3
13	28,680	4,306	13,233	12,361	3	2
Total	341,856	51,785	170,915	135,757	37	28

Key: The results of the 2017 Maltese general elections for each constituency. The number of registered votes in each constituency is shown in the second column, while the third column indicates the quotas that candidates must reach in order to be elected per electoral district. The MLP and the PN votes for the relevant electoral district are represented by columns four and five, and the numbers of parliamentary seats obtained by the parties are displayed in columns six and seven, respectively.

One should note that not every disproportionality necessarily means rigging elections, or any gerrymandering signs, while any disproportional translation of the votes into seats would not mean malapportionment by nature. That is, reshaping of legislative constituencies by jurisdiction is not always intended to gain an advantage for the political parties, or the advantage might have been gained unintentionally [Chen and Rodden (2013)]. In other words, boundary delimitation may yield some advantages for the political parties unintentionally, due to the need to adjust population scales across constituencies, as the population will change in time. Thus, one should bear in mind that investigating the asymmetries in the outcomes of the election should be carried out with care. In the next section, we will describe data and statistical tests that we use to analyze aforementioned disproportionality and asymmetry in the Maltese general election outcomes.

3.3 Data and Statistical Tests

For the purpose of investigating Malta's general elections, we use the data set which contains the election results at electoral constituency level for each candidate and political party between 1960 and 2017 for each general election. We obtained the data we used from the archive of the University of Malta and the Electoral Commission of Malta.[11]

First, we pose two questions; *i*) What results would have been achieved by majoritarian and minority political parties if other electoral formulae had been applied to the 2017 Maltese general elections for the respective constituencies? *ii*) If the boundary restriction had not been valid for the general elections, what would the election outcome have been? That is, we consider the possible consequences of invoking one vast constituency scenario. In these two questions, which have been argued over for a long time by political scientists, the answer to the question of which electoral formula ensures fairer representation is sought. As the implemented electoral system and electoral formulae differ, the resulting election outcome will also vary. The remaining question is whether the alternative seat allocation formulas will grant representation to other parties both in the same electoral districts and in a possible single national constituency. In addition, given the STV outcome of the 2017 general elections, we do not have information based on the electors' second, third, etc. preference votes on the ballot paper. Therefore, one of the reasons for applying the two exercises below is also to validate the closest method of replicating the STV outcome of the general elections, as we use this information in one of the gerrymandering tests that we carry out in the following sections. To put it differently, the mentioned test is based on computer simulations and thus requires an electoral formula that gives the best replication of the STV outcome.

Second, and most importantly, we apply three gerrymandering tests introduced in Wang (2016a,b) for the analysis of general elections. We have investigated whether there was any partisan bias or a possible gerrymandering scenario in the 2017 parliamentary election results. More precisely, we have conducted a first test, *the lopsided-outcomes test*, which examines the constituencies by party and analyzes the two parties' average margins of victory to determine whether any difference between those averages could be attributable to chance. A party which always wins its seats with large margins is probably victim of a gerrymandering technique known as packing. With a second gerrymandering test, *the consistent-advantage test*, we examined the differences between the average seat share and the median seat share of both political parties. With this approach, also known as the skewness test, a possible gerrymandering scenario can be

[11]Parliamentary election data are available at The University of Malta and The Electoral Commission of Malta.

monitored if a party's average seat share is considerably higher than its median seat share, as an asymmetry in the distribution of seat share for that party will be observed. The third and last test (which, as the previous one, will require computer simulations), is called *the excess seat test*. The test aims to check whether one party's share of seats won deviates unexpectedly from national norms. As a consequence, it reveals to what extent the predicted result achieved in the election will vary from actual outcomes of the election. The following sections will first discuss the two questions referred to above, and then apply the above-mentioned gerrymandering tests to the political parties.

3.3.1 Other allocation methods at the district level

Concerning the first question, we tried to find any possible difference regarding the electoral results in each of the thirteen electoral divisions under alternative seat allocation methods other than the STV. Needless to say, it is also important to investigate how non-STV systems can affect the representation of other small parties, as different electoral systems will produce different election outcomes [Cox (1997)]. Consequently, a total of 14 seat allocation formulas were applied to the given vote shares of the political parties in both Matlab©and R programs.[12][13] The gap between the implemented m-th electoral formula and the actual number of the STV seats is given by the mean absolute error. More precisely, the mean absolute error can be defined as

$$MAE_m \equiv \frac{\sum_{d=1}^{13} \sum_{p=1}^{6} |S_{p,d}^{STV} - S_{p,d}^{m}|}{13 \times 6}, \tag{3.2}$$

where $S_{p,d}^{STV}$ represents the actual number of seats obtained by the p-th political party in the d-th district under the single transferable vote system; and $S_{p,d}^{m}$ stands for the seats obtained by the p-th political party in the d-th district under the m-th electoral method. Therefore, the obtained difference, $S_{p,d}^{STV}$- $S_{p,d}^{m}$, will give the deviations from the actual

[12] Computations for the seat allocations have been made with R version 4.0.5 and the **electoral** package, see CRAN. As for Matlab©, the **apport** function has been used for the same purpose of allocating vote shares into seat distribution, see apport.

[13] A total of 14 electoral formulae were applied. However, the conclusion to be drawn from them was far removed from what was obtained under the STV system, thus not included. That is, the obtained errors in Eq. (3.2) were more than those represented in the article. In sum, implemented highest average seat allocation methods are: *Adams, D'Hondt, Danish, Dean, Equal proportions, Hill-Huntington, Imperiali, Modified Sainte-Lague,* and *Sainte-Lague.* Regarding the largest remainders, the performed methods are: *Droop, Hangenbach-Bischoff, Hare, Imperial, and modified Imperial.* For detailed information on the seat allocation methods, see [Herron et al. (2018), Van Eck et al. (2005)].

seats obtained under the STV system. Consequently, this process was applied to all political parties in Malta (in total 6). Regarding the electoral formulas, and for the sake of space saving, only those which attained the lowest MAE in Eq. (3.2) were considered. Table 3.2 presents the results of the 2017 general elections in the same constituencies when the other seat allocation methods are applied.[14] The second and third columns display the actual number of seats obtained under the STV, while the fourth and fifth columns indicate the outcome of the d'Hondt method being applied to votes cast in the same districts. Similarly, columns six and seven show the results achieved with the use of the largest remainder electoral formula, and lastly columns eight and nine show the parliamentary seats that would have been obtained with the implementation of the Sainte-Lague method.

[14]Similarly, outcome of the same experiment for the 2013 general elections can be monitored in Table 3.6 in Appendix 3.B.

Table 3.2: Distinct electoral formulas applied to each electoral constituency

	STV		***D'Hondt***		***Largest Remainder***		***Sainte-Lague***	
District	*MLP*	*PN*	*MLP*	*PN*	*MLP*	*PN*	*MLP*	*PN*
1	3	2	3	2	3	2	3	2
2	4	1	4	1	4	1	4	1
3	4	1	4	1	4	1	4	1
4	4	1	4	1	***3***	***2***	4	1
5	3	2	***4***	***1***	3	2	***4***	***1***
6	3	2	3	2	3	2	3	2
7	3	2	3	2	3	2	3	2
8	2	3	2	3	2	3	2	3
9	2	3	2	3	2	3	2	3
10	2	3	2	3	2	3	2	3
11	2	3	2	3	2	3	2	3
12	2	3	2	3	2	3	2	3
13	3	2	3	2	3	3	3	3
Total	37	28	38	27	36	29	38	27
MAE[15]	-		0.026		0.026		0.026	

Key: In addition to the actual seats obtained under the STV system, implemented electoral formulas d'Hondt, largest remainder, and Sainte-Lague for the given constituency. The last row indicates the mean absolute error as obtained in Eq. (3.2). Finally, the underlined numbers show the differences compared to the actual STV in that constituency. Note that results are presented without taking into account possible bonus seats.

There are differences in quantitative terms only in the fifth constituency as the d'Hondt electoral formula awards the MLP one additional seat given the first-preference votes in that constituency. As for the largest remainder formula with the Hare quota, the PN receives one more parliamentary seat compared with the actual STV results in the fourth electoral district. Finally, the Sainte-Lague electoral formula yields the same results as the d'Hondt. In fact, all these findings are consistent with previous research [Balinski and Young (1984), Lijphart (2012)]. Lijphart (2003) discusses the proportionality of the translation of vote shares into seats, concluding that the highest proportionality can be achieved under the STV, the d'Hondt, the largest remainder, and the Sainte-Lague formulae. He additionally concludes that the d'Hondt electoral formula favors the majoritarian parties in the electoral systems, while the largest remainder gives a representation chance to other smaller parties as well.

[15] There were other seat allocation methods that happened to attain the same MAE [see Eq. (3.2)], such as the Imperiali and the Dean.

The difference in the resulting seat distribution comes from the methods that electoral formulas implement for the given vote shares. That is, as discussed in Section 3.2, STV uses the Droop quota for translating the parties' vote shares into parliamentary seats, whereas the d'Hondt and the Sainte-Lague formulas divide the given vote shares into *1, 2, 3*, etc. divisors (the latter divisors involve the odd integers such as *1, 3, 5*), and allocate parliamentary seats to the political parties with the highest averages. Hence they do not need to specify a quota in order to elect the candidates in a certain district. As for the largest remainder, it uses the so called Hare quota (similar to the STV's Droop quota) for choosing the candidates, so that after electing the candidates in the first counting, the remaining surplus votes are transferred to the other candidates.[16] The main difference between the Droop and Hare quotas is that the latter allows smaller parties to obtain MPs in the initial allocation of seats, while the former makes it harder for smaller parties to obtain seat representation since its formula requires more votes for MPs to be elected. Consequently, the largest remainder formula raises the representation of smaller parties or candidates with smaller vote shares, thereby increasing the proportionality of the electoral system.

The conclusion from this section can be written as follows. For the given vote shares (first-preference votes) of the political parties in Malta, different electoral formulas can yield variations in the election outcomes. Applying distinct seat allocation methods, as previously stated out, might give advantage either to the majoritarian or minority parties depending on the methods chosen, so that implementing methods other than the STV may increase the representation of the minority groups, or on the contrary, may favor the majoritarian political parties in the electoral system. The most important consequence, as will be explained in detail after the second experiment, is that the d'Hondt method approximates the STV system best in terms of converting parties' vote shares into government seats in the 2017 Malta general elections.

