UNIVERSITAT DE BARCELONA

Essays on Public Economics and Public Policy Evaluation – Methods and Applications

Ferran A. Mazaira-Font

2023

PhD in Economics | Ferran A. Mazaira-Font

UNIVERSITAT DE BARCELONA

PhD in Economics

Essays on Public Economics and Public Policy Evaluation – Methods and Applications

Ferran A. Mazaira-Font

PhD in Economics

Thesis title:

Essays on Public Economics and Public Policy Evaluation – Methods and Applications

PhD candidate:

Ferran A. Mazaira-Font

Advisors:

Germà Bel

Daniel Albalate

Date:

April 2023

UNIVERSITAT DE BARCELONA

Essays on Public Economics and Public Policy Evaluation –

Methods and Applications

PhD. in Economics

Ferran A. Mazaira-Font

Supervisors: Profs. Germà Bel and Daniel Albalate

Acknowledgements

I would like to thank my supervisors Dr. Germà Bel and Dr. Daniel Albalate for all their help and advice with this PhD. I would also like to thank all the collaborators I have worked with during these years.

Table of contents

Chapter 1: Introduction 9

Chapter 2: Decoupling synthetic control methods to ensure stability, accuracy and meaningfulness 13

2.1. Introduction 13

2.2. The synthetic control method: an evaluation of its stability, consistency and meaningfulness 16

2.2.1. The working of the SCM 16

2.2.2. An evaluation of the SCM as a bilevel problem 18

2.3. The decoupled SHAP-distance-based synthetic control method...... 20

2.3.1. Optimization function 21

2.3.2. SHAP-based distance 22

2.3.2.a. Outcome evolution model.. 22

2.3.2.b. Shapley additive explanation values................................... 23

2.3.2.c. Feature importance and SHAP-based distance 24

2.3.2.d. Choice of the size of the restricted pool 25

2.4. Empirical illustration: the economic effects of the government formation deadlock in Spain, 2016.. 27

2.4.1. Spain's politic deadlock 28

2.4.2. Results with the standard synthetic control method.................. 29

2.4.3. Results with the decoupled SHAP-distance synthetic control ... 36

2.5. Robustness check: the German reunification and the effect of tobacco control programs in California 41

2.5.1. German Reunification 41

2.5.2. California's tobacco control program 44

2.6. Conclusion 48

2.7. Appendix 51

Chapter 3: The effect of economic, political and institutional factors on government's policy resposes to COVID-19 crisis 53

3.1. Part 1: The effect of health and economic costs on governments' policy responses to COVID-19 crisis under incomplete information 53
3.1.1. Introduction 54
3.1.2. Modeling the decision of the policy response to the crisis 56
3.1.2.a. Base model: Benevolent government with incomplete information 57
3.1.2.b. Extension 1: Different types of decision-maker 61
3.1.2.c. Extension 2: Emotions, beliefs, and political survival......... 62
3.1.3. Variables, data, and sources 63
3.1.4. Empirical model and results 71
3.1.4.a. Base model 71
3.1.4.b. Extension 1: Types of decision-maker 74
3.1.4.c. Extension 2: Emotions, beliefs, and political survival......... 74
3.1.4.d. Robustness check and final model interpretation 83
3.1.5. Discussion and policy implications.. 85
3.1.6. Conclusion 87
3.2. Part 2: Ideology, political polarization, and agility of policy responses: Was weak executive federalism a curse or a blessing for COVID-19 management in the US? 89
3.2.1. Introduction 89
3.2.2. COVID and federalism: Related literature................................ 91
3.2.3. Modelling the subnational policy response to the crisis 93
3.2.3.a. Dynamics of the decision-making process 94
3.2.3.b. Political factors involved in the decision-making process .. 95
3.2.3.c. Utility function of the decision-maker 96
3.2.3.d. Hypotheses derived from the model 99
3.2.4. Variables, data and sources 100
3.2.5. Empirical model and results 107
3.2.5.a. Effect of political affiliation under incomplete information 107

3.2.5.b. Effect of political affiliation with complete information... 112
3.2.6. Discussion and policy implications.. 115
3.3. Appendix 118
Chapter 4: Geography and Regional Economic Growth: The High Cost of Deviating from Nature 125
4.1. Introduction 125
4.2. Regional economic model 128
4.2.1. Capital markets 129
4.2.2. Household maximization problem 131
4.2.3. Equilibria with migration costs 133
4.2.4. Equilibria with spillovers and regional interdependences 134
4.2.5. Implications.. 135
4.3. Data... 136
4.4. Geographic endowment.. 139
4.4.1. Estimation 139
4.4.2. Interpretation.. 141
4.5. Analysis of population distribution by country............................ 143
4.5.1. Capital cities 143
4.5.2. Population misadjustment 145
4.6. Economic consequences of population misadjustment 150
4.6.1. Regional conditional convergence 151
4.7. Discussion... 155
4.8. Conclusion 157
4.9. Appendix 159
Chapter 5: Paying for protection: Bilateral trade with a leader and minor guys' defense spending. 161
5.1. Introduction 161
5.2. A theory of bilateral trade and military spending.......................... 165
5.2.1. Probability of war.. 166

5.2.2. Utility maximization problem 167
5.3. Data and Methods 169
5.4. Results 175
5.5. Discussion / Conclusion 181
5.6. Appendix: 183
Chapter 6: Conclusions and Policy Implications 185
Bibliography 189

Chapter 1: Introduction

According to the World Economic Forum, the total amount of data acumulated worlwide has grown exponentially during the last decade, from 2 zetabytes in 2010 to 44 in 2020[1]. It means that there are 40 times more bytes than there are stars in the observable universe. This growth in data produced and collected has caused disruption to economies, businesses, and lives around the world. It has never been easier for businesses and institutions of all sizes to collect, analyze, and interpret data into real, actionable insights. Though data-driven decision-making has existed in one form or another for centuries, not until recently has existed such an increasing pressure on decision-makers to make better use of evidence.

Governments and public institutions are not exempt from this pressure. Influencing decision-makers and contributing to policy formulation and implementation with research requires an explicit focus on power and politics, and the evaluation of its effects.

Economic policies and institutional design and decision-making vary greatly accross countries. Germany, the US and Canada, are federal states, where decision-making and economic policies are highly decentralized, while France and Greece are highly unitary countries. Belgium has had the two largest government formation deadlocks in Europe in the last 20 years, while other countries such as Portugal or Ireland have experienced none. During the COVID-19 crisis, some countries, like New Zealand, applied lockdowns with an incidence rate of 20 cases per milion inhabitants, while others like Spain, delayer their response until the incidence rate was higher than 130 cases per million.

Do differences in institutional design lead to differences in economic policies? Can these differences be explained? Is the agility of government decision-making influenced by common patterns across countries?

The aim of this dissertation is to contribute to the existing literature on public policy evaluation, with a particular focus on the role of institutions, providing new methodological, theoretical, and empirical results, to provide answers to questions such as the ones stated before.

---

[1]https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

The studies included in the dissertation are structured as follows. Firstly, I define a relevant research question and summarize the existing related literature covering topics related with the research question. Secondly, I build a new methodological or theoretical framework to understand the question under analysis. This formalization allows to identify hypothesis to be empirically tested. Thirdly, I apply different statisical techniques (according to the experimental design and the availability of data) to contrast the hypothesis. A wide variety of econometric tools are used in this dissertation, such as the synthetic control method, Bayesian models, OLS, Panel models, and GMM. To make it easier to follow, each chapter presents a short introduction of the different techniques used in the analysis, its pros and cons, and an extensive explanation on how to interpret their results. All the studies present at least two different techniques, so as to check that results are not dependant on a technical choice. Finally, I discuss the results and its policy implications.

In Chapter 2, I analyze one of the most seminals questions that could be asked about governments and economic outcomes: Do government formation deadlocks affect the economy in the short term? From the methodological point of view, I develop a proposal to improve current methodologies to evaluate causal effects on quasi-experimental designs; concretely, the Synthetic Control Method. I illustrate the main advantages of the proposal evaluating the causal economic effects of the ten-month-long government formation impasse in Spain, after the December 2015 elections, as well as reproducing two previous studies: the impact of German reunification (analyzed in Abadie et al. 2015) and the effect of tobacco control programs in California (Abadie et al. 2010). In line with the results obtained by Albalate and Bel (2020) for the 18-month government formation deadlock in Belgium, my estimates indicate that the growth rate in Spain was not affected by the government deadlock, ruling out any damage to the economy attributable to the institutional impasse.

Chapter 3 focuses on how governments decide in a context of high uncertainty and different degrees of information. Concretely, I build a theoretical model to assess the agility of government response to the COVID pandemic and evaluate the model empirically using data from OCDE and European countries. I find solid evidence that during the first outbreak, in a context of incomplete information, the agility of policy response was highly conditioned by a cost-benefit analysis where the perceived healthcare capacity to deal with the outbreak, and the associated economic costs of lockdown measures, significantly delayed the response. Institution design also played a role: federal states reacted faster than unitary ones. Higher competition in multilevel systems with collaborative governance between different levels of government and non-state institutions - (Scavo, Kearne, and Kilroy, 2008; Schwartz and Yen, 2017; Downey and Myers, 2020; Huang, 2020) provided incentives for more agile and effective responses. However, federal states could be dysfunctional in terms of internal coordination and suffer from high inequality in terms of agility within themselves. For the concrete case of the US, I find that Republican-controlled states reacted later and implemented softer contingency measures, which were associated with higher growth in the number of COVID-19 cases (Hallas et al., 2020; Shvetsova et al., 2022). The highly polarized context of the US provided incentives for Republican governors to align with President Trump's preferred policy, which was to avoid lockdowns. These incentives vanished during the vaccination process, when information about the severity of COVID-19 was complete, and governors, no matter whether Republicans or Democrats, implemented the roll-out of the vaccination program with a similar level of agility.

In Chapter 4, I suggest a new approach to assess the effect of institutional and policy developments (i.e. capital city) on economic growth that distort the natural equilibrium of the geographical distribution of the labor market. I propose a theoretical model of the way in which features of geography and nature can account for population density and distribution within a country. The model is empirically examined using data from comparable European regions. This allows us to detect deviations produced by the forces of human action, led mainly by institutions, and to evaluate the consequences in terms of relative economic performance. The results suggest that deviating from nature's outcomes has a significant negative effect on economic growth and regional convergence. Hence, societies that choose to exploit the opportunities of the best locations, according to the natural endowment, rather than promoting a different distribution of the population across regions by means of institutional intervention, achieve better economic performance.

In Chapter 5, we focus on the most relevant government expenditure until the twentieth century: military expenditure. We examine the effects of military and trade alliances in military expenditure. We develop a theoretical model to understand why these alliances could influence military expenditure. In short, when countries build military and trade alliances with military leaders such as the US, they make themselves more valuable to the leader, and hence increase the likelihood of the leader providing military aid in case of an agression. This increases the military costs of a potential agresor, reduces the probability of war and let the non-leader country reduce its military expenditure. To empirically test the hypothesis derived from the model we employ data of 138 countries for the period 1996-2020. Our results show that trade relation with a military leader is a highly significant driver of military expenditure. For each percentage point in US GDP in trade between a certain country and the US, the military expenditure of the country reduces 0.5 percentage points. Moreover, when the trade balance is particularly beneficial for the US, the effect is even larger.

Finally, Chapter 6 presents the conclusions of the different studies and their policy implications.

Chapter 2: Decoupling synthetic control methods to ensure stability, accuracy and meaningfulness

This chapter is a joint work with Germà Bel and Daniel Albalate, and consists of a paper published in the SERIEs - Journal of the Spanish Economic Association: Albalate, D., G. Bel, and F.A. Mazaira-Font. (2021). Decoupling Synthetic Control Methods to ensure stability, accuracy and meaningfulness. SERIES, Journal of the Spanish Economic Association 12(4), 549-584 doi.org/10.1007/s13209-021-00242-8

Abstract: The synthetic control method (SCM) is widely used to evaluate causal effects under quasi-experimental designs. However, SCM suffers from weaknesses that compromise its accuracy, stability and meaningfulness, due to the nested optimization problem of covariate relevance and counterfactual weights. We propose a decoupling of both problems. We evaluate the economic effect of government formation deadlock in Spain-2016 and find that SCM method overestimates the effect by 0.23 pp. Furthermore, we replicate two studies and compare results from standard and decoupled SCM. Decoupled SCM offers higher accuracy and stability, while ensuring the economic meaningfulness of covariates used in building the counterfactual.

2.1. Introduction

Since the seminal works of Abadie and Gardeazábal (2003) and Abadie et al. (2010), the synthetic control method (SCM) has been increasingly adopted as a technique to evaluate causal effects under quasi-experimental design (see, among others, Montalvo 2011; Billmeier and Nannicini 2013; Cavallo et al. 2013; Kleven et al. 2013; Bohn et al. 2014; Percoco 2015; Acemoglu et al. 2016; Kreif et al. 2016; Albalate and Bel 2020; Sun et al. 2019). The method provides a practical solution to the evaluation of case studies in which either only a single unit or very few aggregate units are treated (countries, regions, cities, etc.) and it is considered one of the most influential recent contributions to empirical policy evaluation (for instance, Athey and Imbens 2017, p. 9). The SCM creates a hypothetical counterfactual (the synthetic unit) by taking the weighted average of pre-intervention outcomes from selected donors (control units). The impact of treatment is quantified by the simple difference between the treated unit and its synthetic cohort after the treatment (post-treatment period).

As discussed in a series of papers by its pioneering authors (see Abadie and Gardeazábal 2003; Abadie et al. 2010, 2015; Abadie and L'Hour 2019), the SCM has two main advantages over other methods, such as regression-based counterfactuals or nearest-neighbor matching. First, by being constrained to nonnegative weights that need to sum one, it does not impose a fixed number of matches and ensures sparsity, while avoiding negative weights or weights greater than one that would imply an unchecked extrapolation outside the support of the data and complicate the interpretation of the estimate. Second, weights are calculated to minimize the discrepancies between the treated unit and the synthetic control in the outcome and the values of certain matching variables or covariates. Thus, the SCM is intended to ensure that the synthetic unit reproduces the control unit not only in terms of the outcome, but also in terms of the drivers that explain the evolution of the outcome of the treated unit before treatment.

In spite of the influential contribution made by the SCM, the method suffers from some weaknesses that, if not properly addressed, may erode the reliability and robustness of its causal estimates and, consequently, of its policy implications. For instance, Ferman et al. (2020) have highlighted that lack of guidance on how to choose covariates gives researchers specification-searching opportunities that directly influence the choice of comparison units and therefore the signification of the results. Abadie (2020) also pointed out that even assuming a proper set of covariates and a counterfactual that matches the treated unit, interpolation biases may arise if this matching is obtained by averaging donors that have large differences in covariates but compensate each other to match the treated unit. As stated by Albalate et al. (2020), the bilevel optimization design of the SCM and its NP-hard nature helps to explain why quasi-experimental methods for estimating covariate importance under the SCM are unstable and highly dependent on the donor pool, thus affecting weight estimation.

The contribution of this paper is twofold. First, we develop a proposal of decoupling synthetic control methods, to overcome the limitations of the bilevel design of the SCM. Our approach is simpler and more operational, since it breaks down the NP-hard problem of the nested optimization into two independent problems of quadratic optimization with linear constraints. The method we propose ensures robustness of the estimation of both covariate importance and the weights. By decoupling the estimation of covariate importance from that of weights, it minimizes interpolation biases and guarantees economic sense. To estimate covariate importance, we use a new methodology for estimating feature importance suggested by Lundberg and Lee (2017; 2019): SHapley Additive exPlanation (SHAP) Values. This method allows us to analyze the marginal effects and average contribution of the different features of a model, even in the case of nonparametric models. Thus, we can obtain sound estimates for each unit of the relation of the different covariates with the outcome and define a distance between the donor pool and the treated unit in terms of how covariates influence the outcome. To estimate weights, the procedure we use minimizes quadratic error in the pre-treatment outcome, restricting the donor pool to the most similar units to the treated unit. Roughly speaking, we obtain a synthetic control that is the benchmark that best reproduces the pre-treatment outcome and whose behavior is explained by the same factors that explain the treated unit.

Second, to illustrate the main advantages of our proposal, we apply both methods to an evaluation of the causal economic effects of the 10-month-long government formation impasse in Spain, after the December 2015 elections. In line with the approach taken by Albalate and Bel (2020) for the 18-month government formation deadlock in Belgium, we use the SCM to build an appropriate counterfactual to identify and isolate the gap between Spain's actual GDP per capita growth rate and the rate at which it would have grown without a government formation deadlock. Our results indicate that the growth rate was not affected by government deadlock, ruling out any damage to the economy attributable to the institutional impasse. Moreover, as a robustness check of the advantages of the decoupled synthetic control method, we use our methodology to reproduce two previous studies: the impact of German reunification (analyzed in Abadie et al. 2015) and the effect of tobacco control programs in California (Abadie et al. 2010).

The rest of this paper is organized as follows: First, we describe the standard SCM and we evaluate its stability, consistency and economic meaningfulness. In light of the limitations identified, in Sect. 2.3 we propose a new decoupled SHAP-distance synthetic control method (DSD-SCM) that overcomes the limitations of the standard SCM. In Sect. 2.4, we apply both methods to the estimation of the causal economic effects of a long government formation deadlock in Spain between December 2015 and

October 2016. We discuss the findings, focusing on the magnitude of the differences between the two methods (SCM vs. DSD-SCM), the advantages of the DSD-SCM, and the economic implications of the impasse. In Sect. 2.5, we present the replication of two case studies, as a robustness check of the improvements of our methodology with respect to the original synthetic control. Concretely, we replicate the analysis of the impact of German Reunification and of the effect of the tobacco control program in California. In Sect. 2.6, we offer our main conclusions and considerations about the new method proposed.

2.2. The synthetic control method: an evaluation of its stability, consistency and meaningfulness

The synthetic control method builds a counterfactual of a specific treated unit as a weighted average of a number of control units (the so-called donor pool), to reproduce what would have been its performance if it had not been exposed to the treatment and to identify, by its difference with respect to reality, the causal effect of the policy. In this section, we first describe the main features of the SCM, and then, we evaluate its consistency and stability.

2.2.1. The working of the SCM

The SCM assumes there are J control units and observations during T periods (pre-treatment). Let X_{TU} be a $(K \times 1)$ vector of the outcome growth predictors of the treated unit (the covariates). Let $X = (X_1, \dots, X_J)$ be a $(K \times J)$ matrix which contains the values of the same variables for the J possible control units. Both X_{TU} and X could include pre-treatment observations of the dependent variable. Let V be a diagonal matrix with nonnegative components reflecting the relative importance of the different growth predictors. Let Y_{TU} be a $(T \times 1)$ vector whose elements are the values of the outcome of the treated unit for the T periods, and $Y = (Y_1, \dots, Y_J)$ a $(T \times J)$ matrix whose elements are the values of the outcome of the control units. Then, the counterfactual is built as YW^*, where $W^* = (w_1^*, \dots, w_J^*)$ is a $(J \times 1)$ vector containing the weights of the control units in the counterfactual. W^* is chosen to minimize the objective function $D(W) = (X_{TU} - XW)'V(X_{TU} - XW)$, subject to $w_i \geq 0$ and $\sum_{i=1}^{J} w_i = 1$.

V is chosen as $V = argmin_{V \in \mathrm{V}}(Y_{TU} - YW * (V))'(Y_{TU} - YW * (V))$, where V is the set of all nonnegative diagonal $(K \times K)$ matrices, whose Euclidean norm is one. Notice that V is the key element for determining W^* and avoiding interpolation biases, since it defines the relative importance of the adjustment of each covariate in the counterfactual.

Several contributions have recently been made aimed at extending the scope of use of the SCM and improving its accuracy and robustness. As regards the former, Powell (2018) suggested a way to estimate policy effects when the outcomes of the treated unit lie outside the convex hull of the outcomes of the other units. Since the treated unit may be part of a synthetic control for a non-treated unit, the post-treatment outcome differences for these units are informative of the policy effect. In recent studies, the SCM has been extended to contexts with disaggregated data, where samples contain large numbers of treated and untreated units, and interest lies in the average effect of the treatment among the treated (see Abadie and L'Hour 2019). Building synthetic controls for each of the treated units as opposed to a synthetic control for the average treated unit has been proposed in order to minimize interpolation biases.

To increase SCM accuracy and robustness, studies have addressed three issues: the role of covariates, the estimation of weights, and the best way to gauge the uncertainty of the estimated treatment effect. As regards the first of these, Doudchenko and Imbens (2016), Gobillon and Magnac (2016) and Kaul et al. (2015) showed that high accuracy can only be achieved if lagged outcomes are included as covariates. However, by so doing, other covariates may become irrelevant, which could lead to interpolation bias if the set of pre-treatment outcomes is not long enough (Botosaru and Bruno (2019)), or if there is an imperfect pre-treatment fit (Arkhangelsky et al. 2018). Ferman et al. (2020) have also highlighted that this lack of guidance on how to choose covariates gives researchers specification-searching opportunities that directly influence the choice of comparison units and the signification of the results. Indeed, they showed that with few pre-treatment periods (between 10 and 30), a researcher would have substantial opportunities to select statistically significant specifications even when the null hypothesis is true. Moreover, Klobner et al. (2015) showed that the current SCM suffers from high numerical instability in covariate importance and weights.

Studies of the estimation of weights have proposed different strategies to reduce interpolation biases. Hastie et al. (2009), and Hastie et al. (2015)

combined a Lasso and Ridge regularization to capture a preference for a small number of nonzero weights, as well as for smaller weights. Likewise, Abadie and L'Hour (2019) introduced a penalization parameter that trades off pairwise matching discrepancies with respect to the characteristics of each unit in the synthetic control against matching discrepancies with respect to the characteristics of the synthetic control unit as a whole.

Finally, to gauge the uncertainty of the estimated treatment effect, the SCM compares the estimated treatment effect with the "effects" estimated from placebo tests in which the treatment is randomly assigned to a control unit (see Abadie and Gardeazábal 2003). Building multiple synthetic controls by leaving countries out of the optimal control has also been proposed (Abadie and L'Hour 2019). In this regard, Xu (2017) proposed a parametric bootstrap procedure to obtain confidence intervals of the estimates of the treatment effect.

Following Albalate et al. (2020), in the next subsection, we show that the bilevel design of the SCM is at the root cause of the main concerns related to its stability, consistency and meaningfulness.

2.2.2. An evaluation of the SCM as a bilevel problem

The SCM is characterized as a bilevel problem. Such problems are optimization problems (upper level) that contain another optimization problem as a constraint (lower level).

Definition 1

For the upper-level objective function $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and lower-level objective function $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ the bilevel problem is given by

$$\min_{x_u \in X_u,\ x_l \in X_L} F(x_u, x_l)$$

$$x_l \in \operatorname*{argmin}_{x_l \in X_L} \{f(x_u, x_l) : g_j(x_u, x_l) \leq 0, j = 1, \dots, J\}$$

$$G_k(x_u, x_l) \leq 0, k = 1, \dots, K$$

where $G_k\colon X_U \times X_L \to \mathbb{R}, k = 1, \dots, K$, denote the upper-level constraints, and $g_j\colon X_U \times X_L \to \mathbb{R}$ represent the lower-level constraints, respectively. Equality constraints may also exist that have been avoided for brevity.

Figure 2.1 illustrates a general bilevel problem. Given a x_u vector, x_l^* is the optimal lower-level vector for the lower-level optimization. But, as seen in the figure, the solution (x_l^*, x_u) is not optimal for the upper-level optimization given x_l^*.

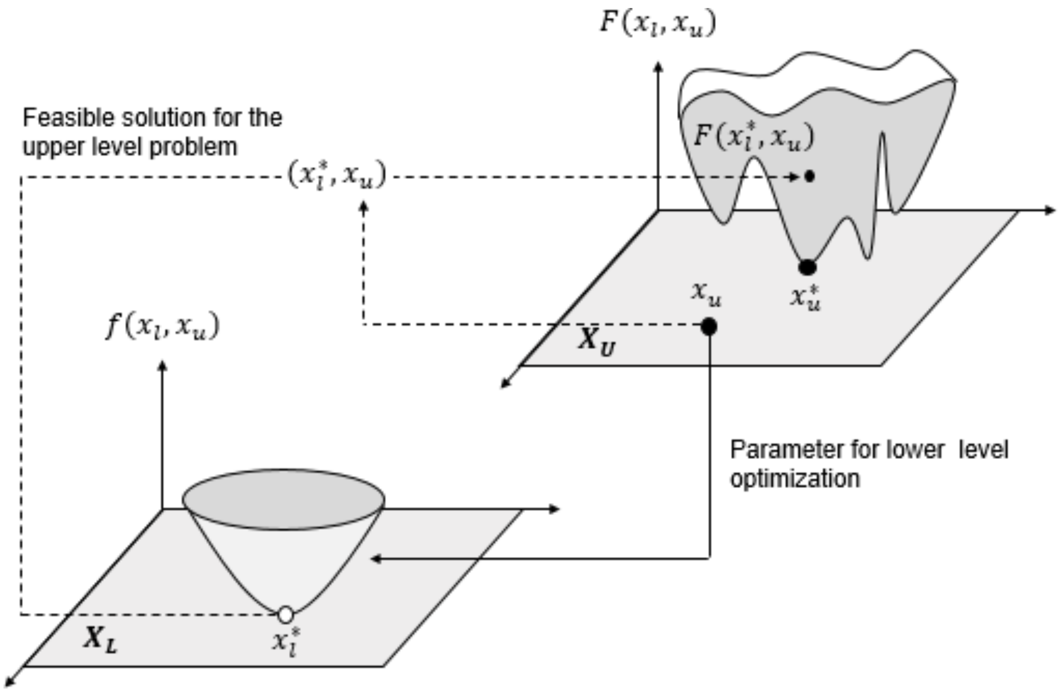

Figure 2.1. A general sketch of a Bilevel problem

The SCM proposed by Abadie et Gardeazabal is a bilevel optimization problem of the form:

$$\min_{V,W}(Y_{TU} - WY)'(Y_{TU} - WY)$$

$$W \in \operatorname*{argmin}_{W} \{(X_{TU} - WX)'V(X_{TU} - WX)\colon w_j \geq 0, \sum_{i=1}^{J} w_i = 1\}$$

$$v_j \geq 0, \sum_{i=1}^{K} v_j = 1$$

Bilevel programming is known to be strongly NP-hard (Hansen et al. 1992), and it has been proven that merely evaluating a solution for optimality is also a NP-hard task (Vicente et al. 1994). Moreover, the hierarchical structure may introduce difficulties such as non-convexity and disconnectedness (that is, that the solution set can be separated into two disjoint sets) even for simpler instances of bilevel optimization, which may cause solutions to be highly unstable to small perturbations and the algorithm to converge to different local optima.

In the particular case of the SCM method, the flaws of bilevel optimization imply that the solution V can be completely arbitrary and highly unstable to small perturbations. As a result, weights are also unstable and V does not offer reliable insights in terms of economic meaningfulness since it can be driven by interpolation biases. In Appendix I (supplementary materials), we illustrate the aforementioned flaws with two simple examples.

In Sect. 2.6, we present an empirical assessment of the numerical instability for the case study of Spain's government deadlock. As seen there, current implementation of the SCM method can lead to very unstable results just by removing or adding to the donor pool units that are given no weights in the synthetic control. Although this is clearly counterintuitive and should not be possible, it is due to the fact that the implementation of the optimization problem is done using an interior point method (Abadie et al. 2011). That is, weights can be given values close to zero, but not zero. Thus, although the final result is presented as 0, the real value for the algorithm could be of the order of 10^{-7} or 10^{-8} (depending on the margin parameter given to the function). Therefore, removing units with zero weight in the solution is equivalent to introduce a very small perturbation, which, as we have showed, can be devastating in terms of optimal parameters and goodness of fit.

2.3. The decoupled SHAP-distance-based synthetic control method

The aim of this section is to propose and present a modification of the SCM that can guarantee economic meaningfulness and the stability of feature importance, at the same time as it increases the robustness of the estimation of weights and treatment effect. Our proposal is coined as the decoupled SHAP-distance synthetic control method (DSD-SCM) and is designed as an operational alternative to the use of the SCM that involves less complexity than the standard approach due to the NP-hard nature of bilevel optimization and guarantees higher stability and economic sense.

In the previous section, we showed that the minimization problem of SCM is defined over covariates and that feature importance estimation is nested to weights, potentially leading to considerable instability and a lack of economic meaningfulness. Therefore, we propose decoupling feature importance from weight estimation by defining the optimization problem of the SCM as a minimization of the error in the pre-treatment outcome adjustment, conditional to using units that are as similar as possible to the treatment unit. As hightlighted by Abadie (2020), donors' similarity to the treated unit is one of the most critical requirements for the synthetic method to be an appropriate tool for policy evaluation. Hence, we also present a concrete methodology for feature estimation and unit similarity that guarantees economic sense and stability, the regularized SHAP-based distance. However, other distances or previous expert knowledge on the feature importance would also be worth considering.

2.3.1. Optimization function

Let us note by $d(X_{TU}, X_i)$ a distance between the treated unit and the unit i, dependent on their respective vector of covariates X_{TU}, Xi. The vector of weights W^* in our modified method is chosen as

$$W^* = \underset{W}{\operatorname{argmin}} \sum_{t=1}^{T} \left(Y_{TU}^t - W\tilde{Y}^t\right)^2 \} \tag{2.1}$$

subject to $W > 0$ and $\sum_{i=1}^{l} w_i = 1$. $\tilde{Y}^t$ is a vector that contains the outcome in time t of the top L most similar units to the treated unit. This optimization problem is a minimization of a quadratic and positive-definite function with linear constraints. The number L of units is chosen to balance the potential trade-off between a pure minimization of the adjustment error and the similarity to the treated unit. We require all the units entering the synthetic control to be similar to the treated unit so as to minimize interpolation biases (as suggested in Abadie et al. (2015), Abadie (2020)). Roughly speaking, this is equivalent to saying that, for example, what most resembles a medium-size house is not the average of a small and a big house, but the average of two medium-size houses.

As we will see next, in our procedure the distance function is not linked to the weights, which in the SCM contributed to increasing instability and reducing economic meaningfulness, but determined independently through an econometric model that involves another quadratic minimization.

Notice that the choice of L is not uniquely determined and depends on several conditionings. For example, the stronger the relation between the covariates and output evolution, the more sense it makes to choose a lower value of L. In the next section, we present a method for assessing the importance of L and for choosing an adequate value.

2.3.2. SHAP-based distance

Intuitively, we would like to consider that a unit is similar to the treated unit if their outcomes evolved in a similar way before the treatment and for similar reasons. For example, a 99% correlation in the evolution of GDP per capita between two countries would tell us nothing about their similarity if one has an economy based on natural resources that grew because of a hike in petrol prices, whereas the other's growth was attributable to manufacturing exports. In short, to define a distance between units it is critical we understand the relationship between their outcome and their covariates. To do so, we propose the following methodology. First, we build a model of the evolution of the outcome using the covariates as explanatory variables. Second, we use one of the newest and most popular methods of model interpretation to estimate the average marginal contribution of each feature to each prediction of the model: the SHapley Additive exPlanation or SHAP values. Finally, by estimating the SHAP values, we are able to define a distance based on feature importance and average contributions to outcome evolution.

2.3.2.a. Outcome evolution model

Let us note the growth rate of unit i by $g_i^t = (Y_i^t - Y_i^{t-1})/Y_i^{t-1}$, where $i \in \{1, \ldots, J, TU\}$. Recall that Y_{TU} is the $(T \times 1)$ vector containing the values of the outcome for the treated unit, and $Y = (Y_1, \ldots Y_J)$ the $(T \times J)$ matrix with values of the outcome for the control units.

Let us consider $G(X_i^s \mid s \in \{1, \ldots, t\})$ a model for g^t, that is $G(X_i^s \mid s \in \{1, \ldots, t\}) = g_i^t + \varepsilon_t$, , where ε_t is the error term at time t. Notice that G is a model that depends on covariates, and for which no concrete functional form is required. It could be a linear model, but also a nonlinear and even a nonparametric model, such as a gradient boosting tree[2]. It may also include past information from covariates. It is important to highlight that the robustness, stability and consistency of the variable importance are inherited by the properties of the modeling technique G. For instance, if we use OLS, under the hypothesis of homoscedasticity, no autocorrelation, and normality of the error terms, estimates are consistent, efficient and non-biased. As we will see in the empirical illustrations in Sects. 2.4 and 2.5, the decoupling

[2] Gradient boosting trees are models that combine into a single prediction a sequence of models, called base learners, in which each subsequent base learner focuses on the residual error of the previous base learners. Often, these base learners are decision trees or stumps.

ensures that the importance attributed to each variable is stable and practically does not change even if we add or remove countries from the donor pool, while it is highly unstable for the original SCM.

2.3.2.b. Shapley additive explanation values

SHAP values have been proposed as a unified framework for assigning feature importance to parametric and nonparametric models (Lundberg and Lee 2017 and Lundberg and Lee 2019). Roughly speaking, given an instance x, the SHAP value of feature i on x corresponds to the marginal impact of feature i on the output of the model, with respect to other instances that share some of the features with x but not i. Formally, let us consider the subset S of the set of input variables V and $G_x(S) = E[G(x)|x_S]$ the expected value of the model G conditioned on the subset of input features S. Then, SHAP values are the combination of these conditional expectations:

$$\phi_i(x) = \sum_{S \subset V \setminus \{i\}} \frac{|S|!\,(|V| - |S| - 1)!}{|V|!} [G_x(S \cup \{i\}) - G_x(S)]$$

where the combinations are needed because for nonlinear functions the order in which features are introduced matters. For a linear model $G(x) = \sum_{i=1}^{k} \alpha_i x_i$, the SHAP value is straightforward: $\phi_j(x) = \alpha_j(\hat{x}_j - E(x_j))$. Notice that model estimation, which is required for the SHAP value, is based on a quadratic minimization, which consists of the second quadratic problem of the DSD-SCM.

Figure 2.2 shows how the SHAP values explain the output of a function f as a sum of the effects ϕ_i of each feature being introduced into a conditional expectation.

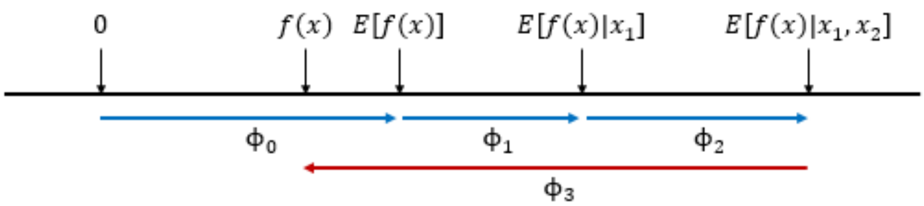

Figure 2.2. SHAP values explain the output of a function f as a sum of the effects ϕ_j of each feature being introduced into a conditional expectation

2.3.2.c. Feature importance and SHAP-based distance

Let us note by $\phi_m(X_1^t)$ the SHAP Value of the covariate m for the treated unit at time t. Then, we can estimate the relative importance of the covariate m in the outcome evolution of the treated unit, RI_m, as:

$$RI_m = \frac{\sum_{t=1}^{T} |\phi_\mathrm{m}(X_1^t)|}{\sum_{k=1}^{K} \sum_{t=1}^{T} |\phi_\mathrm{k}(X_1^t)|}$$

where J is the total number of covariates and T the total number of observations.

Therefore, we can define V as the diagonal matrix such that $V_{m,m} = RI_m$. This matrix has economic sense, because it is exactly estimating the importance of each covariate on the outcome evolution of the treated unit before the treatment. It is also stable in the sense that it relies on the stability of parameter estimation or model inclusion of the different variables. Thus, features whose relation with outcome is less robust will tend to not be considered (for example, discarded in a linear model if their p value is lower than 0.1 or 0.05) or be assigned lower relevance.

Having estimated V, let define AC_m^i the average contribution of feature m in outcome evolution for unit i:

$$AC_m^i = \frac{\sum_{t=1}^{T} \phi_\mathrm{m}(X_1^t)}{T}$$

Then, we can define the SHAP distance between non-treated unit U_i and the treated unit TU as:

$$d_S(U_i, TU) = (AC^i - AC^{TU})'V(AC^i - AC^{TU}) \tag{2.2}$$

where $AC^i = (AC_1^i, \dots, AC_J^i)$ and $AC^{TU} = (AC_1^{TU}, \dots, AC_J^{TU})$ are the vectors containing the average contributions of the covariates for unit U_i and the treated unit TU.

2.3.2.d. Choice of the size of the restricted pool

Given L, let us note by $W(L)$ the solution of (1) and by $R^2(L)$ as the R−squared[3] of the synthetic control $Y(L) = W(L)\tilde{Y}$. Let us note by $l \leq L$ the number of units that have a positive weight in $W(L)$. Let us define the covariate distance between the treated unit and $Y(L)$ as

$$d(L) = \frac{\sum_{i=1}^{l(L)} d(X_{TU},\ X_i)}{l(L)}$$

Since units are ordered by similarity, given $L_1 > L_2$,, we have that $d(L_1) \geq d(L_2)$ and $R^2(L_1) \geq R^2(L_2)$. The higher the number of units, the larger the distance and the higher the goodness of fit. Hence, $d(J) \geq d(L)$ and $R^2(J) \geq R^2(L)$ for $L \in 1, \ldots, J$. Notice that, in particular, $W(J)$ is equivalent to the constrained regression synthetic method suggested in Doudchenko and Imbens (2016).

Let us define also the error loss of L as the ratio

$$EL(L) = \frac{1 - R^2(L)}{1 - R^2(J)}$$

and the similarity gain as

$$SG(L) = \frac{d(J) - d(L)}{d(J)}$$

The error loss is the ratio between the error of the counterfactual $Y(L)$ and the best potential counterfactual in terms of goodness of fit, $Y(J)$. The similarity gain captures the relative increase in similarity between $Y(L)$ and the treated unit, with respect the similarity between $Y(J)$ and the treated. The lower the EL the better, since this means the goodness of fit is near to the maximum possible. Likewise, the higher the SG the better, since this means that the countries in the synthetic control are closer to the treated unit. As we stated in Sect. 3.2.1, L sets a threshold in terms of goodness of fit with respect to similarity loss (or, conversely, in terms of error loss with respect similarity

3 R−squared is defined as in a linear model: $R^2(W) = 1 - \frac{(Y_{TU}-WY)'(Y_{TU}-WY)}{(Y_{Tu}-\bar{Y}_{TU})'(Y_{TU}-\bar{Y}_{TU})}$, where $\bar{Y}_{TU}$ is the mean value of the outcome of the treated unit in the pre-treatment period. Notice that R−squared is not necessarily defined between 0 and 1, since there is no constant and W is nested to V to solve the covariate adjustment. Hence, R−squared could even be negative, since the adjustment of Y_{TU} by WY could be worse than using $\bar{Y}_{TU}$.

gain) for a new unit to be considered in the counterfactual. Therefore, we propose finding the optimal value of L as the value L^*, which minimizes the ELSG (error loss–similarity gain) ratio. That is:

$$L^* = \underset{L \geq L_v}{\operatorname{argmin}} \frac{EL(L)}{SG(L)}$$

where L_v is the minimum L such that $R2(L_v) > \nu R^2(J)$. That is, values of L that ensure at less a certain level of goodness of fit, to prevent degenerated cases where the distance is almost no related with outcome (for $L = J$, the ELSG ratio is defined as ∞). We recommend using ν=0.9 or 0.95. By doing so, it is guaranteed that the goodness of fit of the DSD-SCM will be equal to or higher than that of the SCM, unless the SCM uses all pre-treatment outcomes as the only covariates, and control units that do not resemble the treated unit in the underlying drivers of the outcome. As highlighted in Abadie (2020), the asymptotic bias of the SC estimator should be small in situations where one would expect to have a close-to-perfect fit for a large pre-treatment period. Hence, ensuring that pre-treatment fit is (in general) at least as in the original SC ensures also that any bias in the estimation of the treatment effect is expected to be lower.

Notice that, in comparison with other extensions that limit the donor pool for regularization purposes, such as the best subset selection procedure described in Doudchenko and Imbens (2016), our restriction is primarily linked to similarity. Therefore, in our method, the role of the distance is to offer practical guidance for the applied researcher on the reliability of the estimates, specially in cases where the pre-treatment period is not that large, and in which could be biases in the estimation. For example, if the goodness of fit of the outcome model is low, similarity between units is expected to be less reliable, the number L is expected to be closer to J, and therefore, the researcher may include additional covariates to minimize the risk of interpolation biases, or be more cautious about the conclusions derived from the estimation of the treatment effect. In particular, this reduces the risk of using specification search, because it makes it clear whether the similarity function between units is accurate or not.

Although we have presented here the SHAP distance, any other distance that ensures economic and statistical sense might be used. The main contribution of our proposal is the decoupling of the minimization involved in the original SCM, which helps to prevent instability and biases in the estimates.

2.4. Empirical illustration: the economic effects of the government formation deadlock in Spain, 2016

Although hardly new, lengthy government formation processes in parliamentary regimes after a general election are becoming more usual in Europe. In the last two decades, there have been seven cases of government formation deadlocks lasting more than threemonths: six months in Belgium after the June 2007 election, eighteen months after the June 2010 election and sixteen months after the May 2019 election, both of them again in Belgium, ten months in Spain after the December 2015 election and ten months after the April 2019 election, seven months in the Netherlands after the March 2017 election, and six months in Germany after the September 2017 election.

Contrary to widespread claims that government deadlocks and the associated political instability harm a country's growth by disrupting economic policies that might otherwise promote better performance (Alesina et al. 1996; Angelopoulos and Economides 2008; Aisen and Veiga 2013), studies of recent impasses provide evidence that this might not always be the case. Using the SCM to build an appropriate counterfactual to reproduce Belgium's economic growth if it had had a full-powered government, Albalate and Bel (2020) reported a nonnegative effect on economic growth during the 18 months of government deadlock in that country following the June 2010 election. The study suggests that certain characteristics peculiar to Belgium could be behind this (perhaps) counterintuitive result. First, the country's highly decentralized multilevel governance, which assigns a considerable number of functions and powers to the communities and regions, at the same time as the European Union's institutions have absorbed some of the core functions performed by conventional Member States (Bouckaert and Brans 2012; Hooghe 2012). Second, the existence of robust, efficient institutions, outside government, that played a positive role in protecting the economy from the difficulties of the impasse. Third, the delay in fiscal consolidation that could have caused higher short-term economic growth than might otherwise have been expected.

2.4.1. Spain's politic deadlock

The general election held in Spain on December 20, 2015, resulted in a fragmented political landscape following the emergence of two new political parties: Podemos left-wing) and Ciudadanos (Cs) (right-wing). In spite of winning the election, the Partido Popular (PP) (right-wing), who ruled Spain with an absolutemajority between 2011 and 2015, lost 63 seats and got 123 seats, far from the 176 needed for the majority. Due to the numerous corruption cases in which leading members of the PP were then embroiled, the other main right-wing party, Cs, refused to facilitate a rightwing government and offered their votes to the Partido Socialista Obrero Española (PSOE). Together the two parties controlled 130 of the chamber's 350 seats and needed either the support of Podemos (69 seats) or the abstention of the PP. Neither of the two requirements was met, and fresh elections were held in June 2016. The 2016 election results reinforced the position of the PP, which won 14 additional seats, totaling 137. However, it was still not enough to form a government. After twomonths of negotiations, Cs (along with the Coalición Canaria, a right-wing regional party in the Canary Islands) announced their support for Mariano Rajoy, PP's candidate for the Presidency. With 170 votes and the controversial abstention of the PSOE, Rajoy was re-elected President of Spain on October 29, 2016, ending a ten-month-long deadlock.

Despite this period of impasse and the limited powers of a caretaker government, Spain's economic performance did not appear to suffer greatly. Indeed, even the Spanish Central Bank (Banco de España) published an article in 2017 estimating the negative effect of the political uncertainty of the previous year at just 0.1% of GDP, although this result was not statistically significant (see Gil, Pérez and Urtasun, 2017). If we observe the GDP growth rate (Fig. 2.3), Spain's performance during 2016 was slightly higher than the EU average, and better than the euro area average, as it had been in 2015. However, as Albalate and Bel (2020) discuss in their evaluation of the 18-month government deadlock in Belgium, this comparison tells us only how Spain's performance compared to that of the other countries of Europe, but it offers no causal insights as to how it might have performed had it had a full-powered government.

Thus, we need to build a counterfactual to reproduce how Spain would have performed in the absence of its government formation deadlock.

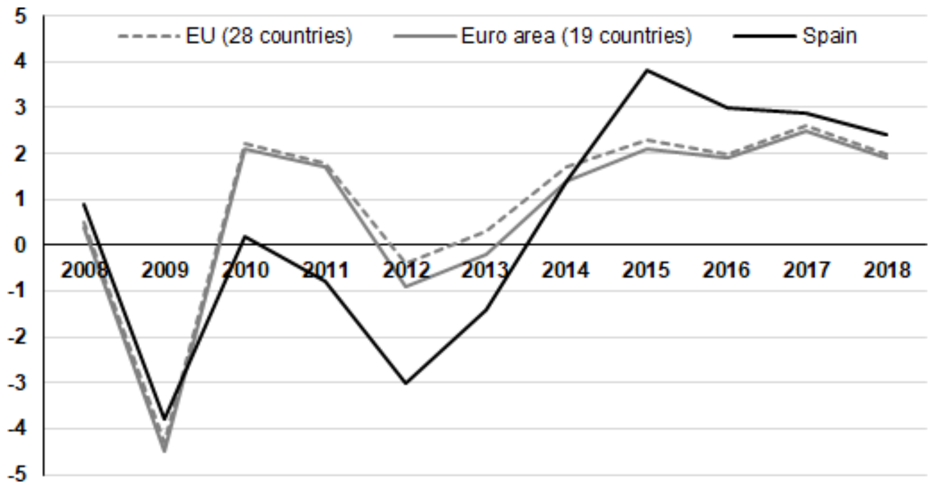

Figure 2.3. Real Gross Domestic Product per capita growth rate (2008-2016)

2.4.2. Results with the standard synthetic control method

To evaluate the robustness and meaningfulness of the SCM and the advantages of implementing our proposedDSD-SCMalternative, we compare the estimates provided by the two methods of the causal effects of this political deadlock. First, we apply the standard SCM, with and without outcome lags in the covariates, to show that in both cases covariate importance is highly unstable, highly dependent on the donor pool and lacking in economic meaningfulness. Second, we implement our proposed SHAP-distance based synthetic control method to show how this approach addresses and avoids the main weaknesses of SCM, providing more stable, accurate and meaningful estimates.

The donor pool used in the comparison includes a sample of the EU-28 countries. Malta and Luxembourg had to be excluded given the amount of missing data for some of the key predictors used in the analysis. Belgium was excluded since it was also affected by a lengthy government deadlock between 2010 and 2011, and Ireland because of the marked change in GDP pc in 2014 (26.3% growth rate) due to the reallocation of the intellectual property of large multinational firms.

Tables 2.1 and 2.2 report the pre-treatment values of several variables typically associated with a country's growth potential and used as covariates, as well as their relative importance, for the casewithout and with lagged outcomes. Table 2.3 presents the weight matrix for the donor pool, where the synthetic weight is the country weight assigned to each country. When the lagged outcomes are not included, the synthetic Spain is made up of the four main contributors: Portugal (33.5%), France (30.7%), Greece (23.3%), and Italy (12.0%). Finland also plays a role, but only a minor one (0.4%). When using this counterfactual to predict Spain's GDP per capita from 2001 to 2015, R2 is 92.60% and the mean absolute percentage error (MAPE) is 0.64%.When initial and final outcomes are included, the results are quite similar. The main contributors remain the same, although their relative importance changes. The minor role played by Finland disappears, and instead, Denmark (3.7%) and Sweden (2.6%) enter the synthetic control. The goodness of fit improves slightly (R2 = 93.44%) and the MAPE remains unchanged at 0.64.

Figure 2.4 shows the GDP per capita evolution of the real and synthetic Spain built with and without lagged outcomes. In both cases, the growth rate during the 2016 deadlock was around 1.8 percentage points (p.p.) higher than expected, while in 2017 and 2018 the gap was reduced to 0.4 and 0.1 p.p., respectively.

Table 2.1: Covariate means and importance without including lagged outcome

Covariate	Spain	Synthetic	Pool	Importance
Openness	57.03	60.13	98.00	54.1%
Low education	50.12	48.12	28.04	26.0%
High education	27.28	18.46	21.44	1.9%
Trade surplus	-1.37	-3.74	-0.51	12.8%
Unemployment	15.90	10.78	9.21	4.2%
Investment	24.02	20.81	22.31	1.1%
Debt	61.41	98.88	56.68	0.0%

Average 2001-2015

Table 2.2: Covariate means and importance including initial and final outcome

Covariate	Spain	Synthetic	Pool	Importance
GDP per capita2001	22,190.00	22,120.28	19,306.36	57.2%
GDP per capita2015	23,080.00	22,523.79	22,645.00	30.1%
Low education	50.12	48.83	28.04	7.5%
High education	27.28	18.41	21.44	0.0%
Openness	57.03	62.88	98.00	3.7%
Unemployment	15.90	10.81	9.21	0.8%
Trade surplus	-1.37	-3.79	-0.51	0.4%
Investment	24.02	20.75	22.31	0.1%
Debt	61.41	97.39	56.68	0.1%

Average 2001-2015

Table 2.3: Percentage weight vector with and without including lagged outcome

Country	Weight without lags	Weight with lags	Country	Weight without lags	Weight with lags
Austria	0	0	Hungary	0	0
Bulgaria	0	0	Italy	12.0	8.8
Croatia	0	0	Latvia	0	0
Cyprus	0	0	Lithuania	0	0
Czech. R.	0	0	Netherlands	0	0
Denmark	0	3.7	Poland	0	0
Estonia	0	0	Portugal	33.5	37.6
Finland	0.4	0	Slovakia	0	0
France	30.7	21.8	Slovenia	0	0
Germany	0	0	Sweden	0	2.6
Greece	23.3	25.4	UK	0	0

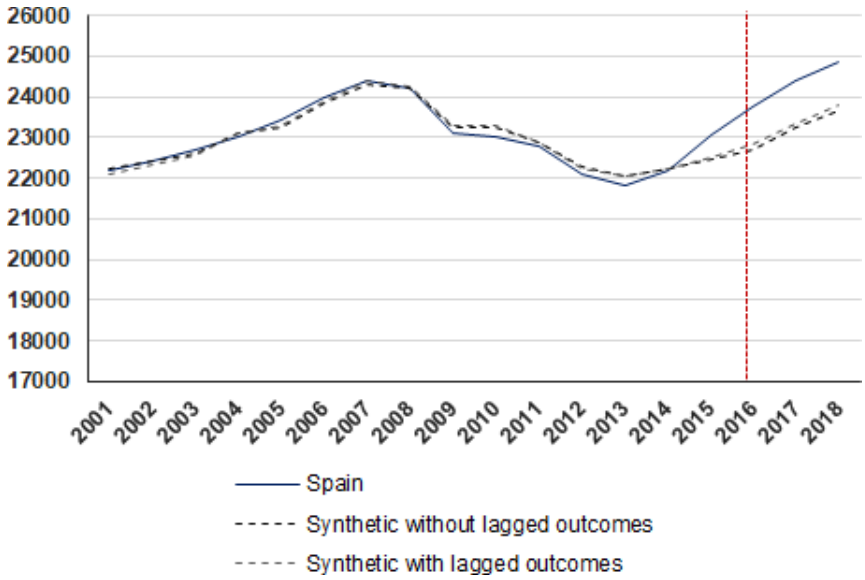

Figure 2.4: GDP per capita evolution: real vs synthetic Spain

To evaluate the robustness of the SCM, two placebo tests have been widely used: in-time and in-space. In the former, the SCM is applied considering that the treatment occurred in an earlier timeframe (i.e., the treatment is reassigned to occur during the pre-treatment period) and so the control is built using observations up to this new moment in time. The test examines the uncertainty associated with making a prediction after the last observation considered for the estimation. In the in-space test, the SCM is applied to the control units as if they too had been treated at the same moment of time as the treated unit. Hence, it tests the uncertainty associated with the volatility of outcomes of the control units during the treatment.

However, neither of these two tests evaluates the stability of covariate importance and weights, and their economic foundations, which are the main ideas on which the SCM is based. Even if the methodology passes the in-time and in-space placebo tests, it would be difficult to rely on its results if, for example, V had no relation with economic theory. Moreover, if a placebo test fails to confirm robustness, it is unable to tell us whether it is because the treatment had no significant effect or because the methodology was not properly applied and its accuracy could be improved (for example, by adding new covariates).

Here, therefore, we analyze the stability and economic meaningfulness of covariate importance and weights and show that SCM does not guarantee any of them, even in those cases when the methodology passes the placebo tests. As Fig. 2.4 shows, in this particular case, including lagged outcomes has almost no impact in terms of the goodness of fit and the estimation of the treatment effect. Nonetheless, as recognized elsewhere, including Doudchenko and Imbens (2016), Gobillon and Magnac (2016) and Kaul et al. (2015), variable importance is greatly affected, and the covariates become almost irrelevant (Tables 2.1 and 2.2). This might not, however, be important in terms of the economic foundations and robustness of V. If the covariates reflect the economy drivers, they could also be influencing lagged outcomes in the sense that countries with more similar values for these covariates would also have similar outcomes. This means that the importance of the covariates could be hidden behind the lagged outcomes; the only problem being that the inclusion of lagged outcomes makes it almost impossible to gain any economic insights from V and to judge whether the estimations are the result of an interpolation bias.

However, an analysis of V shows that its estimation is neither consistent with the economic foundations nor is it stable. First, if we turn our attention to the SCM without lagged outcomes, the most important variables are openness (54.1%) and low education (26.0%), while unemployment, investment and debt have almost no influence (4.2%, 1.1% and 0%, respectively). The Spanish economy's cumulative growth per capita in real terms from 2000 to 2007 was 2.8 p.p. higher than that of the euro area, driven mainly by exceptionally high levels of investment due to the housing bubble (Akin et al. 2014). Total investment in Spain averaged 27.7% of GDP from 2000 to 2007, 5.2 p.p. higher than in the euro area. Once the crisis began, investment dropped significantly, reaching a minimum of 17.4% in 2013, almost 13 p.p. lower than its maximum in 2006. In the euro area, the fall in investment was much lower: from a maximum of 23.4% in 2007 to a minimum of 19.7% in 2013. Unemployment more than tripled, from 8.2% in 2007 to a maximum of 27% in the first quarter of 2013, the highest level in the euro area. As a result, Spain's public debt almost tripled, growing from 35.8% of GDP in 2007 to 99.3% in 2015. In the euro area, however, the increase was much lower, from 65.9 to 90.8%.

Thus, it has no solid economic foundations to devise a similarity measure with respect to Spain that assigns no importance to debt, unemployment and investment, while at the same time assigning almost 70% of the importance to the degree of openness and the percentage of the population with a low education. Openness, for example, remained largely stable before and during the crisis. In the period 2000–2007, imports and exports accounted for 56.2% of GDP, while in 2008–2015 they represented 58.2%. In conclusion, openness and low education levels are not assigned a high level of importance because they are the main drivers of Spain's economy, but rather because a number of countries whose real GDP per capita evolution correlated highly with Spain's presented similar levels of openness and low education.

Secondly, because of the interpolation bias, covariate importance, weights and goodness of fit are highly unstable and dependent on the donor pool (as in Klobner et al. (2015)). Table 2.4 shows the average importance and standard deviation for 100 simulations after removing three countries from the donor pool that were assigned no weights in the synthetic Spain. The standard deviation is higher than 50% of the average importance estimation for almost all covariates, both with and without lagged outcomes. As a result, the distance between the synthetic and real Spain is modified and, so, the weights are adjusted accordingly (Table 2.5). Yet, weight instability may not necessarily compromise the SCM. Indeed, it might just be the result of the fact that the donors are so similar to each other that a small perturbation in V modifies the selection of one of them into the control. However, as Table 2.6 shows, the goodness of fit is significantly affected for the SCM without lagged outcomes and slightly affected for SCM with lagged outcomes.

Table 2.4. Variable importance stability

	No lagged outcomes		With lagged outcomes	
Covariate	Mean	St. Dev	Mean	St.Dev
Debt	2.40 %	4.14%	0.28 %	0.53%
Unemployment	5.57%	2.83%	1.86%	1.85%
Openness	41.3%	22.54%	11.40 %	12.16%
Investment	3.38%	2.85%	0.73%	1.18%
Trade	14.36%	15.52%	3.51%	4.91%
Low education	29.12%	18.06%	12.30 %	13.57%
High education	3.85%	3.13%	0.87%	2.02%
GDP per capita 2001	-	-	40.99	22.11%

			%	
GDP per capita 2015	-	-	28.03 %	10.50%

Notes: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool.

Table 2.5. Unit weights stability

	No lagged outcomes		With lagged outcomes	
Country	Mean	St. Dev	Mean	St.Dev
Austria	0	0	0	0
Bulgaria	0	06	0	0
Croatia	1.17	4.04	1.44	3.27
Cyprus	0	0	0	0
Czech R.	0	0	0	0
Denmark	0	0	1.11	3.47
Estonia	0.47	1.95	0	0
Finland	12.14	14.15	0.35	1.36
France	22.49	16.97	24.16	14.27
Germany	0	0.01	0.09	0.69
Greece	23.38	5.38	22.65	5.68
Hungary	0	0	0	0
Italy	3.90	7.27	11.89	11.84
Latvia	0	0.02	0	0
Lithuania	0	0	0	0
Netherlands	0.05	0.41	0	0
Poland	0.05	0.41	0.03	0.24
Portugal	36.33	10.04	34.07	4.64
Slovakia	0.02	0.13	0.47	2.11
Slovenia	0	0	0	0
Sweden	0	0.02	3.13	4.86
United Kingdom	0	0	0.60	3.45

Notes: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool

.

Table 2.6. Goodness of fit stability

	No lagged outcomes		With lagged outcomes	
Measure	Mean	St. Dev	Mean	St.Dev
R^2	81.42%	14.03%	91.17%	2.84%
MAPE	1.16%	0.43%	0.77%	0.17%

Notes: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool.

Given the high instability of the goodness of fit without lagged outcomes, the SCM does not pass the in-time placebo test using 2012 as the treatment (Fig. 2.5). However, the same does not hold true for the SCM with lagged outcomes. Yet, in both cases, the placebo test fails to provide any information as to why the methodology works properly or not, or whether it can be improved.

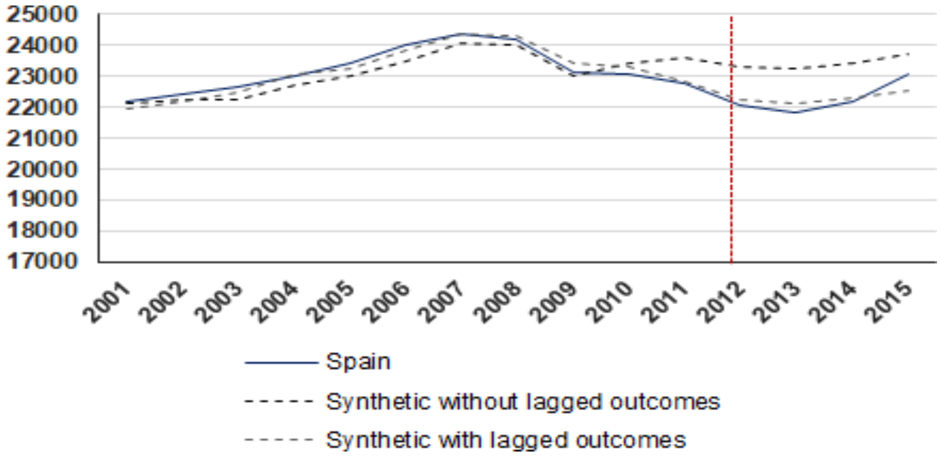

Figure 2.5. Placebo test in time

In conclusion, we have shown, first, that covariate importance may not be consistent with economic theory and provide no meaningful insights; second, that this lack of meaning is due to interpolation biases that make estimations highly unstable and dependent on irrelevant countries (i.e., countries with no weight) in the donor pool; and, third, that although including lagged outcomes may make the results more robust in terms of goodness of fit, it does not solve the problem of meaning and stability of covariate importance. Moreover, it also tends to make the other covariates irrelevant, thus compromising the main idea behind the SCM. Finally, we have also shown that standard robustness checks, such as the in-time placebo test, may be unable to identify these flaws and to suggest any strategy to improve the results.

2.4.3. Results with the decoupled SHAP-distance synthetic control

In this subsection, we build the synthetic Spain adhering to the strategy described in Section 2.3, that is we build a model of GDP per capita growth, define a distance using SHAP values, select a regularization parameter and estimate optimal weights. We consider a linear model of the GDP per capita growth rate from 2001 to 2015, using as our explanatory variables the covariates used in the previous subsection and all the countries in the donor pool including Spain. The results are presented in Table 2.7 (variables with gr indicate growth rates of the covariate). Note that while the covariates are able to explain 73.33% of the variation in economic growth, around 25% of the variation remains unexplained. Thus, a synthetic control that relies solely on the covariates would not be sufficiently accurate or robust.

Table 2.7. Real GDP per capita growth model (OLS)

Variable name	Estimate	Std. Error	t value	Pr(> |t|)
Constant	3.722e-02	4.128e-03	9.016	< 2e-16
invest gr	3.569e-03	1.052e-03	3.391	0.00078
openness gr	1.692e-03	1.704e-04	9.930	< 2e-16
gdp pc lag1	-9.058e-07	9.574e-08	-9.461	< 2e-16
debt gr	-1.705e-03	2.070e-04	-8.236	4.24e-15
unemp gr	-6.046e-03	8.767e-04	-6.897	2.73e-11
educlow	-1.945e-04	6.552e-05	-2.969	0.00321
educhigh	2.786e-04	1.331e-04	2.093	0.03709

Notes: Multiple R-squared: 0.7333. F-statistic: 129.2 on 7 and 329 DF, p-value: < 2.2e-16

Table 2.8 shows the feature importance of the different covariates of economic growth in Spain and the donor pool. According to the DSD-SCM results, Spain's economic evolution has been characterized primarily by high levels of unemployment and debt growth. Conditional convergence has had a much lower impact on Spain than it has had on the donor pool, mainly because Spain's GDP was already very close to the average of the selected countries (a 6% difference, on average, during the period). Using covariate importance, we define the distance between countries in the donor pool and Spain as in (2.2), but normalizing to be between 0 and 1. The corresponding results are presented in Table 2.9.

Table 2.8. Covariate importance

Covariate	Importance Spain	Importance pool
Unemployment	30.03	17.43
Debt	27.78	15.74
Openness	11.94	21.48
Investment	11.48	11.09
Low education	11.13	4.50
High education	4.19	4.39
Conditional convergence (GDP lag1)	3.44	25.37

SHAP feature importance during 2001-2015

Table 9. Donor pool ordered by distance (normalized to 100)

Country	Distance	Country	Distance
Cyprus	3.22	Hungary	26.22
Greece	8.53	Germany	31.71
Italy	9.78	Sweden	39.03
France	11.98	Latvia	40.05
Portugal	14.52	Czech R.	40.15
United Kingdom	16.76	Denmark	51.24
Slovenia	18.61	Estonia	52.98
Croatia	22.15	Lithuania	54.20
Austria	22.76	Poland	54.82
Finland	24.18	Slovakia	64.58
Netherlands	25.25	Bulgaria	100.00

Based on the ELSG ratio (as described in 2.3.2.), the optimal L is 6. Restricting the donor pool to the six most similar countries reduces the number of units in the counterfactual from six (case with no restriction) to four, almost halves the distance with respect to the treated unit (from 0.21 to 0.11) and implies a loss of only 0.45 p.p. in R^2 with respect to the counterfactual that uses all the units in the donor pool (96.84% vs 96.39%). It is worth pointing out that even in the case of no regularization, the average distance of countries in the synthetic control is much lower than the average distance of those in the donor pool (0.21 vs. 0.33). This means that the more similar countries are to Spain, the more likely they are to be selected.

When using the DSD-SCM with L^*=6, the counterfactual consists of Portugal (34.5%), UK (27.5%), Italy (19.4%), and Greece (18.2%). The R^2 is 96.39% and the MAPE 0.42%. Notice that this counterfactual uses fewer countries than the standard method and obtains between 3 and 4 additional percentage points in R^2. Hence, the DSD-SCM ensures greater economic meaningfulness of feature importance and achieves better results while reducing the number of parameters.

Figure 2.6 shows the different counterfactuals we have built. In all cases, growth in 2016 was higher than expected, lying in a range between 1.58 p.p. (DSD-SCM) and 1.81 p.p. (SCM without lagged outcomes). Thus, SCM overestimates the effect of the Spanish deadlock by a 0.23 p.p. However, our primary goal was to provide a more robust method. As Table 10 shows, the covariate importance estimates are highly stable, with standard deviations of 2 p.p. As a result, in all simulations, the same countries are selected and assigned the same weights.

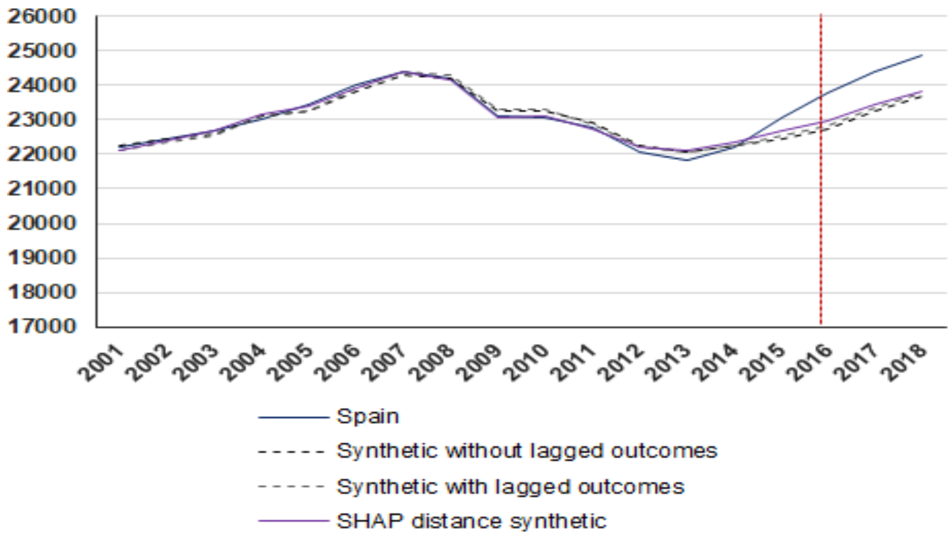

Figure 6. GDP per capita evolution: real vs. synthetic Spain

Table 10. Feature importance stability

Covariate	Mean	St. Dev
Unemployment	29.81%	2.42%
Debt	28.29%	1.43%
Openness	11.59%	1.49%
Investment	11.30%	1.86%
Low education	10.90%	1.38%
High education	4.61%	0.47%
Conditional convergence (GDP lag1)	3.49%	2.11%

Note: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool.

Finally, Fig. 2.7 shows the results of the in-time placebo test. In the case of the in-space placebo test, we excluded countries whose MAPE for 2001–2015 was three times higher than Spain's. Thus, countries with an MAPE greater than 1.2% were excluded when we compared the base model to the best placebos (with eight countries surviving). The comparison showed a difference in the average treatment effect in 2016 for placebo countries of -0.006 p.p. in the growth rate and of 0.92 p.p. in the standard deviation. The treatment effect for Spain is estimated at 1.58, which is higher than 0 at a 7.8% confidence level, assuming a normal distribution of the placebo estimates.

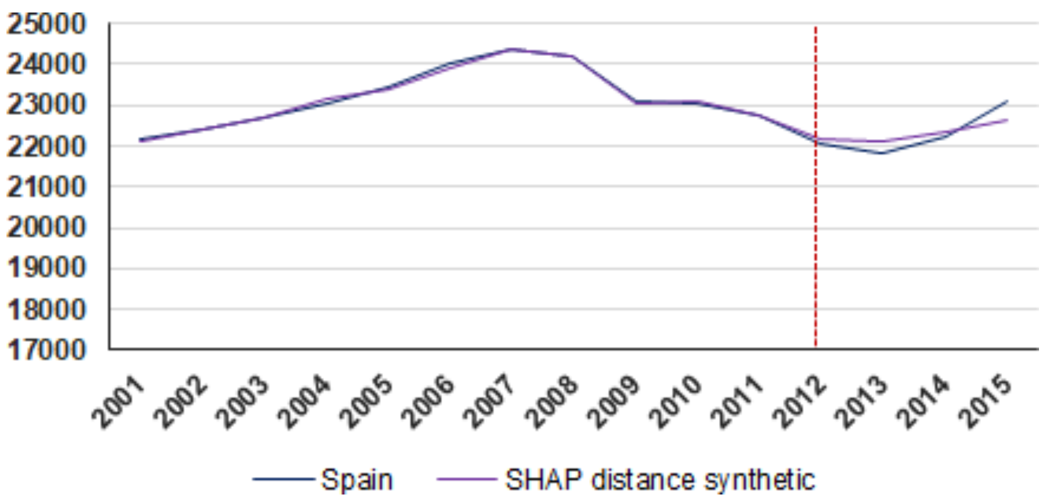

Figure 2.7. Placebo test in time for SHAP distance synthetic Spain

2.5. Robustness check: the German reunification and the effect of tobacco control programs in California

2.5.1. German Reunification

Abadie et al. (2015) evaluate the impact of the German Reunification in 1991 on GDP per capita, considering 1971 to 1980 as the pre-treatment period and 16 OECD countries as the donor pool. They considered five covariates: trade openness, inflation rate, industry share, schooling levels, and invest ment rate. We replicate the analysis using the DSD-SCM.

As investment rate and schooling levels are given on 5-year basis, we build the model of GDP growth in 5-year basis, starting from 1970. As can be seen in Table 2.11, only inflation and initial GDP (conditional convergence) are relevant. Thus, including other covariates would not ensure capturing similarity in economic growth dynamics, since those drivers are not statistically relevant. The relative importance for West Germany for inflation is 11.4% and for conditional convergence 88.6%. This explains why GDP has to be used as a covariate in Abadie et al. (2015): only by including conditional convergence a reasonable fit in terms of outcome can be achieved.

Table 2.11. Real GDP per capita growth model (OLS)

Variable name	Estimate	Std. Error	Pr(> $\|t\|$)
Constant	1.120e-01	2.645e-02	7.88e-05
Inflation	-1.522e-03	4.613e-04	0.00162
Trade	-2.356e-05	6.440e-05	0.71574
Schooling	-1.976e-04	1.407e-04	0.16532
InitialGDP	-4.202e-06	5.899e-07	1.43e-09
Industry	3.118e-04	4.526e-04	0.49348
Investment	4.739e-02	3.687e-02	0.20351

Notes: Multiple R-squared: 0.6834, F-statistic: 21.95 on 6 and 61 DF, p-value: 1.432e-13

The best ELSG ratio is achieved with $L = 7$. The counterfactual consists of 40.4% Austria, 33.3% USA, 19.8% Netherlands, and 6.5% France. The MAPE is 0.64%. In comparison, the counterfactual in the original paper consists of 6 countries (42% Austria, 22% USA, 16% Japan, 11% Switzerland, and 9% Netherlands) and its MAPE is 0.85%. Notice that the counterfactual with the decoupled method has less units and higher goodness of fit. Both counterfactuals share 3 units, which account for 93.5% of the importance in the decoupled and 73% in the original synthetic method. Remarkably, France, which is the 6th most similar country to West Germany according to the similarity distance in Abadie et al. (2015), does not belong to the original counterfactual, while it has a 6.5% weight in the DSD-SCM. Japan and Switzerland, which have a positive weight in the original counterfactual, are among the less similar countries (10th and 13th, respectively). Thus, the original counterfactual is build with countries that the method considers not to be similar to the treated unit. If we applied the decoupled method with the original distance, we get that the best ELSG ratio is achieved with L=9 units, instead of seven, and the counterfactual would be exactly the same as the one built with the SHAP distance. As explained in Sect. 2.3.2., the larger the number of similar units needed for the second step of the decoupled method, the lower the reliability of the similarity measure. Hence, decoupling the problem allows to identify the lower economic sense of the distance built in the nested optimization.

Figure 2.8 shows the results. Both methods estimate a clear negative impact, but they differ in the estimation. Concretely, GDP per capita in West Germany in 2003 was 4283.5 USD lower that it would have been without the Reunification according to the DSD-SCM, while the difference accounted for 3379.3 USD in the original estimation. It is worth mentioning that other extensions of the synthetic control method, such as the constrained regression or the best–subset (Doudchenko and Imbens 2016) discussed in Sect. 2.2, also find a negative impact, although lower than the estimated by the decoupled synthetic. Concretely, the impact in year 1995 for these extensions lays between 790 and 1019 USD, while in our case is 1301 USD.

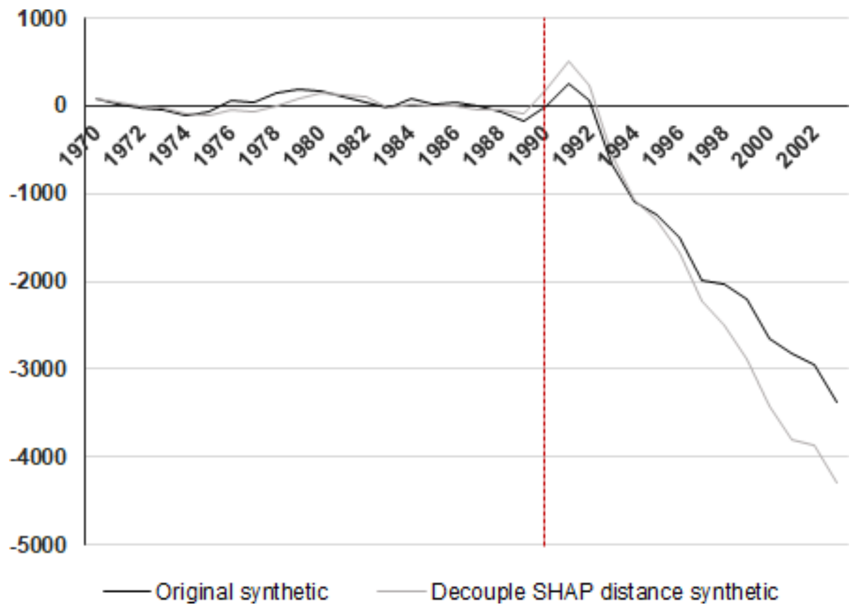

Figure 2.8. Per capita GDP gap between West Germany and Synthetic West Germany

Moreover, results obtained with the decoupled synthetic method are also robust to placebo test in time and space, as can be seen in Figs. 2.9 and 2.10. The placebo test in time considered a placebo re-unification effect from 1985 to 1990. For the placebo in space, all units with a MAPE lower than three times that of the decoupled synthetic were considered.