3.3.2 Uniqueness of the electoral constituency

In this section, we explore what might happen if Malta had a single national electoral constituency. Instead of dividing the country into numerous constituencies, one might ponder the potential consequences of having one large constituency. A single electoral district scenario, which frequently comes up in electoral system debates, is crucial because in some societies, the representation rights of minority groups are neglected due to electoral district separation. As a result, the single constituency system might enhance the representational chances of at least some minority national groups Reynolds (2008).

[16] More precisely, the Hare quota is defined as $HQ \equiv \frac{total\ valid\ votes}{total\ number\ of\ seats}$.

Furthermore, regarding proportionality, a national-level election scenario can increase proportionality enormously, as in the cases of Israel and the Netherlands Gallagher and Mitchell (2006), Lijphart (2012).

Thus, we questioned how the representation of parties in parliament would change if Malta's 2017 general elections were held in one single electoral district. Accordingly, we adopted the electoral formulas mentioned in the previous section in a single electoral constituency scenario.[17] Table 3.3 depicts the number of parliamentary seats that would be achieved as a result of adopting the STV, the d'Hondt, the largest remainder, and the Sainte-Lague electoral formulas.[18] Note that the bonus seat concept was not taken into account in the calculations, so that each electoral formula was applied to 65 representatives. The second column in the Table shows the observed seat numbers under the STV system for both parties, while the third, fourth and fifth columns indicate the number of representatives that can be obtained by applying the d'Hondt, largest remainder, and Sainte-Lague electoral formulas, respectively.

As can be seen, if one vast constituency scenario were applied hypothetically, d'Hondt and Sainte-Lague electoral formulas would appoint only one extra MP to the PN. When evaluated in this respect, d'Hondt and Sainte-Lague are again the electoral formulas that most closely replicate the STV system. On the other hand, under the largest remainder method the right of a new party to be represented in the parliament would be ensured by increasing the chance of representation for small parties. Thus, Alternattiva Demokratika Party (AD) would be able to obtain its representation in the government if its votes were not concentrated in a particular electoral district under the largest remainder method.

Table 3.3: Electoral formulas under the hypothesis of unique constituency

Parties	***STV***	***D'Hondt***	***Largest Remainder***	***Sainte-Lague***
MLP	37	36	36	36
PN	28	29	28	29
AD	-	-	1	-
Total	65	65	65	65

Key: The hypothetical scenario applied where Maltese general elections of 2017 are held under one single constituency on a nationwide level.

[17] As in the previous section, the seat allocation methods were also simulated with the R program in the same way. It is worth mentioning that R gives more efficient (i.e. shorter computing time) results than Matlab©.

[18] See the case for the 2013 general elections in Table 3.7 in Appendix 3.B.

In summary, the following conclusions can be drawn from the two exercises applied above. First of all, the seat distribution across parties is not robust to the electoral district design (one single nationwide district vs the actual 13 district map). A single constituency scenario, in particular, will be crucial to enhancing the rights in terms or representation of minority groups as shown in Table 3.3, and will also regulate the proportionality of the electoral system.

Secondly, since there is no information about the second or third preference votes of the electors, different electoral formulae were applied to the distribution of "first-preference votes" (the only available in the data set) amongst political parties in order to find the closest system to the STV in terms of lowest MAE [see Eq. (3.2)]. After applying the d'Hondt formula to first-preference vote distribution, the simulated seat distribution gives the closest number of MPs compared to the observed distribution of seats under the STV. In addition to our findings, there is also consensus that the d'Hondt method most closely represents the STV system for translating the given vote shares into the parliamentary seats [Buhagiar and Lauri (2009), Grofman and Lijphart (2003)]. Consequently, we will perform the simulations needed for the excess-seat test (the third gerrymandering test that we will run in Subsection 3.4.3), using the d'Hondt formula.

3.4 Gerrymandering statistical tests

3.4.1 Lopsided-outcomes test

To begin with, we first run the *lopsided outcomes test,* an analysis of intents for detecting partisan gerrymandering. This test is used to compare the average vote margins in the electoral districts, to check if the difference is due to chance in any sense. In other words, if a political party obtains representatives in certain constituencies by high vote margins, it is likely to be a victim of packing, a gerrymandering strategy [Wang (2016a,b)]. Accordingly, the party that designs the electoral districts seeks as many members as possible from constituencies with small margins. Thus, while the opposing party's votes are heavily concentrated in particular constituencies, its chances in the others are diminished. More precisely, we compare the difference between the average vote shares of the MLP and the PN in the electoral constituencies. That is to say, we question the cost in votes per seat for a particular district for the corresponding political party.

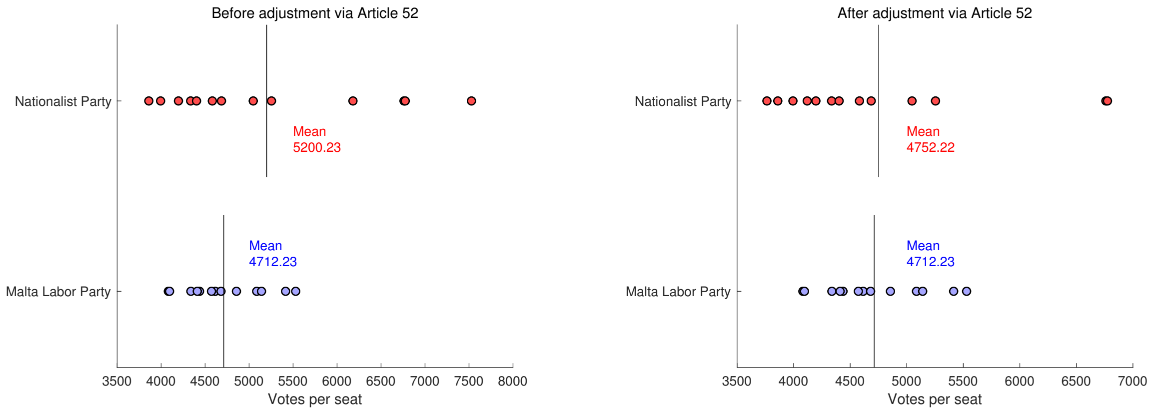
Figure 3.6 depicts how many votes both parties would have received on average for the

MPs they will return to the parliament from each electoral district. As noted in the Section 2, as a consequence of the 1987 amendment, the party that wins the majority of the popular vote across the country or for the purpose of adjusting seat-vote proportionality, is awarded bonus seats, which in the current case added two bonus seats to the PN to balance the seat-to-vote ratio, regardless of the majority of votes in the 2017 general elections. Therefore, the first chart in Figure 3.6 shows the absence of these bonus seats, whereas the second chart shows the situation after the bonus seats have been added to the PN.

Figure 3.6: Analysis of Intents: Lopsided outcomes test

(a) Votes per seat and political party with no bonus seats

(b) Votes per seat and political party with bonus seats



Key: Calculated votes per seat for each party in the 2017 general elections. The red points represent the Nationalist Party's effort to obtain a particular seat in each electoral constituency, and the blue points likewise show the same case for the Malta Labour Party. Before [panel (a)] and after the bonus seats were added to the Nationalist Party [panel (b)] respectively.

In both graphs, the red dots show the number of votes per seat for the Nationalists, while the blue dots display the number of votes per seat for the MLP in each of the electoral districts. In other words, while the average number of votes per seat for the MLP was 4,712, this figure corresponds to an average of 5,200 votes per seat for the Nationalists. Regarding the second graph in Figure 3.6, that is after two bonus seats were granted to the PN under the Article 52, the average number of votes per seat for the PN decreases to 4,752, while for the MLP it remains the same. At first glance, obtaining a particular seat in the constituencies seems to be more costly for the PN than for the MLP. However, the bonus seats concept seems to adjust the average vote per seat for the PN as demonstrated in the second graph in Figure 3.6. In order to determine whether the difference in average votes may be attributed to chance, and following Wang (2016a,b), we compute the corresponding t-statistic:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3.3}$$

where

$$s_1^2 \equiv \frac{\sum_{i=1}^{n_1}(x_i - \overline{x}_1)^2}{n_1 - 1}, \qquad s_2^2 \equiv \frac{\sum_{j=1}^{n_2}(x_j - \overline{x}_2)^2}{n_2 - 1}$$

and where $\overline{x}_1$ and $\overline{x}_2$ are the sample means, s_1^2 (s_2^2) is the sample variance of x_1 (x_2), n_1 (n_2) is the sample size of the x_1 (x_2) sample.[19]

More formally, the null hypotheses can be expressed as $H_0\colon \mu_1 = \mu_2$, i.e. the data provide convincing evidence that there is no significant difference between the population mean vote shares of the MLP and the population mean vote shares of the PN, the alternative hypothesis being $H_1\colon \mu_1 \neq \mu_2$ (in other words, the difference between the two means is statistically significant).

The obtained t-statistic as a solution to the Eq. (3.3) was -1.35 and the p-value of the test was 0.19, which is higher than any reasonable significance level. After the 2 bonus seats adjustment in favor of the PN, the t-statistic equals -0.13 with the corresponding p-value of 0.90. Therefore, we fail to reject the null hypothesis in both cases (before and after granting bonus seats), and conclude that the data do not provide convincing evidence of the difference between the average vote shares of the two political parties. That is to say, given the vote shares of the two political parties, the lopsided outcomes test does not provide any evidence of a possible gerrymandering scenario due to the 2017 boundary changes of the electoral constituencies.

3.4.2 Consistent-Advantage test

As a second test, we examined the difference between average and median seat shares for the two political parties. The mean-median difference is a statistical approach used to test the skewness of the distribution of a random variable, and to explore the asymmetry

[19] Note that there are two options for the t-test, assuming equal variances and unequal variances. Here only the unequal variances case will be shown since the conclusions of both were identical. This is the so-called Welch-Satterthwaite correction and refers the resulting statistic to the t-distribution instead of the standard normal with a random number of degrees of freedom, see Armitage and Berry (1994), Lehmann and Romano (2005).

in the seat share distribution. The case is usually conducted as analyzing the mean-median difference in vote shares for the political parties such as in FPTP with single-member districts. However, the Maltese case differs from this system as it features multi-member constituencies, as previously mentioned in Section 2. Therefore, if the average seat share obtained by a given political party across districts is statistically significantly higher than that party's median seat share, the party in question is most likely to succumb to the gerrymandering strategy [Wang (2016b)]. The main idea behind the *consistent-advantage test* is that if a party's seat share is intentionally concentrated in a few constituencies with very high margins, the party's average seat share will eventually increase, while the median value will be low because the party's seat share in the *majority* of constituencies would be reduced.

Before proceeding, consider the seat share distribution across the 13 electoral districts in 2017 shown in Table 3.4.

Table 3.4: Seat share distribution (%)

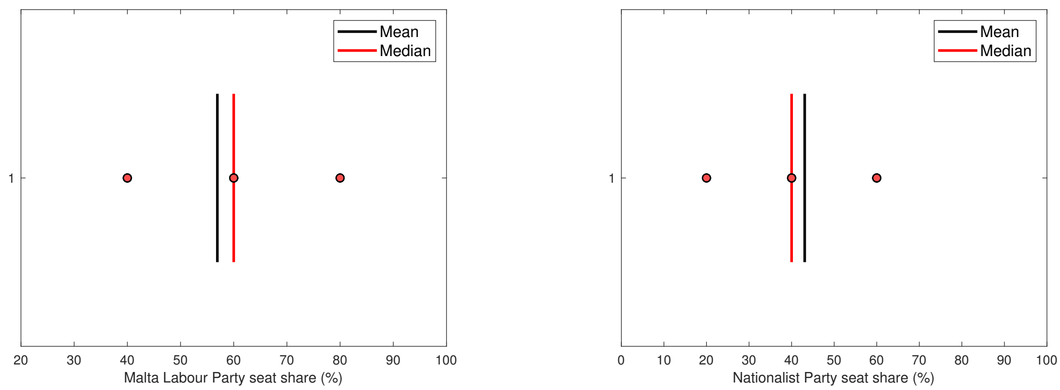
MLP	40	40	40	40	60	60	**60**	60	60	80	80	80	80
PN	20	20	20	20	40	40	**40**	40	40	60	60	60	60

The median value for the MLP seat share is 56.92%, while the average is 60.0% (lower). The case is the opposite, however, for the PN: the median falls short of the mean, 40.0% and 43.08% respectively. In short, according to these comparisons, it seems that the electoral district distribution favors the MLP against the PN.

Visualization of the data mentioned above is displayed in Figure 3.7. The graph on the left shows the distribution of the seat share for the MLP across the 13 electoral districts, the black and the red vertical lines denoting the mean and the median respectively. Thus, the median seat share is slightly higher than the average seat share which implies that no evidence of gerrymandering against the MLP can be found. Likewise, the second graph on the right displays the same calculations for the Nationalists. Unlike the MLP, the average seat share for the Nationalists appears to be slightly higher than the corresponding median. An open question is whether the mean is statistically significantly higher than the median or not. We turn to this next.

Figure 3.7: Analysis of Intents: Mean-median difference in seat shares

(a) Seat share distribution for the MLP (b) Seat share distribution for the PN



Key: Seat share distributions across electoral districts. In both figures black (red) bars represent the average (median) seat shares.

In order to statistically analyze the mean-median difference of the parties' seat shares, we applied the skewness test introduced in Wang (2016a,b) which, in turn, follows Lemma 3 in Cabilio and Masaro (1996), p. 351, for testing the symmetry of a distribution function relating to an unknown median.[20] Thus, assuming a random variable X which follows cumulative distribution function F, it can be shown that if its mean and its median are the same, then (asymptotically)

$$S_k = \frac{\sqrt{n}(\overline{X} - m)}{s} \to N(0, \sigma_0^2(F)) \tag{3.4}$$

where $\overline{X}$ and m represent the sample mean and the sample median respectively, s denotes the sample standard deviation, and n stands for the sample size. Regarding the variance of the test statistic, $\sigma_0^2(F)$, and this is a key point, it depends on the *parent* distribution F (normal, uniform,...) from which the sample is drawn [see Cabilio and Masaro (1996) for details].[21] We assume that the seat shares of the parties follow uniform distributions, $U(a, b)$, where a and b denote the minimum and maximum seat shares of the party in question: SS_{PN}^{min} and SS_{PN}^{max} for the Nationalist Party, and SS_{MLP}^{min} and SS_{MLP}^{min} for the Malta Labour Party.

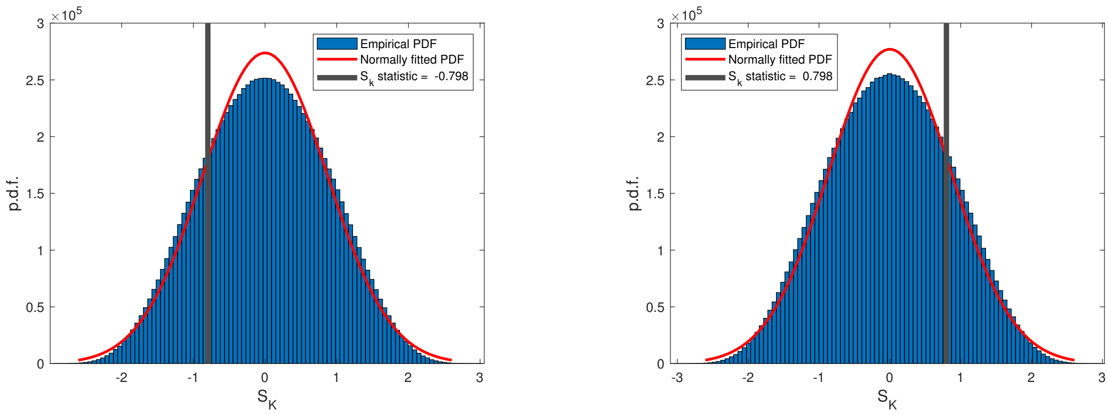
[20] Note that the *true* median of the seat share distribution is known for neither of the two parties.

[21] Wang (2016a,b) assume normality for the vote shares of the political parties, and eventually this would allow the vote share to be negative in the simulated (theoretical) distribution function.

We next simulate the distribution of the S_k statistic in Eq. (3.4) under such an assumption. Thus, after running 10 million random samples of size 13 (the total number of electoral districts) out of uniform distributions for the MLP and the PN seat shares, recall $U(SS_{MLP}^{min}, SS_{MLP}^{max})$ and $U(SS_{PN}^{min}, SS_{PN}^{max})$, we obtain the empirical distributions of S_k for both parties, S_k^{MLP} and S_k^{PN}. The result is shown in Figure 3.8, where the two empirical distributions are represented by the histograms colored blue. For completeness, the figure also shows the corresponding fitted normal distributions in red: the computed mean and variance for S_k^{MLP} are -0.25×10^{-3} and 0.7541, respectively, while those of S_k^{PN} are 0.2×10^{-5} and 0.7978. Finally, the figure also shows the obtained values for the test statistics on the black bar: S_k^{MLP}= -0.7978 and S_k^{PN}= 0.7975, so that the implied p-values are 0.81 and 0.19 respectively.

Figure 3.8: S_k distributions for the MLP and the PN

(a) S_k^{MLP} : Malta Labour Party (b) S_k^{PN}: Nationalist Party

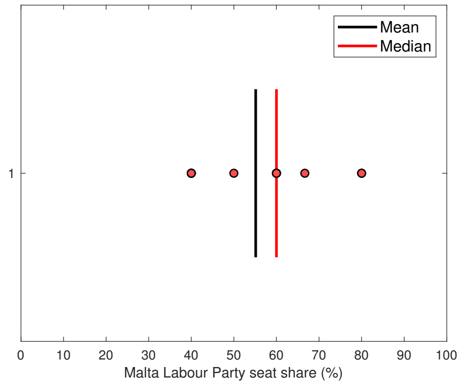


Key: Skewness test statistic S_k (black) for both parties [See Eq. (3.4)]. Histograms (blue) show the empirical probability density functions for the MLP [panel (a)] and for the PN [panel (b)]. The red lines represent the normally fitted probability density functions with mean and variance equal to -0.25×10^{-3} and 0.7541, respectively, for the MLP, and 0.2×10^{-5} and 0.7978 for the PN.

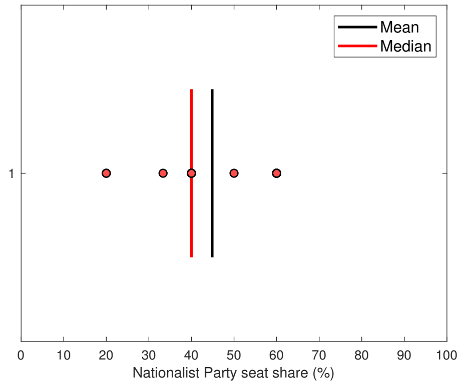
For completeness, we also consider the case with bonus seats granted to the PN. Once again, the mean and median seat shares are represented in Figure 3.9 for both political parties. At first glance, the mean seat share is smaller than the median seat share for the MLP, 55.13% and 60.0% respectively. As regards the PN, mean seat share (44.87%) still exceeds the median seat share (40.0%) after the bonus seats were introduced in order to adjust for the balance between the seat and the vote shares at the national level.