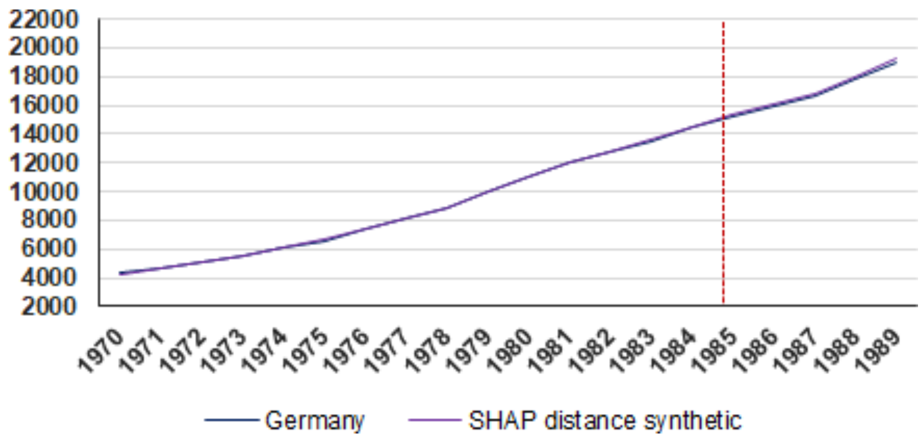

Figure 2.9. Placebo test in time for the German re-unification

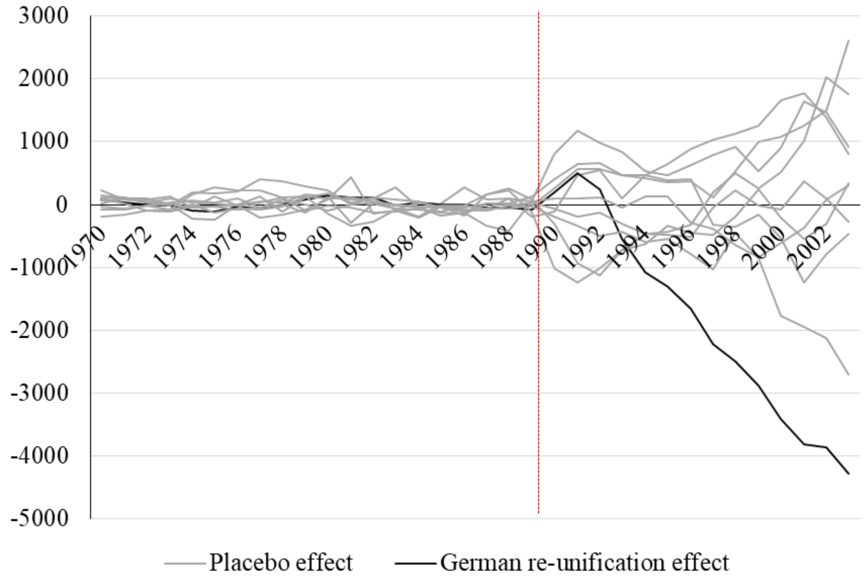

Figure 2.10. Placebo test in space for the German re-unification

In conclusion, the comparison between the original synthetic control method and our decoupled version in the German re-unification shows that while both methods get a similar conclusion, the decoupled ensures higher economic sense and stability in the similarity measure between countries (for instance, using the distance metric of the original synthetic would require the top 9 most similar countries instead of the top 7 to get the same goodness of fit) and achieves higher goodness of fit.

2.5.2. California's tobacco control program

Abadie et al. (2010) estimate the effect of California's Tobacco Control Program implemented in 1988, in terms of per capita cigarettes sales. They use annual state-level panel data from the period 1970 to 2000. They consider the following covariates: income per capita (in natural logarithm), beer consumption, percentage of population aged 15–24, and the average retail price of a cigarettes pack. In order to increase the goodness of fit of the counterfactual, they are also forced to include three lagged outcomes: cigarettes sales in 1975, 1980 and 1988. The donor pool consists of 38 states where no control program or cigarettes' tax raise was implemented during the period of analysis. The counterfactual build using the SCM consists of 5 states: Utah (33.4%), Nevada (23.4%), Montana (19.9%), Colorado (16.4%), and Connecticut (6.9%). Remarkably, 95% of the covariate importance is given to previous lagged outcomes. Indeed, authors highlight that the counterfactual does not reproduce covariates such as GDP per capita because they are given a very small weight, meaning that it does not have substantial power predicting the per capita cigarette consumption. Actually, none of the covariates is given more than 3% weight, which compromises the main idea of covariates. That is why, for instance, the two most relevant states in the counterfactual, Utah and Nevada, are among the less similar to California, according to the distance estimated in the study. Concretely, they are ranked the 34th and 35th out of 38, respectively. Hence, a counterfactual is built where covariates are not relevant (only lagged outcomes) and the units are not similar to the treated unit.

We reproduce the analysis by using the DSD-SCM. Table 2.12 shows the estimated parameters for the model of cigarettes consumption per capita (in natural logarithm), using only non-lagged outcome covariates. The relative importance based on SHAP values for the covariates is: 55.7% for retail price, 31.8% for income per capita, 10.4% for the percentage of young people, and 2.0% for beer consumption.

Table 2.12. Cigarettes consumption per capita evolution model (OLS)

Variable name	Estimate	Std. Error	Pr(> $|t|$)
Constant	2.981060	0.963366	0.002270
lnincome	0.329826	0.090611	0.000351
Beer	0.009114	0.002922	0.002099
Age15to24	-4.090836	1.732094	0.019199
Price	-0.009874	0.001264	3.78e-13

Notes: Multiple R-squared: 0.35, F-statistic: 25.58 on 4 and 190 DF, p-value: < 2.2e-16

The best ELSG ratio is achieved with a restricted pool of the 30th most similar states. Notice that a large number of donor states is needed, indicating that the distance metric might not be accurate. The counterfactual consists of 40.3% Utah, 21.0% Nevada, 16.5% Montana, 9.0% Colorado, 9.0% Nebraska, and 4.2% New Hampshire. The MAPE is 0.98%, compared to a 1.05% in the original synthetic. Both counterfactuals share 4 states, which account for around 90% of the total relevance, although the relative weights are slightly different. However, the choice of the states in this second counterfactual is more meaningful. For instance, Utah is the 15th most similar according to the SHAP-based distance, instead of the 34th. As can be seen in Fig. 2.11, both methods estimate that by the year 2000 annual per capita cigarette sales in California was about 26 packs lower than what they would have been in the absence of the program, although the impact is slightly larger (0.55 packs) according to the decoupled synthetic.

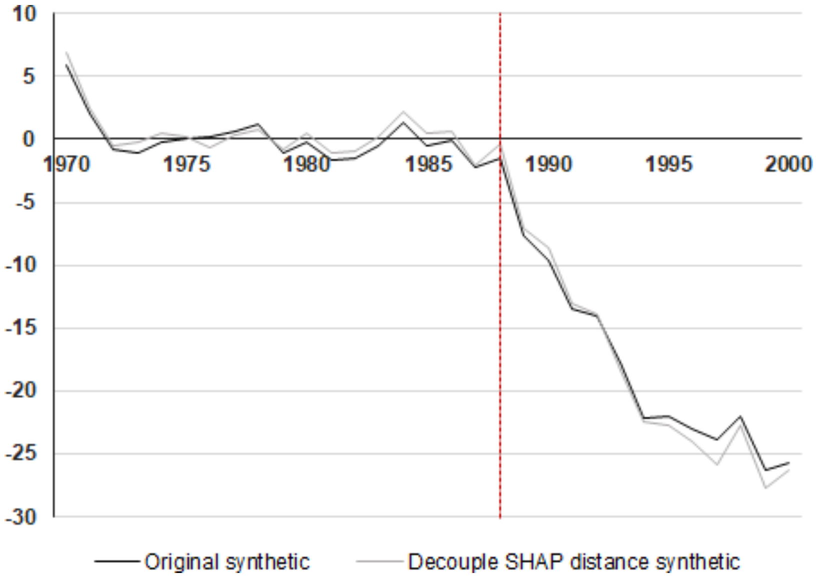

Figure 2.11. Per capita cigarettes consumption gap between California and Synthetic California

Finally, figures 2.12 and 2.13 show the results of the placebo tests. While the placebo in space shows that the decrease in California was clearly larger than in the other states, as in the original study and other extensions (see Doudchenko and Imbens 2016), the placebo test in time shows a 7% decrease in 4 years, even when no legislation was passed. Hence, it suggests that, at least partially, the effect observed in California, although significantly larger than in the rest of states, might be caused by other factors than the tobacco program. It is important to highlight that in the original paper by Abadie, Diamond and Hainmueller, no placebo test was provided. However, when applying the original methodology, the placebo test also fails, and even more than in the decoupled version (10% decrease, instead of 7% with the decoupled), due to the poor performance of the similarity measure, as discussed before. Figure 2.14 show the results.

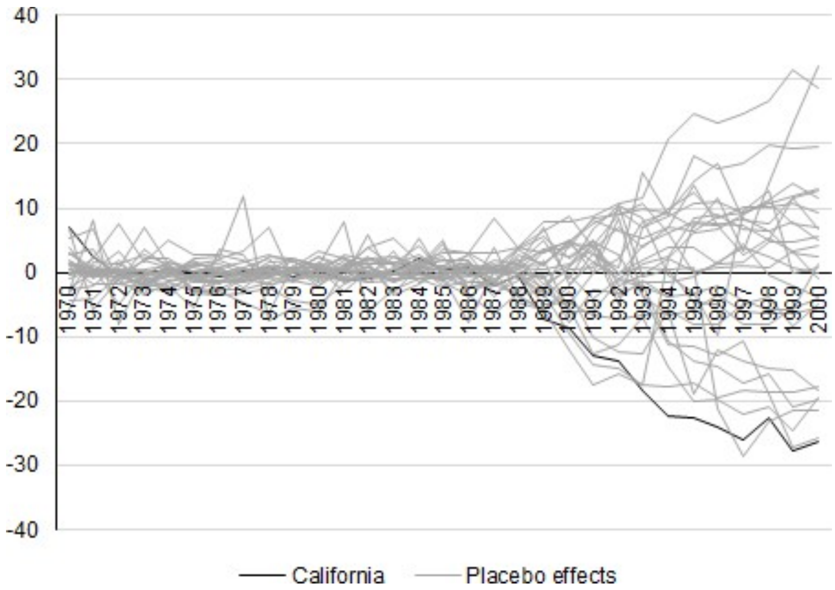

Figure 2.12. Placebo test in space for the California tobacco program

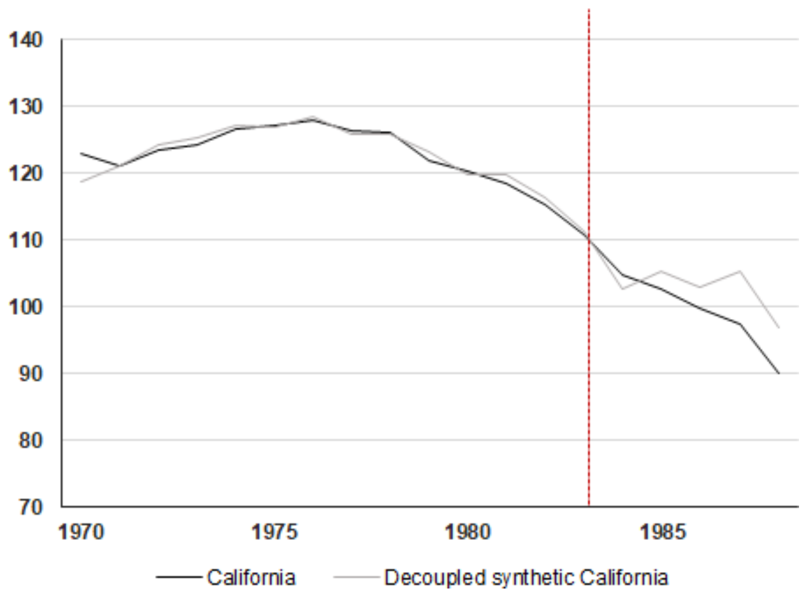

Figure 2.13. Placebo test in time for the California tobacco program

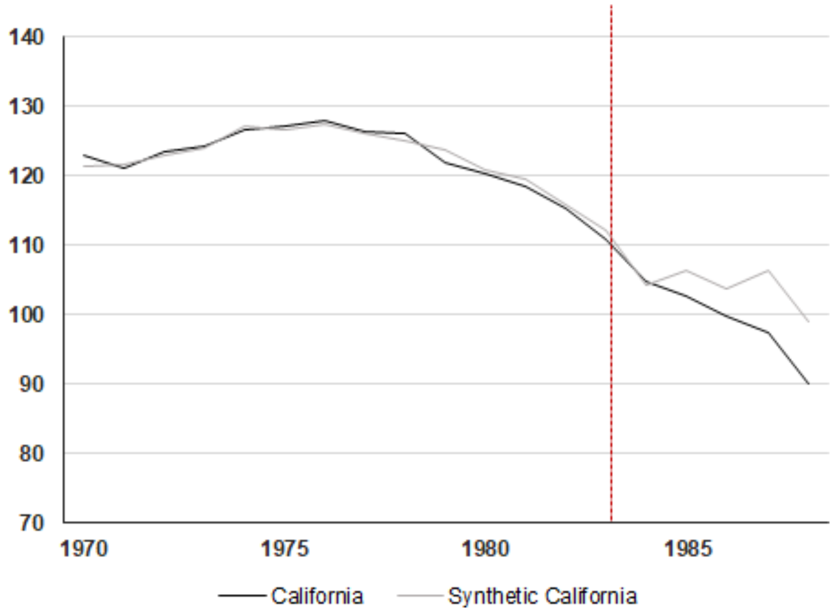

Figure 2.14: Placebo test in time for the California tobacco program with the original Synthetic Control Method

As in the German reunification case, the decoupled synthetic method ensures higher economic sense and stability in the similarity measure between countries and achieves higher goodness of fit. Furthermore, it gives a clear guidance on why results should be taken cautiously, as seen in the placebo test in time. Because the similarity metric is not highly accurate, units that are not similar to the treated unit have to be considered as part of the counterfactual to get a proper goodness of fit of the pre-treatment period.

2.6. Conclusion

The synthetic control method has been an influential innovation in quasi-experimental design, combining as it does elements of matching and difference-in-differences, and providing a systematic approach to building a counterfactual. Similarly, it offers new opportunities for evaluating causal treatment effects in single—or in very few—aggregate units of interest. The method's impact on the empirical policy evaluation literature has been far-reaching and continues to grow, with its application in an increasing number of disciplines, including economics, political science, epidemiology, transportation, engineering, etc.

The SCM is credited with many advantages, including its transparency, sparsity and interpretability. Nevertheless, we have shown that it also suffers from a number of critical drawbacks and limitations, some of them directly derived from its bilevel nature. In short, we have shown that (1) the covariate importance may not be consistent with economic theory, thus eroding the model's meaningfulness; (2) estimates are unstable—due to the interpolation bias and the nested nature of the optimization problem—and overly dependent on irrelevant countries in the donor pool; and (3) including lagged outcomes does not solve the problem of meaning and the stability of covariate importance—even if the goodness of fit improves—but rather it makes other covariates irrelevant, compromising the main idea underpinning the SCM.
As an alternative to the SCM, we have proposed the decoupled SHAP-distance synthetic control method (DSD-SCM), which overcomes the main limitations of the standard method by decoupling feature importance from weight estimation and by providing a new methodology for feature estimation and unit similarity that ensure meaningfulness and stability.

Here, both methods were used to evaluate the effects on GDP growth of a ten-month government formation deadlock in Spain and to re-estimate the impact of German Reunification in West Germany GDP per capita (Abadie et al. 2015) and the effect of the tobacco control program in California (Abadie et al. 2011). Regarding the first case study, we provide evidence, consistent with Albalate and Bel (2020), refuting the negative economic effects of lengthy impasses in government formation. Thus, not only did Spain's economy not suffer any damage, but it actually benefited by 1.58 p.p.; however, and more importantly in the context of this paper, the SCM overestimates these causal effects by 0.23 p.p. with respect to the DSD-SCM. Moreover, we have demonstrated that the DSD-SCM is a more stable, accurate and meaningful method than the standard SCM. Concerning the second and third cases of study, we show that the DSD-SCM provides a better counterfactual, both in terms of fitting and similarity of the units with respect to the treated unit. Both methods provide similar conclusions. Namely, that German Reunification had a significant and negative impact in West Germany GDP per capita and that the tobacco control program reduced tobacco consumption. For the German reunification, the gap estimate with the DSD-SCM is 27% larger (904.3 USD) than with the original method, while for the tobacco consumption the difference between the two methods is less than 5% (0.5 packs).

After almost two decades of being first proposed, the SCM has shown to be a very useful method for policy evaluation. Some weaknesses of the originally proposed version has been diagnosed and corrections suggested, so that results obtained can be more precise, robust and meaningful. This has been the main objective of this research. Future research should try to further improve SCM, by focusing on how to assess the similarity between units in the donor pool and the treated unit

2.7. Appendix

Country	GDPpc	Investment	Low educ.	High educ.	Trade surp.	Openness	Unempl.	Debt	Correl.w/ Spain[1]
Austria	34799.3	23.0	23.2	17.8	3.3	96.9	5.0	74.1	32.2
Bulgaria	4684.0	23.3	28.2	19.6	-8.2	107.5	11.3	26.8	24.7
Croatia	10212.7	24.3	26.1	15.0	-5.5	81.9	13.6	53.3	73.5
Cyprus	22330.0	20.0	31.9	30.3	-1.9	117.0	7.5	69.8	88.2
Czechia	14212.7	27.6	15.6	13.6	2.7	126.4	6.8	33.3	42.7
Denmark	44235.3	20.5	28.2	27.7	5.8	94.7	5.8	41.3	68.7
Estonia	11397.3	29.0	19.1	28.8	-1.8	140.9	9.6	6.6	47.5
Finland	34439.3	22.6	24.9	30.6	3.1	75.5	8.3	45.3	70.4
France	30627.3	22.0	33.0	25.3	-0.3	56.0	9.0	76.3	43.5
Germany	31506.7	19.9	21.8	21.9	5.3	76.0	7.6	70.1	7.1
Greece	19633.3	19.6	39.3	19.9	-7.9	56.0	14.7	131.8	86.8
Hungary	9894.7	22.6	26.4	16.3	2.0	147.7	8.1	69.0	51.9
Italy	27232.7	19.9	48.4	12.0	0.4	51.9	8.9	115.6	55.8
Latvia	8695.3	26.4	21.7	20.7	-9.0	103.4	12.1	25.8	43.2
Lithuania	8906.7	21.1	18.6	25.1	-4.7	123.2	11.4	27.5	26.7
Netherlands	37693.3	20.6	32.5	26.2	8.2	131.5	5.2	55.7	47.7
Poland	8680.7	20.1	19.9	17.0	-1.5	78.9	12.7	48.1	5.6
Portugal	16602.7	20.9	68.3	13.4	-5.6	69.5	10.4	91.3	88.5
Slovakia	11454.0	23.8	17.7	13.4	-0.2	155.0	14.6	42.6	21.0
Slovenia	17049.3	23.7	22.0	19.1	1.4	126.8	7.1	40.7	63.8
Spain	22962.7	24.0	50.1	27.3	-1.4	57.0	15.9	61.4	100.0
Sweden	38762.0	23.1	23.7	27.5	5.0	83.7	7.3	43.6	36.5
UK	30013.3	16.9	26.4	30.4	-1.8	55.5	6.1	58.1	55.8

1. Correlation between Spain's GDP and each country. Note: Average 2001-2015

Chapter 3: The effect of economic, political and institutional factors on government's policy resposes to COVID-19 crisis

3.1. Part 1: The effect of health and economic costs on governments' policy responses to COVID-19 crisis under incomplete information

This part is a joint work with Germà Bel and Óscar Gasulla, and consists of a paper published in Public Administration Review: Bel, G., Gasulla, O., and Mazaira-Font, F. (2021). The effect of health and economic costs on governments' policy responses to COVID-19 crisis under incomplete information. Public Administration Review 81(6): 1131-1146.

Abstract: The COVID-19 pandemic has become an unprecedented health, economic, and social crisis. The present study has built a theoretical model and used it to develop an empirical strategy, analyzing the drivers of policy-response agility during the outbreak. Our empirical results show that national policy responses were delayed, both by government expectations of the healthcare system capacity, and also by expectations that any hard measures used to manage the crisis would entail severe economic costs. With decision-making based on incomplete information, the agility of national policy responses increased as knowledge increased and uncertainty decreased in relation to the epidemic's evolution and the policy responses of other countries.

Keywords: COVID-19; crisis management; public policy; policy response: organizations.

JEL CODES: D81, H12, I18

Highlights:

- Governments had incomplete information when they responded to COVID-19.
- Confidence in healthcare-system capacity and expected costs delayed their responses.
- Federal countries were more agile than unitary countries in developing policy responses.
- Healthcare-system capacity does not fully guarantee epidemic management.

3.1.1. Introduction

The coronavirus outbreak has produced an unprecedented health, economic, and social crisis, developing into a transboundary crisis, as characterized in Boin (2019). Global leaders, including Antonio Guterres (Secretary General of United Nations) and Angela Merkel (Chancellor of Germany), have compared its impact to World War II.

In a crisis, authorities must engage in coherent analysis and search for proper responses, despite time limitations, uncertainty, and intense pressure (Boin et al., 2005); this has been the case during the COVID-19 crisis (Van Dooren and Noordegraaf, 2020). The rapid spread of the pandemic has forced countries to take unprecedented measures. More than 90% of the world's population lives in countries that have placed restrictions on people arriving from other countries. Many of these countries have closed their borders completely to non-citizens and non-residents, according to the Pew Research Center (see Connor, 2020). Quarantines, social distancing, and isolating infected populations can contain the epidemic.

There is no clear consensus on the specific impact of each measure used to mitigate propagation (see Anderson et al., 2020; Koo et al., 2020). At present, the literature includes few policy analyses related to COVID-19. Among these, Moon (2020) has analyzed the policy response in Korea; Huang (2020) has shown that collaborative governance (cooperation between different levels of government and non-governmental organizations) was a key factor in Taiwan's fight against COVID-19; and Gupta et al. (2020) have analyzed behavioral responses to policies mandated in the US. Any analysis of COVID-19 policy is restricted, given the provisional character and limitations of the existing data (Stock, 2020).

Despite this, there is a widespread consensus among researchers and international organizations that early prevention and response are critical (Grasselli, Pesenti, and Cecconi, 2020), especially given the acute effect of pandemics on disadvantaged sectors of the population (Cénat et al., 2020; Deslatte, Hatch, and Stokan, 2020; Furceri, Loungani, and Ostry, 2020; Kapiriri and Ross, 2020; Menifield and Clark, 2020; Scott, Crawford-Browne, and Sanders, 2016).

The available information allows us to analyze why some national policy responses have been more agile than others. Within the domain of policy decision-making and implementation, agility is defined as "speed in responding to variety and change" (Gong and Janssen, 2012: S61). Lai (2018: 459) defines agility as the "iterative, successive process of adjustment and routine-breaking actions." Agility is related to policy-response quality (Lai, 2018); it is also an aspect of robustness in policy design (Howlett, Capano, and Ramesh, 2018). It has been a key factor in countries like South Korea, which have dealt with the COVID-19 crisis successfully (Moon, 2020). Agility is thus a relevant policy issue, as the time dimension is central to crisis management. The policies that governments have implemented to deal with COVID-19 have followed distinct national (rather than consensual international) standards, in line with policy responses to previous epidemic crises (Vallgårda 2007; Baekkeskov, 2016).

This article investigates why some countries took longer to institute lockdown measures than others. We present a model that characterizes the drivers of coronavirus reaction time, namely, the number of known diagnosed cases per million people (incidence rate) when the government approved hard measures (partial or complete lockdowns). Our base model includes three main factors: the expected capacity of each health system to deal with the outbreak, the expected economic costs of hard measures, and the level of information available to governments forming these expectations. We extend our analysis to account for differences in governance and political regimes, emotional beliefs and biases affecting the assessment of pandemic-related risk, and political survival factors.

We estimate an equation derived from our modeling. Using data from the OECD and European countries, we find that three main factors are statistically relevant. First, the government's expected capacity to fight the outbreak, measured as total healthcare expenditure per capita (adjusted for purchasing power parity), is a factor that delays policy response, accounting for 26.6% of the total delay. The higher a government's healthcare expenditure, the more it is likely to believe it can handle the outbreak—hence the longer delay in responding.

When it comes to preventing economic costs, the more a country is exposed to globalization and trade, the more (relatively) affected it will be by hard measures, such as border closures. We use total trade (% GDP) and the total travel and tourism contribution to GDP as proxies for the expected cost of hard measures. Both are highly significant; together, they account for 37.0%

of the total predictive power of the model. As expected, the higher the cost, the slower the reaction.

To represent the level of information, we use the number of countries that instituted hard measures before a country experienced her first coronavirus cases. As expected, countries that experienced their first coronavirus cases when other countries already had lockdowns in place anticipated their responses. The level of information is responsible for 19.5% of the model's explanatory power. The evidence also confirms of the relevance of decision-making processes and types of decision-makers. Concretely, federal states are more agile than unitary states.

In regard to emotional and perception-related factors, proximity bias—represented by the distance from Wuhan to the capital city of each country—accounts for 5.9% of response agility. Finally, we extend our analysis by testing several variables related to values, ideological biases, and the political survival hypothesis, finding no systematic role for any of these factors.

The rest of the article is organized as follows. First, we outline the theoretical framework used to model the speed of response during the COVID-19 outbreak and to formulate empirical predictions, according to our model. Next, we discuss the data and present empirical results derived from our base equation. We extend the analysis by considering several additional hypotheses. We then conduct robustness checks. Finally, we draw our main conclusions and discuss some policy implications.

3.1.2. Modeling the decision of the policy response to the crisis

We present a theoretical model developing an empirical strategy that we later follow to analyze the drivers of policy-response agility. We begin with a basic model, representing a cost-benefit analysis carried out by a rational, benevolent government, which cares only about social welfare and has incomplete information on the pandemic. We then present two extensions. First, we allow for different types of decision-makers (governments and political systems). Second, we consider the possibility that governments are: 1) not entirely rational and potentially emotionally biased; and 2) not fully benevolent, but driven by self-interest (i.e. stay in office).

3.1.2.a. Base model: Benevolent government with incomplete information

At the start of the pandemic, a set of nature features, such as the density of population Wong and Li, 2020), the share of population above 65 years old and with pre-existing comorbidities (Knight et al., 2020; Álvarez-Mon et al, 2021), temperature, and humidity (Mecenas et al., 2020), determine the virus reproductive number under no contention measures, ρ, and the death rate, *d*. The strategies used to fight the outbreak can be modelled as a sequential decision-making process with incomplete information, where governments, instead of observing the true parameters involved in decision-making, achieve only partial estimations. As noted in the Introduction, even after seven months, we lack clear knowledge of how the virus is propagated. We do not know how effective the various mitigation measures are (Stock, 2020). Indeed, the very first response guidelines issued by the WHO in January 2020 were mainly addressed to communication and clinical management (WHO, 2020a; WHO, 2020b), and did not consider specific recommendations on contention measures, since due to the lack of information it was not even clear whether the virus was transmitted between humans.

In every time period, a government can decide to implement either hard or soft measures to contain the virus. If the government implements soft measures (SM) at time t (e.g. temperature control at airports or testing people with symptoms coming from affected countries) the transmission rate is reduced to $\rho_t = \delta^S \rho$. If it implements hard measures, it loses π units of utility (lost production) but reduces transmission rate to $\delta^H \rho$, with $\delta^H < \delta^S < 1$. It is worth highlighting that, according to cross-country estimates (Hilton and Keeling, 2020; Katul et al., 2020), all countries in our sample, no matter nature determinants, had reproductive numbers far above 1 (different methodologies lead to estimates ranging from 2 to 6.5), which imply that the pandemic would collapse their healthcare system unless massive tracking and severe contention measures were taken.[1]

Let n_{t-1} be the number of infected people at the end of time t − 1. At the beginning of period t, the virus infects $\rho_t n_{t-1}$ people, who are then treated. Let the capacity of the healthcare system be c. If $n_{t-1} < c$, then no infected people die at t and all are cured. Otherwise, the number of fatalities at t is $f_t = d(n_t - c)$, and the rest are cured. The capacity of the healthcare system, while relevant in the direct sense of treating patients and avoiding fatalities, also influences the transmission rate by identifying and correctly diagnosing patients, thus breaking propagation chains. Hence, $\boldsymbol{\rho_t}$ must be seen as a function of contention measures, healthcare-system capacity, and other potential country-specific effects (e.g., hand-washing habits).

Let us consider a 4-period process, as shown in Figure 3.1. At t = 0, nature determines an initial number of infected people n_0 and the transmission rate ρ. At t = 1, infected people transmit the virus to others and then receive treatment. Therefore, $n_1 = \rho n_0$, and the number of fatalities at t = 1 is $f_1 = d \max\{n_0 - c, 0\}$. The government estimates the transmission rate $\rho_1 = \hat{\rho}$ and the total number of infected people, $\widehat{n_1}$. Based on that information, the government estimates the expected transmission rate, death rate, and healthcare-system capacity during the following periods, as well as the impact and cost of various measures ($\hat{\rho}_{t+1} = E_t(\rho_{t+1})$, $\hat{d}_{t+1} = E_t(d_{t+1})$, $\hat{c}_{t+1} = E_t(c_{t+1})$, $\widehat{\delta^S} = E_t(\delta^S)$, $\widehat{\delta^H} = E_t(\delta^H)$, $\hat{\pi} = E_t(\pi)$). Based on these estimations, the government decides whether to implement soft or hard measures. The process continues until t = 4, when a vaccine is discovered and propagation drops to 0. Figure 3.1 shows how the government expects the pandemic to evolve, at t = 1.

Let us note, using $\hat{f}_{t+i}(\hat{n}_t) = \hat{d}_{t+i} \max\{\hat{n}_t - \hat{c}_{t+1}, 0\}$, the expected fatalities at time $t+i$, given the death-rate and capacity expectations, and by l the cost per fatality. Given the information available to the government at $t = 1$, the expected costs (at $t = 1$) of various available strategies are as follows:

$$EC(HM,HM) = l\{\hat{f}_1(\hat{n}_0) + \hat{f}_2(\hat{\rho}\hat{n}_0) + \hat{f}_3(\hat{\rho}^2\delta^H\hat{n}_0) + \hat{f}_4(\hat{\rho}^3(\delta^H)^2\hat{n}_0)\} + 2\pi \tag{3.1}$$

$$EC(HM,SM) = l\{\hat{f}_1(\hat{n}_0) + \hat{f}_2(\hat{\rho}\hat{n}_0) + \hat{f}_3(\hat{\rho}^2\delta^H\hat{n}_0) + \hat{f}_4(\hat{\rho}^3\delta^H\delta^S\hat{n}_0)\} + \pi \tag{3.2}$$

$$EC(SM,HM) = l\{\hat{f}_1(\hat{n}_0) + \hat{f}_2(\hat{\rho}\hat{n}_0) + \hat{f}_3(\hat{\rho}^2\delta^S\hat{n}_0) + \hat{f}_4(\hat{\rho}^3\delta^S\delta^M\hat{n}_0)\} + \pi \tag{3.3}$$

$$EC(SM,SM) = l\{\hat{f}_1(\hat{n}_0) + \hat{f}_2(\hat{\rho}\hat{n}_0) + \hat{f}_3(\hat{\rho}^2\delta^S\hat{n}_0) + \hat{f}_4(\hat{\rho}^3(\delta^S)^2\hat{n}_0)\} \tag{3.4}$$

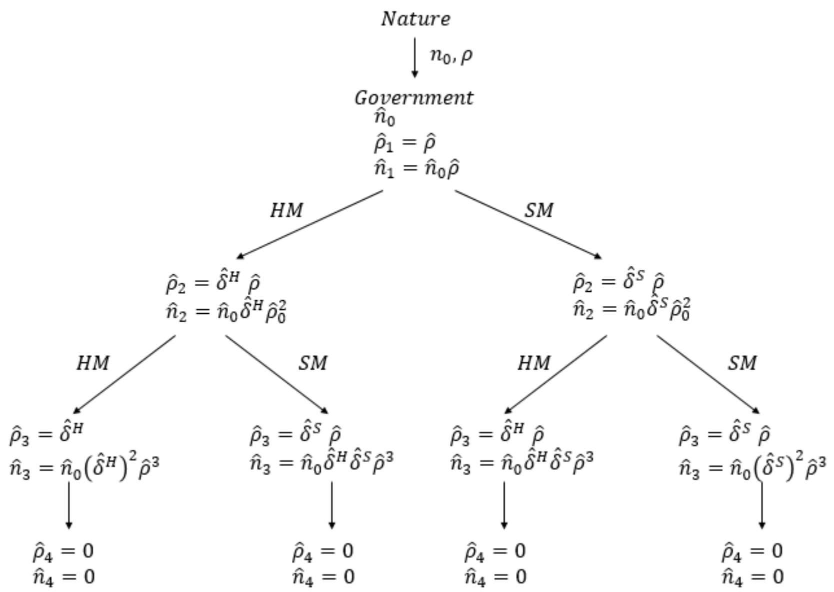

Figure 3.1: The 4-period decision-making process, with government expectation of transmission rates and numbers of infected people at t = 1.

First, note that *EC (HM, SM)* $\leq$ *EC(SM, HM),* with strict inequality, if the healthcare system collapses under soft measures. Delaying the adoption of hard measures is a weakly dominated strategy if the government expects a collapse. Therefore, under the assumption of rationality, a government will only delay implementation if it underestimates the risk or is overconfident about its healthcare-system capacity, due to incomplete information. The latter reasoning is consistent with the offsetting behavior hypothesis, put forward by Peltzman (1975), which implies that risk is compensated for: agents adjust their behavior in response to perceived levels of risk and behave less carefully when they feel more protected. This hypothesis has been frequently tested, for instance, in car-safety studies (Chirinko and Harper, 1993; Peterson, Hoffer, and Millner, 1995).

H1: The less information, the higher the probability of a delayed response

Second, let us analyze what determines whether a government decides to implement hard or soft measures. The dynamics of government action or inaction during crises do not imply that action is always beneficial or functional (Rosenthal and Kouzmin, 1997). Hence, governments must consider the costs and benefits of action (Comfort, Waugh and Cigler, 2012). A government will apply hard measures (at least once) if and only if the expected economic costs and fatalities are lower than they are predicted to be under soft measures [public debate on these interactions was well in the air already by mid-March in the European countries most badly affected by the pandemics, Spain and Italy. (e.g. Bel, 2020; Ferro, 2020)].

It is sufficient to compare a case in which the government applies hard measures once. Noting via ΔC the difference between *EC(HM, SM)* and *EC(SM, SM)*, we have:

$$\Delta C = l\{\hat{f}_3(\hat{\rho}^2\delta^H\hat{n}_0) + \hat{f}_4(\hat{\rho}^3\delta^H\delta^S\hat{n}_0) - \hat{f}_3(\hat{\rho}^2\delta^S\hat{n}_0) - \hat{f}_4(\hat{\rho}^3(\delta^S)^2\hat{n}_0)\} + \pi \quad (3.5)$$

The more production a country expects to lose, because of hard measures, π, the fewer incentives the government has to implement hard measures, since ΔC increases as π increases. Note that fatality costs are positive only if the government believes that the system will collapse under soft measures. In that case, the incentives to implement hard measures increase.

H2: The greater the expected capacity of the healthcare system, the fewer incentives there are to implement hard measures.

H3: The higher the expected economic costs of hard measures, the fewer incentives there are to implement those hard measures.

Third, even if the system collapses, the government may decide not to implement hard measures. Let us assume that the system will collapse under soft measures at t = 3 and t = 4 and will never collapse under hard measures. Then:

$$\Delta C = -l\{\hat{f}_3(\hat{\rho}^2\delta^S\hat{n}_0) + \hat{f}_4(\hat{\rho}^3(\delta^S)^2\hat{n}_0)\} + \pi \quad (3.6)$$

The government will implement hard measures if and only if the total number of fatalities multiplied by the cost per fatality is higher than the penalty cost of the hard measures. Therefore, the larger the process (all other things being equal), the higher the probability of hard measures.

Overall, this theoretical description of the decision-making process, which assumes a welfare-centered cost-benefit analysis, allows us to identify two main insights. First, the decision about which strategy to follow depends on the seriousness of the pandemic and the economic and fatality costs expected by the government. Governments may decide to follow different strategies because they expect different associated costs. Second, if a healthcare collapse is expected, it is better for a government to anticipate hard measures than to delay their implementation. In the present crisis, governments that instituted hard measures only after diagnosed coronavirus cases escalated would had been better off anticipating that policy response.

For the sake of simplicity, we assume that there is one type of hard measure. However, if a range of hard measures existed, the conclusions would be the same. The only difference would be that governments would choose a set of hard measures that minimized expected costs, according to the available information.

3.1.2.b. Extension 1: Different types of decision-maker

So far, we have assumed that there is only one type of decision-maker, the government, which operates with the same constraints and efficiency in every country. This is clearly not the case. Parliamentary systems and regimes have different decision-making processes, both in terms of who makes decisions (the national government or both national and sub-national governments) and how they are made and approved. For instance, a presidential regime is less dependent on approval from parliament to institute measures. For this reason, it may be able to react faster than a more institutionally complex governmental system. These differences may directly affect decision-making agility and the implementation of measures. Debates have arisen over whether authoritarian governments have an advantage in responding to crises (Schwartz 2012; Kleinfeld, 2020).

Let us use g to indicate the type of government and assume that there are two types: agile and slow. Agile governments (AG) resemble those modelled above: once they decide, the decision is approved and implemented during the next period. By contrast, slow governments (SG) face a more complex decision-making and/or implementation process. They either need additional time to approve the measures (slow in decision-making) or they fail to reduce the coronavirus transmission rate within a single time period, requiring two periods after the decision to apply hard measures (slow in implementation). Due to this delayed implementation—and coming back to the example in Figure 2.1—these governments slow down the transmission rate at $t = 3$ by applying hard measures at $t = 1$ and $t = 2$. For such governments, there is less incentive to implement hard measures. In conclusion, slow governments are expected to implement hard measures later than agile governments, when the decision-process is more complex. They may also have fewer incentives to implement hard measures when their lack of implementation agility will reduce the expected benefit of such measures.