Figure 3.9: Analysis of Intents: Mean-median difference in seat shares

(a) Seat share distribution for the MLP



(b) Seat share distribution for the PN

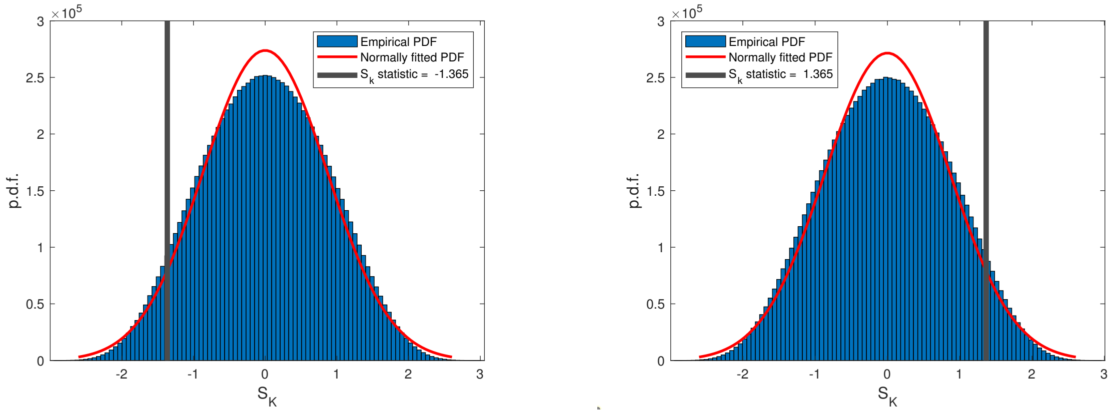


Key: Seat share distributions across electoral districts. In both figures black (red) bars represent the average (median) seat shares.

For completeness, we also apply the test statistic S_k once more considering the two bonus seats granted to the PN. The results are presented in Figure 3.10 where blue histograms represent the two empirical distributions, and red lines show the fitted normal distributions. In this case, the mean and variance for S_k^{MLP} are 0.29×10^{-3} and 0.7539, respectively, whereas for those of S_k^{PN} are 0.87×10^{-4} and 0.7539. Lastly, the black lines depict the test statistics: S_k^{MLP}= -1.3649 and S_k^{PN}= 1.3644, so that the implied p-values are 0.94 and 0.06 respectively. To conclude: the equality of the mean and the median seat shares for both parties across the 13 electoral districts, i.e. the absence of gerrymandering according to the *consistent-advantage test*, can be rejected regardless of whether the bonus seats are considered or not.

Figure 3.10: S_k distributions for the MLP and the PN

(a) S_k^{MLP} : Malta Labour Party (b) S_k^{PN}: Nationalist Party



Key: Skewness test statistic S_k (black) [See Eq. (3.4)]. Histograms (blue) show the empirical probability density functions for the MLP [panel (a)] and for the PN [panel (b)]. The red lines represent the normally fitted probability density functions with mean and variance equal to 0.29×10^{-3} and 0.7539, respectively, for the MLP, and 0.87×10^{-4} and 0.7539 for the PN.

3.4.3 Excess-seats test

Following Wang (2016a,b), we run a third test, the *excess-seats test,* with the purpose of checking whether the vote share obtained by a party given the current 13 electoral district setting unexpectedly deviates from the vote share that this party would have obtained at the national level. More precisely, the experiment that we run is as follows. Imagine that we could build up a set with a large enough number of replicas for the results of the general elections in 2017, so that in each of these we had 13 electoral districts randomly chosen.[22] And, next, among all the possible combinations in this set, we could select only those *fantasy* (or *synthetic*) Maltas in that the difference between *i*) the observed distribution of votes for the six parties participating in the electoral process, and *ii*) the synthetic distribution of votes (at the national level) were equal to zero (or low enough up to some arbitrary point). Finally, we will compare the seats that each party obtained in the 2017 general elections with the average number of seats

[22] A similar analysis was carried out by Wang (2016b) for the case of the state of North Carolina in the United States. In each potential replica, thirteen electoral districts (the number of North Carolina electoral districts) were randomly chosen from the whole (i.e nationwide) set of electoral districts. Among such potential replicas, only those with a vote share distribution "equal" to that of North Carolina were kept: the "synthetic" or fantasy North Carolinas. Finally, a comparison was made between the observed seats obtained by each party in North Carolina, and the averages that those parties would have obtained in the synthetic North Carolinas.

which that party would have obtained in those fantasy Maltas. The question that we will answer is: what is the probability that a given party would have obtained a higher number of seats than the ones effectively won if the electoral district distribution had been different?

The first step consists of calculating all the possible combinations of the 13 electoral districts. In general, we have that the number of combinations with repetition of m elements, taken n at time, $CR_{m,n}$, is given by

$$CR_{m,n} \equiv \binom{m+n-1}{n} = \frac{(m+n-1)!}{(m-1)! \times n!}. \tag{3.5}$$

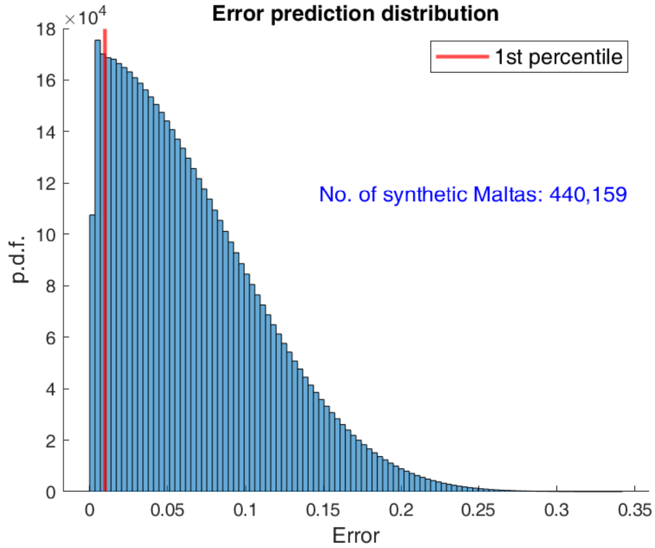
In our case, $m = n = 13$, so that we obtain a potential number of 5,200,300 synthetic Maltas.[23] For each i-th combination of these, we compute the (norm of the) difference between the vote share distribution of the i-th potential synthetic Malta, $\boldsymbol{V}_i \in \mathbb{R}^6_{\geq 0}$, and the observed vote share distribution, $\boldsymbol{V}^{obs} \in \mathbb{R}^6_{\geq 0}$, where

$$\boldsymbol{\Delta_i} \equiv \parallel \boldsymbol{V}^{obs} - \boldsymbol{V_i} \parallel . \tag{3.6}$$

And from this set, we keep only those combinations of electoral districts for which the difference between the vote share distribution of the potential synthetic Malta and the observed vote share distribution is small enough, more specifically, less than the first percentile. As a result, we are finally left with 440,159 fantasy Maltas. The result is shown in Figure 3.11, where the histogram for $\boldsymbol{\Delta_i}$ is shown in blue color, and a red vertical line represents the 1st percentile of the distribution.

[23]This is an instance of the so-called bootstrapping sampling method. Computations were made with the Matlab©**combinator** function (for more information, see Dekking et al. (2005), Efron and Tibshirani (1994)).

Figure 3.11: Error prediction distribution obtained for the simulated general elections of Malta



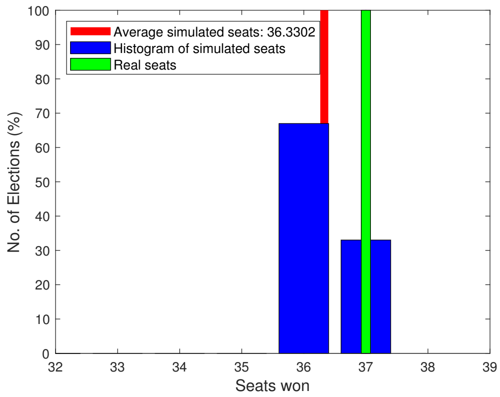
Key: Histogram of Δ_i in blue color [See Eq. (3.6)]. The vertical red line represents the 1^{st} percentile (0.0023).

As a final step, we next calculate the seats that each party (in practice, only the two majoritarian parties at play) would obtain in each of these fantasy Maltas, thereby obtaining two seat distributions [See Figure 3.12]. Consider first the MLP. As the figure shows, the MLP would obtain 36 seats in 66.98% of the elections and 37 seats in the remaining 33.02%, or 36.33 on average, the actual number of seats being higher, 37. In short, the probability for the MLP to obtain a higher number of seats than the one effectively obtained is 0. In other words, this third test does not provide empirical evidence of gerrymandering in favor of the Malta Labour Party.

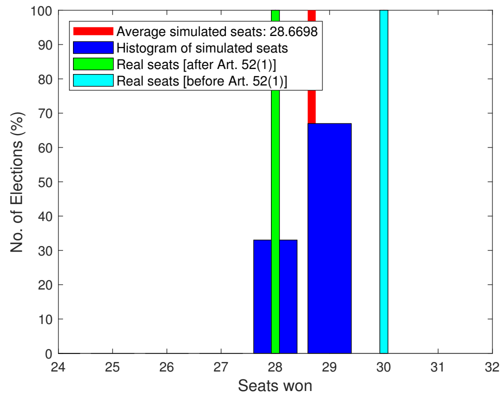
As the natural counterpart, a similar analysis can be carried out for the Nationalist Party. The PN would have obtained 28 seats and 29 seats in 33.02% and 66.98% of the synthetic Maltas, respectively, implying an average of 28.67 seats. On this occasion, the comparison with the actual number of seats that the PN obtained in 2017 requires some qualification though. Recall that, as a consequence of the constitutional amendment of Article 52, the PN was granted two bonus seats, so that its total number rose from 28 to 30. Thus, before the amendment was implemented, the probability for the PN to obtain a higher number of seats would have been 66.98%. Or, in other words, the electoral district set up would imply the existence of gerrymandering against the Nationalist Party which was corrected by the adjustment for the bonus seats. The bonus seats adjustment, however, would more than offset the initial partisan bias against the PN.

Figure 3.12: The excess-seats test: simulated and actual seats

(a) Malta Labour Party



(b) Nationalist Party



Key: Observed and simulated seat numbers. In both figures, the blue bars represent simulated seat numbers, while the green bars show the actual seat. In addition, the turquoise bar in the second figure shows the actual seat number after the bonus seats were granted to the PN under Article 52.

3.4.4 The 2013 general elections

Additionally this section briefly presents the analysis of the 2013 general elections, with the same purpose of applying the above mentioned gerrymandering tests. First, the results of the *lopsided outcomes test* are shown in Figure 3.13 in Appendix 3.B, with the same conclusion as in the 2017 general elections. In other words, by adding four extra bonus seats to the PN in the 2013 general elections, the discrepancy in the average vote shares was reduced, thus eventually regulating the proportionality of the votes-seat ratio. Applying the t-test for both cases (with/without bonus seats), we do not find any significant difference between the political parties' vote shares, hence no evidence of possible gerrymandering scenarios according to the first test.

Regarding the second test, the *consistent-advantage test* of the mean-median difference of the seat shares, the difference was 0 for both cases [see Figure 3.14 and Figure 3.15 in Appendix 3.B]. Thus, the test statistic (S_k) results for the MLP and the PN were 0, implying p-values of 0.50 and 0.59, respectively. As for the bonus seat case, the average seat share of the MLP exceeds the median seat share after four additional seats are added to the PN [see Figure 3.16 and Figure 3.17 in Appendix 3.B]. Consequently, the test statistic (S_k) turned out to be 1.6092 for the MLP and -1.6088 for the PN with the corresponding p-values of 0.03 and 0.97, respectively. That is, granting additional seats to ensure votes-seat proportionality revealed a statistically significant asymmetry in the mean-median seat share of the MLP. One might, therefore, question the credibility of the bonus seat concept for this situation, as it eventually might not be necessary. In sum, the second test does not reveal any evidence of gerrymandering in the 2013 general elections of Malta, thereby reinforcing the results obtained with the first test.

Finally, the results for the third test, the *excess-seats test*, are shown in Figure 3.18, Appendix 3.B. However, some explanation is required here; the MLP would obtain 36 seats in 53.87% of the elections and 37 seats in the remaining 46.13%, or 36.46 on average. That is, the probability of the MLP obtaining more seats than the actual ones is 0. The actual seat numbers obtained by the MLP would differ from the simulated seat numbers with 2.6 seats. Nonetheless, this 2.6 seat discrepancy from the actual results would have no change the overall outcome of the 2013 general elections. For completeness, the PN would obtain 28 seats and 29 seats on 46.13% and 53.87% of the synthetic Maltas with an average of 28.54 seats. Note that the PN was granted four additional seats due to the Article 52 so that its total number of seats rose from 26 to 30. Hence, before the amendment was implemented, the probability for the PN to obtain a higher number of seats than the one actually obtained is 1. One would, therefore, expect distortions in the outcome of the 2013 general elections if the PN did not receive additional bonus seats; yet, the bonus seat concept once again rectified the proportionality of translating the PN's given vote shares into parliamentary seats. To conclude, once again we find no possible gerrymandering scenarios upon examination of the 2013 Maltese general elections.

3.5 Concluding Remarks

Whilst the pace of governments collecting data rises, usage of knowledge is also becoming more significant in the rapidly changing technical environment. Many insights can also be collected for democratic campaigns, and various computer systems or algorithms can be used to gain benefits or to undertake various political initiatives. As a result of these factors, in elections the question of fairness is becoming more and more relevant. Political scientists have devised many methods and statistical tests to discourage such politics and, in particular, have sought to reduce the possible anomalies caused by redrawing electoral constituencies.

In the light of the above information, we have examined Malta's two recent general elections in this article. On the one hand, Malta is open to potential gerrymandering possibilities by means of the two-party system; on the other, the single transferable vote system proves itself to be one of the most proportional among electoral systems in terms of converting votes cast into seat distribution. As a first step, we have used statistical tests to detect possible attempts at gerrymandering by analyzing the average vote margins of the MLP and the PN. That is, we have questioned how far political parties go in their attempts to obtain parliamentary seats in the corresponding electoral districts. Consequently, we have posed the hypothesis test to determine whether the average vote shares of each political party in the electoral constituencies were statistically different, and we have measured its significance with the t-test. According to the results, we have found that the difference between the average vote shares of the two parties was not statistically significant. As a second analysis of intents, we have measured the mean-median difference in national-level seat shares of both political parties as a skewness test in order to detect partisan asymmetries that might have been caused by political maneuvering. Once again, we have come to the conclusion that the mean-median difference, which we have subjected to significance tests, was not statistically significant. At the last stage, and with the information obtained from the two tests mentioned, we have obtained statistical inference with bootstrap simulations, taking into account the votes of all parties that entered the 2013 and the 2017 general elections. Accordingly, we have attempted to identify possible anomalies or partisan bias by comparing the observed average vote shares of the political parties with the vote shares estimated via the bootstrap simulations. As in the first two tests, we did not find any signs of gerrymandering in the final test, and we can conclude that Malta's two last governmental elections were held in fair election processes.

In general, three more important conclusions can be drawn from this article. Firstly, and frequently discussed in political science literature, the STV favors the majoritarian parties, especially as regards the droop quota, which makes it harder for the smaller parties to return representatives to the government. Thus, the system reinforces governability by reducing the chance of having coalitions in the parliamentary. Secondly, as argued by the political elite in Malta, there would be no change in the overall outcome by having a single constituency for the entire nation. However, the electoral district magnitude makes it difficult for small parties to enter parliament. The proportionality of the electoral system can still be adjusted if the number of delegates from each district is increased (which will lower the droop quota). Third and finally, implementing the bonus seat concepts by the constitution, the STV system in Malta regulates proportional casting of given vote shares of the political parties into the seat distribution for the Maltese legislature.

APPENDICES

3.A Appendix: How the STV works

This appendix illustrates the way the STV works on an empirical base. In Table 3.5 below, seven candidates from the first electoral division in Malta's 2017 general elections are displayed together with the votes they received. Note that although there are a total of 23 candidates in constituency 1, only 7 are included in the example for simplicity. The second column in the table shows the aforementioned first-preference votes. That is, it corresponds to the total of the first-preferences received by the electors on the ballot paper. At this stage, firstly, the amount of quota candidates have to reach in order to be elected from this constituency will be determined as $(24,196/(5+1))+1 = 4,033$, as stated in Eq. (3.1). In other words, the listed candidates must have reached a total of 4,033 votes in order to be elected. As reflected in the second column, only two of the candidates (Jose Herrera and Mario De Marco) at the first counting of the votes will reach the quota and will be automatically elected from the constituency. Therefore, according to the number of second preference votes on the ballots, the number of votes cast by these candidates as surplus votes after exceeding the quota is transferred to other candidates.

Table 3.5: Election in the first electoral constituency in Malta, 2017

Electorate: 25,598, Total Valid Poll: 24,196, Seats: 5, Quota: 4,033								
Candidates	*First count*	*Second count*		*Third count*		*Fourth count*		
Debattista, Deo (MLP)	2,378	+194	2,572	+357	2,929	+1,104	*4,033*	
Farrugia, Aaron (MLP)	3,600	+149	3,749	+284	*4,033*	-	*4,033*	
Herrera, Jose' (MLP)	4,630	-597	*4,033*	-	*4,033*	-	*4,033*	
Parnis, Silvio (MLP)	1,385	+121	1,506	-1,506	-	-	-	
De Marco, Mario (PN)	4,721	-688	*4,033*	-	*4,033*	-	*4,033*	
Grech, Claudio (PN)	2,606	+215	2,821	+738	3,559	+474	*4,033*	
Mifsud, Paula (PN)	2,000	+338	2,338	+127	2,465	-2,465	-	

Source: The Electoral Comission Malta

In the Table, the amount of votes to be added to other candidates is indicated in the third column. For example, since Jose Herrera surpassed the quota with 597 votes and Mario De Marco with 688, these surplus votes were distributed among the other candidates in proportion to the electors' second choice preferences on the ballot paper, using the formula

$$\left[\frac{second\, preference\, votes}{total\, votes\, of\, the\, candidate}\right] \times surplus\, votes. \tag{3.7}$$

This transfer procedure continues until sufficient candidates (5 in the current example) have met the quota to fill all the seats to be elected. In certain cases, for example, after calculating the number of second preference-votes for candidates, if any candidate fails to reach the determined quota, then the candidate with the fewest votes is excluded and his/her votes are transferred to the voters' second preferences among the candidates. Accordingly, as seen in column three, Silvio Parnis is eliminated at this stage as he has the lowest amount of votes (1,506), and his votes are distributed among other candidates to be shared. After the third count votes are calculated, Aaron Farrugia is elected as he reaches the quota. Paula Mifsud will be eliminated at this stage since she has the lowest number of votes (2,465). Her votes are therefore transferred to the fourth count. As a consequence, a total of five delegates will be appointed as members of the parliament from this district when the fourth count is completed.

3.B Appendix: the 2013 general elections

Table 3.6: Distinct electoral formulas applied to each electoral constituency

	STV		***D'Hondt***		***Largest Remainder***		***Sainte-Lague***	
District	*MLP*	*PN*	*MLP*	*PN*	*MLP*	*PN*	*MLP*	*PN*
1	3	2	3	2	3	2	3	2
2	4	1	4	1	4	1	4	1
3	4	1	4	1	***3***	***2***	4	1
4	4	1	4	1	***3***	***2***	4	1
5	4	1	4	1	***3***	***2***	4	1
6	3	2	3	2	3	2	3	2
7	3	2	3	2	3	2	***2***	***3***
8	3	2	***2***	***3***	***2***	***3***	3	2
9	2	3	2	3	2	3	2	3
10	2	3	2	3	2	3	2	3
11	2	3	2	3	2	3	2	3
12	2	3	2	3	2	3	2	3
13	3	2	3	2	***2***	***3***	***2***	***3***
Total	**39**	**26**	**37**	**28**	**34**	**31**	**37**	**28**
MAE	-		0.051		0.128		0.051	

Key: In addition to the actual seats obtained under the STV system, implemented electoral formulas d'Hondt, largest remainder, and Sainte-Lague for the given constituency in the 2013 general elections. The last row indicates the mean absolute error as obtained in Eq. (3.2). Finally, the underlined numbers show the differences compared to the actual STV in that constituency. Note that results are presented without taking into account possible bonus seats.