H4: The more presidential/executive a governance system is, the greater its ability to implement hard measures quickly.

3.1.2.c. Extension 2: Emotions, beliefs, and political survival

Alongside the uniformity of each country's decision-making process and legal constraints, two other hypotheses can also be questioned. First, it may be wrong to assume that expectations will be rational; decision-making can be highly influenced by emotions and beliefs, especially when information is lacking (Kahneman, 2011). Akerloff and Shiller (2009) argue that emotions play a role in economics and are a key driver of market failures and financial crises. When decision-makers confront a crisis with incomplete information, especially where policy responses involve unprecedented restrictions on human rights, emotional biases and beliefs related to risk-aversion, information processing, and the role of government can affect response speed. For instance, the greater the geographic proximity of the crisis, the more it provides an incentive for policy action, based on heightened fear and attention (Nohrstedt and Weible, 2010).

H5: Emotional biases and beliefs related to risk-aversion, information processing, and the role of government affect response speed

So far, we have assumed that policymakers only care about maximizing social-welfare functions. However, an abundant literature shows that politicians behave as both citizens and candidates. In other words, while they do care about maximizing social welfare, they are also motivated by self-interest, e.g. winning or staying in office (Osborne and Slivinski, 1996; Besley and Coate, 1997). Applying the logic of political survival (Bueno de Mesquita et al., 2003) to crisis management, it follows that, since voters punish governments for improper crisis responses, risk-averse governments will implement proactive policies, especially within highly competitive contexts and close to elections (Baekkeskov and Rubin, 2014). In terms of modeling, we can modify the utility function to include a political reward $\varphi > 0$, which reduces the cost of hard measures, resulting in $\pi - \varphi$.

H6: Highly competitive contexts provide incentives for more agile policy responses.

3.1.3. Variables, data, and sources

Sample

To ensure a certain homogeneity between countries, our model considers the 36 OECD countries. We provide a robustness check by increasing the sample to include the five non-OECD EU states (Bulgaria, Romania, Cyprus, Malta, and Hungary) and four EU-candidate states (Albania, Montenegro, North Macedonia, and Serbia). Next, we discuss the variables used, based on the theoretical model. We explain how they are specified and what sources they are obtained from.

Variables

Incidence rate when policy response began. We define the 'incidence rate when the policy response began' as the number of coronavirus cases (according to the Johns Hopkins Coronavirus Resource Center) adjusted per total population at the point when the government began to implement hard measures. This variable captures the amount of time each government waited before implementing hard measures. Hard measures severely restrict the free movement of citizens (partial or total lockdowns). They include: closing borders; closing schools, universities, and public places; prohibiting public events and public gatherings; closing most or all non-essential shops; imposing curfews; and forcing people to work from home. It can be argued that these hard measures are of different intensity, either because of its nature or because they may not be applied nationwide.

To establish a more homogeneous criteria, at least such two measures must be in place for a country to be categorized as implementing hard measures. Table A.3.1 presents the first hard actions taken by each country. The data were obtained from the IMF database of policy responses to COVID-19 (https://www.imf.org/en/Topics/imf-and-COVID-19/Policy-Responses-to-COVID-19) and the Think Global Health timeline (www.thinkglobalhealth.org/), in addition to official government websites and press briefings. Although a perfectly homogenous criterium may be not possible to establish, with the two measures threshold we ensure, for instance, that at the time of policy response, 60% of the countries had implemented a nationwide closure of non-essential shops, and 85% closed educational institutions.[2]

Fatality costs and the capacity of the healthcare system to fight the outbreak. Total healthcare resources per capita (purchasing power parity-ppp) in 2017, the last available year, are used as a proxy for the government's expected healthcare-system capacity, including fatality costs. Several statements made by political leaders have highlighted the relevance of healthcare-system capacity in decision-making—and expenditure as the primary proxy for healthcare-system capacity. For instance, both Spanish Prime Minister Pedro Sánchez and French President Emmanuel Macron made public statements on (casually) the same day, March 10th, presenting their countries' robust healthcare systems as the best possible preparation for fighting the pandemic when they both were still sustaining that lockdown measures were not needed. Similarly, the Leader of the U.K. Labour Party, Keir Starmer, made a statement on May 7th, establishing a direct causal link between the U.K.'s higher incidence of coronavirus (in comparison to other European countries) and the Conservative government's cuts to healthcare expenditure. Data on healthcare expenditure per capita (ppp) have been obtained from the World Bank database (https://data.worldbank.org/indicator/SH.XPD.CHEX.PP.CD). While nominal expenditure can be strongly associated with different costs, adjusting for ppp allows makes it possible to control for cost differences. The results have been checked using alternative variables (healthcare expenditure as a % of GDP, a relative measure; and public healthcare expenditure as a % of GDP) to account for the direct capacity of public healthcare systems.

Although expenditure is a key indicator of a healthcare system's overall capacity and performance, and used as such by political leaders, this proxy may be, at least partially, inaccurate, as it does not reflect expenditure efficiency or reveal whether the expenditure has targeted areas relevant to fighting the pandemic. For this reason, we have carried out an additional robustness check by considering the Global Healthcare Security Index health indicator (https://www.ghsindex.org/) as an alternative measure of healthcare-system capacity. Built by The Economist Intelligence Unit, the Johns Hopkins Center for Health Security, and the Nuclear Threat Initiative, the indicator measures a healthcare system's capacity to fight pandemic outbreaks, in terms of personnel deployment, hospital beds, capacity in clinics and community-care centers, healthcare assessments, infection-control practices, available equipment, and the ability to test and approve new medical countermeasures.

In accordance with H2, derived from the theoretical model and the offsetting behavior hypothesis (Peltzman, 1975), we expect stronger healthcare-system capacity to be negatively associated with policy-response agility.

Economic costs. When determining policies, governments consider their costs and benefits. The hard measures used to confront the COVID-19 crisis, given their intrinsic characteristics, inevitably slow down business activity, damaging the economy. Trade and tourism are particularly damaged by measures that strongly restrict mobility. For instance, the Prime Travel Technology Index, which measures the performance of global-technology companies in the travel and tourism industry, fell by more than 50% between mid-February and mid-March. By the end of July, prices were around 25% lower than in February (https://www.primeindexes.com/). By comparison, the MSCI World Index, which represents a broad cross-section of global markets in all sectors, fell by 30% between mid-February and mid-March; by the end of July, prices were only 5% lower than in February (https://www.msci.com/). We therefore use two indicators to consider the relevance of economic costs: the total direct and indirect contribution of travel and tourism, and total trade (imports and exports), both as a percentage of total GDP in 2018. Both indicators have been obtained from the World Bank Database (https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS, https://tcdata360.worldbank.org/indicators/tnt.tot.contrib.gdp).

We hypothesize (H3) that the higher the economic cost of adopting hard measures, the less agile the government adopting them will be.

Uncertainty and information. We use the number of countries that had announced or were implementing hard measures when their governments first began dealing with the pandemic (the first case diagnosed within the country) as the main indicator for the level of government information. We also use two alternative specifications. First, we use the number of countries previously affected by the COVID-19 pandemic. Second, we restrict previously affected countries to those that share borders with the country in question, or are connected to it by less than 250km of sea, with the exception of Japan, Australia, South Korea, and New Zealand, which are considered neighbors, due to their historical ties and strong economic relationships.

Any time that elapses after a crisis erupts gives the government a chance to adjust its response and reduce the risk of problems, such as cognitive overload or panic (Moynihan, 2008). For this reason, countries in which the first case occurred relatively late are expected to have had more accurate information and a greater understanding of the risks involved, allowing policymakers to reduce the gap between planning and practice (Comfort, 2007). As the theoretical model (H1) predicts, we expect them to have acted relatively quickly, taking advantage of the extra information and clearer calls for urgent action before the crisis escalated (Farazmand, 2007).

Types of decision-makers

As the theoretical model states, different types of decision-makers implement different policy responses to the COVID-19 outbreak. We operationalize these differences with the following three variables:

Political regime. Scores ranging from -1 (Parliamentary system) to 1 (Presidential system) represent various types of government. Semi-presidential countries, such as France and Lithuania, are ranked as 0. To be defined as "presidential," systems must have an executive presidency that is separate from the legislature. Semi-presidential countries have both an executive presidency and a separate head of government, who leads the remaining executive; this individual is appointed by the president and accountable to the legislature. Parliamentarian governments have no executive presidency or head of state. The head of government leads the executive and must maintain the confidence of the legislature to remain in power. Data have been obtained from the institutional web pages of each country.

Multi-level governance. The dummy variable equals 1 when the country has a unitary system and 0 when the system is federal (source: https://www.britannica.com/topic/political-system/Federal-systems). We have no clear expectation for this variable. While more vertical and hierarchical systems may respond more quickly (Yan et al., 2020) and federal systems can be highly dysfunctional (see for the US, Maxeiner, 2019), as hypothesized in H4, decentralization may also lead to more agility and effectiveness (Christensen, Lægreid, and Rykkja, 2016). Multilevel systems with collaborative governance between different levels of government and non-state institutions - (Scavo, Kearne and Kilroy, 2008; Schwartz and Yen, 2017; Downey and Myers, 2020; Huang, 2020) provide incentives for more agile and effective responses, as noted in H6.

Authoritarianism. This study rates the level of authoritarianism in each country on a scale of 0 to 100, based on the Political Rights and Civil Liberties Index from Freedom House (https://freedomhouse.org/countries/freedom-world/scores); the scale ranges from 0 (no political rights and civil liberties) to 100 (full political rights and civil liberties). Thus, a country with a score of 80 in the Freedom House Index receives a score of 20 for authoritarianism. We do not have a clear expectation for this variable, as in the former case.

Tenure of the Prime Minister. We use the number of days since the PM took office as a proxy for her experience and decision-making determination. Data have been obtained from the institutional web pages of each country. Experienced decision-makers are expected to be more agile, as they are more aware of electoral punishment (Bechtel and Hainmueller, 2011).

Coalition government. The dummy variable equals 1 when a country's national government is formed by two or more parties, and 0 otherwise (source: institutional webpages). We expect collation-based governments to be less agile, given the transaction costs of crossed monitoring and control between different parties in government (Thies, 2001).

Emotions, beliefs, and political survival

We consider several variables related to emotional biases, beliefs, and the logic of political survival, following the discussion in the theoretical section above.

Proximity bias on information processing. We consider the distance in *kilometers from Wuhan, China* to the capital city of each country (source Google Maps API), as a proxy for geographic-proximity bias. For decision-makers affected by emotional biases (H5), we expect countries closer to Wuhan to demonstrate more agile policy responses. The variable is included as the logarithm of the distance needed to capture a concave dissipation effect.

Gender bias on risk aversion. The second indicator corresponds to the gender of the Prime Minister. The question of whether female prime ministers have taken faster and more executive action has been widely discussed (e.g. *CNN,* April 16, 2020; *The Guardian*, 25 April 2020). One possible explanation is that women are more risk-averse than men and value safety more highly, as Barnes and Beaulieu's (2018) survey experiment on women and risk aversion argues. We specify the variable *Gender PM* as a dummy that takes value 1 for women and 0 otherwise (source: countries' official web pages). We expect female prime ministers to demonstrate more agile policy responses.

Ideology. To account for the possibility that different ideologies or beliefs about the role of government can influence how crises are viewed and managed (Dror, 1994), we consider the ideology of the main political party in the national government, as this party has the primary role in the decision-making process, even (to some extent) in federal countries. A scale ranging from -1 (left) to 1 (right) is used to represent the ideological position of the Prime Minister's party. Center parties are ranked as 0 (main sources: the World Bank Database of Political Institutions https://datacatalog.worldbank.org/dataset/wps2283-database-political-institutions, and international alliances that include governing parties). Where ideological beliefs play a role (H5), we expect left-wing parties to demonstrate more agile policy responses, as they tend to be more concerned with inequality. Pandemics attack the most disadvantaged segments of society with particular intensity (Kapiriri and Ross, 2020; Deslatte, Hatch and Stokan, 2020).

Days to next election. Applying the logic of political survival (Bueno de Mesquita et al., 2003) to disaster management suggests the following: since voters punish governments for improper crisis responses, risk-averse governments will implement proactive policies, especially within highly competitive contexts and close to elections (Baekkeskov and Rubin, 2014). Among the hypotheses presented here, one has particular interest for our research: the relationship between policy responses and the electoral cycle. As our theoretical model (H6) suggests, the closer a government is to its next election, the more comprehensive its policy response will be (Bechtel and Hainmueller, 2011). The variable "days to next election" corresponds to the logarithm of the number of days between the first diagnosed case of coronavirus in the country and the next scheduled or expected nationwide election date (sources: National Democracy Institute database https://www.ndi.org/ and countries' official websites).

Table 3.1 describes the variables and their sources. Table 3.2 presents descriptive statistics.

Table 3.1. Variables: Description and sources

	Description	Source
Dependent variable		
Incidence rate	The number of diagnosed cases adjusted per million inhabitants when the government began implementing hard measures.	IMF & Think Global Health
Covariates		
Health expenditure per capita (ppp)	Logarithm of total healthcare expenditure per capita in 2017 (ppp).	World Bank
Tourism	Logarithm of total travel and tourism contribution to GDP.	World Bank
Trade	Logarithm total trade -imports & exports- as % GDP.	World Bank
Previously locked countries	Total # of countries that had begun to implement hard measures when pandemic hits the country.	Own elaboration
Political regime	Score representing from -1 (Parliamentary system) to 1 (Presidential system)	Institutional webs
Unitary	Dummy variable that equals 1 if the state is Unitary and 0 if it is Federal	Encyclopedia Britannica
Authoritarianism	Score from 0 (full political rights & civil liberties) to 100 (no political rights & civil liberties) on level of authoritarianism of country	Freedom House Index
Tenure of the Prime Minister (LN)	Logarithm of # of days since the PM took office	Institutional webs
Coalitional government	Dummy variable that equals 1 if the national government is a coalition	Institutional webs
Km from Wuhan	Logarithm of the distance in kilometers between Wuhan and the capital city of the country	Google maps API
Gender of the Prime Minister	Dummy variable that equals 1 if the Prime Minister is a female	Institutional webs
Ideology	Score from -1 (left) to 1 (right) of the political orientation of the political	World Bank Database of

	party of the PM. Center parties are given a 0. The classification is based on international political alliances.	Political Institutions and institutional webs
Days to next election	Logarithm of the number of days between the first diagnosed case in the country and the next scheduled or expected relevant election date	National Democracy Institute and institutional webs
Alternative covariates		
Health expenditure % GDP	Logarithm of the health expenditure as % GDP	World Bank
Public health expenditure % GDP	Logarithm of public health expenditure as % GDP	World Bank
GHS Health capacity	Health capacity score (0-100) to fight pandemic outbreaks	GHS Index
Previously affected countries	Total number of countries that had diagnosed cases when the pandemic hits the country	Own elaboration
Previously affected neighbors	Total number of neighboring countries that had diagnosed cases when pandemic hits the country	Own elaboration

Table 3.2. Descriptive statistics

	Min	Max	Mean	St Dev
Incidence rate when policy response began	.01	379.90	68.13	89.53
Health expenditure per capita (ppp)	7.03	9.28	8.24	.53
Tourism (LN)	1.46	3.54	2.23	.48
Trade (LN)	3.31	5.96	4.52	.54
Previously locked countries	0	8.00	1.03	1.40
Political regime	-1	1	-.61	.73
Unitary state	0	1	.78	.42
Authoritarianism	0	68.00	10.25	12.87
Tenure of the Prime Minister (LN)	3.91	8.55	6.75	1.12
Coalitional government	0	1	.50	.51
Km from Wuhan (LN)	6.92	9.47	8.90	.53
Gender of the Prime Minister	0	1	.19	.40
Ideology	-1	1	.11	.92
Days to next election (LN)	4.45	7.51	6.62	.80
Health expenditure % GDP (LN)	1.44	2.84	2.14	.27
Public health expenditure % GDP (LN)	1.04	2.22	1.77	.32
GHS Health capacity (LN)	3.45	4.30	3.89	.23
Previously affected countries	1	45.00	21.22	12.41
Previously affected neighbors	0	5.00	1.78	1.49

3.1.4. Empirical model and results

Our empirical analysis is based on the theoretical model presented. First, we estimate the base model: a benevolent government. We then test potential extensions of the model, estimate a final model, carry out robustness checks, and interpret the results.

3.1.4.a. Base model

Agility in taking action (cases adjusted by total population when hard measures are taken) is affected by the healthcare system's ability to avoid fatalities and reduce the transmission rate, the cost of hard measures, and information accessible to the government on expected coronavirus deaths and transmission rates. As the previous section explains, the following variables are used to capture these drivers: healthcare expenditure per capita, tourism, trade, and previously locked-down countries. We thus estimate a base model in the form:

$$Cases = f(population, healthcare, tourism, trade, locked - down\ countries) \quad (3.7)$$

A discrete modeling approach is appropriate, given the non-negative discrete nature of the problem. A GLM with negative binomial distribution is used in this empirical approach. The negative binomial allows us to capture over- and under-dispersion, providing more robust estimates of the parameters and standard errors than a Poisson distribution. We also use OLS to adjust an alternative specification of the model. To do this, we transform the target into the logarithm of the incidence rate. Although, for a general discrete problem, this approach may lead to non-normality of residuals and fail to solve the relationship between variance and mean associated with counting problems (e.g., Long, 1997; Lindsey, 2000), in this case, once the transformation residuals can be considered normal (the p-value is 0.2020 for the Shapiro-Wilk test and 0.1364 for the Anderson-Darling test) and homoscedastic (the White test for heteroscedasticity yields p-value = 0.3346), the average variance-inflation factor (VIF) is 1.34 and no individual VIF is above 2.

Table 3.3 presents the results using both modeling techniques. The two methods yield similar estimations of the parameters. In both cases, the theoretical hypotheses cannot be rejected for all parameters. Confidence that existing healthcare-system capacity can deal with the crisis is associated with a higher incidence rate and thus negatively associated with policy-response agility. In this regard, our result is consistent with the offsetting-behavior hypothesis. Expectations of economic impact, if hard measures are delayed, are also negatively related to policy-response agility. By contrast, increased information and reduced uncertainty are associated with more agile policy responses, as long as more countries have adopted hard measures.

Table 3.3. Estimated parameters of the models

	Negative Binomial (1)	OLS Robust (2)
Constant	-35.7629***	-24.6835***
	(3.2399)	(3.9229)
Healthcare capacity	1.8814***	1.9741***
	(.3199)	(.3779)
Tourism	1.7654***	2.0864***
	(.3086)	(.3806)
Trade	1.4632***	1.6666***
	(.2760)	(.3972)
Locked countries	-.6307***	-.6597***
	(.1359)	(.2448)
N. Observations	36	36
R-Squared		.8167
F-Test		5.174e-11***
Residual/Null deviance	.6833	

Standard errors in brackets. Significance: *** p<0.01, ** p<0.05, * p<0.1

Note**:** The estimations are robust to the exclusion of Sweden, which followed a recommendation-based approach, rather than a lockdown strategy. They are also robust to the exclusion of the U.S., which can be considered an outlier, given its system of multi-level governance and high expenditure on healthcare. The estimated value of the coefficients, when these countries are excluded, varies less than 10%, relative to the estimation in table 3. The significance levels remain the same.

The negative binomial distribution avoids transforming the target and guarantees a proper fitting for the counting outcome, without the assumption of residual normality. We therefore take it as our base model. Next, we check the results (table 3.4) using alternative specifications for healthcare-system capacity and level of information.

Estimations using alternative specifications for healthcare-system capacity and level of information yield results that are almost identical to those obtained with the base model—Estimation (1). The same thing happens when we run OLS Robust estimations (results available on request). When healthcare-system capacity is measured in relative terms (Estimations 3 and 4), goodness of fit is slightly lower, revealing that the absolute level of healthcare resources (adjusted by ppp) is more relevant than the relative level. When the level of information is measured in previously affected countries (Estimations 6 and 7), the level of significance changes from $p<0.01$ to $p<0.10$ and $p<0.05$. This shows that governments obtain more information from the strategies adopted by other governments than from any other source. It is worth noting that *Affected neighbors* provide more explanatory power than *Affected countries*, revealing a proximity effect, which will be discussed later.

3.1.4.b. Extension 1: Types of decision-maker

Starting from our base model, estimated using the negative binomial distribution (which avoids transforming the target and guarantees a proper fitting for a counting outcome without assuming the normality of residuals), we test the relevance of variables affecting the type of decision-maker. The results are shown in table 3.5. While unitary states are less agile than federal states, there is no sound evidence that presidential systems react faster. In addition, there is no evidence that other factors, coalitions, tenure of the PM, or authoritarianism influence government-response agility. Notice that, although we find no evidence that authoritarianism influences agility, it may influence policy-response severity. Indeed, Sweden, the only country able to sustain a recommendation-based strategy, has the lowest authoritarianism score (0 out of 100).

3.1.4.c. Extension 2: Emotions, beliefs, and political survival

Finally, we test several hypotheses involving emotions, beliefs, and political survival, as discussed in the previous section. We test these hypotheses from our base model extended via the type of player. We consider the *unitary* variable only, since the parliamentary system is not relevant when included together with the *unitary* dummy. The results are presented in table 6.

It is clear that the distance from Wuhan is a significant factor in determining policy-response agility (Estimation 13). The further a country is from Wuhan, the slower its reaction, consistent with the geographic-proximity hypothesis. By contrast, the variable for prime-minister gender (Estimation 14) does not significantly affect policy-response agility. This result is consistent with Pondorfer, Barsbai, and Schmidt (2017), who found no actual gender differences in risk preferences, but rather a perception based on stereotypes.

Estimation (15) shows that *ideology* has no significant influence on policy-response agility in relation to the COVID-19 crisis. Alongside the main decision-maker's ideology, which can be thought of as a conjunctural belief, we have tested the relevance of more structural beliefs about the role of the state in relation to egalitarianism. We have used the World Bank GINI index (https://data.worldbank.org/indicator/SI.POV.GINI) and the population head count ratio at national poverty lines to operationalize this test. No significant evidence has been found (results are available in table A.3.3, in the Appendix).

Finally, our findings on political survival (Estimation 16) are consistent with the hypothesis that the closer a government is to the next election, the more agile its policy response will be.

Note that in Estimations 8–16, all variables in the base model keep the same sign and level of significance. We can therefore conclude that the basic results are very stable throughout all estimations conducted in this section.

Table 3.4. Estimated parameters of the models with alternative specifications

	Base model (1)	(3)	(4)	(5)	(6)	(7)
Constant	-35.7629***	-28.3715***	-23.5908***	-26.5582***	-38.9770***	-41.9789***
	(3.2399)	(3.2480)	(2.4836)	(4.6241)	(3.3164)	(3.0826)
Healthcare capacity	1.8814***				2.1605***	2.4985***
	(.3199)				(.3694)	(.2795)
Tourism	1.7654***	1.4700***	1.4176***	1.5804***	1.9634***	1.9613***
	(.3086)	(.3563)	(.3682)	(.3883)	(.3473)	(.3241)
Trade	1.4632***	2.1681***	1.9171***	1.8864***	1.6058***	1.5988***
	(.2760)	(.3646)	(.3430)	(.3725)	(.3507)	(.3273)
Locked countries	-.6307***	-.8000***	-.9081***	-1.0642***		
	(.1359)	(.1776)	(.1657)	(.1568)		
% GDP health		2.7408***				
		(.8381)				
% GDP public health			1.4059**			
			(.6372)			

GHS index				1.3924*		
				(.8457)		
Affected countries					-.0345*	
					(.0197)	
Affected neighbors						-.2883**
						(.1135)
N. Observations	36	36	36	36	36	36
Residual/Null deviance	.6833	.5652	.5311	.5046	.6179	.6406

Standard errors in brackets. Significance: *** p<0.01, ** p<0.05, * p<0.1

Note**:** In addition to testing alternative specifications of the main drivers in the base model, we also tested the relevance of additional second-order effects related with the distribution of the costs, in accordance with reviewer suggestions. We tested whether the percentage contribution of MSME to the economy (% of employment generated by MSMEs) or the percentage unemployment were relevant as a second-order economic factor, and whether the percentage of the population over 65 was relevant as a second-order fatality-cost factor. These variables were not relevant. Including them in the model did not change the significance or order of magnitude of the other estimates. Data on the MSME contribution to employment were taken from Eurostat (https://ec.europa.eu/eurostat/statistics-explained/pdfscache/45509.pdf) and institutional web pages for Australia, Canada, Mexico, and South Korea. No data were available for New Zealand, Israel, or Chile, due to differences in classification criteria. Data on the percentage of unemployment were obtained from the World Bank (https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS). Data on the percentage of the population over 65 were obtained from the World Bank (https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS). Results available in table A2, in the Appendix.

Table 3.5. Estimations of extensions of the model with types of decision-maker

	Base model					
	(1)	(8)	(9)	(10)	(11)	(12)
Constant	-35.7629***	-33.6977 ***	-37.3742 ***	-33.2863***	-36.8153 ***	-35.4222 ***
	(3.2399)	(3.5063)	(3.0024)	(4.0559)	(3.3536)	(3.2469)
Healthcare capacity	1.8814***	1.7326***	1.9742***	1.6942***	1.8645***	1.8936***
	(.3199)	(.3261)	(.2973)	(.3733)	(.3164)	(.3189)
Tourism	1.7654***	1.6977***	1.9011***	1.6793***	1.8550***	1.6932***
	(.3086)	(.3160)	(.2792)	(.3149)	(.3108)	(.3276)
Trade	1.4632***	1.2396***	1.4032***	1.3224***	1.4956***	1.3679***
	(.2760)	(.3068)	(.2530)	(.3057)	(.2737)	(.3064)
Locked countries	-.6307***	-.5632***	-.6677***	-.5699***	-.6579***	-.5933***
	(.1359)	(.1352)	(.1284)	(.1683)	(.1383)	(.1376)
Political regime		-.3818				
		(.2319)				
Unitary state			1.0113***			
			(.3135)			
Authoritarianism				-.0172		
				(.0208)		

PM Tenure					.1281	
					(.1257)	
Coalition						.2170
						(.3101)
N. Observations	36	36	36	36	36	36
Residual/Null deviance	.6833	.7032	.7443	.6881	.6929	.6873

Standard errors in brackets. Level of Significance: *** p<0.01, ** p<0.05, * p<0.1

Note: Following a referee suggestion we tested also whether the size of the country, measured as the total population (LN), was relevant. The parameter was found not significant. Data was obtained from the World Bank (https://data.worldbank.org/indicator/SP.POP.TOTL). Results available upon request.

Table 3.6. Estimations of extensions of the model with emotions, beliefs and political survival

	Base model extended with type of decision-maker (9)	(13)	(14)	(15)	(16)
Constant	-37.3742***	-44.4938***	-37.8580***	-37.5440***	-41.7941***
	(3.0024)	(3.1912)	(3.3412)	(2.9558)	(2.6196)
Healthcare capacity	1.9742***	2.0338***	2.0301***	1.9462***	2.0790***
	(.2973)	(.2712)	(.3323)	(.2937)	(.2507)
Tourism	1.9011***	1.9030***	1.9299***	2.0037***	1.8751 ***
	(.2792)	(.2568)	(.2979)	(.2789)	(.2386)
Trade	1.4032***	1.3059***	1.3981***	1.4463***	1.2296***
	(.2530)	(.2377)	(.2526)	(.2502)	(.2180)
Locked countries	-.6677***	-.6670***	-.6568***	-.7285***	-.6133***
	(.1284)	(.1216)	(.1288)	(.1371)	(.1041)
Unitary state	1.0113***	1.2142***	1.0036***	1.0219***	1.2222***
	(.3135)	(.3004)	(.3139)	(.3129)	(.2717)
Km form Wuhan		.7701***			

		(.2308)			
Gender PM			-.1167		
			(.3586)		
Ideology				.1741	
				(.1420)	
Days to next election					.6194***
					(.1402)
N. Observations	36	36	36	36	36
Residual/Null deviance	.7443	.7923	.7450	.7539	.8104

Standard errors in brackets. Level of Significance: *** p<0.01, ** p<0.05, * p<0.1

Note: Although "trust in government" reflects public perception, rather than government's beliefs, it might also inform governments' beliefs on the potential acceptance of hard measures by population (Robinson et al., 2020). We investigated its relevance, using a ranking provided by the World Bank database, *Public Trust in Politicians* (https://govdata360.worldbank.org/). The variable is not relevant to response agility (results available in table A3, in the Appendix). This is consistent with findings in Mizrahi, Vigoda-Gadot and Cohen (2021) that during crises citizens value more transparency and responsiveness than trust. Like authoritarianism, however, trust may be relevant to response severity and a topic for further research. For example, Sweden was the only country able to sustain a recommendation-based strategy; it may be significant that Sweden has one of the highest scores for "trust in government" (5.24 over 7 vs. an average of 3.59 for other countries) and the lowest score for authoritarianism (0 out of 100).

Table 3.7. Robustness check including additional countries in the sample.

	Base model OECD (1)	Base Model (17)	Extended model OECD (18)	Extended model (19)
Constant	-35.7629***	-33.8892***	-44.2591***	-41.4189***
	(3.2399)	(2.6753)	(2.8718)	(3.2929)
Healthcare capacity	1.8814***	1.8423***	2.0619***	1.9154***
	(.3199)	(.2713)	(.2438)	(.2591)
Tourism	1.7654***	1.3735***	1.8647***	1.4459***
	(.3086)	(.2676)	(.2326)	(.2394)
Trade	1.4632***	1.3094***	1.2145***	1.1470***
	(.2760)	(.2671)	(.2150)	(.2481)
Locked countries	-.6307***	-.6258***	-.6399***	-.6441***
	(.1359)	(.1211)	(.1055)	(.1101)
Unitary state			1.2965***	1.1224***
			(.2711)	(.3483)
Km from Wuhan			.3808*	.6654**
			(.2266)	(.2786)
Days to next election			.5083***	.0929
			(.1489)	(.1569)
Num. observations	36	45	36	45
Residual/Null deviance	.6833	.6897	.8316	.7565

Note: Standard errors in brackets. Level of Significance: *** p<0.01, ** p<0.05, * p<0.1

3.1.4.d. Robustness check and final model interpretation

We conduct two robustness checks and estimate the final model. First, we check whether the base model and significant extensions are robust to the inclusion of new countries. We introduce to the sample the five non-OECD EU states (Bulgaria, Romania, Cyprus, Malta, and Hungary) and four EU-candidate states (Albania, Montenegro, North Macedonia, and Serbia)

As table 3.7 shows, the base model is robust to the inclusion of additional countries (Estimation 17). Both the type of player extension and proximity bias are also robust (Estimation 19). However, the policy-survival factor is not significant when additional countries are included.

Next, we carry out an additional robustness check by conducting a Bayesian estimation of the model. Low sample size can lead to less robust estimations of parameters and standard errors, thus compromising the GLM significance test, which relies on asymptotic properties of the estimators (Western and Jackman, 1994). We perform the Bayesian estimation using the *brms package* available in R (Bürkner, 2017) and using no prior to avoid introducing any bias. Since the *days to election* variable is not robust to the inclusion of additional countries, we include only the *Unitary* dummy and the *kilometers from Wuhan* extension. As Figure 3.2 shows, all parameters are robust to the Bayesian estimation.

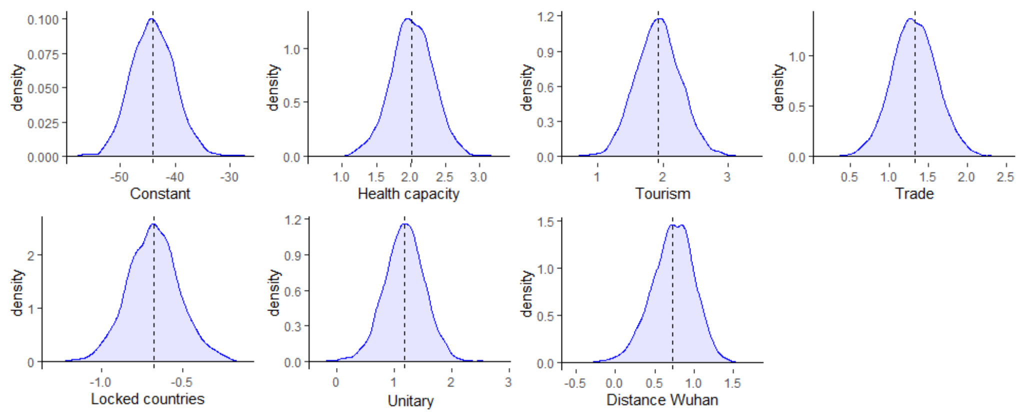

Figure 3.2. Distribution of the model parameters using a Bayesian estimation

Finally, to gain a complete understanding of the model beyond the significance of the parameters, we estimate the relative importance of each variable included in the model, using a new methodology for model interpretation suggested by Lundeberg and Lee (2017, 2019): SHAP (SHapley Additive ExPlanation) values. On synthesis, given an observation $x = (x_1, \ldots, x_J)$, the SHAP value of feature *j* on instance *x* corresponds to the way in which the concrete value of feature *j* on *x* modifies the output of the model with respect to other instances that share some features with *x* but not *j*. For a parametric model $F(x) = g(\sum_j \alpha_j x_j)$, where g is a function of the weighted features of *x,* the SHAP value corresponds to: $\varphi_j(x) = \alpha_j(x_j - E(X_j))$ where *X* is the set of observations and $E(X_j)$ is the average value of the *j* feature on *X*. Then, noting as *N* the total number of observations, we can estimate the relative importance of feature *j* in the model as:

$$RI_j = \frac{\sum_{i=1}^{n} |\varphi_j(x_i)|}{\sum_{k=1}^{J} \sum_{i=1}^{n} |\varphi_k(x_i)|}$$

Table 3.8 presents the relative importance of each variable in the final model, estimated using the Bayesian approach.

Table 3.8: Final model

	Final model (12)	Bayesian estimate [a]	Relative importance
Constant	-44.4938 ***	-44.0198***	
	(3.1912)	(4.0020)	
Healthcare capacity	2.0338***	2.0172***	26.6%
	(.2712)	(.3113)	
Tourism	1.9030***	1.9227***	20.9%
	(.2568)	(.3456)	
Trade	1.3059***	1.3228***	16.1%
	(.2377)	(.2884)	
Locked countries	-.6670***	-.6773***	19.5%
	(.1216)	(.1606)	
Unitary state	1.2142***	1.1824***	11.0%

	(0.3004)	(0.3554)	
Km from Wuhan	.7701***	.7288***	5.9%
	(.2308)	(.2737)	
Num. observations	36	36	
Residual/Null deviance	.7923	.7945	

Note: Standard errors in brackets. Level of Significance: ***p<0.01, **p<0.05, *p<0.1

[a] The Reset test for functional form or omitted variables with a polynomial fitting of degree 4 does not reject the null hypothesis (p-value 0.6049). Therefore, the functional form is correct, and the estimates do not suffer from omitted variables.

3.1.5. Discussion and policy implications

All governments have been overwhelmed by the pandemic and forced to implement hard measures to avoid a complete healthcare-system collapse and its associated fatalities, which would have led to an even more negative valuation of their policy responses. According to our theoretical model, once a government has a clear expectation that it will have to implement hard measures, the choice to enact them immediately strictly dominates the choice to delay them. For this reason, the fact that healthcare-system capacity and cost-related variables have a significant influence on reaction time has a very relevant implication: they negatively affected government strategy.

Because initial expectations did not match reality (otherwise, governments would have had not taken hard measures), governments with strong healthcare systems were overconfident about their ability to fight the outbreak and did not immediately implement hard measures. The associated economic costs created a fear of excessive economic damage. Both overconfidence and economic fears delayed the implementation of hard measures, increasing overall costs. Notice that implementing ‘hard measures’ as a result of a ‘rational’ (cost-benefit based) decision process with incomplete information might not have been ‘optimal’ in all countries. Whether hard measures had been or not optimal in each case would need an ex-post evaluation of effects, which is beyond the scope of our study, and estimating the actual impact of the pandemic on the fatality rates and economic costs associated with agile and slow policy responses is a question for future research, as complete data will not be available until the COVID-19 crisis is over.

Our results are empirically robust and supported by frequent public statements made by political leaders. Indeed, overconfidence in healthcare-system capacity has been described as one of the main causes of policy-response delays by global healthcare experts, including Pedro Alonso, Director of the World Health Organization's Malaria Program, who said on May 6th that Western pride prevented most advanced countries from reacting quickly.

As the pandemic triggered a decision-making process based on incomplete information, variables related to additional information (e.g., the policy responses of other countries) and valuation of risk (proximity bias) are key factors, directly accounting for 25% of the total.

Finally, multi-level governance is also relevant. Federal countries, which are more decentralized and better at fostering political collaboration, were more agile than unitary states.

We are aware that our identification strategy cannot draw strong claims of causal relations from these empirical results; this is a limitation of the present research. However, we believe that our theoretical model (built on a very simple hypothesis), when combined with many statements and observations that support a causal relationship—made by policy leaders and healthcare experts—can reduce this limitation.

There is a wide consensus that strong healthcare-system capacity improves social welfare, while high levels of trade and tourism are important engines of economic growth. However, these benefits risk biasing governments, particularly in the context of crisis management under incomplete information. Ballesteros and Kunreuhter (2018, p. 9), in their analysis of organizational decision-making in the face of uncertainty shocks, warn that: "the riskification of uncertainty leads to the delusion that increasing formal insurance take-up is a sufficient mechanism to reduce vulnerability against uncertainty shocks." An important policy implication emerges from this analysis. The COVID-19 pandemic has generated frequent demands to increase health expenditure. Indeed, such expenditure may improve health-system performance on a regular day-to-day basis, as long as the additional capacity meets positive social cost-benefit requirements. However, it will not provide full insurance for managing future pandemics, as strong healthcare-system capacity can induce governments to make riskier decisions, particularly under incomplete information.