Table 3.7: Electoral formulas under the hypothesis of a single constituency

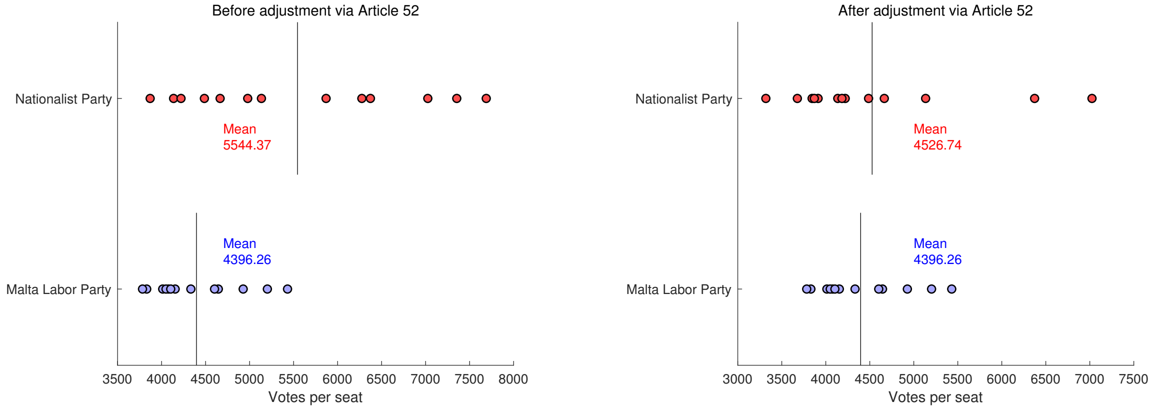
Parties	***STV***	***D'Hondt***	***Largest Remainder***	***Sainte-Lague***
MLP	39	36	36	36
PN	26	29	28	29
AD	-	-	1	-
Total	65	65	65	65

Key: The hypothetical scenario applied where the Maltese general elections of 2013 are held under one single constituency on a nationwide level.

Figure 3.13: Analysis of Intents: Lopsided outcomes test

(a) Votes per seat and political party with no bonus seats

(b) Votes per seat and political party with possible bonus seats

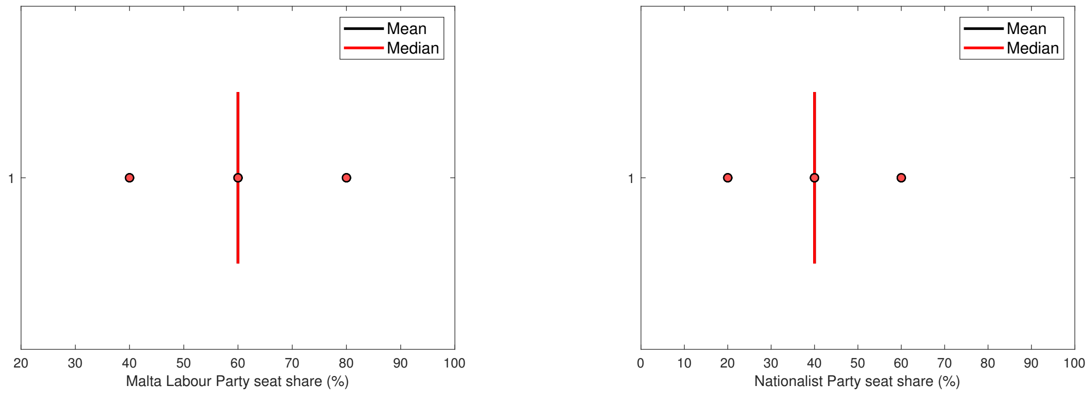


Key: Calculated votes per seat for the each party in the 2013 general elections. The red points represent the Nationalist Party's effort to obtain a particular seat in each constituency, and the blue points likewise show the same case for the Malta Labour Party. Before [panel (a)] and after the bonus seats added to the Nationalist Party [panel (b)] respectively.

Figure 3.14: Analysis of Intents: Mean-median difference in seat shares

(a) Seat share distribution for the MLP

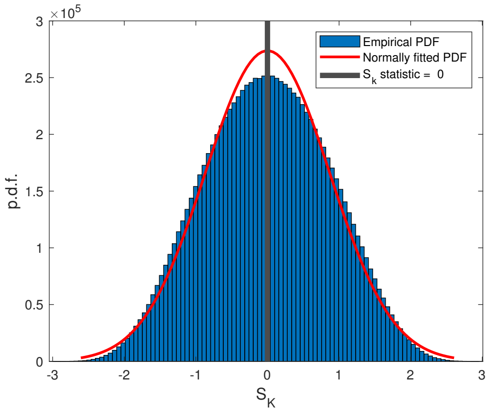
(b) Seat share distribution for the PN



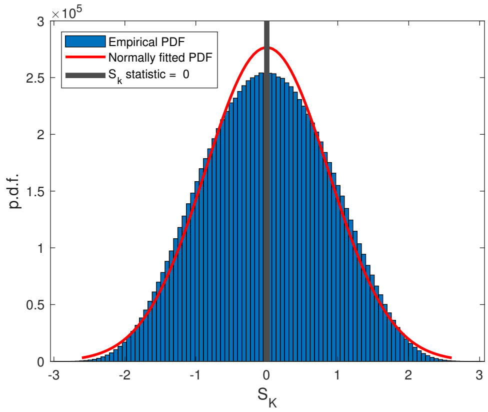
Key: Seat share distributions across electoral districts. In both figures black (red) bars represent the average (median) seat shares.

Figure 3.15: S_k distributions for the MLP and the PN

(a) S_k^{MLP} : Malta Labour Party



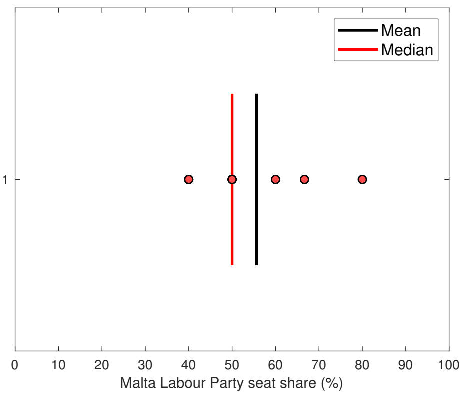
(b) S_k^{PN}: Nationalist Party



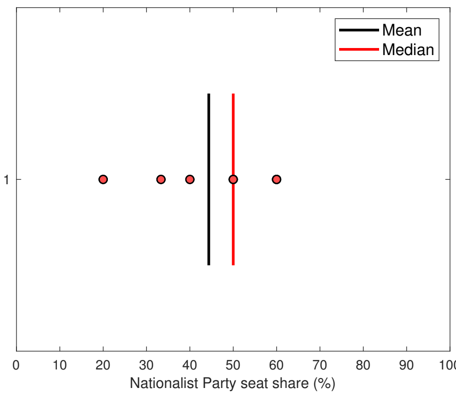
Key: Skewness test statistic S_k (black) for both parties [See Eq. (3.4)]. Histograms (blue) show the empirical probability density functions for the MLP [panel (a)] and for the PN [panel (b)]. The red lines represent the normally fitted probability density functions with mean and variance equal to 0.31×10^{-4} and 0.7544, respectively, for the MLP, and 0.32×10^{-4} and 0.7546 for the PN.

Figure 3.16: Analysis of Intents: Mean-median difference in seat shares

(a) Seat share distribution for the MLP



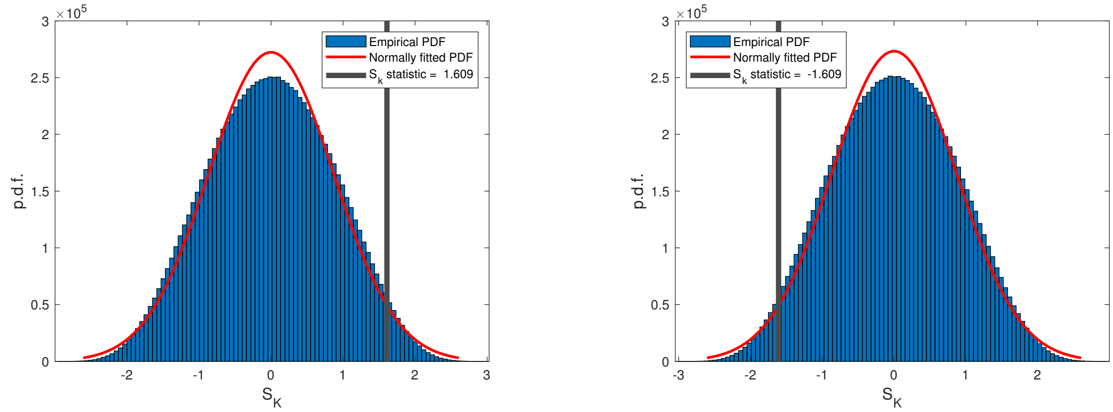
(b) Seat share distribution for the PN



Key: Seat share distributions across electoral districts. In both figures black (red) bars represent the average (median) seat shares.

Figure 3.17: S_k distributions for the MLP and the PN

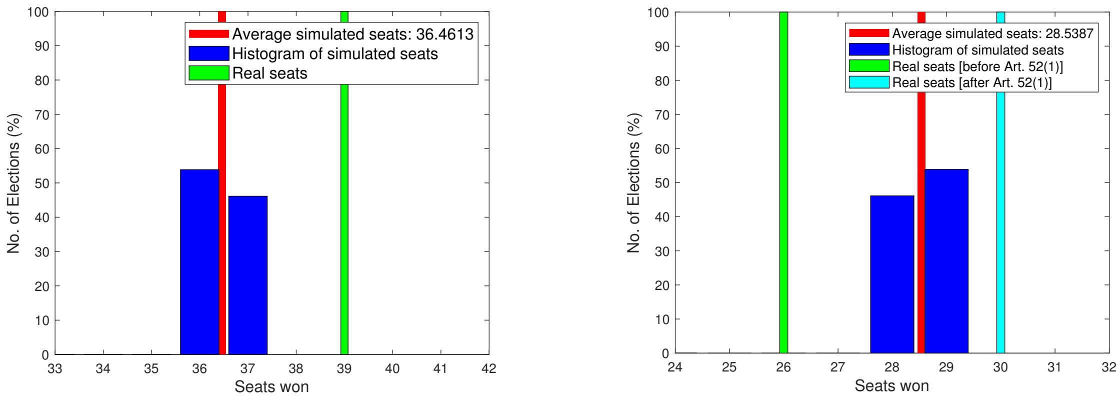
(a) S_k^{MLP} : Malta Labour Party (b) S_k^{PN}: Nationalist Party



Key: Skewness test statistic S_k (black) for both parties [See Eq. (3.4)]. Histograms (blue) show the empirical probability density functions for the MLP [panel (a)] and for the PN [panel (b)]. The red lines represent the normally fitted probability density functions with mean and variance equal to -0.1×10^{-3} and 0.7541, respectively, for the MLP, and 0.4×10^{-3} and 0.7545 for the PN.

Figure 3.18: The excess-seats test: simulated and actual seats

(a) Malta Labour Party (b) Nationalist Party



Key: Observed and simulated seat numbers for the corresponding political parties in the 2013 general elections. In both figures, the blue bars represent simulated seat numbers, while the green bars show the actual seat numbers. In addition, the turquoise bar in the second figure shows the actual seat number after the bonus seats were granted to the PN under Article 52.

Conclusions of the Thesis

This thesis examines key topics in the disciplines of Economics and Political Sciences from a comprehensive stance. On the one hand, it demonstrates the significance of policy analysis and modeling by using empirical case studies and the latest methodological approaches. The first chapter explores the economic consequences of the Islamic revolution by blending synthetic control methods with diverse optimizations in an environment where causality analysis is becoming increasingly important. In addition, in the second chapter on Macao's integration process, we use both the panel data approach and the synthetic control estimator to examine how the aforementioned process affects Macao's macroeconomic development. As a result of this empirical application, we sought to advise the reader in the application points of alternative approaches in causality analysis.

On the other hand, in the third chapter, we examine the matter of gerrymandering, which is regularly and intensely discussed in the political science literature, in order to control electoral boundaries and prevent any unfair outcomes that may occur. To put it simply, the process of drawing constituency boundaries in a way that benefits certain political parties is known as "*gerrymandering*". As an example, if a party nominates the boundaries of specific regions in its favor, it might manipulate the election outcomes against the free will of the people, thereby breaching democracy's concept of representation. Therefore, as political maneuvering continues to grow in recent years, so do the research and models on this topic. Consequently, summary results of each chapter are briefly explained in the following paragraphs.

The economic impact of the Islamic Revolution on Iran's economy is examined in Chapter 1 of the thesis. Essentially, this chapter is the first study to employ synthetic control estimator to evaluate the impact of the revolution on the Iranian economy, which makes it a significant contribution to the field. In addition, we examine the performance of alternative conventional optimizations and different programming (Matlab©, R, and Stata©), and pick the one that provides the most accurate outcomes in terms of goodness of fit. As a conclusion of this study's findings, the yearly per capita GDP in Iran dropped by around 20.15 percent on average and by 47.70 percent cumulatively through 1978 to 1980. Thus, Iran's cumulative per capita GDP loss after the revolution amounted to 6,479 US dollars, or an average yearly loss of about $2,159. We perform various exercises to evaluate the statistical significance, and the results indicate that our conclusion is fairly robust in identifying the negative impact of the Islamic Revolution on Iran's economic development.

Chapter 2 investigates the economic progress of Macao as an empirical case, in the context of its integration into mainland China. The accords established in 1999 brought Macao under the administration of the People's Republic of China while it was still a Portuguese colony. Macao's economy boomed in the subsequent years, mostly due to the deregulation of the gaming industry. Both the synthetic control estimator and the panel data approach were used to evaluate the policy evaluation that underlies this economic development. The quantitative evaluation of Macao's integration process can be specified as follows. First, the aforementioned process in Macao has a positive and statistically significant impact on both the level of GDP per capita and its long-term growth rate in the period 2000-2012. Despite the fact that both approaches produce qualitatively strong results, the average gap obtained differs depending on the methodology employed to construct the counterfactual unit. This implies that methods (SCM vs. PDA) and model selection criteria (AICc vs. LASSO) are essential. As a result, an average of $42,321 increase in per capita GDP and an 11.4% difference were obtained between the observed growth rate and counterfactual units, both of which were significant. Second, we found that foreign investment had a statistically significant and positive effect. Net foreign investment per capita in Macao grew by $6,358 per year after intervention. Third, we examined the impact of this integration process on the unemployment rate. As a first view, we observed that this procedure led to an overall drop in long-term unemployment. However, we found that this impact was not statistically significant regardless of the methodology employed. Fourth, we focused at the exports and imports of goods and services, which are important macroeconomic indicators. When evaluating the former, we cannot infer that the impact obtained is significant, as the significance tests provide different findings. In the latter case, the significance of this effect is disputed, as the positive effect obtained is highly reliant on the methodology adopted. General conclusions include comprehending the relevance of causality analysis methodologies, as well as the requirement of evaluating the significance of results obtained in a variety of ways.

Chapter 3 switches our focus to a different field, Political Sciences. Ensuring a fair election process, which is one of the most fundamental aspects of democracy, is unquestionably vital. Especially in today's fast-evolving technology era, governments are finding it simpler to modify their information gathering and electoral constituencies in ways that will give them an edge. What becomes important at this stage is that we can assess how fair this process is and any potential biases in the election outcome. Consequently, the third chapter provides a statistical analysis of Malta's two most recent general elections as an empirical case study. In other words, we seek to assess whether the redistricting of electoral districts implemented via legislative amendments prior to the 2013 and 2017 general elections was subject to a possible gerrymandering scenario based on the general election outcomes. To begin, we examine the average vote margins of the two major political parties in constituencies. In other words, we use the t-test to see if there are any significant differences in the average number of votes received by political parties in order to obtain particular parliamentary seats. The *lopsided-outcomes test* indicated that there was no statistically significant difference in the average vote shares of the political parties, implying that no gerrymandering scenario could be proven according to the first test. Possible partisan asymmetry was attempted to be discovered with the mean-median test of the parties' seat shares in the second test, the *consistent-advantage test*. The second test did not disclose any asymmetry as a result of the Skewness test statistic, indicating that no potential gerrymandering scenario was found. Finally, the third test, the *excess-seats test*, is based on computer simulations. That is, after replicating the vote shares of the political parties that contested in the 2013 and 2017 general elections using millions of possible constituency combinations, we compared the resulting number of seats to those actually obtained. The significant variations that may occur would potentially result in partisan asymmetry in the election outcomes. Nonetheless, as in the previous two tests, we did not encounter a possible bias or gerrymandering situation in the 2013 and 2017 general elections, according to the last test. Overall, Malta's electoral system of single transferable vote leads to a fair election process for political parties. Furthermore, the notion of bonus seats established by the amendment is fundamental in minimizing uneven election outcomes.

Future research will rely on the use of machine learning techniques in causality analysis. In recent years, many methodologies have been adopted with machine learning or artificial intelligence algorithms, and various economic disciplines are starting to benefit from corresponding techniques. In this direction, both the credibility of the models are increased and the researcher is assisted in issues such as model selection criteria. When it comes to the third chapter, the concept of gerrymandering is still not known in many countries, and therefore the literature lags a little in this regard. It should be noted that since all electoral systems can be manipulated during the election process, investigations on this issue should proceed in the same direction. I hope this doctoral thesis will serve as a guide for future researchers.

Bibliography

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

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.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics*, 121(2):368–391.

Alesina, A., Özler, S., Roubini, N., and Swagel, P. (1996). Political instability and economic growth. *Journal of Economic Growth*, 1(2):189–211.

Amuzegar, J. (1992). The Iranian economy before and after the revolution. *Middle East Journal*, pages 413–42.

Arjomand, S. A. (1986). Iran's Islamic Revolution in comparative perspective. *World Politics*, pages 383–414.

Armitage, P. and Berry, G. (1994). *Statistical methods in medical research*. Blackwell Scientific, Oxford [u.a.], 3. ed. edition. Literaturverz. S. 581 - 594.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Balinski, M. L. and Young, H. P. (1984). Fair Representation: Meeting the Ideal of One Man, One Vote. *Michigan Law Review*, 82(4):1028.

Becker, M. and Klößner, S. (2017). Estimating the economic costs of organized crime by synthetic control methods. *Journal of Applied Econometrics*, 32(7):1367–1369.

Becker, M. and Klößner, S. (2018). Fast and reliable computation of generalized synthetic controls. *Econometrics and Statistics 5*, 5:1–19.

Bickerstaff, S. (2020). *Election Systems and Gerrymandering Worldwide*. Studies in Choice and Welfare. Springer.

Bilgel, F. and Karahasan, B. C. (2017). The economic costs of separatist terrorism in Turkey. *Journal of Conflict Resolution*, 61(2):457–479.

Billmeier, A. and Nannicini, T. (2013). Assessing Economic Liberalization Episodes: A Synthetic Control Approach. *The Review of Economics and Statistics*, 95(3):983–1001.

Botosaru, I. and Ferman, B. (2019). On the role of covariates in the synthetic control method. *Econometrics Journal*, 22(2)(2):117–130.

Bowler, S. and Grofman, B. (2010). *Elections in Australia, Ireland, and Malta under the single transferable vote: Reflections on an embedded institution*. University of Michigan Press.

Buhagiar, A. and Lauri, J. (2009). STV in Malta: A Crisis? *Voting Matters*, 26:1–12.

Cabilio, P. and Masaro, J. (1996). A simple test of symmetry about an unknown median. *Canadian Journal of Statistics*, 24(3):349–361.

Chen, J. and Rodden, J. (2013). Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quarterly Journal of Political Science*, 8(3):239–269.

Cox, G. W. (1997). *Making Votes Count*. Cambridge University Press.

Cox, G. W. and Katz, J. N. (2002). *Elbridge Gerry's Salamander*. Cambridge University Press.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

De Miño, W. H. and Lane, J. C. (1996). STV in Malta: Some surprises. *Representation*, 34(1):21–28.

De Miño, W. H. and Lane, J. C. (2010). Malta: STV in a two-party system. In *Elections in Australia, Ireland, and Malta under the Single Transferable Vote: Reflections on an Embedded Institution*, pages 178–204. University of Michigan Press.

Dekking, F., Kraaikamp, C., Lopuhaä, H., and Meester, L. (2005). *A modern introduction to probability and statistics. Understanding why and how.*

Demirci, S. (2013). The Iranian Revolution and Shia Islam: The role of Islam in the Iranian Revolution. *International Journal of History*, 1(2):37–48.