3.1.6. Conclusion

In this part of the study, we have built a theoretical model and used it to design and implement an empirical strategy, analyzing why some countries took longer than others to implement lockdown measures. In other words, we set out to discover the drivers of policy-response agility during the COVID-19 outbreak.

Our findings show that welfare variables, involving a cost-benefit analysis of policy responses, were the most significant drivers. Together, healthcare capacity and expected economic costs accounted for around 65% of the total importance. If governments have had complete information, we would have expected these factors not to be relevant; once governments know for certain that they must implement hard measures, they clearly prefer to anticipate rather than to delay. The importance of these variables therefore indicates that governments may have been biased in their risk assessment of the pandemic by healthcare-system capacity and the fear of direct economic costs.

In addition, information about the progress of the pandemic was a key driver, accounting for around 25% of total relevance. The more information governments had access to, the more agile they were in their policy responses. Last but not least, we found empirical evidence that decision-making processes and individual actors were also relevant. Decentralized federal states, which promote political competition, were more agile than unitary states.

While we found no evidence that concerns related to inequality, poverty or trust in government shaped policy-response agility, they may have influenced the severity of instituted measures. Hence, these topics deserve future research. Future studies should analyze in depth the wide range of policy responses in the U.S., given its complex governance, institutional design, and comparatively high level of political and ideological polarization.

Endnotes

[1]. Simulations done using the R *EpiModel* package for SIR models, with a population of size N = 1000, 1 initial case, and an initial reproductive number

of 1.85 (assuming 6.1 days of effective transmission, before the individual is quarantined, hospitalized, recovered or died, and an effective contact rate of 0.3), lead to achieving the pick of the pandemic after around 35 days, with daily new cases reaching the 3% of the total population.

[2.] We are aware that choosing a national scale may mask sub-national governments' activity, and this is a relevant issue in federal countries, as discussed in Downey and Myers (2020) when comparing the US and Australia. Unfortunately, most data we use in our empirical analysis (e.g. incidence rate) are only available at the national level, and this is a constraint to consider also subnational levels in our empirical exercise. This might be particularly relevant for the case of the US, where debate on executive federalism (e.g. Eleazar, 1993; Bulman-Pozen, 2016) has emphasized that the absence of a formal coordinating institution has impact on sub-national policy adoption in the US (Downey and Myers 2020). Because of this, on the one hand, we have included a categorical variable on whether the response was Unitarian or Federal (see the subsection "Type of decision-maker"). On the other hand, we re-run our estimations in table 3 also excluding the US, and the results remained the same.

3.2. Part 2: Ideology, political polarization, and agility of policy responses: Was weak executive federalism a curse or a blessing for COVID-19 management in the US?

This part is a joint work with Germà Bel and Óscar Gasulla, and consists of a (forthcoming) paper published in the Cambridge Journal of Regions, Economy and Society: Gasulla, O., Bel, G., and Mazaira-Font, F. (2022). Ideology, political polarisation and agility of policy responses: was weak executive federalism a curse or a blessing for COVID-19 management in the USA?, Cambridge Journal of Regions, Economy and Society, https://doi.org/10.1093/cjres/rsac033

Abstract: We investigate whether weak executive federalism was beneficial or damaging for COVID-19 management in the US. We formulate a policy response model for subnational governments, considering the national government's preferred policy, in addition to other factors, with incomplete and with complete information. The hypotheses derived are tested using econometric techniques. Our results suggest that ideological and political biases were more influential in a situation of incomplete information than in one of complete information. As such, weak executive federalism allowed more agile policy responses in Democrat-led states when information was incomplete, thus reducing the rates of incidence and mortality. When information was complete, ideological and political biases were found to be of no relevance at all.

Keywords: COVID-19; crisis management; public policy; policy response; federalism.

JEL CODES: D81; H12; H77; H118

3.2.1. Introduction

The COVID-19 crisis has intensified the debate regarding the respective effectiveness of centralized and decentralized responses to emergency situations. Indeed, recent studies have analysed differences in the policy responses to the pandemic of federal and unitary countries (e.g. Chattopadhyay et al., 2021). However, if we look beyond these specific case studies, multivariate empirical cross-country analyses have tended to conclude that federal countries adopted more agile (Bel et al., 2021) and more effective (Toshkov et al., 2021) policy responses. Yet, in line with long-standing debates about the relative strengths and weaknesses of executive federalism in the US (e.g. Eleazar, 1993; Bulman-Pozen, 2016), the early COVID-19-related literature largely attributes the US's mediocre performance in the crisis in 2020 to failings in these processes of intergovernmental negotiation (e.g. Kettl, 2020).

Thus, our primary research question, here, is whether a weak executive federalism was a curse or a blessing for US management of the COVID-19 crisis. Our study builds on the policy response with the incomplete information model proposed in Bel, Gasulla and Mazaira-Font (2021). First, we modify this model to reflect the co-existence of national and subnational governments and then extend it to a subsequent situation characterised by complete information. Our main hypothesis is that the high degree of ideological and political polarization in the US caused inter-state differences in the agility and effectiveness of their policy response, and that this effect may have differed in scenarios with and without complete information. We test these hypotheses with data from the US measuring the intensity of policy response: first, with the initial hard measures taken against COVID-19 and, second, with the vaccination rollout.

Our study makes two contributions to the extant literature. First, rather than the timing of the response (i.e. who acted first), we evaluate the agility of response, relative, that is, to COVID-19 incidence rates and regional factors which might influence that policy response in a context of incomplete information. More specifically, we do not contribute by showing who acted first (which has been established in the literature); we contribute by showing who was more agile in the policy response (as distinct from 'being first'), which requires establishing a relation between policy response and rates of incidence of the virus, something that has not, to date, been attempted for the US.

Second, we also make an original contribution by evaluating the agility of response in the vaccination phase when the information on COVID costs was complete. In this regard, we expect to find, adjusting by the incidence rate, that Democratic-led states reacted quicker and with greater stringency with incomplete information (outbreak of the crisis), but did not do so with complete information (vaccination rollout).

Our empirical results indicate that political and partisan factors were more influential with incomplete information, but that their influence disappeared when information was more complete. Hence, with respect to our main research question, we can conclude that had there been more executive federalism available to the Trump administration, its performance in the initial stages of the COVID crisis would have been worse. In other words, the weakness of executive federalism in the US was a blessing rather than a curse for its COVID-19 management.

3.2.2. COVID and federalism: Related literature

The centralized vs. decentralized response to crises debate is long-standing. Christensen, Lægreid and Rykkja (2016) argue that decentralization can lead to greater agility and effectiveness, and Congleton (2021) claims that decentralization allows policy responses that are better tailored to environmental conditions and preferences, and favours innovation. However, Janssen and van der Voort (2020) conclude that the more agile policy response provided by decentralized management should be balanced with the fact that centralized management allows for better adaptive governance, especially the management of shared resources and assets (Dietz, Ostrom and Stern, 2003). Yet, on balance, multilevel systems in which different levels of government and non-state institutions engage in collaborative governance seem to provide incentives for more agile and effective responses (Scavo, Kearne and Kilroy, 2008; Downey and Myers, 2020).

The COVID-19 outbreak has sparked an intense debate on the potential differences in policy response to the crisis manifest by federal and unitary countries. To date, narrative discourses and case studies provide either contradictory or mixed results. For example, Kennedy, Sayers and Alcantara (2022), in an empirical analysis of political accountability and federalism in crisis management, find citizens unable to assign responsibility to the correct level of government in Canada; yet, Wehde and Choi (2021), in a study conducted in Oklahoma, US, find just the opposite. Interestingly, a narrative cross-country analysis comparing the COVID-19 management of federal and centralized countries tends to conclude that it was not whether countries had a federal or unitary structure, but rather whether they had better or worse governance, which influenced management of the COVID crisis (Cameron, 2021). Yet, beyond specific case studies, multivariate empirical cross-country analyses seem to find that federal countries had more agile (Bel et al., 2021) and more effective (Toshkov et al., 2021) policy responses.

Having said that, considerable diversity has been recorded in the COVID crisis management of federal countries. Thus, Hegele and Schnabel (2021) report predominantly centralized decision making in Austria and Switzerland but predominantly decentralized decision making in Germany, although Desson et al. (2020) conclude that flexible governance in all three instances contributed to comparatively better performance. Overall, a common recommendation that emerged during the crisis was to that of the need to improve intergovernmental relations and coordination (Chattopadhyay and Knüpling, 2021).

Political polarization has become more and more extreme in the US in recent decades (Nolette and Provost, 2018), and it seems this polarization, and its associated ideologies, played a significant role in the mediocre performance of COVID-19 management in the country (Jacobs, 2021). This situation tended to be exacerbated by increasingly disconnected Federal-State relations (Benton, 2020); in contrast, State-Local relations and coordination resulted in a much better performance (Benton, 2020; Mallinson, 2020). More specifically, various studies report that Republican-controlled states reacted later, re-opened sooner (Warner and Zhang, 2021), and implemented softer contingency measures, which were associated with a higher growth in the number of COVID-19 cases (Hallas et al., 2020; Shvetsova et al., 2022). However, none of these studies standardized the comparison by incidence rates – and as such may have generated misleading results – given that agility and severity would have depended on the risk level faced by each state.

The weakness of executive federalism in the US has been blamed for its mediocre performance in addressing the COVID-19 crisis in 2020 (see, for example, Bowling, Fisk and Morris, 2020; Kettl, 2020; López-Santana and Rocco, 2021; Rocco, Béland and Waddan, 2020). However, when other metrics are considered, such as the speed of vaccination rollout, the US led the rankings until summer 2021, and its efforts were, as of November 2021, comparable to those of such countries as Germany and Australia (Ritchie et al., 2021), typically considered as exemplifying federal countries with relatively good COVID management records (Rozell and Wilcox, 2020).

Given these differences in performance metrics, any hypothesis that seeks to link the weaknesses of executive federalism and a poor policy response to

COVID-19 is controversial. Indeed, Kincaid and Leckrone (2021: 243) conclude that "Executive federalism has been contentious, but federal and state agencies' bureaucratic relations continued to be largely cooperative, except when the Trump administration interfered with some federal agencies' functioning". Likewise, Cigler (2021: 674) argues that it was not the lack of federal powers that undermined performance in the US, but rather "the President's failure to accept responsibility and exercise existing authority quickly and fully, decisively and competently". Against this backdrop, it was state partisanship, rather than federalism, that shaped state public health interventions and resulted in differences in outcomes (Birkland et al., 2021; Neelon et al., 2021).

Our research here seeks to determine whether weak executive federalism is to be blamed for the relatively poor performance of COVID-19 management in the US. We compare the policy response of Republican- and Democrat-led states to the outbreak of the COVID-19 crisis, controlling for the risk factors in each state, and extend this analysis to the first stage of the vaccination rollout, so as to compare the response with and without complete information.

3.2.3. Modelling the subnational policy response to the crisis

We present a theoretical model that develops an empirical strategy which we then use to analyse the impact of political affiliation on policy-response agility. We build on the model proposed by Bel, Gasulla and Mazaira-Font (2021), representing a cost-benefit analysis undertaken by a rational government that cares about social welfare and which has incomplete information about the pandemic. Different strategies to manage the pandemic are analysed, which can be constrained by institutional characteristics, emotional biases, and the pursuit of self-interest.

We extend the basic model by inserting subnational leaders into a sequential decision-making process with incomplete information that translates into partial estimates of the parameters involved in the decision (including, for example, the effectiveness of their measures) and full disclosure of the preferences of the national leader, who is also involved in the process albeit at the national level. We assume two main types of measure: soft and hard. Soft measures (SMs), which are of the same nature at both the national and subnational level (the only difference being where they are applied), describe measures intended to contain transmission but without severely affecting human rights and freedom of movement (e.g. information campaigns, temperature controls at airports, etc.). Hard measures (HMs) refer to measures that do affect human rights and freedom of movement, such as lockdowns and border closures. As national and subnational leaders have different powers, their respectively imposed hard measures differ, and, as such, we can assume that two types of HM exist: subnational hard measures (SHM) and national hard measures (NHM).

3.2.3.a. Dynamics of the decision-making process

At the start of the pandemic, a set of exogenous factors, including the share of population above 65 years old and with pre-existing comorbidities (see Álvarez-Mon et al., 2021; Montserrat et al., 2021), determined the virus reproductive number under no containment measures, ρ, and the death rate, d, both at the national and subnational levels. With the information available at that time, transmission rates well above 1 were estimated for all countries and regions (Hilton and Keeling, 2020; Katul et al., 2020), and the overall fatality rate was estimated at between 0.4 and 1.4% (Verity et al., 2020). A total of 2,200,000 deaths was predicted for the US during the first outbreak if no contention measures were implemented (Ferguson et al., 2020).

In each time period, national and subnational governments could decide whether to implement either hard or soft measures to contain the virus, based on their powers. Moreover, the national leader could also urge subnational leaders to adhere to a specific strategy. Four scenarios in terms of measures implemented are, therefore, possible: First, both national and subnational leaders implement soft measures at time t and the transmission rate is somehow reduced. Second, the national leader implements hard measures but the subnational leaders adopt soft measures and the rate of transmission falls more than in the first scenario, but costs in terms of production increase. Third, only the subnational leaders implement hard measures but the national leader adopts soft measures and, here, the effects are (with respect to the first scenario) as in the second scenario, that is, a lower transmission rate and higher production costs. Finally, both national and subnational leaders implement hard measures, as a result of which the transmission rate is lower than in all the previous three scenarios and production costs are higher.

These scenarios can be expressed more precisely as follows:

1. If both national and subnational leaders implement soft measures at time t, the transmission rate is reduced to $\rho_t = \delta^{SS}\rho$.
2. If the national leader implements hard measures but the subnational leaders implement soft measures, at the subnational level there is a loss of π_0 units of utility (lost production) but the transmission rate is reduced to $\delta^{HN}\rho$, with $\delta^{HN} < \delta^{SS} < 1$.
3. If the subnational leaders are the only ones to implement hard measures, at a subnational level there is a loss of π_1 units of utility and a reduction in the transmission rate to $\delta^{HS}\rho$, with $\delta^{HS} < \delta^{SS} < 1$.
4. If both national and subnational leaders implement hard measures, there is a loss of $\pi_2 > \pi_0, \pi_1$ units of utility and a reduction in the transmission rate to $\delta^{HH}\rho$, with $\delta^{HH} < \delta^{HS}, \delta^{HN}$.

Notice that the efficiency of the measures depends on the measures themselves, and on the degree of compliance with them. Hence, the $\boldsymbol{\delta^i}$ factors have also to be interpreted by taking into consideration the degree of compliance expected from the population in relation to these measures.

3.2.3.b. Political factors involved in the decision-making process

As discussed at the beginning of this section, it can be assumed that decision-makers had to conduct a cost-benefit analysis when deciding which measures to implement and when to implement them: that is, they sought to maximize healthcare outcomes (keeping the number of deaths as low as possible) while incurring the minimum economic cost. Thus, they found themselves having to evaluate the different actions that might be taken in terms of both healthcare and economics.

Additionally, they might also have pursued their own self-interests, like staying in office, for example. Thus, it can be assumed that the economic costs of applying hard measures were slightly reduced, since in this way they avoided the political costs of voter punishment at the ballot box for their improper response to the crisis, above all in highly competitive contexts and in a period close to elections (Baekkeskov and Rubin, 2014).

For each scenario, we can consider that each subnational leader is subject to a penalty k if they do not follow the national leader's preferred strategy. We can assume the penalty to be small or negative even (a reward, in fact) if the subnational and national leaders belong to opposing political parties. In contrast, the penalty is expected to be positive if both the national and subnational leaders belong to the same party. Hence, we would expect subnational leaders to lean in the same direction as that of their leader (Kahneman, 2011; Levy Yeyati et al., 2020). We can also assume that the greater the political polarization, the higher the expected value of this penalty will be (see Goelzhauser and Konisky, 2020, for recent evidence of punitive federalism in the US).

3.2.3.c. Utility function of the decision-maker

Within this setting, let n_{t-1} be the number of infected people at the end of time t–1 in a given subnational region. At the beginning of period t, the virus infects $\rho_t n_{t-1}$ people, who are then treated. Let us denote by c the perceived capacity of the healthcare system to deal with the pandemic, which is assumed to be equal (in relative terms) for all regions. Notice that it is reasonable to assume that the perceived healthcare capacity is equal for all states at the subnational level, since all states operate under the same national healthcare system. If $n_{t-1} < c$, then no infected people die and all are cured at t. Otherwise, the number of fatalities at t is $f_t = d(n_t - c)$, and the rest are cured.

To illustrate how the process works, we present a four-period process of decision-making (Figure 3.3). At the outset, nature determines the initial number of infected people n_0 and the transmission rate ρ for each region. For the sake of simplicity, we do not include an index for each region, but these parameters are expected to vary across regions. However, the transmission rate is expected to be much higher than 1 in all regions (Hilton and Keeling, 2020; Katul et al., 2020). At t = 1, the infected people transmit the virus to others and then receive treatment. Therefore, $n_1 = \rho n_0$, and the number of fatalities at t = 1 is $f_1 = d$ max $\{n_0 - c, 0\}$. Both the subnational and national governments estimate the transmission rates $\rho_1 = \hat{\rho}$ and the total number of infected people, $\widehat{n_1}$.

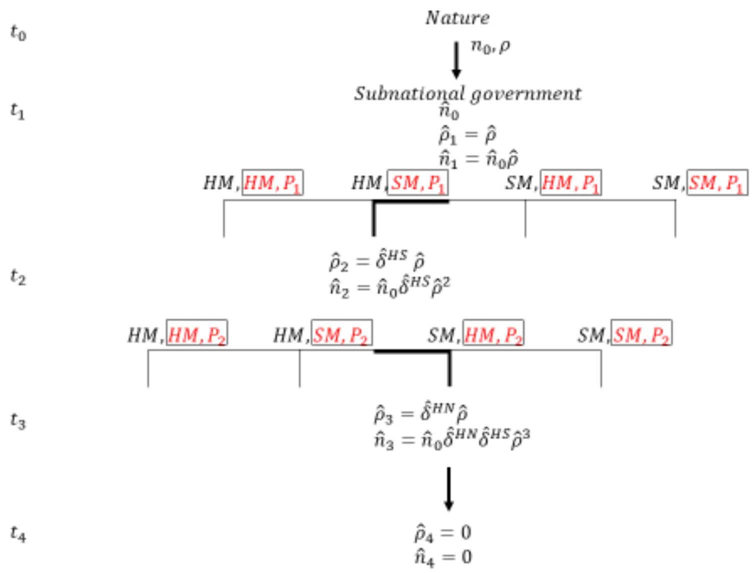

Figure 3.3. The 4-period decision-making process example for a subnational government. In boxes, decision taken by the national government.

Each subnational government estimates its own parameters and the national government estimates any additional ones. These estimates can be expected to be influenced by the advice of experts and national institutions, such as the White House COVID-19 Response Team. Based on this information, both sets of decision-makers estimate the expected transmission and death rates during the following periods, as well as the impact and cost of the various measures they might simultaneously implement: $\hat{\rho}_{t+1} = E_t(\rho_{t+1})$, $\hat{d}_{t+1} = E_t(d_{t+1})$, $\hat{c}_{t+1} = E_t(c_{t+1})$, $\widehat{\delta^{SS}} = E_t(\delta^{SS})$, $\widehat{\delta^{HN}} = E_t(\delta^{HN})$, $\widehat{\delta^{HS}} = E_t(\delta^{HS})$, $\widehat{\pi_i} = E_t(\pi_i)$).

Total fatalities are expected to be the product of the expected death rate and the total number of infected persons minus those that can be treated: $\hat{f}_{t+i}(\hat{n}_t) = \hat{d}_{t+i}\max\{\hat{n}_t - \hat{c}_{t+1}, 0\}$, with a total cost of $l\hat{f}_{t+i}(\hat{n}_t)$, where l represents the cost per fatality. Based on these estimates, the national government decides whether to implement soft or hard measures at the national level (in boxes in Figure 1), and which policy it prefers its subnational leaders to adopt ($P_j \in \{SM, SHM\}$).

The process continues until t=4, when a vaccine is discovered and, thanks to herd immunity, propagation falls to a stationary transmission rate, which is, on average, well below 1. Figure 1 shows how the subnational government expects the pandemic to evolve, at t=1, conditional to its deciding to implement hard measures at t=1 and soft measures at t=2, with the national government opting for the reverse strategy, and preferred subnational policies P_1 and P_2. Let us assume, for instance, that the national leader prefers to implement soft measures at the subnational level ($P_1 = P_2 = SM$), and that the national and subnational leaders belong to the same party. Then, the expected healthcare costs of the strategy shown in Figure 1 for the subnational leader in the first period correspond to the costs of the deaths of the infected population at the beginning of the pandemic. Thus, no measures were applied at the onset of the pandemic and the virus spread at the maximum transmission rate, which means at time t = 1 there are $\hat{\rho}\hat{n}_0$ infected persons. However, if the subnational leader implements hard measures during the first period while the national leader opts for soft measures, there is also an economic loss π_1. Moreover, the fact that the subnational leader is not following the preferred policy of the national leaders incurs a penalty k. Hence, the total expected cost in period 1 is

$$EC(HM, SM) = l\hat{f}_1(\hat{n}_0) + \pi_1 + k$$

Following the same reasoning for the subsequent periods, we find that the total expected cost of the strategy shown in Figure 1 is:

$$EC(HM, SM) = l\{\hat{f}_1(\hat{n}_0) + \hat{f}_2(\hat{\rho}\hat{n}_0) + \hat{f}_3(\hat{\rho}^2\hat{\delta}^{Hs}\hat{n}_0) + \hat{f}_4(\hat{\rho}^3\hat{\delta}^{HN}\hat{\delta}^{HS}\hat{n}_0)\} \\ + \pi_0 + \pi_1 + k$$

The decision whether to apply hard or soft measures at the subnational level depends on the trade-off between the expected number of lives saved and economic costs, as well as the potential political cost. However, in the subnational case, the trade-off is altered by the national government in two ways. First, by applying hard measures at the national scale, it reduces the incentive for hard measures at the subnational scale, since transmission rates are expected to decrease without the need for additional costs. Subnational governments would only implement hard measures if they expected – with the information available to them – that the benefit of applying subnational hard measures would be higher than their cost; for instance, if they expected the healthcare system to collapse even with the national hard measures in place.

Second, there is an additional political cost (or reward) in the equation, which stems from following or deviating from the national leader's preferred subnational policy. For instance, all things being equal, a subnational leader belonging to the same party as the national leader would be expected to have more incentives to apply subnational hard measures if this was the national leader's preferred policy.

3.2.3.d. Hypotheses derived from the model

Taking these differences into account, three main hypotheses emerge from the model at the subnational level:

H1: The higher the expected economic costs of subnational hard measures, the fewer the incentives for the subnational leader to implement these hard measures, especially if national hard measures are in place.

H2: Highly competitive contexts provide incentives for more agile policy responses at the subnational level.

H3: Highly polarized contexts provide incentives for subnational decision-makers to align with the national leader's preferred policy if they belong to the same party.

Finally, recall that for simplicity's sake we have assumed that at t=4 propagation falls to a stationary transmission rate due to the discovery and rollout of a vaccine. However, this also forms part of the decision process as policymakers have to decide on the percentage of the population to be vaccinated and the speed at which this target should be met. Both objectives are also subject to a cost-benefit analysis, but in this case complete information is available about vaccination costs, the reduction in the propagation of the virus and number of fatalities, and the costs avoided from continuing to implement hard measures. Notice, also, that in the vaccination process, variables related to the awareness and willingness of the population to be vaccinated might also play a role.

H4: The higher the costs of hard measures and the greater the efficiency of vaccination, the higher are the incentives for subnational decision-makers to implement a massive and rapid vaccination campaign regardless of policy competition or any other factor.

3.2.4. Variables, data and sources

Sample

Our empirical analysis is conducted for the US. To ensure homogeneity in terms of the implications of policy response, we present our estimates considering the 49 mainland states, excluding Hawaii on the grounds that it is isolated at sea, more than 2,100 miles from continental US. Below, we discuss the variables used and explain how they are specified, in relation to the theoretical model, and identify the sources from which the data were drawn.

Variables

Targets

Incidence rate when policy response began. We define the '*Incidence rate when policy response began*' as the number of coronavirus cases (based on the Johns Hopkins Coronavirus Resource Centre) adjusted per total population of a state, when the states' governors began to implement hard measures. This variable captures the agility of the policy response at the subnational level, as it identifies the stage of the pandemic when decision-makers reacted.

The *Stringency Index of The Oxford Covid-19 Government Response Tracker* (Hale et al., 2021; https://doi.org/10.1038/s41562-021-01079-8) was used to determine the moment when hard measures can be considered to have been implemented. This index records the strictness of 'lockdown style' policies that primarily restrict people's behaviour, including the closing of schools and workplaces, mandatory curfews, and border closures. It ranges from 0 (no measures at all) to 100 (maximum level of stringency).

When governments applied soft measures, the index ranged between 0 and 20; however, when restrictions of movement were imposed, it increased well above 30. The Federal government started applying hard measures on 16 March, five days after the World Health Organization (WHO) officially declared COVID-19 a pandemic. Borders were closed to non-essential travel, home schooling was recommended, as was avoiding social gatherings of more than 10 people, discretionary travel, and eating and drinking in bars, restaurants, and public food courts. These measures corresponded to a stringency index of 37.96. By that date, international pressure was considerable and many other countries had implemented even harder measures. For instance, the average stringency index of the measures implemented by EU countries was 59.39. Here, we define the moment when a state governor applied ‘hard measures’ as the day when the stringency of those measures was at least as high as that of those applied by the Federal Government (i.e. 37.96).

Early vaccination rate. We define the early vaccination rate as the percentage of vaccinated people amongst those eligible for vaccination in the first 60 days after the vaccine became available in the US (i.e. 11 February 2021). We used the vaccination rate as our main proxy for evaluating governor agility once they had complete information about the seriousness and costs of COVID-19 and experience in managing the pandemic. We used data from the subnational *Oxford Covid-19 Government Response Tracker*.

Covariates

Political affiliation. We analysed each governor’s political affiliation (source: National Governors Association). The variable equals 1 if the Governor is Republican, and 0 otherwise. We took into account that the Montana governorship flipped from Democrat to Republican following the November 2020 election.

As explicitly stated by President Trump, during the first COVID-19 outbreak the federal strategy was to impose hard measures at the national level as regards, that is, international travel and trade, while being much less restrictive at the subnational level. For instance, on 15 March, the Trump administration restricted all international travel while continuing to allow domestic flights.

Evidence supporting the hypothesis that the Trump administration sought to impose hard measures at the national level, but softer measures sub-nationally, can be found in Trump’s statements calling on various states to soften their lockdowns and to ‘liberate’, specifically, Michigan, Minnesota and Virginia (*New York Times,* 17 April 2020; all three states with a Democratic governor at that time). In the case of the vaccination campaign, newly elected President Biden, who began his term in January 2021, urged Americans to get their shot and enforced massive vaccination.

In line with these events and based on *H3*, we expect to find that Democrat governors reacted with greater agility during the first COVID-19 outbreak, while their Republican counterparts were more likely to follow Trump's strategy and apply subnational hard measures later. However, based on *H4*, we expect to find no differences in terms of agility attributable to political affiliation during the vaccination campaign.

Unemployment. To evaluate the economic baseline of a state, and the potential economic cost of the subnational measures, we gathered information on the unemployment rate in each state in January 2020 for the model with incomplete information and on the change in unemployment between January and November 2020 for the model with complete information. The data were obtained from the Bureau of Labor Statistics. We expected to find greater resistance to implementing hard measures under incomplete information at a higher unemployment rate, the economic fabric thus being more vulnerable to disruptive measures.

Proportion of elderly people. We estimated the perception of potential health costs attributable to COVID-19 at the subnational level for each state as the logarithm of the percentage of population 65 years or older (US Census Bureau). We expected governors of states with a higher proportion of elderly people to act faster and in a more effective way due to the greater vulnerability of that population segment to COVID-19 infection.

Days to next election. As voters may punish governments for improper crisis responses (Bueno de Mesquita et al., 2003), risk-averse administrations will implement proactive policies, especially within highly competitive contexts and close to elections (Baekkeskov and Rubin, 2014). The variable 'days to next election' corresponds to the logarithm of the number of days between the first diagnosed case of coronavirus in the state and the next scheduled state election date (National Governors Association and states' official websites) in the model without complete information (first outbreak). In the model with complete information (vaccination), it corresponds to the logarithm of the number of days between the first vaccination and the next scheduled state election date.

Covariates primarily affected by nation-wide measures

Evidence that healthcare capacity at the national level (e.g. health expenditure as % GDP) and tourist- and trade-related economic costs (%

contribution to GDP of tourism and trade) were relevant drivers of the agility of government policy responses to the COVID-19 outbreak has been reported by Bel, Gasulla and Mazaira-Font (2021) in their cross-country analysis. Given that perceptions of healthcare capacity refer primarily to the national level (see modelling section) and tourist- and trade-related economic costs are associated primarily with national measures, we did not expect these factors to be as influential at the subnational level as when employed in cross-country analyses. Nonetheless, we operationalized three variables to take them into account.

Number of beds. We included in our model for the first outbreak the variable 'number of beds', as a measure of state health system standalone capacity in terms of hospitalizations (Becker's Hospital Review, 2021).

Trade and tourism. We considered the relevance of the economic costs of nation-wide border closure using two indicators: total travel contribution and total trade (imports and exports), both as % of total GDP. The first indicator was obtained from the US Travel Association and the second from the US Census Bureau.

Covariates specifically related to the vaccination phase, and related to the capacity of the healthcare system to inoculate vaccines and the population's willingness to vaccinate.

Number of nurses. As a measure of the health system`s capacity to vaccinate the population, we included in our model the number of nurses per million inhabitants. Data were obtained from the US Census Bureau and the National Council of State Boards of Nursing's electronic information system.

Minority status. Evidence points to different levels of participation of ethnic minorities in medical research (e.g. Scharff et al., 2010), reflecting the history of federal medical studies conducted on vulnerable population groups (e.g. the Tuskegee syphilis study, see Tobin, 2022). Thus, we sought to determine whether greater mistrust among ethnic minorities with regards to medical research affected their vaccination dynamics. To this end, we included as a variable the percentage of non-white population in the state (with data being obtained from the US Census Bureau).

Education. The level of educational attainment is likely to be a factor in the vaccination decision, as the more educated are likely to have more and better information about the dynamics of the vaccination process and the availability of the vaccine. They are also more likely not to fear medical applications. Therefore, we included as a control the percentage of population with a bachelor's degree or higher educational attainment (with data again being obtained from the US Census Bureau).

Table 3.9 describes the variables and their sources. Table 3.10 shows the descriptive statistics, while Table 3.11 reports the average value of the variables for Democrat and Republican states.

Table 3.9. Variables: Description and sources

	Description	Source
Dependent variables		
Incidence rate at policy response	Number of diagnosed cases per 100,000 inhabitants when the State's Governors began to implement hard measures.	Oxford Covid-19 Government Response Tracker
Early vaccination rate	Percentage of vaccinated people amongst the eligible for vaccination group after the first 60 days the vaccine was available in the US	Oxford Covid-19 Government Response Tracker
Covariates		
Political affiliation	Political affiliation (Republican or Democrat) of each Governor	Official webpages
Unemployment	Unemployment rate in January 2020 for every US state	US Bureau of Labor Statistics
Unemployment change	Change in unemployment rate between January 2020 and November 2020	US Bureau of Labor Statistics
Rate of elder people	Logarithm of the percentage of population 65 years or older (2018)	Population Reference Bureau

Days to next election	Logarithm of the number of days between the first diagnosed case in the state and the next scheduled state election date for the first model, and between the vaccine was available in the US and the next scheduled state election date for the second model (complete information)	National Governors' Association and States' institutional webs
National-affected covariates		
Number of beds	Number of hospital beds per 1,000 population in a US State	Becker's Hospital Review.
Trade	Logarithm of the trade (imports and exports) contribution as % of total GDP in 2018	US Census Bureau
Tourism	Logarithm of the travel contribution as % of total GDP in 2018	US Travel Association
Vaccine-related covariates		
Number of nurses	Number of registered nurses per 1,000,000 population in a US State (2019)	US Census Bureau, Nurses
Minority status	Percentage non-white population (2019)	US Census Bureau
Education	Percentage population 25 or older with Bachelor's Degree or higher (2019)	US Census Bureau

Table 3.10. Descriptive statistics

	Min	Max	Mean	St Dev

Incidence rate at policy response	0.10	45.63	3.43	7.00
Early vaccination rate	7.03	9.28	8.24	0.53
Political affiliation	0.00	1.00	0.51	0.51
Unemployment	-3.77	-2.81	-3.38	0.23
Unemployment change	-1.10	7.50	2.53	1.85
Rate of elder people	-2.20	-1.58	-1.81	0.12
Days to next election (1st outbreak)	5.35	7.19	6.53	0.63
Days to next election (vaccination)	5.76	7.27	6.69	0.37
Death rate November 2020 (x 100,000 inhabitants)	15.22	199.85	86.50	43.46
Number of beds (x 1,000 inhabitants)	1.70	4.80	2.58	0.69
Trade	-1.33	-0.38	-0.82	0.21
Tourism	-1.49	-0.90	-1.27	0.13
Number of nurses (x 1,000,000 inhabitants)	1.19	3.30	1.76	0.40
Minority status (percentage of non-white population)	0.07	0.75	0.31	0.16
Education (% Bachelor's degree or higher, pop. 25 or older)	0.21	0.07	0.33	0.60

Table 3.11. Mean of the variable according to the political affiliation of the governor

	Republicans	Democrats
Number of States	25	24
Incidence rate at policy response	4.63	2.16
Early vaccination rate	17.88	17.83
Unemployment	-3.43	-3.30
Unemployment change	1.95	3.18
Rate of elder people	-1.78	-1.83

Days to next election (1st outbreak)	6.52	6.55
Days to next election (vaccination)	6.74	6.63
Death rate November 2020 (x 100,000 inhabitants)	85.71	90.34
Number of beds (x 1,000 inhabitants)	2.72	2.43
Trade	-0.82	-0.80
Tourism	-1.25	-1.28
Number of nurses (x 1,000,000 inhabitants)	1.77	1.74
Minority status	0.28	0.34
Education (% Bachelor's degree or higher, 25 or older)	0.31	0.34

3.2.5. Empirical model and results

The empirical analysis we conduct is based on the theoretical model presented. First, we estimate the difference between Republican- and Democrat-led states in terms of their agility of response. Recall that, for the scenario with incomplete information, this agility of response corresponds to the number of cases of infection adjusted by the total population when subnational decision-makers started to apply hard measures (that is, with a level of stringency at least as high as the measures implemented at the federal level); while for the vaccination process, it corresponds to the percentage of eligible population vaccinated in the first 60 days after the vaccine became available in the US. Second, we test whether the differences are relevant or not, after adjusting for the cost-benefit analysis presented in the model. Finally, we test whether the differences are relevant or not, again, after adjusting for all other extensions of the model.

3.2.5.a. Effect of political affiliation under incomplete information

As is well known, the Democrat-led states reacted earlier (in time) and with greater stringency than the Republican-led states (See Figure 3.4.). After the federal government started applying hard measures (16 March), the stringency index of subnational measures in Democrat states was 41.4 vs. 36.2 in Republican states. It would take the Republican-led states a further four days to achieve a stringency of 41.4. However, this does not imply that the Democrat states were more agile (in the sense of acting at an earlier stage of the pandemic), because they were hit earlier by the disease (Figure 3.5).

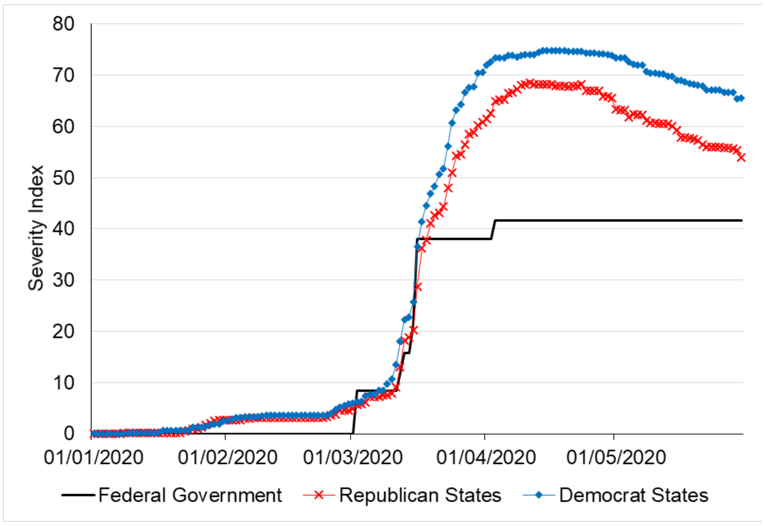

Figure 3.4. Severity of containment measures applied by the Federal US Government, Democrat-led States and Republican-led States; first COVID-19 outbreak

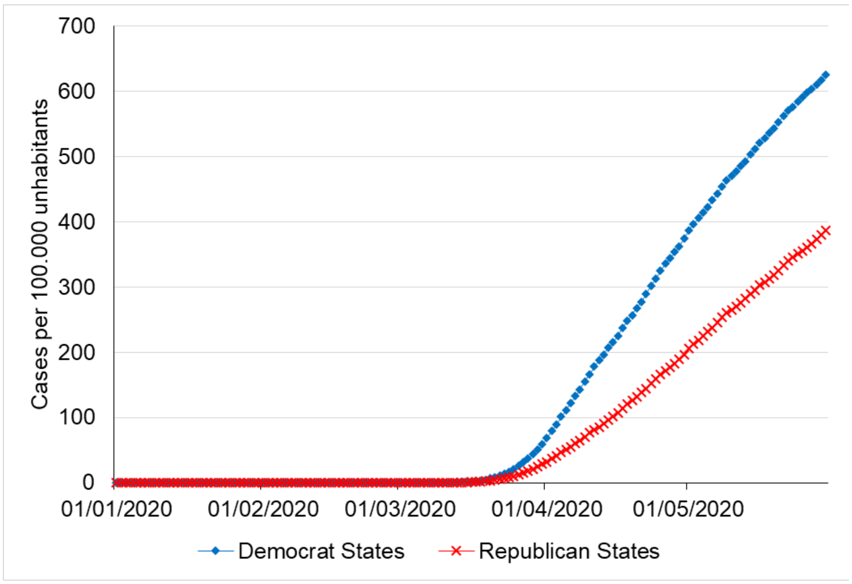

Figure 3.5. Average incidence rate of the states, by the political affiliation of the Governor, during the first outbreak of COVID-19

To consider this difference, we have defined agility in the policy response to COVID-19 as the number of cases of infection adjusted by the total population from the time hard measures were adopted. To test the effect of political affiliation on agility we began with a simple model of the form:

$$Cases = f(population, political\ affiliation\ )$$

As in Bel, Gasulla, and Mazaira-Font (2021), we used a negative binomial distribution, given the problem's non-negative discrete nature. Alternative techniques, such as OLS, might also be used; however, they would require transforming the target and the variance of the problem (for instance, considering the natural logarithm of the cases per million population), which make them less suitable. Nevertheless, we considered also a Bayesian robustness check, with no prior information on the distribution of the parameters, to avoid any potential bias of the estimates due to assumptions about the distribution of the parameters. According to our theoretical model, with incomplete information, agility at the subnational level is also expected to be affected by the relative costs of subnational hard measures. We estimated a base cost-benefit model as:

$$Cases = f(population, unemployment, old\ people, political\ affiliation\ )$$

Next, we checked whether the inclusion of costs primarily related to national-level measures, such as tourism and trade, and healthcare capacity was relevant (estimation 3). We finally tested the robustness of the political affiliation effect when including other political competition effects (estimation 4).

Table 3.12. presents the different estimates. The political affiliation effect was highly significant in all estimations. Republican governors responded with less agility than their Democrat counterparts to the first outbreak of COVID-19, even when adjusting by cost-benefit and political competition effects. Had Democrat-led states reacted in the same way as the Republican-led states, their average number of cases per 100,000 inhabitants at the moment of policy response would have risen from 2.16 to between 4.36 and 4.74; that is, the rate would have more than doubled. This provides a sound rationale for the fact that Republican-led states ended up with more cases in the subsequent outbreaks.

Table 3.12. Estimated parameters under incomplete information (GLM negative binomial)

	(1)	(2)	(3)	(4)
Constant	-10.722***	-11.999***	-11.244***	-12.086***
	(0.219)	(3.248)	(3.891)	(3.751)
Republican governor	.745**	.761**	.813**	.757***
	(.304)	(.307)	(.323)	(.306)
Unemployment		1.741***	1.851***	1.731***
		(.675)	(.691)	(.689)
% Old people		-3.865***	-4.076***	-3.845***
		(1.244)	(1.298)	(1.248)
Trade			-.127	
			(.719)	
Tourism			.417	
			(1.221)	
Number of beds			-.131	
			(0.227)	
Days to next election				.014
				(.240)
N. Observations	49	49	49	49
Residual/Null deviance	.911	.741	.733	.740

Standard errors in brackets. Level of Significance: *** p<0.01, ** p<0.05, * p<0.1

Note: We tested whether results are robust to the inclusion of Hawaii. In all cases, the political affiliation effect (*Republican governor*) is highly significant at 5% or 10% (results available upon request). We also tested whether geographical factors were determinant (dummy variables corresponding to East Coast, West Coast, and South; and Density of population, to account for rural versus urban dynamics). These are not significant and including them does not alter the significance of the other variables. The Republican governor effect for estimates including geographical factors lies in the range .767 to .804.

However, the fact that, Republican governors, on average, responded with less agility than their Democrat counterparts does not mean that all Republican governors responded with less agility. For example, Republican Mike DeWine (Ohio) applied hard measures with an incidence rate of 0.26

cases per 100,000 inhabitants, while Democrat Tony Evers (Wisconsin), whose state had lower levels of unemployment than Ohio (3.5 vs 4.1%) and a similar percentage of old people making up the population (17.0 vs 17.1%) applied these measures with an incidence rate of 1.25. Another example can be found in the comparison between Republican Eric Holcomb (Indiana) and Democrat J.B. Pritzker (Illinois): the former, with higher expected costs due to higher unemployment (3.1 vs 2.8%) and a higher share of old population (15.8 vs 15.6%) applied hard measures with an incidence rate of 0.37, while the latter did so with 2.15. In fact, Republican governors Mike DeWine (Ohio) and Larry Hogan (Maryland) were considered among the five most aggressive governors in fighting the pandemic outbreak (Scher, 2020).

Estimations (2), (3), and (4) show that subnational costs were relevant for agility. As expected, the higher the rate of unemployment (the higher the expected economic costs), the lower the agility; and the higher the percentage of old people (the higher the expected healthcare costs), the higher the agility. Notice also that results from estimation (3) showed no significant effect of health capacity and trade/tourism, in line with our expectations as explained when presenting our theoretical approach.

Our estimates rely on 49 data points. A small sample size can lead to a less robust estimation of parameters and standard errors, thus compromising the significance test of GLM, which relies on asymptotic properties of the estimators (Western and Jackman, 1994). Therefore, we conducted a robustness check by means of a Bayesian estimation of our model, which we performed using the *brms package* available in R (Bürkner, 2017), and using no prior information to avoid introducing any bias. Since the covariates primarily affected by national measures and the *days to election* variable are not relevant, we only estimated identification (2). As Figure 3.6 shows, all parameters were robust to the Bayesian estimation and close to the GLM estimates. Hence, there is no evidence of our results having been compromised by small sample size.

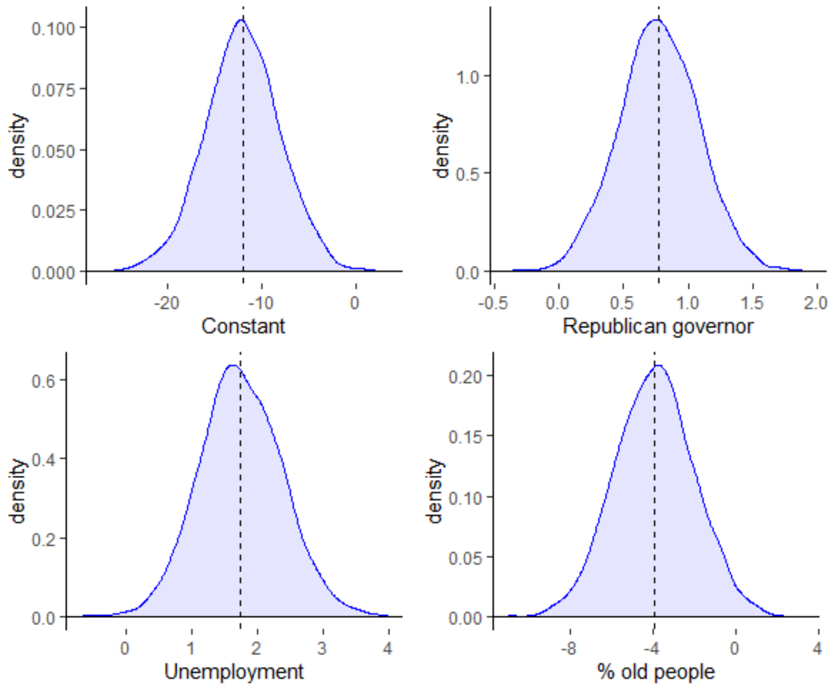

Figure 3.6. Distribution of the parameters of the model using a Bayesian estimation

3.2.5.b. Effect of political affiliation with complete information

After three major waves of COVID-19 in the US during the course of 2020, the great costs and losses attributable to the pandemic, in terms, that is, of premature deaths, long-term impairments, mental health losses and direct economic costs, were painfully evident. Indeed, Cutler and Summers (2020) estimate the costs at around $16 trillion. More specifically, the US GDP fell by 3.5% in 2020 (source: World Bank), while economic predictions for that year, made before the pandemic, were for 2.0% growth (Source: International Monetary Fund).

After the effectiveness of the Moderna, Pfizer, and Jansen vaccines had been demonstrated (94, 95, and 70%, respectively), the US Food and Drug Administration (FDA) issued an emergency use authorization to expedite their availability. In this way, the Pfizer-BioNTech COVID-19 Vaccine was approved on 11 December 2020, Moderna on 18 December 2020, and Jansen on 27 February 2021.

Once these data about the effectiveness of mass vaccination were made public and the information required for decision-making was complete, there were no incentives for subnational leaders not to implement a mass vaccination strategy as rapidly as possible. As such, we would expect to observe no statistically significant difference in terms of early vaccination rate attributable to the political affiliation of a State's governor.

As Figure 3.7 shows, there was almost no observable deviation between Republican- and Democrat-led states. On day 30, the vaccination rates for Republican and Democrat states were 3.95 and 3.72%, respectively; by day 45, they had risen to 9.35 and 9.15%, respectively; and, by day 60, they stood at 17.88 and 17.83%, respectively.

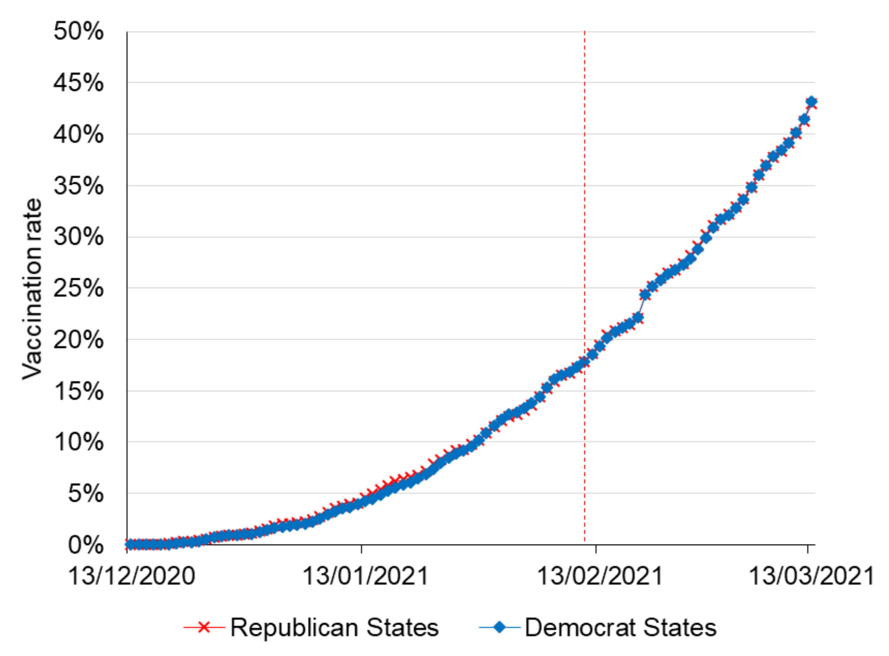

Figure 3.7. Average vaccination rate of Democrat-led States and Republican-led States during the first 3 months of vaccination. The dashed line corresponds to the day considered as the early vaccination rate (February 11th)

However, other drivers, including economic costs, healthcare costs or days to election (as a proxy of political competition costs), may have potentially influenced the agility of the vaccination program. Hence, we conducted three further estimations, following the same strategy as above. First, we estimated the effect of political affiliation on the vaccination rate of US states without considering any other covariate (estimation 5). Then, we checked whether the inclusion of subnational costs (unemployment change during the pandemic and percentage of population that died due to COVID-19 complications) was relevant (estimation 6). Finally, we included the number of days to the next election as a potential driver (estimation 7).

As Table 3.13 shows, political affiliation had no significant effect on delaying or accelerating the vaccination campaign, even when adjusting by economic costs, health costs, political competition factors, standalone healthcare capacity, minority status and education. This means that once there was complete information about the optimality of this policy and its outstanding social benefits, no differences according to political affiliation existed between the strategies implemented by the states, nor were they conditioned by other factors, consistent with *H4* herein. Since the health and economic costs of COVID-19 were extremely high for all states (Cutler and Summers, 2020), they all had great incentives to act as swiftly as possible. Moreover, and with respect to the lack of significance of the control variables specifically included in this last estimation (only the number of nurses has some significance -even if weak), it might well be that the huge dimension of the Covid-19 crisis and the information available about its effects reduced the relevance of differences in the variables affecting the willingness to be vaccinated for which we have controlled.

Table 3.13 Estimated parameters under complete information (OLS)

	(5)	(6)	(7)	(8)
Constant	17.824***	19.098***	14.793**	16.668***
	(.587)	(1.145)	(7.046)	(3.240)
Republican governor	.059	-.127	-.178	-.129
	(.805)	(.859)	(.871)	(.845)
Unemployment change		-.113	-.103	
		(.250)	(.252)	
% Deaths		-.010	-.009	
		(.010)	(.010)	
Days to next election			.629	
			(1.163)	
Number of nurses				1.598

				(1.159)
Minority status				1.029
				(3.244)
Education				5.698
				(6.428)
N. Observations	49	49	49	49
R-squared	0.00	0.04	0.04	0.05

Standard errors in brackets. Level of Significance: *** p<0.01, ** p<0.05, * p<0.1.

Note: Results are robust to the inclusion of Hawaii. They are also robust If we consider 30 or 90 days, instead of 60. In all cases, variables are non-significant (p>0.1). Results available upon request.

3.2.6. Discussion and policy implications

We have assessed the impact of political polarization on the agility of response to the COVID-19 crisis (adjusted by incidence rate) in two scenarios: First, during the first wave, under incomplete information; and, second, at the start of the vaccination rollout, when the severity and costs of COVID-19, as well as the effectiveness of the vaccines developed, were well known.

Our results provide robust evidence that, even when considering the inter-state differences in the initial evolution of the pandemic and differences in the risk and cost-related factors across states, Republican governors were – overall – less agile than their Democrat counterparts in responding to the health crisis. This provides a sound rationale for the fact that Republican-led states presented more cases of infection in the subsequent outbreaks, which is consistent with Neelon et al. (2021), who found that, adjusted by population and other factors such as the proportion of elderly people in the population, Republican-led states had lower COVID-19 incidence and risk rates than Democratic-led states from March 2020 to early July 2020, but that this association was then reversed.

Subnational cost considerations were relevant factors in explaining the agility of policy response, which is consistent with the results obtained for national decision-makers in Bel, Gasulla and Mazaira-Font (2021) in their cross-country analysis. However, unlike the results reported in this cross-country analysis, we did not find health capacity, tourism, and trade to be relevant. As discussed when formulating our theoretical model, we did, in fact, expect the perception of healthcare capacity to be similar for all states; hence, the subnational perception of this capacity was not expected to play a relevant role.

It seems reasonable to assume that the costs primarily affected by nation-wide measures, and although not equal for all states, were not given special consideration by governors when making their decisions, precisely because they were measures that were not completely under their control. Indeed, several governors asked domestic passengers arriving from other US states to self-quarantine, but they did not (or could not) order a border closure. Finally, estimates show that the policy survival variable "days until next election" was not significant.

Interestingly, political bias in the policy response, which was such a relevant factor at the time of the outbreak of the crisis, ceased to be important in the vaccination phase, when we found no difference in the agility of policy response between Republican- and Democrat-led states. The primary difference between the two phases was that information was much more limited in spring 2020 than it was by winter 2021, when information on the health and economic costs of COVID was much more robust, as was information on the efficacy of the vaccines. The evidence that ideological and partisan differences in policy response disappeared in the vaccination phase suggests that such biases had much greater potential to influence policy responses when information was incomplete than when information was more complete.

So, what can be concluded about the weak executive federalism in the US and the country's management of the COVID crisis? Lack of coordination has been blamed for shortcomings and overpricing in procuring medical supplies (Kettl, 2020: 599), but this criticism has likely been overemphasized. Spain's experience in this regard is highly illustrative: the Spanish government centralized all decision-making concerning the purchase of medical supplies, but little more than ten days later most regional governments began transgressing central procedures and implemented their own purchasing policies, prompted by the lack of efficacy of a central government that lacked experience in the practices of purchasing medical supplies, both nationally and internationally (Bel and Esteve, 2022).

If we look for a broader perspective on how crisis management was coordinated worldwide during the COVID-19 crisis, Dougherty et al. (2020)

have shown that centralization was, in fact, a key feature, with recentralization being twice as frequent as decentralization across OECD countries. A focus on the world's ten largest countries by population (and, hence, those with the most complex governance) shows that seven of them are federations (India, the US, Pakistan, Brazil, Nigeria, Russia, and Mexico), while three are unitary states (China, Indonesia, and Bangladesh). Here, the case studies published in Chattopadhyay and Knüpling (2021) reveal that all the federal states, with the exception of the US, centralized management. Therefore, the most likely counterfactual of executive federalism in the US, as a means of coordinating the crisis, would have been centralization.

In such a scenario, it is reasonable to conclude that had the US centralized crisis management, the Trump administration would have had greater latitude to impose its preferred policy on the Democrat-led states in spring 2020. Based on our empirical exercise, we estimate that if the Democrat states had responded more slowly (in line, that is, with those of the Republican states), the incidence of COVID-19 at the time of adopting hard measures would have increased from 2.16 to between 4.36 and 4.74. For the Democrat-led states this would have meant a much higher number of infections and deaths than they actually experienced, thanks to the greater agility of their responses.

The results of our study are consistent with Cigler's (2021) claim that it was not the lack of federal powers that undermined performance during the COVID crisis in the US, but rather Trump's mismanagement and his administration's general incompetence in exercising existing federal powers. In this regard, weak executive federalism proved to be beneficial for the agility of policy responses in the US, making it possible for the Democrat-led states to set their own priorities, based on their own specific health situation and policy preferences, and so contribute to decreasing rates of infection and, ultimately, saving lives.

3.3. Appendix

Table A.3.1.: List of date of first hard measures

Country	Date hard measures	Description
Australia	3/19/2020	Border closure; closure some non-essential shops; 4 Square meter rule
Austria	3/15/2020	Nationwide lockdown (including closure of schools), closure of all non-essential shops, ban of public gatherings
Belgium	3/12/2020	Closure of schools (but not universities), discos, cafes and restaurants, and the cancellation of all public gatherings for sporting, cultural or festive purposes
Bulgaria	3/13/2020	Closure of non-essential shops and workplaces, mandatory quarantine for all people coming from most affected countries
Canada	3/16/2020	Border closure, states of emergency including closure of non-essential shops, ban of public gathering, etc. in all Canadian states but Manitoba, New Brunswick and Nova Scotia
Chile	3/16/2020	Border closure, state of emergency, partial lockdowns in affected cities and regions, closure of schools with at least one case.
Croatia	3/17/2020	Closure of most non-essential shops, schools and universities; 14-days mandatory quarantine for people coming from affected countries, border closure
Cyprus	3/13/2020	Border closure, ban of public gatherings
Czech Rep.	3/12/2020	Border closure, nationwide curfew, schools suspended, closure of non-essential shops
Denmark	3/11/2020	Closure of schools and universities, banning of public gatherings, home-work public sector, border closure

Estonia	3/13/2020	Border closure, closure of schools, ban of public gatherings, closure of recreation and leisure shops
Finland	3/16/2020	Closure of schools and universities, banning of public gatherings, shut-down of most government-run facilities (libraries, etc.)
France	3/16/2020	Closure of most non-essential shops, ban of public gatherings, closure of schools and institutes of higher education
Germany	3/16/2020	Closure of education institutions, ban of public gatherings, closure of non-essential shops in some states
Greece	3/13/2020	Closure of education institutions, ban of public gatherings, closure of cafes, bars, museums, shopping centers, sports facilities and restaurants, border closure with limiting countries and affected countries
Hungary	3/15/2020	Closure of education institutions, bars, restaurants, cafes, public events, border closure
Iceland	3/13/2020	Closure of educational institutions, ban of public gatherings and events
Ireland	3/24/2020	Closure of education institutions, bars and public houses
Israel	3/14/2020	Closure of education institutions, most non-essential retail, ban of public gatherings
Italy	3/8/2020	Complete lockdown north Italy, ban public gatherings
Japan	3/5/2020	Closure of education institutions and extension of the law's emergency measures for an influenza outbreak to include COVID-19
Korea	2/20/2020	Border closure with China, massive testing and surveillance, partial lockdowns on more affected areas
Latvia	3/14/2020	Closure of educational institutions, ban of public events

Lithuania	3/12/2020	Closure of educational institutions, ban public gatherings, borders closure, closure of non-essential shops
Luxembourg	3/15/2020	Closure of non-essential shops, ban of public gatherings, closure educational institutions
Mexico	3/26/2020	Closure of non-essential shops and non-essential activities, ban of public gatherings, closure educational institutions
Netherlands	3/15/2020	Closure of educational institutions; closure of cafés, restaurants, sports clubs, saunas, sex clubs, coffeeshops, museums; ban of public events
New Zealand	3/23/2020	Border closure, ban of public gatherings, closure of all venues and enforcement of telework whenever possible
Norway	3/12/2020	Closure of kindergartens, schools, universities, and some none-essential shops (bars, restaurants, pubs, clubs, among others)
Poland	3/11/2020	Closure of all schools and universities, gathering restrictions and closure of cultural institutions, such as philharmonic orchestras, operas, theatres, museums, and cinemas
Portugal	3/12/2020	State of emergency; closure of establishments in the hospitality sectors such as restaurants, pubs, bars; public gathering restrictions; closure of all education institutions (from kindergartens to universities)
Romania	3/9/2020	Border closure with affected regions; all schools, kindergartens and universities closed
Slovak Rep.	3/15/2020	Implementation of state of emergency with all non-essential stores closed, closure of all schools and 14 days quarantine for people arriving from Slovakia from Italy, China, South Korea
Slovenia	3/15/2020	Closure of all educational institutions, bars and restaurants, and gathering restriction

Spain	3/14/2020	State of emergency declared, with closure of all educational institutions, hospitality sector establishments. People are to remain locked down in their homes except for essential activities
Sweden	3/27/2020	Reunion right restriction to 50 people
Switzerland	3/13/2020	Closure of all educational institutions and gathering restriction of more than 100 people, cancelation of all sport events
Turkey	3/12/2020	Closure of all schools and universities, travel bans and border closure with affected countries
U. Kingdom	3/18/2020	Closure of all schools, restaurants, pubs/clubs & indoor leisure facilities
United States	3/15/2020	State of emergency >25 states with closure of education institutions, curfew population, borders closure (main affected areas, including EU)
Serbia	3/15/2020	Closure of all education institutions from kindergartens to universities, ban public gathering, border closure
N. Macedonia	3/11/2020	Closure of all education institutions from kindergartens to universities, border closure and ban of public gatherings
Albania	3/8/2020	Closure of education institutions, gyms, bars and restaurants
Malta	3/12/2020	Closure of all schools, university and childcare, bars, restaurants and gym, mandatory quarantine to travelers from any country
Montenegro	3/13/2020	Closure education institution, bars & borders; ban on public gatherings

Note: Sweden never applied a lockdown strategy, preferring to follow a recommendation-based approach. The moment of policy response was March 27th, when the government banned public gatherings of more than 50 people and imposed up to 6-month prison sentences on those who broke the ban. This was the hardest measure approved in Sweden, 10 days after all European countries had closed their borders. At that point, two restrictions were in place: Sweden was isolated by its neighbors and public gatherings were prohibited.

Table A.3.2: Estimated parameters of the base model with second-order costs

	Base model (1)	(A1)	(A2)	(A3)
Constant	-35.7629***	-40.5855***	-36.5993***	-35.7945***
	(3.2399)	(4.1372)	(3.2741)	(3.2279)
Healthcare capacity	1.8814***	2.0900***	1.9468***	1.8941***
	(.3199)	(.3419)	(.3226)	(.3201)
Tourism	1.7654***	2.4624***	1.6669***	1.9072***
	(.3086)	(.2932)	(.3129)	(.3062)
Trade	1.4632***	2.0031***	1.4973***	1.5725***
	(.2760)	(.2700)	(.2718)	(.2746)
Locked countries	-.6307***	-.6779***	-.6067***	-.7077***
	(.1359)	(.1441)	(.1296)	(.1417)
% contribution MSMEs		.0143		
		(.0173)		
% unemployment			.0544	
			(.0400)	
% population >65				-.0454
				(.03726)
N. Observations	36	33	36	36
Residual/Null deviance	.6833	.7369	.6950	.6929

Standard errors in brackets. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.3.3: Estimated parameters of alternative specifications for ideology and trust

	(13)	(A4)	(A5)	(A6)
Constant	-44.4938***	-44.6347***	-47.2778***	-44.3957***
	(3.1912)	(4.2258)	(4.1158)	(3.6499)
Healthcare capacity	2.0338***	2.0284***	2.2444***	1.9620***
	(.2712)	(.3075)	(.3301)	(.2577)
Tourism	1.9030***	1.9105***	1.9475***	1.9121***
	(.2568)	(.2638)	(.2577)	(.2577)
Trade	1.3059***	1.3131***	1.3302***	1.3184***
	(.2377)	(.2801)	(.2427)	(.2392)
Locked countries	-.6670***	-.6723***	-.6108***	-.6794***
	(.1216)	(.1218)	(.1279)	(.1256)
Unitary state	1.2142***	1.2175***	1.3274***	1.2263***
	(.3004)	(.3106)	(.3276)	(.3015)
Km from Wuhan	.7701***	.7851***	.8057***	.8040***
	(.2308)	(.2317)	(.2296)	(.2438)
GINI index		-.0073		
		(2.6575)		
% poverty			.0216	
			(.0243)	
Trust				.0289
				(.1443)
N. Observations	36	36	36	36
Residual/Null deviance	.7923	.7924	.7926	.7911

Standard errors in brackets. Significance: *** p<0.01, ** p<0.05, * p<0.1

Chapter 4: Geography and Regional Economic Growth: The High Cost of Deviating from Nature

This chapter is a joint work with Daniel Albalate and Germà Bel, and consists of a paper published in the Journal of Regional Science: Albalate, D., Bel, G., and Mazaira-Font, F. (2022). Geography and Regional Economic Growth: The High Cost of Deviating from Nature. Journal of Regional Science 62 (2): 360-388

Abstract. We analyze the role of nature and geography in determining economic and social outcomes. We propose a theoretical model relating geography and nature to economic growth, and examine that model using data from NUTS 2 European regions. By doing this, we identify the predictive power of first-nature variables to explain regional population distribution. Then we analyze the effects of misadjustment between the actual and predicted distribution of populations on economic performance. Our results indicate that deviating from first-nature outcomes has a significant negative effect on economic growth. The main policy implication emerging from our analysis is that strategies that harmonize with nature and geography yield better social welfare than those policies that conflict with them.

Keywords: Geography; Population; Growth; Conditional convergence.

4.1. Introduction

Throughout history, humanity has made unimaginable progress, not only overcoming the obstacles or conditioning factors of nature and geography, but even putting them to use in the service of its aspirations and interests. Communities have been able to survive, develop and flourish in settlements with hostile climates, to overcome the limits set by water – the barrier effects of large rivers and oceans – and the terrain, with its harshness and geological complexity. The vast knowledge acquired and applied over time has led to technological advancement and hence to the possibility of mankind adapting to its environment and vice versa.

This progress has inspired claims on the *end of geography* (O'Brien 1992) and the *death of distance* (Cairncross 1995; 2001), and Friedman (2005) has described the new era as that of a *placeless society*, a *shrinking, flat world*, all due to advances in transportation, information and communication technologies. At the same time, the most isolated places in the world have become easily accessible. This has allowed the development of prosperous settlements, now integrated into the global economy, that would otherwise have been challenged by the limiting features attributable to their natural environments. Human life has apparently become liberated from the constraints of space and frictional effects of distance (Graham 1998).

Neglecting the ever-present role and restraints of nature and geography may be premature. Regarding the so-called *death of distance*, Rietveld and Vickerman (2004) pointed out that many economic activities have not become that 'footloose', due to transaction costs and other reasons. Proximity to higher-tiered urban centers continues to be an important positive determinant of local job growth, despite the alleged death of distance (Partridge *et al.* 2008). International conflicts, past and current economic dilemmas and challenges worldwide also exhibit a strong relationship with geography and nature (Senese 2005; Starr 2005; Kaplan 2012).

The role of nature and geography has been and still is crucial to understanding many of the social, political, and economic outcomes and prospects of human settlements. From a historical perspective, the capacity of the environment to support human life has commonly been considered a major restraint on population growth, density, and prosperity. This idea was discussed, for instance, by Machiavelli (1519), Botero (1588) and Montesquieu (1748), and formed part of the famous demographic theory of Malthus (1798). Earlier, Plato and Aristotle expressed concerns about overpopulation and limited resources. Thus, thinkers and scholars alike have long spoken of the importance of place and natural constraints and endowments for the location and density of human settlements and their effects on economic prospects.

In economic geography models, agglomerations are expected to be located and to develop according to a set of first-nature and second-nature determinants (Krugman 1993). Among first-nature determinants, geography and nature play the most crucial role. Economic activity depends on the physical landscape, climate, access to the sea and to navigable rivers - among other factors-. These natural factors are exogeneous to the economy (Rodríguez-Pose et al., 2013). Human action and incentives define the second-nature determinants that lead to increasing returns, due to scale and density economies, knowledge spillovers, etc. (Krugman 1991). Labor migrations between regions are responses to market signals and they determine the balance between agglomeration and dispersion forces (Krugman and Venables 1990).

While migrations had a basic survival objective for a long time, modern migrations have become more voluntary and selective, and basic survival has been often replaced with the objective of improving the migrant's own economic and social position (Faggian, Rajbhandari and Dotzel 2017, p. 130). In that regard, contemporary migration, particularly between developed economies, can be seen as a supply reaction to the creation of job opportunities in a region (Rajbhandari, Faggian and Partridge, 2020), involving migrants with higher human capital than the population that stays in their own origins. Moreover, locational advantages –attributable to geography and nature– favor the concentration and mobility of human and economic settlements (Black and Henderson 1998; Ellisson and Glaeser 1999; Glaeser and Shapiro 2003), leading to both the concentration of populations and the growth of productivity (Beeson 2001; Mitchener and McLean 2003).

The most extensive strand in the literature exploring this relationship is the one concerned with the role played by geography in relation to economic growth and development (Diamond 1997; Gallup, Sachs, and Mellinger 1999; Sachs and Warner 2001). Mitton (2016) evaluates the determinants of economic development in 1,867 subnational regions of 101 countries, focusing on within-country effects of geography and institutions. Several geographic factors had significant explanatory power for within-country differences in per-capita GDP, including terrain ruggedness, tropical climate, ocean access, temperature range, storm risk and natural resources such as oil, diamonds, and iron. Beyond the constraints imposed on economic prosperity, some authors have also argued that geography may have an impact on institutions, another relevant set of economic determinants (Acemoglu, Johnson, and Robinson 2001, 2002; Easterly and Levine, 2003). Moreover, this strand of the literature points out that even though it is still a determining factor, once institutions are accounted for, the contribution of geography as a determinant of economic growth partially diminishes (Acemoglu, Johnson, and Robinson 2001, 2002; Easterly 2001; Rodrik, Subramanian, and Trebbi 2004).

Geography and nature provide an endowment that may facilitate the location, concentration and growth of some settlements or make these more difficult in the case of others. This dependence can be tempered or even completely reversed by the intervention of human capital accumulation that translates into knowledge and technical advances (Glaeser *et al*. 2004; Bhattacharyya 2009) and/or by means of institutions (Acemoglu, Johnson, and Robinson 2002), which use a framework of incentives, regulations and investments.

In this article, we propose a theoretical model of the way in which features of geography and nature can account for regional economic growth, due to their effects on population density and distribution. This model is empirically examined using data from comparable European regions. We identify the strong predictive power of first-nature variables in explaining regional population density and capital city location, to the extent that we can estimate the degree of geographic harmonization of the actual distribution compared to the predicted distribution. This allows us to detect deviations produced by the forces of human action, led mainly by institutions, and to evaluate the predicted consequences in terms of relative economic performance. Our results indicate that deviating from nature's outcomes has a significant negative effect on economic growth and may also increase inequalities. This suggests that societies that opt to accommodate to the provisions of nature, and consequently, to exploit the opportunities of the best locations, rather than forcing a different distribution of the population across regions, perform better. A relevant policy implication emerges: policies that harmonize with nature and geography yield better social welfare than those policies that conflict with them.

The remainder of the article is organized as follows. First, we propose the theoretical model. After that we present the data and the empirical model that allows calculating deviations by considering actual versus predicted population density and capital city location, across regions and countries. Then, we interpret the results focusing on the countries that deviate the most. The economic cost of deviations is estimated in the next section, where an econometric convergence growth model is estimated. Finally, we discuss our main results and conclude.

4.2. Regional economic model

The aim of this section is to provide a conceptual model for understanding how population distribution and geography could impact economic growth.

Let us consider a closed economy formed by M regions. Let us assume that all of them occupy an equal area (equal to 1 to normalize), have equal access to technology A and a neoclassical production function of the form:

$$Y_i = F(g_i, K_i, D_i) = G(g_i)AK^{\alpha}D^{1-\alpha}$$

where K_i is the capital of a region i, D_i the total population living in the region (since area is equal to one, this is density of population), g_i its geographic endowment and $G(g)$ a function of geographic endowment such that $\frac{\partial G}{\partial g} > 0$. Total production of the economy is

$$Y = \sum_{i=1}^{M} Y_i$$

Firms' maximization problem

Let us consider that a large number of firms face the classic problem of profit maximization under conditions of competitive labor and capital markets in every region *i:*

$$\prod_i = \max_{K,D} F_i(g_i, K, D) - R_i K - w_i D$$

First-order conditions imply that capital and labor are paid their marginal contributions:

$$w_i = \frac{\partial F_i}{\partial D}, \; R_i = \frac{\partial F_i}{\partial K}$$

4.2.1. Capital markets

For there to be an equilibrium, assuming that financial markets are competitive and there are no externalities, we need the return on capital to be equal in all regions. Proposition 1 shows that there is equilibrium only in allocation of capital between the different regions.

Proposition 1

Let us note by K the total amount of capital in the economy. Let us also note by $K_i = t_i K$, with $0 < t_i < 1$ and $\sum_{i=1}^{M} t_i = 1$, the capital of the region i. Then, given $D_1, \ldots, D_M$, there exist unique values $t_1^*, \ldots, t_M^*$ such that $R_i = R_j$ for all i *and* j. These values correspond to a proportional allocation of capital with respect to production, that is:

$$t_i^* = \frac{Y_i}{Y}$$

Proof

Let us consider $M = 2$. By (1) and (4), return on capital is

$$R_i = \alpha \frac{Y_i}{K_i}$$

Let us note that $Y_i(t) = Y_i(g_i, tK, D_i)$. Then,

$$R_1 = R_2 \Leftrightarrow \alpha \frac{Y_1(t_1)}{t_1 K} = \alpha \frac{Y_2(t_2)}{t_2 K} \Leftrightarrow \frac{Y_1(t_1)}{t_1 K} = \frac{Y_2(1-t_1)}{(1-t_1)K}$$

Let us note that $X_1(t) = \frac{Y_1(t)}{t}$. Notice that $X_1(t)$ is decreasing, since

$$X_1'(t) = \frac{\frac{\partial Y_i(t)}{\partial K} tK - Y_1(t)}{t^2} = \frac{\alpha Y_1(t) - Y_1(t)}{t^2} = (\alpha - 1)\frac{Y_1(t)}{t^2} < 0$$

Using the same reasoning, $X_2(t) = \frac{Y_2(1-t)}{1-t}$ is increasing in t. Then, since $\lim_{t\to 0} X_1(t) = \infty$ *and* $\lim_{t\to 1} X_2(t) = \infty$, *there exists a unique t^* such that $X_1(t^*) = X_2(t^*)$*

Moreover,

$$\frac{Y_1(t^*)}{t^*} = \frac{Y_2(1-t^*)}{1-t^*} \Leftrightarrow Y_1(t^*) = t^*\big(Y_1(t^*) + Y_2(1-t^*)\big) \Leftrightarrow t^* = \frac{Y_1(t^*)}{Y}$$

Consider now the case of $M > 2$. By induction hypothesis, let us assume that the results hold for M regions. We want to see whether it holds if we consider an economy with $M + 1$. Let us define,

$$X(t) = \sum_{i=1}^{M} \frac{Y_i(tt_i)}{tt_i}$$

Where $t_1, \dots, t_M$ are the unique t_i (which exist by induction hypothesis) such that $\sum_{i=1}^{M} t_i = 1$ and $\beta \equiv \frac{Y_i(tt_i)}{tt_i} = \frac{Y_i(tt_j)}{tt_j}$ for all i and j. Notice that t_i does not depend on t, because the last equality holds for all $t \in (0,1)$.

As before, X is decreasing in t:

$$X'(t) = \sum_{i=1}^{M} \frac{\frac{\partial Y_i(t)}{\partial K} t t_i^2 K - t_i Y_i}{(t_i t)^2} = \sum_{i=1}^{M} (\alpha - 1) \frac{t_i Y_i}{(t_i t)^2} = (\alpha - 1) \frac{\beta M}{t} < 0$$

and $\lim_{t \to 1} X(t) = \infty$. *Note that* $X_{M+1}(t) = \frac{Y_{M+1}(1-t)}{1-t}$ *. As in the case of two regions, there exists* t^* *such that* $X(t^*) = X_{M+1}(t^*)$*, and by the induction hypothesis, there exist* $t_1, \ldots, t_M$*, such that* $X_i(t_i t^*) = X_{M+1}(t^*)$

4.2.2. Household maximization problem

Let us assume that households can choose where to locate and can move without any costs. Let us assume that utility of region i is a function of the form:

$$u_i(w_i, D_i) = f(w_i + \tau_i) + e(D_i)$$

where f is a concave and strictly increasing function, w_i is the net income per capita after taxes in the region i, τ_i represents public transfers per capita at region i and $e(D_i) = aD^2 + bD_i$ represents the externalities associated with density of population. Following theoretical models in urban economics (e.g., O'Sullivan 2007; Duranton and Puga 2020) we assume that increasing density in lowly populated areas has a positive effect (e.g., accessibility of a greater diversity of goods and services), but at a certain point externalities become negative (e.g., congestion or a rise in land prices). Thus, $a < 0$ and $b > 0$.