Diamond, L. J. and Plattner, M. F., editors (2006). *Electoral systems and democracy.* A journal of democracy book. The Johns Hopkins Univ. Press, Baltimore. Includes bibliographical references and index.

Echevarría, C. A. and García-Enríquez, J. (2019). The Economic Consequences of the Libyan Spring: A Synthetic Control Analysis. *Defence and Peace Economics*, 30(5):592–608.

Echevarría, C. A. and García-Enríquez, J. (2020). The economic cost of the Arab Spring: the case of the Egyptian revolution. *Empirical Economics.*

Edmonds, R. L. and Yee, H. S. (1999). Macau: From Portuguese Autonomous Territory to Chinese Special Administrative Region. *The China Quarterly*, 160:801–817.

Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap.* Chapman and Hall/CRC.

Esfahani, H. and Pesaran, H. (2008). The Iranian Economy in the Twentieth Century: A Global Perspective. *Iranian Studies*, 42.

Farrell, D. M. (1997). *Comparing Electoral Systems.* Prentice Hall/Harvester Wheatsheaf London ; New York.

Farrell, D. M. and Katz, R. S. (2014). Assessing the proportionality of the single transferable vote. *Representation*, 50:13–26.

Ferman, B. and Pinto, C. (2016). Synthetic controls with imperfect pre-treatment fit. *Munich Personal RePEc Archive.*

Ferman, B., Pinto, C., and Possebom, V. (2020). Cherry Picking with Synthetic Controls. *Journal of Policy Analysis and Management*, 39(2):510–532.

Firpo, S. and Possebom, V. (2018). Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets. *Journal of Causal Inference*, 6(2):1–26.

Fisher, R. A. (1935). *The Design of experiments.* Oliver and Boyd, Edinburgh [usw.].

Gallagher, M. (2010). *"The (Relatively) Victorious Incumbent under PR-STV: Legislative Turnover in Ireland and Malta", in Elections in Australia, Ireland, and Malta under the Single Transferable Vote: Reflections on an embedded institution edited by Shaun Bowler and Bernard Grofman*. University of Michigan Press.

Gallagher, M. and Mitchell, P. (2006). *The Politics of Electoral Systems*. Oxford University Press.

Gardeazabal, J. and Vega-Bayo, A. (2017). An empirical comparison between the synthetic control method and Hsiao et al.'s panel data approach to program evaluation. *Journal of Applied Econometrics*, 32:983–1002.

Gelman, A. and King, G. (1990). Estimating the Electoral Consequences of Legislative Redistricting. *Journal of the American Statistical Association*, 85(410):274–282.

Gelman, A. and King, G. (1994a). A Unified Method of Evaluating Electoral Systems and Redistricting Plans. *American Journal of Political Science*, 38:514–554.

Gelman, A. and King, G. (1994b). Enhancing Democracy Through Legislative Redistricting. *American Political Science Review*, 88:541–559. , Parts reprinted in California Policy Studies Brief, a publication of the California Policy Seminar, Vol. 7, No. 5 (April, 1995).

Giugăl, A., Johnston, R., Chiru, M., Ciobanu, I., and Gavriș, A. (2017). Gerrymandering and malapportionment, Romanian style: The 2008 electoral system. *East European Politics and Societies*, 31(4).

Goderbauer, S. (2016). Political Districting for Elections to the German Bundestag: An Optimization-Based Multi-stage Heuristic Respecting Administrative Boundaries. In Lübbecke, M., Koster, A., Letmathe, P., Madlener, R., Peis, B., and Walther, G., editors, *Operations Research Proceedings 2014*, Operations Research Proceedings, pages 181–187. Springer.

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

Grofman, B. (1985). Criteria for Districting a Social Science Perspective. *UCLA Law Review*, 33:77–184.

Grofman, B. and Lijphart, A. (2003). *Electoral Laws and their Political Consequences*. Agathon Press, New York, [N.Y.].

Hastie, T., Tibshirani, R., and Wainwright, M. (2016). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press. Taylor & Francis Group. Chapman and Hall.

Herron, E. S., Pekkanen, R., and Shugart, M. S., editors (2018). *The Oxford handbook of electoral systems*. Oxford University Press, New York, NY.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81:945–960.

Horiuchi, Y. and Mayerson, A. (2015). The opportunity cost of conflict: Statistically comparing Israel and synthetic Israel. *Political Science Research and Methods*, 3(3):609–618.

Hsiao, C., Steve Ching, H., and Ki Wan, S. (2012). A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics*, 27(5):705–740.

Hsiao, C. and Zhou, Q. (2019). Panel parametric, semiparametric, and nonparametric construction of counterfactuals. *Journal of Applied Econometrics*, 34(4):463–481.

Hsieh, P.-F. and Wang, S.-M. (2014). The Impacts of Liberalization in Casino-based Economy —The Case of Macao. *Journal of Business and Economics*, 5:1957–1966.

Jagannath, R. and Upadhye, N. (2018). The lasso estimator: Distributional properties. *Kybernetika*, 54(4):778–797.

Johnston, R. (2015). Which Map? Which Government? Malapportionment and Gerrymandering, UK-Style. *Government and Opposition*, 50(1).

Johnston, R. J. (2003). *"Constituency. Redistribution in Britain: Recent Issues." In Electoral Laws and Their Political Consequences, edited by Bernard Grofman and Arend Lijphart.* New York: Agathon.

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2021). Standard synthetic control methods: The case of using all preintervention outcomes together with covariates. *Journal of Business & Economic Statistics*, 0(0):1–15.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 28(5):1356–1378.

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer Series in Statistics. Springer New York, New York, NY.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses*. Springer texts in statistics. Springer, New York, NY, 3. ed. edition. Literaturverz. S. [702] - 756.

Li, K. T. and Bell, D. R. (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics*, 197(1):65 – 75.

Lijphart, A. (2003). *"Degrees of Proportionality of Proportional Formulas." In Electoral Laws and Their Political Consequences, edited by Bernard Grofman and Arend Lijphart.* New York: Agathon.

Lijphart, A. (2012). *Patterns of democracy: Government forms and performance in thirty-six countries.* Yale University Press.

Liu, M. T., Chang, T. T. G., Loi, E. H., and Chan, A. C. H. (2015). Macau gambling industry: current challenges and opportunities next decade. *Asia Pacific Journal of Marketing and Logistics*, 27(3):499–512.

Lo, S. (2009). Casino Capitalism and Its Legitimacy Impact on the Politico-administrative State in Macau. *Journal of Current Chinese Affairs*, 38(1):19–47.

Magar, E., Trelles, A., Altman, M., and McDonald, M. P. (2017). Components of partisan bias originating from single-member districts in multi-party systems: An application to Mexico. *Political Geography*, 57:1–12.

Mair, P. (2003). *"Districting Choices under the Single-Transferable Vote." In Electoral Laws and Their Political Consequences, edited by Bernard Grofman and Arend Lijphart.* New York: Agathon.

Mair, P. and Laver, M. (1975). Proportionality, p.r. and s.t.v. in ireland. *Political Studies*, 23(4):491–500.

Maloney, S. (2015). Iran political economy since the revolution. *Contemporary Review of the Middle East*, 3(2):223–225.

McClelland, R. and Gault, S. (2017). The synthetic control method as a tool to understand state policy. *State and Local Finance Initiative. Urban Institute.*

Moin, B. (1999). Khomeini. Life of the Ayatollah. *London: I.B. Tauris.*

Penadés, A. and Santiuste, S. (2013). La desigualdad en el sistema electoral español y el premio a la localización del voto. *Revista Española de Ciencia Política*, 1(32).

Pinotti, P. (2015). The economic costs of organised crime: Evidence from Southern Italy. *The Economic Journal*, 125(586):F203–F232.

Ratto Trabucco, F. (2019). Gerrymandering Hypothesis in the Italian Constituencies: the Case of Genoa's District. *Oñati Socio-legal Series*, 9(6).

Reynolds, A. (2008). *"Reserved Seats in National Legislatures: A Comparative Approach", in Redistricting in Comparative Perspective edited by Lisa Handley and Bernard Grofman.* Oxford University Press, 2008.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. 66:688–701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. 2:1.

Sheng, M. and Gu, C. (2018). Economic growth and development in Macau (1999–2016): The role of the booming gaming industry. *Cities*, 75:72–80.

Simpson, T. (2018). Neoliberal exception? The liberalization of Macau's casino gaming monopoly and the genealogy of the post-socialist Chinese subject. *Planning Theory*, 17(1):74–95.

Stephanopoulos, N. and McGhee, E. (2015). Partisan Gerrymandering and the Efficiency Gap. *University of Chicago Law Review*, 82:831–900.

Taagepera, R. and Shugart, M. S. (1989). *Seats and votes: The efects and determinants of electoral systems*. New Haven, CT: Yale University Press.

Taylor, P. J., G., G., and J., J. R. (2003). *"The Geography of Representation: A Review of Recent Findings." In Electoral Laws and Their Political Consequences, edited by Bernard Grofman and Arend Lijphart.* New York: Agathon.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Tideman, N. (1995). The Single Transferable Vote. *Journal of Economic Perspectives*, 9(1):27–38.

Van Eck, L., Visagie, S., and De Kock, H. (2005). Fairness of seat allocation methods in proportional representation. *ORiON*, 21(2).

Vega-Bayo, A. (2015). An R Package for the Panel Approach Method for Program Evaluation: pampe. *The R Journal*, 7(2):105–121.

Wan, S. K., Xie, Y., and Hsiao, C. (2018). Panel data approach vs synthetic control method. *Economics Letters*, 164:121–123.

Wang, S. S.-H. (2016a). Three Practical Tests for Gerrymandering: Application to Maryland and Wisconsin. *Election Law Journal: Rules, Politics, and Policy*, 15(4):367–384.

Wang, S. S.-H. (2016b). Three Tests for Practical Evaluation of Partisan Gerrymandering. *Stanford Law Review*, 68:1263–1321.

Wong, S. H. W. (2019). Gerrymandering in Electoral Autocracies: Evidence from Hong Kong. *British Journal of Political Science*, 49(2):579–610.

Yee, H. S. (2001). *Macau in Transition*. Palgrave Macmillan UK.

Index