The household maximization challenge is to choose the region i that maximizes utility. To illustrate which kind of equilibrium will be reached, consider, w.l.g., the case of two regions.

Proposition 2

Let us note by D_i, $i = 1, 2$, the population living in region i. Then, $D = D_1 + D_2$ and $D_1 = tD$, with $t \in [0, 1]$. The necessary and sufficient conditions for the existence of a Nash equilibrium in population distribution between region 1 and 2, given capitals K_1 and K_2 are:

$t = 1$ and $u_1(D) > u_2(0)$, or

$t = 0$ and $u_1(0) < u_2(D)$, or

$t \in (0, 1)$ and $\frac{\partial u_i(D_i)}{\partial D_i} < 0$

When one region is empty, the equilibrium is Pareto optimal if, and only if, household utility of the non-empty region at D is higher than the maximum utility of the other region.

Proof

For there to be an equilibrium, expected utility gain from moving has to be non-positive. First, for i) and ii), let us consider, w.l.g., that $t = 1$. *If* $u_1(D) < u_2(0)$*, then the whole population would move to region 2, so* $t = 1$ *is an equilibrium if, and only if,* $u_1(D) > u_2(0)$. *For iii), let us consider that there is an equilibrium between regions 1 and 2, and people living in both. Since there cannot be any utility gain from moving, utility in both regions has to be the same. Thus,*

$$\hat{u} \equiv u_1(tD) = u_2((1-t)D)$$

Let us consider, w.l.g., that $\frac{\partial u_1(D_1)}{\partial D_1} > 0$. *Then, the marginal gain from moving from 2 to 1 is:*

$$u_1(D_1 + \varepsilon) - \hat{u} \approx u_1(D_1) + \frac{\partial u_1(D_1)}{\partial D_1}\varepsilon - \hat{u} = \frac{\partial u_1(D_1)}{\partial D_1}\varepsilon < 0$$

or all $\varepsilon > 0$, *small enough. Thus, households would be better off moving to region 1, and the distribution would not hold to a Nash equilibrium.*

Figures 4.1 present examples of the two equilibria for an economy with $\alpha = 0.3$, $g_1 = 1$, $g_2 = 1.1$, $K = 200$, $a = -5 \cdot 10^{-3}$ and $b = 3.75 \cdot 10^{-2}$. For the single region equilibrium (Figure 4.1.a), total population is 30. The red dotted line represents the utility people would get if they lived in the worst region. For the two-region equilibrium (Figure 4.1.b), total population is 100. The 10 percent relative difference in geographic endowment translates into a 22 percent relative difference in population density.

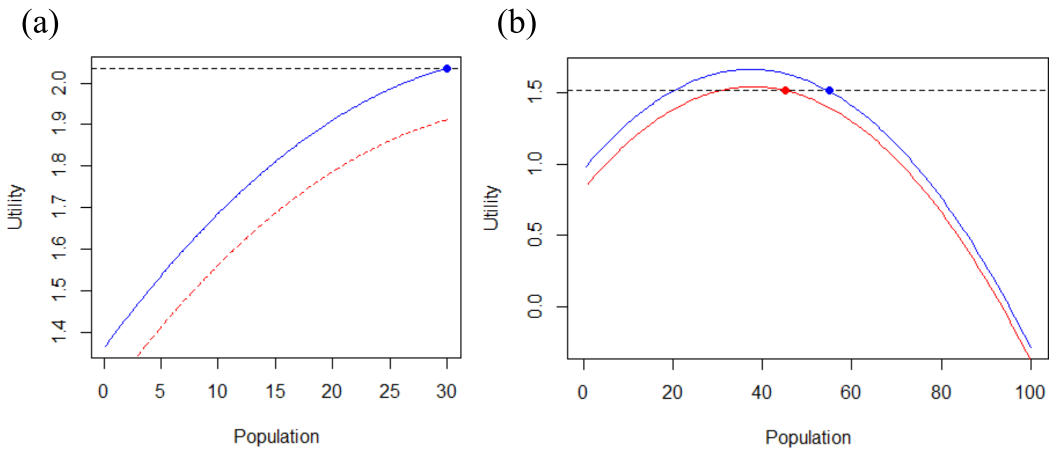

Figure 4.1. Single region equilibrium where the whole population lives in the best region (a) and two-region equilibrium (b).

4.2.3. Equilibria with migration costs

Proposition 2 shows that, as stated in the Introduction, migrations can be seen as consequence of welfare differences between regions (Faggian, Rajbhandari and Dotzel 2017, p. 130; Rajbhandari, Faggian and Partridge, 2020) and lead to a better social outcome, increasing productivity (see, for example, Mitchener and McLean 2003; Peri, 2012) and holdings of capital (Palivos, 2009), while having almost no negative effect on native wages (see, among others, Cortes, 2008; Card, 2009; Ottaviano and Peri, 2012). However, cultural concerns, job insecurity and stagnated disposable income trigger opposition to immigration (for example Malhotra, Margalit, and Mo, 2013; Dal Bó, et al. 2020), leading institutions to severely restrict free movement of human capital, which entails a high cost in terms of foregone GDP (see Clemens, 2011).

Hence, let us now discuss briefly how equilibria would change if there were moving costs c and information asymmetries, so that people from one region could not know exactly what their utility would be if they moved to the other region. If migration flows between regions have associate costs, potential migrants will remain in their regions if the expected increase in utility is lower than the moving costs:

$$u_2 + \varepsilon_1 - u_1 < c, \quad u_1 + \varepsilon_2 - u_2 < c$$

where ε_i represents the information asymmetry, that is, the error of households in region i when trying to anticipate what their utility would be after moving. Notice that derivatives do not play any role in this case. Moreover, as shown in Figure 4.2, two-region equilibrium is no longer guaranteed to be Pareto efficient, since the fact that those who move bear the whole cost of moving could deter people from doing so, even when the social gain from the utility increase produced by reducing the over-population externality could be much higher than the private cost of moving. In this situation, a government intervention could make everybody better off by, for example, subsidizing the moving cost. In other words, a non-Pareto efficient equilibrium with moving cost (left figure) can be Pareto improved (right figure) if moving costs -grey area- are subsidized (or if people are forced to move). Utility gains derived from the movement of people from blue to red region correspond to red and blue rectangles.

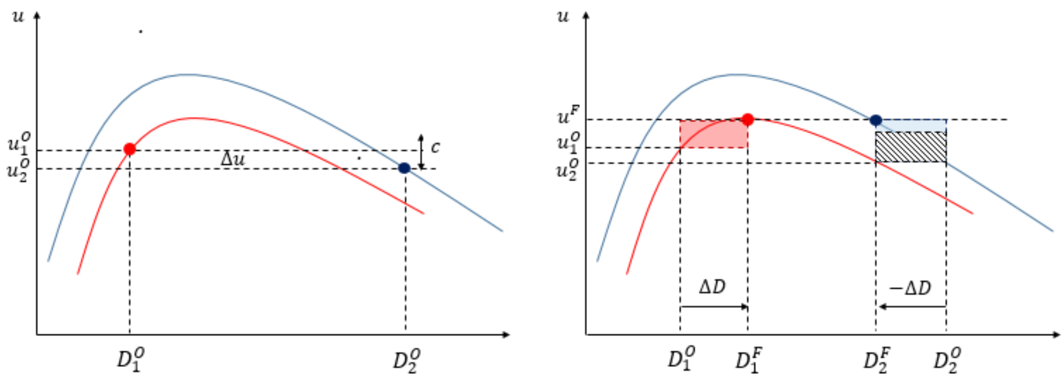

Figure 4.2. Two-region equilibrium and Pareto efficient equilibrium

4.2.4. Equilibria with spillovers and regional interdependences

In our model we have assumed that all regions had equal access to technology and that there were no spillovers or interdependences between them, besides the free movement of capital and population migrations, without and with costs. However, empirical evidence shows that economies' income levels are interdependent (e.g. Fingleton, 2003; Postiglione, Andreano, and Benedetti, 2013). To provide causal mechanisms and a theoretical framework to the hypothesis that the relative location of an economy influences its economic growth and steady-state, theoretical extensions of the Solow-Swan model introducing spatial externalities derived from (physical and human) capital accumulation (López-Bazo, Vayá and Artis, 2004; Fingleton and López-Bazo, 2006), and from technological interdependencies (Ertur and Koch, 2007), have been developed. These extensions preserve the main properties of the Solow-Swan model: the diminishing returns on capital and labor, the existence of a steady-state and the conditional convergence. Hence, either we consider technological interdependencies or externalities from capital accumulation, the consequence for our model is that the net salary of a region depends also on other regions. The properties of the utility function are preserved. It is concave and strictly increasing in the net salary and public transfers and exhibits the same density externalities.

However, due to interdependence with other regions, capital investment or any technological, capital or labor shock affecting one connected region also affects the others. Therefore, equilibria are of the same type than those analyzed before, but migration flows needed achieve and equilibrium after a shock or to move from a non-Pareto equilibrium to the Pareto equilibrium can be stressed or mitigated by regional interdependencies. For instance, let us assume an economy with three regions. Region 1 and Region 2 are highly interdependent, while Region 3 is independent. Let us also assume that at equilibrium all regions have the same density, $D = \frac{1}{3}$.

Now, consider the case where there is a technology shock in Region 1. If all regions have access to same technology, there would be no change in population distribution. If there are spatial spillovers and only Region 2 is interdependent with Region 3, then a migration flow is needed to achieve the new equilibrium. For the sake of simplicity, let us consider that around the equilibrium, the first-order approximation of the utilities is $u_1 = c - \gamma\left(D_1 - \frac{1}{3}\right) + \epsilon$, $u_2 = c - \gamma\left(D_2 - \frac{1}{3}\right) + \rho\epsilon$, and $u_3 = c - \gamma\left(D_3 - \frac{1}{3}\right)$, where c is the utility of the three regions at equilibrium, ϵ the technology shock and ρ the spatial interdependence factor. For utilities to be equal again, we need: $D_3 = \frac{1}{3} - \frac{(1+\rho)\epsilon}{\gamma}$, $D_2 = D_3 + \frac{\rho\epsilon}{\gamma}$, and $D_1 = D_3 + \frac{\epsilon}{\gamma}$. Notice that the higher the interdependence between Region 1 and 2, the larger the migration flow from Region 3 to the other Regions.

4.2.5. Implications

The model has three main implications. First, population distribution tends toward extreme outcomes and over-population. Geographic differences may lead to empty and over-populated regions, with respect to their optimal level. Moreover, relatively small differences in geographic endowment can also lead to much larger differences in population density. Second, the model predicts that areas with better geographic attributes will have higher densities of population (as in Beeson 2001; Mitchener and McLean 2003), unless there is much more public expenditure in worse areas or a historical legacy that cannot be overcome because of moving costs or incomplete information. In both cases, the result will be a non-Pareto equilibrium with lower utility and higher inequality. Third, more densely populated areas will tend to have higher output and capital per capita (as in Krugman 1991).[4]

As a final remark, notice that in our model we have not considered any potential density-related externality regarding production function (such as economies of scale). However, even with a simple neoclassical function, agglomerations emerge as an equilibrium as long as there are concave externalities with a monotonic change on population well-being. Including production externalities with a concave functional form, as for the households, will add an additional agglomeration force to the equilibrium, further reinforcing the implications of the model.

4.3. Data

In this section we present the data used to estimate the model on population density across European regions, which is presented in the following section.

Level of aggregation

As the aim of our article is to identify to what extent geographic drivers explain differences in population density, we need delimited regions to be small enough to consider their geographic attributes as representative of the region, and large enough to preclude specific municipalities or metropolitan areas that are particular agglomerations within a region. Therefore, we choose the level of aggregation denoted by NUTS 2 from the Nomenclature of Territorial Units for Statistics of Eurostat, which have population of between 800,000 and three million. The current NUTS 2016 classification, valid since 1 January 2018, lists 323 regions at NUTS 2 level.

We consider only EU member states, because other countries included in NUTS classifications, such as Switzerland or Norway, are missing key data from Eurostat. We also exclude EU territories located outside Europe,

[4] In Krugman the causality channel is due to externalities in production.

namely five from France (Overseas France), three from Spain (Ceuta, Melilla and Canary Islands), and two from Portugal (Açores and Madeira). Moreover, we include Bremen (DE50) in Lüneburg (DE93), Hamburg (DE60) in Schleswig-Holstein (DEF0), and we group all five London NUTS regions into one single NUTS 2 area. We end up with 258 NUTS 2 specimens. We consider density of population as at 1 January 2019.

Geographic indicators

We consider several potential geographic drivers of population density (as in Mitton 2016): temperature, rainfall, access to navigable waters and unevenness of the land.

Temperature. We use the daily heating and cooling degree days of each region. Degree days are used as an indicator of energy demand for the heating or cooling of buildings by comparing the day's average outside temperature against the optimal threshold of 18°C. The heating degree day (HDD) is the number of degrees below the threshold, the cooling degree day (CDD) the number of degrees above it. We compute the average of the HDDs and CDDs during the period 1977 to 2018, using data from Eurostat.

Daily rainfall. Average daily rainfall (l/m^2) has been calculated as the average of daily rainfall per year, using Copernicus Project data from 1977 to 2018.

Access to navigable waters. We have considered two main variables to capture access to navigable water. The first variable is the distance in kilometers from the boundary of the region to the nearest sea (or ocean). The second variable is direct accessibility to navigable rivers, that is, the kilometers of navigable river per squared kilometer of the NUTS2. We consider navigable rivers as those rivers with more than 100 segments or with a Strahler index higher than 5, according to the European Environment Agency Spatialite file for rivers. Figure 3 shows the map of the navigable rivers considered.

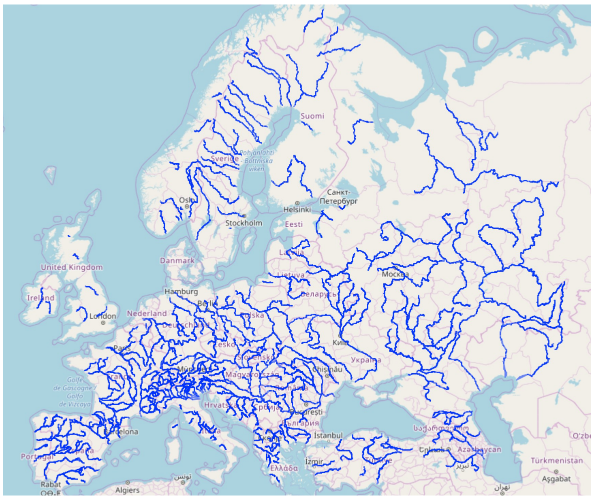

Figure 3. Main rivers in Europe.

Unevenness. We calculated unevenness as the interquartile range of the height of every LUCAS (Land Use and Coverage Area frame Survey) grid point of each NUTS 2 area. LUCAS is a survey carried out by Eurostat every three years to identify changes in land use and coverage. It contains observations from over 1,000,000 points.

Natural resources. We have considered as natural resources the presence of coal mines or oil refineries within the boundaries of the region. Data have been obtained from the Refineries Sites in Europe Database (Concawe Organitzation) and from the European Commission. Table 4.1 shows a description of the geographic data. Table A.4.1 in appendix provides detailed information on all variables and data sources.

Table 4.1. Summary of the data set.

Variable	Mean	St. Dev
Daily HDD	7.93	2.35
Daily CDD	0.16	0.24
Daily rain (l/m2)	2.07	0.60
Distance to nearest sea (km)	116.67	145.55
River density (km river/km2)	17.34	17.83
Unevenness (km)	0.2440	0.2434
Natural resources (%)	39.45	48.97

N = 258

It is important to highlight that in the Section Economic consequences of population misadjustment we build a conditional convergence model starting from 2001. To be consistent with this time span, we re-estimate HDD, CDD and daily rainfall from 1997 to 2000 for that model. Since, on average, the climate is very stable, changes happen on a geologic scale, and tend to affect all the regions simultaneously, the correlation between indicators built using data until 2000 or 2018 is higher than 99.5%.

4.4. Geographic endowment

Our approach to determining the relationship between geographic variables and population density is to build a linear regression model over the logarithm of the density of population, using the geographic indicators described, that is:

$$GE(i) \equiv \log(Density_i) = \sum_{j=1}^{J} \alpha_j \, GI_{ij} + u_i$$

where GE(i) refers to the geographic endowment given to the i-th NUTS 2 region, that is defined as the logarithm of its density of population. $GI_{i,j}$ is the value of the j-th geographic indicator of the NUTS 2 region i, and u_i is the residual.

4.4.1. Estimation

To estimate the geographic endowment, we exclude regions containing capital cities, since their population density can be strongly biased by political intervention. As the theoretical model suggests that small differences in geographic endowment can lead to substantial differences in density of population, the geographic indicators are included in quadratic form. Moreover, we distinguish between warmer and cooler seas by introducing an interaction between distance to the sea and temperature. Table 4.2 shows the results of a model built by weighting regions by their area size, so smaller regions are less important as the model is about density, and by selecting significant variables at 15 percent or higher. We perform two estimations. First, a Bayesian estimation using the brms package available in R (Bürkner 2017), without a prior to avoid introducing any bias. The Bayesian estimation does not assume asymptotic properties of the estimates and therefore is more suitable for significance tests on small sample sizes (Figure 4.4). Secondly, we also provide OLS estimates. As can be seen in Table 4.2, estimates are almost identical.

Notice that CDDs are not important, and only HDDs matter. This is because Europe is a relatively cold area, with an average of only 0.16 CDD versus 7.93 HDD. Notice also that only distance to the sea is important in terms of access to navigable waters. It may seem counter-intuitive that rivers are not relevant, but the explanation is that most regions have a navigable river (72 percent of the regions, corresponding to 88 percent of the total European area). Since the river network is very dense across Europe, it has no significant impact at a regional level. However, as we will see later, rivers do determine distribution within a region, that is, where to place a city within a region.

Table 4.2. Geography endowment model

	Bayesian estimation		OLS	
Variable name	Estimate	Std. Error	Estimate	Robust Std. Error
Constant	1.65874***	0.42509	1.65620**	0.82357
HDD	0.13837***	0.06235	0.13642	0.11530
HDD^2	-0.01687***	0.00289	-0.01680***	0.00645
Rain	2.01921***	0.31475	2.03330***	0.63571
$Rain^2$	-0.36743***	0.06396	-0.37035***	0.13885
$Uneveness^{-1}$	0.02243***	0.00431	0.02244***	0.00503
$Uneveness^{-2}$	-0.00004***	0.00001	-0.00004***	0.00001
Distance to the sea	-0.01028***	0.00214	-0.01031***	0.00255
HDD: Distance to sea	0.00129***	0.00024	0.00129***	0.00029
Natural resources	0.29930***	0.08413	0.29795***	0.09853

Significance codes: ***: p< 0.01; **: p<0.05; *: p< 0.1. N: 233 (capital regions not included) Errors weighted by area. R-squared: 0.7727. F-statistic: 84.2 (p-value: 0.000).

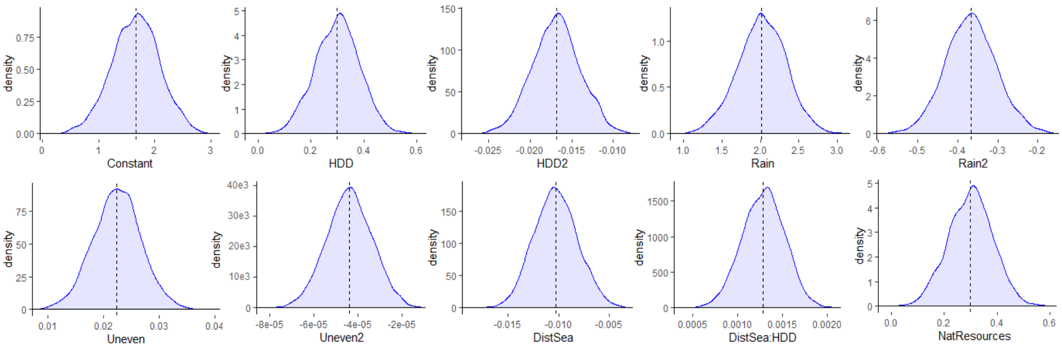

Figure 4.4. Distribution of the parameters of the model using a Bayesian estimation.

4.4.2. Interpretation

Understanding why a model makes a prediction can be as crucial as the prediction's accuracy. It provides insight into how a model can be improved and supports understanding of the process being modeled. To do so, we evaluate feature importance and plot the marginal relation estimated by the model between explanatory features and the target (partial dependence functions).

To evaluate feature importance, we use SHAP (SHapley Additive exPlanation; Lundberg and Lee 2017, 2019) values. Table 4.3 shows the relative importance of each driver.

Table 4.3. Relative importance of the geographic factors.

Factor	Relative importance
Temperature	44.56%
Rain	19.00%
Access to navigable waters	13.12%
Unevenness	11.81%
Natural resources	11.51%

NOTE: SHAPs weighted by area.

Temperature is by far the most important geographic factor explaining population density. Access to navigable waters has a relatively low importance, although that might be a singular characteristic of Europe, because most of the continent is made up of small surface peninsulas, so navigable seas and oceans are relatively close everywhere.

Apart from overall feature importance, it is also pertinent to understand whether the relationship between the target and a feature is linear, monotonic or more complex. Partial dependence functions estimate the marginal effect that features have on the predicted outcome of a model (Friedman, 2001). Therefore, they correspond to SHAP values. As we can see in Figure 4.5, partial dependence plots show that geographic endowment decreases as the HDD increases; that is, the colder the region, the less attractive it is. In terms of rainfall, the curve suggests the more the better, but with diminishing returns.

The distance to the sea impacts differently according to temperature. For warm regions (in Figure 4.5, defined as having an HDD of less than 7, with an average of 4.99), such as Mediterranean countries, the effect is more relevant and positive: the closer to the sea, the better. For cold regions (in Figure 6, those regions with an HDD over 8, with an average of 9.61), the effect is almost not relevant and being closer to the sea does not increase geographic endowment. On the one hand, for cold countries with direct access to the sea, such as the United Kingdom or Finland, it is not relevant because almost all NUTS 2 areas have access. Moreover, the sea may not be always easily navigable, as it freezes. For cold countries without direct access to the sea, such as Austria or the Czech Republic, because they are crossed by navigable rivers, geographic dynamics are largely determined by the other factors.

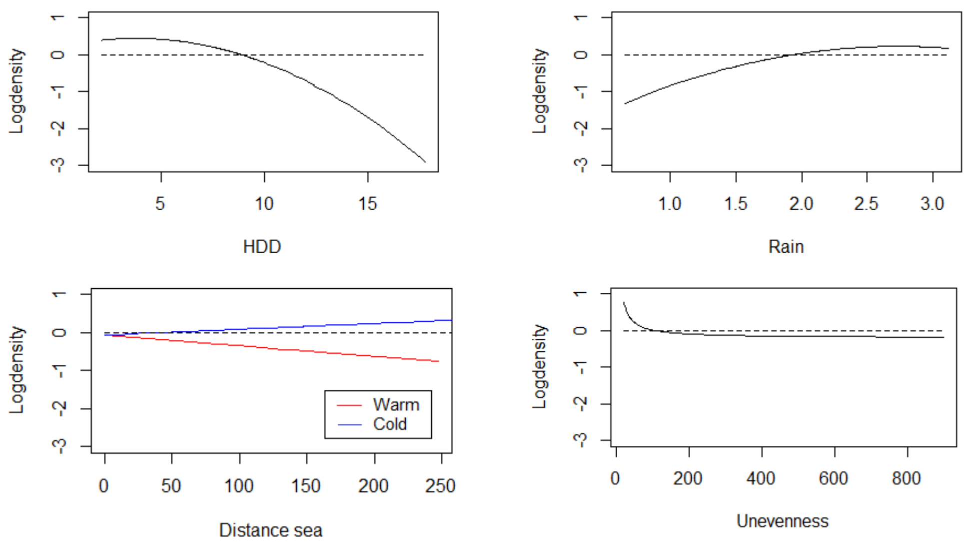

Figure 4.5. Partial dependence plots for non-binary variables.

As predicted by the theoretical framework presented above, geographic endowment tends to produce extreme population density outcomes, as can be seen in the quadratic relationships with respect to temperature and rainfall, or the inverse relationship with respect to unevenness.

Assuming that the greater the number of people who want to live in an area, the higher is its attractiveness, the geographic endowment (*GE)* can be interpreted as a proxy for the attractiveness of a region (to European people) based on its geographic attributes. Obviously, there are several other factors that may influence population density, such as historical events (e.g., wars), cultural and religious differences or public investment. However, the *GE* constitutes an ideal framework for assessing to what extent population is distributed according to geography within a country and identifying the cause of misadjustment between current and expected distribution, and which of those, if any, are a consequence of deviating from geography.

4.5. Analysis of population distribution by country

Our analysis of population distribution by country comprises two parts. First, we analyze whether the choice of capital city is geographically optimal. Second, we estimate the degree of geographic harmonization for each country by the percentage of the population that would have to move to another region within the same country to achieve the expected distribution of population according to the geographic endowment.

4.5.1. Capital cities

To evaluate the choice of the capital location, in terms of geography, we calculated the relative potential of the NUTS 2 region where the capital is located with respect to the maximum potential that could be achieved within the city. The potential of the capital city of country C is defined as:

$$Potential(C) = 100\frac{GE(C)}{\max\limits_{NUTS_i \in C} GE(NUTS_i)}$$

Table 4.4 presents the relative potential of the capitals of those countries considered in the analysis, except for countries that consists of a single NUTS 2 region. It also presents the difference between the geographic factors of the capital and the average of those across the country. Although access to navigable waters is captured in the model by distance, we also included whether the capital city has a river. Rivers play an important role in deciding where to place the city within a region: all but Madrid (Spain) have a navigable river.

As shown by the theoretical model above, capital cities, which are highly populated areas with relatively high output per capita, are generally placed in nearly optimal areas, as their endowment almost reaches the maximum potential of the country (91.8 over 100). Madrid is the most notable exception, scoring only 59.5 over 100. It is placed in the middle of Spain and on the high Spanish plateau, completely isolated from the sea, in a relatively colder and less rainy area.

Table 4.4. Capital attractiveness

Country	Potential	HDD vs avg	Rain vs avg	Sea dist vs avg.	Uneven. vs avg.	River
Germany	100	0.8	0.2	53.2	146.1	Yes
Croatia	100	-0.3	-0.4	-110.8	-131.7	Yes
Portugal	100	0.6	-0.5	37.6	-216.9	Yes
Romania	100	-1.6	0.1	-27.8	-152.1	Yes
Italy	100	-1.1	0.1	-57.3	-286.6	Yes
Ireland	99.4	-1.4	-0.3	0.0	-129.0	Yes
Netherlands	97.7	-1.6	-0.5	-15.7	-86.6	Yes
UK	97.7	-0.3	0.1	-48.2	54.1	Yes
Bulgaria	95.6	0.0	0.2	-20.3	-6.8	Yes
Sweden	95.4	-0.9	0.2	0.0	-154.8	Yes
Slovenia	95.1	-1.1	-0.2	-5.0	-190.7	Yes
Denmark	94.1	-0.1	0.0	0.0	-4.3	Yes
Czechia	92.1	0.0	-0.4	27.7	-219.7	Yes
France	89.4	0.2	-0.1	-22.6	-39.7	Yes
Greece	89.1	0.0	-0.2	0.0	-0.8	Yes
Poland	88.6	-2.4	-1.1	113.2	-496.3	Yes
Austria	88.1	-0.7	-0.1	22.2	-169.8	Yes
Hungary	86.9	-0.1	-0.2	30.7	4.9	Yes
Belgium	86.7	-1.4	0.0	0.0	-18.3	Yes
Finland	85.0	-1.0	-0.2	-87.6	-188.9	Yes
Slovakia	80.2	-0,4	-0,1	-9,7	-32,0	Yes
Spain	59.5	0,4	-0,4	200,5	-171,2	No
Average	91.8	-0.6	-0.2	3.6	-112.3	Yes

4.5.2. Population misadjustment

The GE is an estimation of the logarithm of the density of population. Thus, it can be used to estimate the expected density of population that every region should have if population were distributed according to geographic characteristics. For each country, we estimate the theoretical size of the population that should live in each region, subject to the constraint that total population cannot change. Let us note by *EP(NUTSi)* the expected population of region i, and by $\boldsymbol{A_i}$ its area. Then,

$$EP(i) = 100 \frac{e^{GE(NUTS_i)} A_i}{\sum_{k=1}^{N_C} e^{GE(NUTS_k)} A_k} P_C$$

where country C is the country where the i region is located, $k = \{1, \dots N_C\}$ are the regions that in country C and P_C is the total population of C.

Given the expected population of every region and the real population, we estimate the percentage of total population that would have to move to achieve the expected distribution. We also adjust for the capital bias. Capital bias is the result of differences in how countries classify their capital city. Some countries, such as Belgium, consider the capital city to be a NUTS 2 area. Others, such as Spain or France, consider the capital city with its metropolitan area to comprise the NUTS 2 zone. While others, such as Croatia, include their capital city in a NUTS 2 region that contains not only the city and its metropolitan area, but also a significant part of the country. The larger the territory included in the NUTS 2 sector, the lower the misadjustment due to the capital, since it is diluted among a larger area. To correct for this, we adjust the estimation of population misadjustment by country, delimiting the NUTS 2 area of all capitals according to the boundaries of their metropolitan region, splitting the given NUTS 2 zone into two regions, when needed.

Table 4.5 presents the results of population misadjustment. On average, 24.4 percent of the population would have to move within their country in order to achieve a purely geography-based distribution of population, in a range that extends from 14.4 percent (Bulgaria) to 35.6 percent (Spain). Without the correction for capital bias, results are very similar, with the exception of Croatia and Ireland, whose misadjustment would be underestimated because they include their capital within a NUTS 2 region far larger that the metropolitan area.

Table 4.5. Population misadjustment

Country	% misadjustment adjusted by capital effect	% misadjustment not adjusted by capital effect
Bulgaria	14.4	9.6
Slovenia	15.1	10.5
Czechia	16.0	15.9
Romania	16.3	16.4
Poland	16.9	14.8
Finland	19.2	18.7
Croatia	20.6	7.5
Slovakia	22.9	22.9
Sweden	24.0	24.2
Belgium	24.1	23.1
Germany	24.2	26.3
Italy	24.2	21.3
Austria	24.7	23.4
France	25.3	25.3
Netherlands	26.4	27.0
Hungary	26.4	26.1
Denmark	28.9	28.5
Ireland	31.5	30.0
Portugal	32.4	32.4
UK	34.1	35.3
Greece	34.2	34.2
Spain	35.6	35.6
Average	24.4	23.1

Figure 4.6 shows the comparison between current distribution of population and the expected distribution according to geographic endowment. Figure 4.7 shows the necessary change for regions to transition from current to expected distribution. That is, it shows the percentage increase needed for the resident population of the NUTS 2 area to meet expected population distribution. Green regions are those that are underpopulated with respect to the expected distribution. Therefore, more people should live there so there should be a population increase. Red regions are the opposite. They are overpopulated areas, whose population should decrease so as to achieve the expected

distribution. In Portugal, for example, we can see that there is a highly overpopulated region, Lisbon (the dark red point), which would need to reduce its current population by more than 50 percent, and two moderately overpopulated regions that would need a decrease in current population of between 25 percent and 50 percent: the northern region (Porto) and the southern region (Algarve).

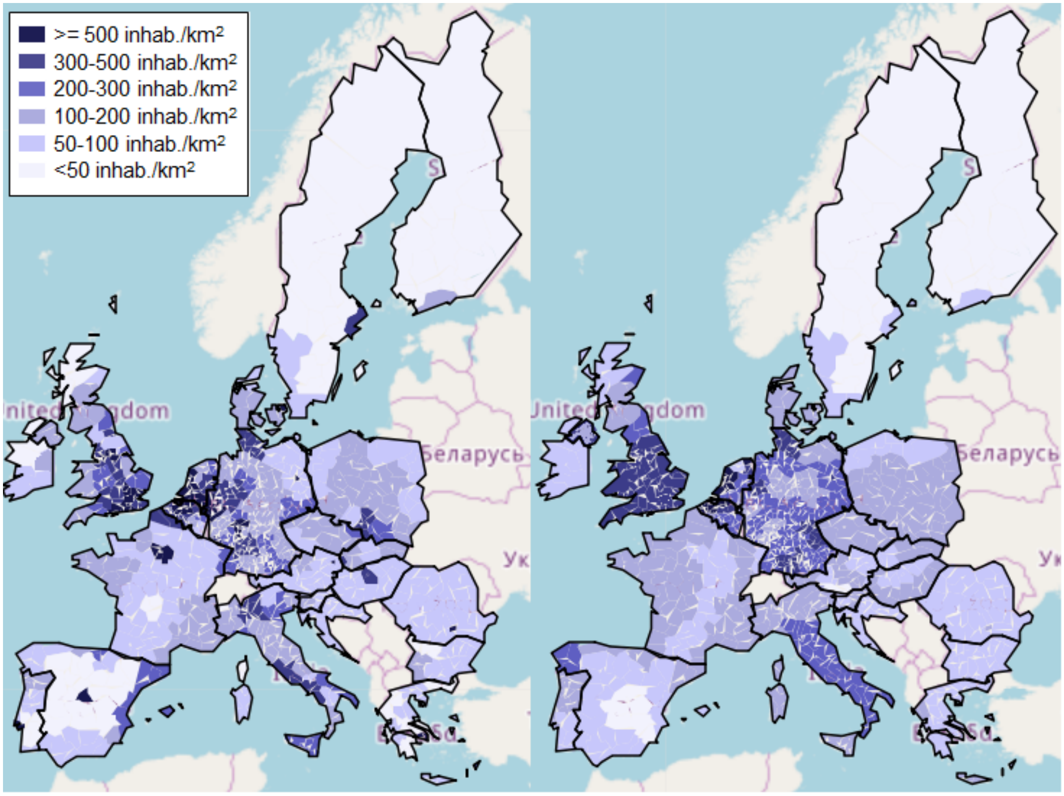

Figure 4.6. Current population distribution in Europe (left) vs. expected according to geographic endowment (right).

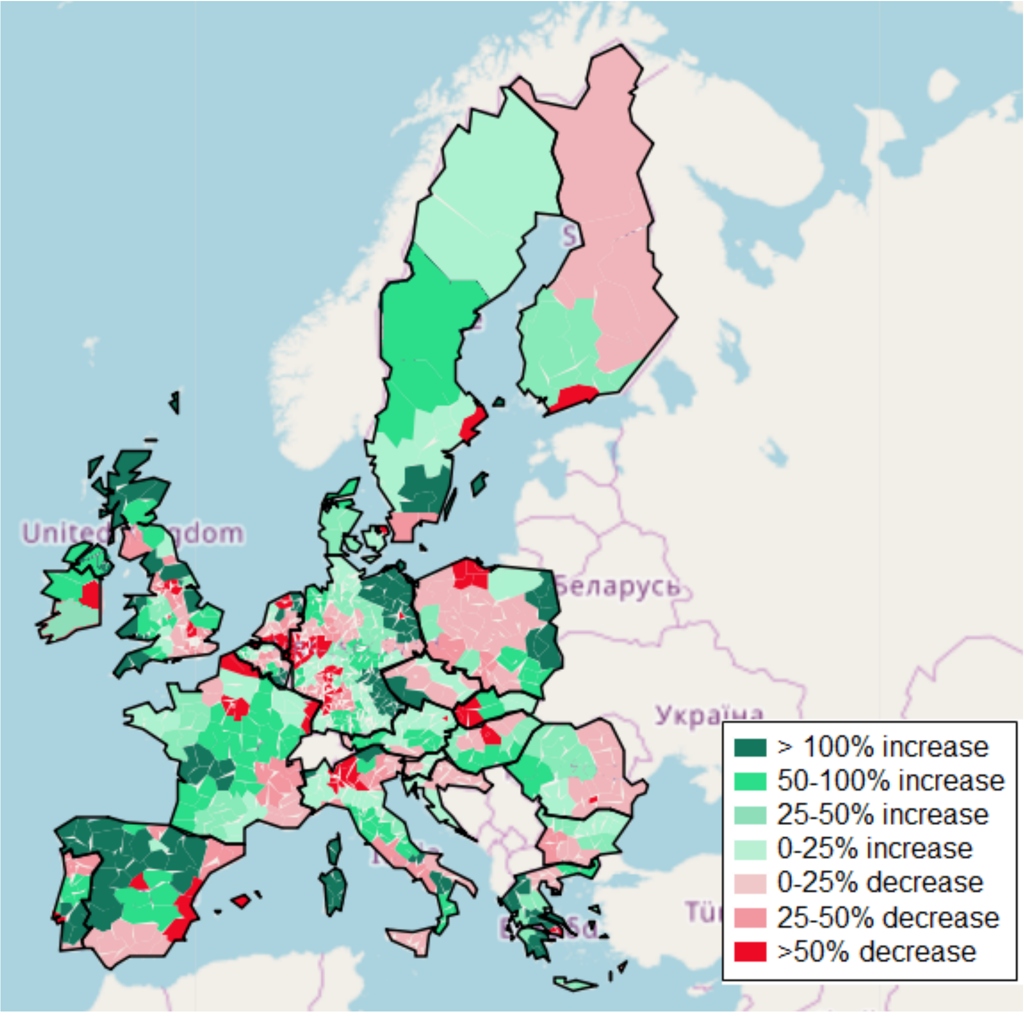

Figure 4.7. Percentage change in population by NUTS2 that would be needed to adjust actual population distribution with expected.

The theoretical model presented above introduced three potential drivers of greater deviations in population distribution in the long run: information asymmetry, moving costs and policy interventions. From results displayed in Table 4.5 and from Figures 4.6 and 4.7, the latter seems to emerge as the most relevant, at least for those countries in which the misadjustment is particularly high (above 30 percent): Spain, Greece, the United Kingdom, Portugal and Ireland.

These five countries share the trait of being peripheral in the context of the European continent. Hence, they are more isolated, which implies that any policy altering population distribution could be less affected by other countries' policies.

In the case of Ireland and the United Kingdom, parallel processes of consolidation took place as global hubs of large firms developed toward the end of the last century. Many multinational firms relocated to the Dublin area in the 1980s and 1990s (Gunnigle and McGuire 2001), following Ireland's entry into the European Union (then the EEC) in 1973, and corporate tax reforms introduced in 1997 and 1999 by the minister of finance, Charlie McCreevy, lowered corporate taxes from 32 percent to 12.5 percent and thus laid the framework for Ireland's base erosion and profit shifting tools (BEPS), considered among the world's largest (Torslov, Wier and Zucman 2020). As a result, the Greater Dublin area has experienced an impressive population growth of 46 percent over the last 30 years (source http://www.greaterdublindrainage.com/) while the rest of the state grew by 30 percent, consolidating a secular trend for the preeminence of Dublin that dates back to the end of the Irish War of Independence (1919-1921) and the Civil War (1922-1923).[5]

In the United Kingdom, the deregulation of financial markets in the 1980s (removal of controls on foreign exchange and fixed rate commissions, entry of foreign companies, switch to electronic trading, etc.), helped to kick off a financial transformation, dubbed the 'Big Bang', that cemented London as the global financial capital. Many financial institutions relocated to London, and its metro-area experienced a population growth like that of Greater Dublin. From 1991 to 2019 it grew by 34 percent, while the rest of the UK population grew by 16 percent.[6] This growth is far larger than that experienced in other large European capital areas such as Paris, Berlin, Rome or Amsterdam. Among the large capital cities, only Madrid has experienced such a large growth of population (33 percent in Madrid vs. 20 percent in the rest of Spain).

[5] After the Civil War, Dublin was consolidated as the political capital. Its population grew from around 500,000 inhabitants in 1925 to 1,025,000 in 1991. Cork, which was the main city opposing the Anglo-Irish Treaty (1921), lost political influence and economic relevance and its population only grew 112.000 inhabitants during the same period.

[6] Data obtained from https://worldpopulationreview.com/ on June 6, 2020.

Spain, Greece and Portugal, besides the common trait of being peripheral countries in the European Union, share another characteristic: the three of them have French civil law systems and legal origin (La Porta *et al.* 2008). This has determined a type of nation-building based on the French model and a very centralized administrative tradition, which drive territorial policies persistently targeted to reinforce the political and economic role of the capital city in the country.[7] The paradigmatic case is that of Spain, which was analyzed in Bel (2011, 2012). Other studies have documented the extreme degree of political centralization in Portugal (Magone 2011) and Greece (Ifantis 2004).[8] One of the probable consequences of these nation-building policies has been a particularly strong promotion of concentration of the population in the capital city of the country, which is the most important factor explaining the extremely intense misadjustment of population in these countries.

4.6. Economic consequences of population misadjustment

According to the theoretical model presented above, deviations from nature in terms of population distribution lead to non-Pareto allocations of population, which may be perpetuated due to the high cost of moving, incomplete information, and overinvestment in overpopulated areas designed to compensate for overpopulation externalities. Hence, no matter the underlying reason (previous migration shocks, public investment, etc.), greater deviations from nature are expected to be associated with lower utility and, potentially, higher inequality (thus contributing to explain the increase of inter-regional inequality in Europe since 1980, emphasized in Cörvers and Mayhew, 2021, and Rosés and Wolf, 2021). To test these two economic consequences of population misadjustment, in this section we present an empirical estimation of the impact of population misadjustment on economic growth.

---

[7] Recall that even though it is not a peripheral country, France has a misadjustment rate of 25.3 percent, slightly above the European average (24.4 percent). The fact that France is more centrally located in the continent may have palliated the effect of this type of nation-building, and derived administrative tradition and territorial policies on population misadjustment.

[8] In Lundell (2004), Spain, Greece and Portugal appear as the countries with the most centralized systems of party candidate selection for elections in the (then) EU15.

4.6.1. Regional conditional convergence

Conditional convergence theory states that an economy grows faster the further it is from its own steady-state value (Barro and Sala-i-Martin 1995), which is conditioned by different covariates such as the saving rate or human capital (Mankiw, Romer, and Weil, 1992). As it allows for different steady states, it is a widely used framework for analysis of the long-term drivers of economic growth (see, Barro 1995; Sala-i-Martin, 1997), and for testing convergence between regions located in different countries (see, Cartone, Postiglione, and Hewings, 2021).

To test whether population misadjustment has an effect on economic convergence, we estimate a model of the form:

$$g_i = \gamma_0 + \gamma_1 \mathrm{q_i} + \gamma_2 s_i + \gamma_3 h_i + \gamma_4 pm_i + \gamma_5 c_i + \varepsilon_i$$

where g_i is the GDP per capita growth rate of the region i over the period 2000-2018, q_i is the natural log of the initial GDP per capita (year 2000), h_i is the human capital, s_i is the natural logarithm of the saving rate, pm_i is the population misadjustment of the country in which the region i is situated (as at the year 2000), c_i is a dummy variable equal to 1 when the capital city is in the region i, and ε_i is the residual. Human capital is measured as the natural logarithm of the percentage of the population aged between 25 and 64 years that has the highest education level (ISCED level 5–8, corresponding to tertiary levels), and the saving and investment rate is calculated as the natural logarithm of the percentage of the gross fixed capital formation a share of GDP. These variables also take values from the year 2000 (See Table A.4.1 in the appendix for details on the description of variables).

We considered a time span of 19 years, from 2000 to 2018, because it is the largest time window at Eurostat with available data of real GDP per capita at NUTS2 level for the 22 countries included in the analysis. Moreover, more institutional homogeneity in economic regulation is ensured due to the Euro already being in place during the period. Indeed, studies of regional convergence are heterogeneous and consider time windows that range from less than 10 years (LeSage and Fischer, 2008) to more than 30 (Mankiw, Romer, and Weil 1992; Ertur and Koch, 2007).

We present three different estimations of the conditional convergence equation. In the first estimation, all regions are weighted equally in the quadratic error minimization. In the second estimation, we weight regions by total population. Therefore, regions with larger populations are considered to be more representative, to contain more information. Finally, we also present an estimation that weights regions according to their relative population within their respective countries. Hence all countries contribute equally to the model.

Table 4.6. Regression estimates for conditional convergence model from 2000 to 2018.

Variable	Equal weights	Weighted by population	Weighted by relative population
Constant	5.17886***	5.61842***	4.46462***
	(0.3598)	(0.2787)	(0.5991)
$\log GDP^{2000}$	-0.42146***	-0.47199***	-0.37592***
	(0.0351)	(0.0237)	(0.0519)
Misadjustment	-1.40105***	-1.24990***	-1.07203***
	(0.1706)	(0.2003)	(0.2656)
LogEducation	0.11069***	0.12838***	0.05603^{+}
	(0.0244)	(0.0265)	(0.0381)
LogSavings	0.05482	0.01127	-0.03508
	(0.0541)	(0.0528)	(0.1102)
CapitalCity	0.16702***	0.14164***	0.09247*
	(0.0506)	(0.0319)	(0.0566)
R^2	0.6313	0.6921	0.5811
F	82.18	107.9	66.6

N = 246. 9 regions (6 from Poland, 3 from UK) excluded because of missingness of data. Robust standard errors in brackets. ***: p< 0.01; **: p< 0.05; *: p< 0.1, +: p<0.15

NOTES: a) Estimations excluding Ireland, that changed their GDP calculations in 2015 which resulted in a >10% increase that year, lead to results with higher predictive power (especially for the first and third estimation, where the R squared increase up to 0.72) and a <1% signification of the capital parameter in the third estimation. The estimated parameters are modified by less than 10% (relatively)

b) Including population growth, depreciation and technological progress as a covariate (as in Cartone, Postiglione and Hewings, 2021) does not modify the results (less than 10%, relatively), since the variable is not significant. However, 8 additional regions would be removed because of missing data.

Table 4.6 presents the results for the three estimations of the conditional convergence model. We only provide OLS results because they are almost identical to Bayesian estimates. Population misadjustment has a negative impact on economic convergence. This result is consistently statistically significant at 1 percent across models. Countries that deviate farther from nature tend to have less growth and lower steady states. Therefore, the greater the deviation the lower the speed of convergence and the inter-regional inequality decrease, both non-weighted and weighted by population (Concept 1 and Concept 2 inequality, as dubbed by Milanovic 2005), and the lower the total welfare of the population. Concretely, for each 10 percentage points of misadjustment, annual growth is reduced by around 0.5 to 0.7 percentage points. Moreover, as expected from the theoretical model, the positive estimate of the capital city parameter confirms that part of this effect consists of a rent transfer to over-populated regions by means of a public intervention.

To gain a complete understanding of the convergence model beyond the significance of the parameters, we estimate the relative importance of each variable included in the model. We use again the methodology suggested by Lundberg and Lee (2017, 2019): SHAP (SHapley Additive ExPlanation) values. Table 4.7 presents the relative importance of each variable in the conditional convergence model. The initial GDP is the most relevant variable and accounts for 50 percent to 60 percent of the total predictive power of the model, depending on the estimation. Remarkably, the population misadjustment and the capital city effect jointly account for around 30 percent of the total importance, being even more relevant than human capital.

Table 7. Relative importance based on SHAP value.

Variable	Equal weights	Weighted by population	Weighted by relative population
$\log GDP^{2000}$	52.5%	56.1%	60.6%
Misadjustment	23.4%	19.9%	23.1%
LogEducation	15.4%	17.0%	10.1%
LogSavings	0	0	0
CapitalCity	8.7%	7.0%	6.2%

Finally, as a robustness check, we re-estimate the model allowing for spillovers and regional interdependences. As discussed in the section Regional economic model, spillover and interdependences preserve the main economic implications of population misadjustments, but may affect its intensity. Following Rey and Montouri (1999), we use a spatial autoregressive (SAR) model, including a spatial lag of growth rates from adjacent regions.[9] Table 4.8 presents the results for the three estimations of the conditional convergence model, spatially augmented. As it can be seen, there is no change in signification of sign of the different coefficients with respect the specifications without spatial effects. In terms of relative importance, notice that conditional convergence is also around two times more important than population misadjustment (see Table 4.9).

Table 4.8. Regression estimates for conditional convergence model from 2000 to 2018 spatially augmented

Variable	Equal Weights	Weighted by population	Weighted by relative population
Constant	2.19304***	3.87390***	3.19753***
	(0.3006)	(0.3344)	(0.3913)
$logGDP^{2000}$	-0.18041***	-0.33077***	-0.27387***
	(0.0250)	(0.0278)	(0.0317)
Misadjustment	-0.67252***	-0.92829***	-0.92638***
	(0.1487)	(0.1810)	(0.2022)
LogEducation	0.06042***	0.09736***	0.03041
	(0.0193)	(0.0236)	(0.0312)
LogSavings	0.00133	-0.04972	-0.07780
	(0.0360)	(0.04675)	(0.0590)
CapitalCity	0.09983***	0.14100***	0.11458***
	(0.0283)	(0.0281)	(0.0287)
Common factor test (LR)	134.59***	50.75 ***	34.94***
R^2	0.8117	0.7567	0.6784
z-value	14.11***	7.66***	5.75***

N = 246. 9 regions (6 from Poland, 3 from UK) excluded because of missingness of data.
Standard errors in brackets. ***: p< 0.01; **: p< 0.05; *: p< 0.1, +: p<0.15
NOTES: There are 5 NUTS2 that correspond to islands, whose adjacencies have been defined as: Balearic Island is adjacent to Catalonia and Valencia; Corse is adjacent to Sardegna; Sardegna is adjacent to Corse and Sicilia; Sicilia is adjacent to Sardegna and Calabria; Åland to Etelä-Suomi, Stockholm, and Östra Mellansverige.

[9] See Elhorst (2014) for a detailed review of spatial econometric models and a critical discussion on when to use these models and how to interpret the results.

Table 4.9. Relative importance of the spatially augmented model based on SHAP value.

Variable	Equal weights	Weighted by population	Weighted by relative population
logGDP^{2000}	26.2%	39.2%	42.2%
Misadjustment	13.1%	14.8%	19.4%
LogEducation	9.7%	13.1%	0
LogSavings	0	0	0
CapitalCity	6.1%	7.0%	7.2%
Spatial effects	44.9%	25.9%	31.2%

4.7. Discussion

Our results provide empirical evidence that large population misadjustments with respect to geographic endowment come at a cost. As expected from our model, the farther a country deviates from the expected population distribution based on its geographic endowment, the lower its regional convergence and the higher its economic inequality will be. Remarkably, part of the effect consists of a rent transfer to the capital city.

These results, together with those obtained in the section "Analysis of the population distribution by country; Population misadjustment", suggest a potential novel causality channel by which institutions affect economic performance.

Recent economic growth literature has emphasized the role of institutions in economic growth and development (see Acemoglu, Johnson, and Robinson 2002; Acemoglu and Robinson, 2005; Rodrik, Subramanian, and Trebbi 2004; Mitton, 2016, among others). 'Institutions' is a broad term that includes the diverse, complex interaction of individuals, firms, states, legislation and social norms which make up a society's social, economic, legal and political organization (see North 1981).[10] According to Acemoglu and Johnson (2005), these institutions are intimately linked to the distribution of political power in society and, as such, regulate the relationship between

[10] The definition of institutions in North (1981: 201-202) is "a set of rules, compliance procedures, and moral and ethical behavioral norms designed to constrain the behavior of individuals in the interest of maximizing the wealth or utility of principals."

ordinary private citizens and elites with access to political power. Rodrik, Subramanian, and Trebbi (2004) propose a taxonomy of four categories of institutions that can impact economic performance. Institutions are 1) market-creating; 2) market regulating; 3) market stabilizing, and 4) market-legitimizing.

Thus, institutions have enough mechanisms to reverse the outcomes of nature and induce or promote the population distribution that best serves specific societal goals. As we showed, by using a framework of incentives, regulations and investments, institutions are the most relevant drivers of a population distribution equilibrium with extreme deviations from nature (as in Acemoglu, Johnson, and Robinson 2002). In turn, population distribution influences economic growth and income distribution: by privileging certain regions with respect to others, institutions not only transfer rents to the privileged region, but also harm overall economic growth.

On the one hand, for the three European continental countries with the most extreme misadjustments (Spain, Portugal and Greece) the deviation is the result of an intentional political intervention by central government, based on the desire to maintain control over the territory by privileging their capital regions (Madrid, Lisbon, and Athens, respectively). Political intervention in the design of policies such as transportation infrastructure, prioritizing objectives related to the administrative and political concentration of power, and largely neglecting productivity-related objectives, has probably prevented the development of an efficient distribution system in the economy, thus damaging potential economic growth. This is shown in the evidence obtained, for example, in the case of Spain, in Albalate, Bel and Fageda (2012) and Bertoméu-Sánchez and Estache (2017). The result of such intense forced deviations from nature is, according to our econometric results, detrimental to economic performance.

On the other hand, the high concentrations of population in London and Dublin, which explain the greater misadjustment of the United Kingdom and Ireland, seem to be a consequence of public policies intended to promote the development of market forces and private industry located in those regions, taking advantage of the role of agglomeration economies with localized accumulated capital. However, this concentration may have come at the expense of other regions. For instance, in the case of the United Kingdom, as stated by Ronen Palan: “The Bank of England consistently pursued policies that favored the City's position as a world financial center, even when such policies were seen as harmful to the UK's mainland manufacturing needs." (Palan 2010: 165). Inner London's GDP per capita was 328% of the European Union average in 2010, compared with 70 percent in West Wales - the biggest gap in any EU state, according to Eurostat.

Rising inequality and economic performance differences between regions have become a relevant policy debate, and a desire to redress the balance is expressed all the way up to the top. In 2014, the UK prime minister David Cameron said that for too long the UK economy had been "too London-focused and too centralised".[11] In 2009 he had already written that *"Over the last century Britain has become one of the most centralized countries in the developed world."*[12]

4.8. Conclusion

The expansion of knowledge and technological innovations in transportation and communication have led to claims of the end of geography; a world in which distance would not play any significant role in decisions about human settlements. In this article, we have analyzed whether the features of nature and geography still play a relevant role in economic and social outcomes, by facilitating or limiting location, concentration, and growth of human settlements.

We have proposed a theoretical model to represent the way in which geography and nature can account for regional economic growth, through their effects on population density and distribution. This model has been empirically examined using data from NUTS 2 European regions. This has allowed us to identify a strong predictive power of first-nature variables to explain the regional distribution of populations, and to estimate the degree of geographic harmonization of the actual distribution of population compared to the predicted distribution.

---

[11] This statement was acknowledged in the news article, "Regions to get £6 billion in government funding", *BBC News/Business,* July 7, 2014. Retrieved on June 6, 2020 (https://www.bbc.com/news/business-28190016).

[12] David Cameron, "A radical power shift" *The Guardian,* February 17, 2009. Retrieved on June 6, 2020 (https://www.theguardian.com/commentisfree/2009/feb/17/cameron-decentralisation-local-government).

After estimating the misadjustments between actual and predicted regional population distribution, we have analyzed their impact on relative economic performance, together with the impact of institution-related factors, such as the conditional of being the capital of the country. Our main results suggest that deviating from nature's outcomes has a significant negative effect on economic growth and regional convergence. Hence, societies that choose to exploit the opportunities of the best locations, according to natural endowment, rather than promoting a different distribution of the population across regions by means of institutional intervention, achieve better economic performance. That is, policies that harmonize with nature and geography yield better social welfare than those policies that conflict with them. To what extent deviating from nature's outcomes has a relevant impact on within-country inequality and on social cohesion is an interesting issue that deserves future research.

4.9. Appendix

Table A.4.1 Data and sources

Variable	Description	Source
Population	Population by NUTS2, in 2000 and 2018	Eurostat
Area	Area of the NUTS2	Eurostat
HDD	Average daily number of degrees below 18ºC from 1977 to 2018	Eurostat
CDD	Average daily number of degrees above 18ºC from 1977 to 2018	Eurostat
Daily rainfall	Average daily rainfall (l/m2) from 1977 to 2018	Copernicus Project
River density	Kilometers of navigable river per km2	EEA
Distance to sea	Nearest distance from the boundary of the region to the sea	Google Maps
Terrain unevenness	Interquartile range of the height (in km) of every LUCAS grid point of each NUTS 2 area	LUCAS-Eurostat
Natural resources	Dummy variable indicating whether there are coal mines or oil refineries within the region	European Comission and Concawe
Regional GDP per capita	Regional GDP per capita (PPP), in 2000 and 2018	Eurostat
Human capital	Natural logarithm of the percentage of the population aged between 25 and 64 years that has the highest education level (ISCED level 5–8), in 2000	Eurostat
Saving rate	Natural logarithm of the percentage of the gross fixed capital formation (as a % of the regional GDP) by NUTS2, in 2000	Eurostat

Chapter 5: Paying for protection: Bilateral trade with a leader and minor guys' defense spending.

This chapter is a joint work with Germà Bel, Daniel Albalate, and Xavier Ros-Oton.

Abstract: Military spending was the main government expenditure until the 20th century, and it continues to represent a significant fraction of most governments' budgets. In this article, we develop a theoretical model to understand the way in which both military and trade alliances with military leaders can impact defense spending. By increasing the costs of military aggression by a non-ally, an alliance reduces the probability of war and allows minor partners to reduce their military spending in exchange for a stronger trade relationship with an alliance leader and a higher trading surplus for the latter. To test the hypotheses derived from our model, we employ data on 138 countries for the period 1996–2020. Our results show that the importance of the trade relationship and the balance of trade with the leader of a military alliance is a significant driver of military spending. Thus, it appears that the greater the weight of trade with the military leader and the higher its trade surplus, the lower is the defense expenditure of the minor partner.

Key words: Military alliances; Trade; Defense spending.

5.1. Introduction

Military spending in Ukraine before the Russian invasion of February 2022 fluctuated between 3.8 (2020) and 3.2% (2021) of its GDP (SIPRI Military Expenditure Database). In relative terms, this rate of expenditure was higher, as a proportion of GDP, than that in any other European country sharing a border with Russia (the case, for example, of Finland, Estonia, Lithuania, and Poland). In fact, with the exception of Russia (and Greece in 2021), no other country in Europe dedicated a higher percentage of its GDP to its defense needs. The limited war waged in the region of the Donbas since 2014 had obliged Ukraine to increase its military spending; however, back in 2008 – before the Great Recession – Ukraine was already dedicating 2.3% of its GDP to defense, with only the United Kingdom and Greece (and Russia) in Europe spending more. The fact that Russia's other European neighbors –

most notably Estonia, Lithuania, Poland and Finland – had much smaller military budgets might reflect the fact that their partners in military and trade alliances afforded them a protective umbrella, with an alliance's main power typically playing a leading role in this respect. For example, in the case of the NATO alliance, U.S. military spending relative to GDP (3.7 and 3.5% in 2020 and 2021, respectively) is well above that of its partners.

The importance attached to forging military alliances is longstanding in the literature. In The Art of War, Sun Tzu instructs the generals of the Kingdom of Wu of the need to help its smaller allies in times of war. Kautilya, in his Arthaśāstra, recommends minor rulers make alliances with kings more powerful than the ruler of their neighboring enemy. In ancient Greece, while opinions diverged on autarchy and openness (Arnopoulos 2001), both Plato and Aristotle wrote of the desirability of entering into defensive military alliances, while Aristotle, a greater proponent of openness, linked contracts for the exchange of commodities with military alliances in Politics (Book III).

Relations based on the provision of protection and safety in exchange for payments or services of a different nature have been a continuum in History. Indeed, the social institutions that regulated these relations constituted the foundations of the feudal system (Brenner 1990). Subsequently, the transition from the 'redistributive world-empire' variant of feudalism to a capitalist world-economy (Wallerstein 1976) would be triggered – among other factors – by the expansion of trade, and empires would be replaced, Adam Smith (1776) argued, by a system of sovereign states and a balance of power organized along trading lines. As such, this relationship between the trade and military alliances forged between sovereign states today lies at the heart of international relations and the international political economy.

Seminal studies by Joanne Gowa (1989, 1994) provide the theoretical basis for the association between military alliances and trade. In testing this theoretical hypothesis, Gowa and Mansfield (1993) found that alliance membership had a positive effect on the bilateral trade of allied countries, while Gowa and Mansfield (1994) reported this influence to be stronger for goods produced under conditions of increasing rather than constant returns to scale. The existence of preferential trading agreements as a booster for trade among allies has been further analyzed by Mansfield and Bronson (1997), who concluded that when military alliances and preferential trading deals are combined the impact on trade is higher than when entered into singularly. In the same vein, a more recent study by Long and Leeds (2006) found that specifying economic agreements in alliance treaties is positively related to trade among partner states.

However, an earlier study by Morrow, Siverson and Tabares (1998) tended to find across several specifications a negative impact of military alliances on trade, this effect being stronger in bilateral than in multilateral alliances. Long (2003) argues that this disparity of results is attributable to the nature of the military alliance: that is, whether the military alliance constitutes a defense (mutual military assistance if one party is attacked) or a non-defense pact (neutrality, non-aggression, or consultation). Thus, defense pacts lead to greater trade among partners than is the case of non-defense pacts. This explanation is consistent with results presented in an empirical study by Survey (1989), who found that U.S. bilateral trade was higher (both imports and exports) with members of NATO.

On the specific costs and benefits of NATO membership, Olson and Zeckhauser's (1966) seminal study examined the divergent economic contribution made by members to the military alliance. The authors analyzed incentives for small state members to contribute proportionately less to the military effort than their larger counterparts and found evidence that the former did indeed enjoy a larger positive security externality (Long 2003). This issue attracted considerable attention during Donald Trump's tenure, with the president constantly demanding that the country's NATO partners increase their military spending, while he reduced the US's contribution to NATO's collective budget. In a recent study, Alley and Fuhrmann (2021) identify the huge budgetary cost to the U.S. of its network of military alliances and wonder whether the financial toll is worthwhile.

However, Poast (2022, 526) argues that "the economic benefits of trade openness and support for the dollar are still notable and, by themselves, likely overwhelm the budgetary costs of maintaining the alliances". Indeed, in terms of the impact on trade, Egel et al. (2016) estimate that a 50% reduction in U.S. security commitments overseas could reduce the country's annual bilateral trade in goods and services by 18 per cent (excluding, that is, trade with Canada and Mexico). As for support for the U.S. dollar, Eichengreen, Mehl, and Chiţu (2019) find that alliances increase the proportion of international units in foreign exchange reserves by about 30 percentage points. The dollar's current dominance as an international unit is supported by the status of the U.S. as a global power, which helps to guarantee the security of its allies. And this status "allows the U.S. government to place dollar-denominated securities at a lower cost because demand from major reserve holders is stronger than otherwise. The cost to the U.S. of financing budget and current account deficits is correspondingly less" (Eichengreen, Mehl, and Chițu 2019, 322).

Overall, it is apparent that major partners in military alliances contribute proportionately more to military spending than do their minor partners, but that the former enjoy benefits from (1) increased trade and (2) from increased demand for their currencies, which together reduce the costs of financing their budget deficit and current account deficits. In return, minor alliance members enjoy greater positive security externalities, which sees them contributing proportionately less to the alliance in terms of military spending.

The aim of this article is to investigate the relationship between the benefits obtained by leading partners in military alliances in terms of trade and current account financing, and the benefits obtained by minor partners in terms of reduced military spending. More specifically, we examine how minor partners benefit a leading partner by increased bilateral trade deals in exchange for a reduction in the risk of war and, hence, of their military spending. To do so, we first develop a theoretical model relating bilateral trade and military spending and formulate our core hypotheses. We then empirically estimate the effect of the relevance of bilateral trade and the bilateral trade surplus with the leading partner on the minor partner's military spending. In conducting this estimate, we draw on a rich empirical literature analyzing the factors that influence military spending.

We find that the greater the minor partner's trade with the leading partner, the lower its military spending tends to be. Likewise, the greater the bilateral trade surplus (the lower the bilateral trade deficit) for the leading partner, the lower is the military spending of the minor partner. In this way, we make a novel and relevant contribution both to the existing literature on the relationship between military alliances and trade and on the determinants of military spending.

5.2. A theory of bilateral trade and military spending

Let us consider two economic and military rival superpower countries, A and B, and a less powerful country both economically and militarily, C. Let us assume that country C, for historical, cultural, economic, and political reasons, is a political ally of A, which means it is under no military threat from A. But, at the same time, is not ally of B. Further, the economic surplus generated by countries i and C because of economic trade shall be denoted by $S_i > 0$.

Following the rationalist view of war (see, among others, Fearon 1995; Powel 1999; Martin et al. 2008; Grossman 2013), we assume that military conflicts may occur as a result of a cost-benefit analysis. As in Martin et al. (2008), we consider that two countries can engage in a negotiation on how to share the trade surplus when at peace. However, during this process, disputes might break out and these might be resolved and end peacefully, or they might escalate into conflict. Even in instances where the negotiation does not fail, wars may occur if either of the two parties believes they could be better off by using military force.

The timing of the negotiation is the following: Country C suggests a deal to country i, which consists of a fraction α_i of S_i that would be captured by i. If α_i is greater than or equal to n_i, which represents the negotiation power of country i , the deal is accepted, and each country secures their share. Otherwise, both countries obtain 0.

Since A and C are allies, we assume that there is no risk of escalation to war. However, when negotiating with B, C acknowledges that there is a risk of military conflict. Let us assume that the expected utility loss for C of a military conflict (and eventual invasion) is W_C. Let us also assume that C estimates the potential gains for B as $\widetilde{W}_B$, where $\widetilde{W}_B$ follows a symmetric triangular distribution with mean W_B and range $[W_B - \theta, W_B + \theta]$, $W_B > S_A + S_B$, and $W_B - W_C < S_A + S_B$. Hence, war is Pareto dominated by peace and the expected gains for B are higher than the total economic surplus from trade. However, these gains may come at a cost. Let us denote by Y the total production of C without trade agreements, and by m the military expenditure of C as a share of Y. It is clear that Y has to be much greater than S_A, S_B, W_B, and W_C. Let us also consider that A, as an ally of C, would provide help to C in case of military conflict. This help might be provided as military aid or as economic and political sanctions on B. For simplicity, we assume that the total help provided by A equals the economic gains it obtains from C. We assume that C knows that a fraction k of the help would be in the form of military aid, and $1-k$ as economic sanctions. The net gains of a military conflict for B are:

$$\widetilde{W}_B - \nu(mY + sk\alpha_A S_A) - (1-k)\alpha_A S_A \tag{5.1}$$

where ν represents the marginal costs for B of the military strength of C so that when $\nu > 1$ military force is more effective than economic sanctions, and $s > 1$ represents the relatively higher efficiency of the military capabilities of A with respect to C.

5.2.1. Probability of war

In line with the rationalist view of war, the leader of B considers military action against C only if she expects a net utility increase with respect to a trade agreement. Therefore, the leader of C estimates the probability of war as the probability of $\tilde{x} = \widetilde{W}_B - \nu(mY + sk\alpha_A S_A) - (1-k)\alpha_A S_A > \alpha_B S_B$. If we then denote by $x = W_B - \nu(mY + sk\alpha_A S_A) - (1-k)\alpha_A S_A$ the expected net gains, as shown in Figure 5.1, the probability is:

$$Prob.\ war \equiv P_W = \begin{cases} \dfrac{(x+\theta-\alpha_B S_B)^2}{2\theta^2} & if\ x \le \alpha_B S_B \\ 1 - \dfrac{(\alpha_B S_B - x - \theta)^2}{2\theta^2} & if\ x > \alpha_B S_B \end{cases} \tag{5.2}$$

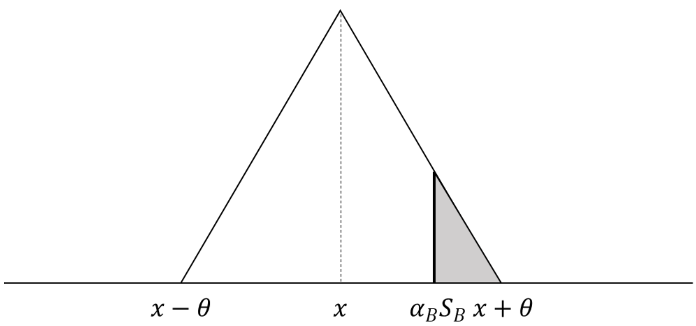

Figure 5.1: Probability distribution of the net gains from war of B, $\tilde{x}$, according to C. The grey area represents the probability of war, that is, the probability that the net gains $\tilde{x}$ are higher than the share of the trade surplus captured by B.

5.2.2. Utility maximization problem

Based on the estimated probability of war, the ruler of C decides which allocations of the trade surplus to offer to A and B, and the military expenditure, to maximize the utility level u of a representative agent of her own country. The maximization problem for the ruler of C is:

$$\max_{\substack{n_A \le \alpha_A \le 1 \\ n_B \le \alpha_B \le 1 \\ 0 \le m \le 1}} Y(1-m) - P_W W_C + (1-P_W)((1-\alpha_A)S_A + (1-\alpha_B)S_B) \quad (5.3)$$

Let us denote by $h = (1-\alpha_A)S_A + (1-\alpha_B)S_B + W_C$. Let us assume that $W_B - S_A(vsk + 1 - k) + \theta - S_B - \frac{\theta^2}{(n_A S_A + n_B S_B + W_C)v} > 0$, and that n_i and W_C are large enough to ensure that the optimum is achieved with $P_W \ll 0.5$. Let us also assume that the uncertainty θ is lower than the maximum benefits of peace for C, which are $\hat{h} = (1-n_A)S_A + (1-n_B)S_B + W_C$.

The partial derivatives of the maximization problem (5.3) are:

$$\frac{\partial u}{\partial m} = Y\left[\frac{v(x+\theta - \alpha_B S_B)h}{\theta^2} - 1\right] \quad (5.4)$$

$$\frac{\partial u}{\partial \alpha_A} = S_A\left[\frac{(vks + 1 - k)(x+\theta - \alpha_B S_B)h}{\theta^2} - 1 + P_W\right] \quad (5.5)$$

$$\frac{\partial u}{\partial \alpha_B} = S_B\left[\frac{(x+\theta - \alpha_B S_B)h}{\theta^2} - 1 + P_W\right] \quad (5.6)$$

Solving for the interior optimum in m, we obtain

$$m = \frac{W_B - \alpha_A S_A(vsk + 1 - k) + \theta - \alpha_B S_B - \theta^2/(hv)}{vY} \quad (5.7)$$

Since we assume that $W_B - S_A(vsk + 1 - k) + \theta - S_B - \frac{\theta^2}{(n_A S_A + n_B S_B + W_C)v} > 0$, and since $Y \gg W_B$, then $0 < m < 1$. Moreover, (5.3) is a quadratic form in m with a negative coeffiecient in the second order term. Therefore, the interior optimum is also global.

By substituting in the expression of the probability of war, we obtain:

$$P_W = \frac{\theta^2}{2(\nu h)^2} \tag{5.8}$$

We also assume that uncertainty is lower than the maximum benefits of peace, that is, $\theta = \frac{\hat{h}}{\mu}$ with $\mu > 1$. Hence, $\frac{\partial u}{\partial \alpha_B} < S_B \left[\frac{1}{\nu} - 1 + \frac{1}{2\mu^2\nu^2}\right]$. Therefore, if $\nu > \frac{\mu + \sqrt{\mu^2+2}}{2\mu}$, (5.6) is negative no matter the values of n_i. For the case of maximum uncertainty, with $\mu \approx 1$, it suffices that $\nu > 1.37$. For more reasonable values, such as $\mu \approx 2$, it suffices that $\nu > 1.11$. As military expenditure is much more efficient than economic sanctions, we can assume that (5.6) will be always negative and therefore the $\alpha_B = n_B$.

Finally, substituting the previous results in (5.5) we get:

$$\frac{\partial u}{\partial \alpha_A} = S_A \left[\frac{\nu ks + 1 - k}{\nu} - 1 + \frac{\theta^2}{2(\nu h)^2}\right] \tag{5.9}$$

As we can see, if military aid or the relatively higher military efficiency of the superpower are sufficiently large, (5.9) is always positive and the optimum is $\alpha_A = 1$. More specifically, the conditions are $k > \frac{\nu-1}{\nu s-1}$ or $s > \frac{\nu-1+k}{\nu k}$. In contrast, if $\frac{\nu ks+1-k}{\nu} < 1$, then (5.9) equals 0 for $\alpha_A = 1 - \frac{1}{S_A}\left[\sqrt{\frac{\theta^2}{2\nu(\nu-\nu ks+1-k)}} - (1-\alpha_B)S_B - W_C\right]$. If $\alpha_A \in (n_A, 1)$, then the optimum is interior. Otherwise, (5.9) is always negative and the optimum is $\alpha_A = n_A$. Table 5.1 presents example solutions of the problem when using different parameters.

Table 1. Results of the bilateral trade optimization problem for different parameters of k, ν, s, and θ, with $Y = 100$, $S_A = 6$, $S_B = 5$, $W_C = W_B = 10$, and $n_A = n_B = 0.5$

Input parameters				Results				
k	ν	s	θ	m	α_A	α_B	P_W	u
0.5	1.3	1.1	4.5	5.7%	0.5	0.5	2.5%	99.5
0.8	1.3	1.1	4.5	2.1%	1.0	0.5	3.8%	100.0

0.5	1.5	1.1	4.5	4.8%	0.5	0.5	1.9%	100.4
0.5	1.3	1.5	4.5	1.5%	1.0	0.5	3.8%	100.6
0.5	1.3	1.1	6.5	6.4%	0.5	0.5	5.2%	94.8

In conclusion, three main implications can be derived from the model. First, as can be seen in expression (5.7), the greater the benefits from trade, the lower the military expenditure will be and the lower the probability of war, which is consistent with reports in the extant literature (see, among others, Copeland 1996; Martin et al. 2008). Second, the higher the efficiency of military force ν, the lower the military expenditure will be. Finally, as discussed in the introduction, the higher the military commitment k or the relatively higher military capabilities s, the lower the military expenditure by C and the higher the benefits from trade for military leader A, since country C renounces direct benefits from trade in order to be more valuable to A and to receive more protection. For instance, C may accept to incur trade deficits with A in exchange for a reduction in the risk of war.

More specifically, the empirical analysis conducted below tests the following hypotheses:

H1: The greater the volume of trade with the military leader, the lower the minor partner's military spending

H2: The greater the military leader's trading surplus, the lower the minor partner's military spending

5.3. Data and Methods

To test our hypotheses empirically, we employ data for the period 1996–2020 for 138 countries, for which information is fully available for both the dependent variable – military spending as a percentage of GDP (%MSpending) – and all covariates. Our dependent variable is obtained from the SIPRI Database. Naturally, we exclude the U.S. from the sample, given that our primary purpose is to evaluate how bilateral trade between the U.S. and other countries affects the latter's military spending. Hence, our dependent variable is the defense spending of these countries as a percentage of their GDP (in constant terms).

Since our main hypothesis is that countries with stronger commercial ties with the U.S. have fewer incentives to spend on the military, due to the protection expected and received from the largest, most powerful army in the world, our key variables are related to existing bilateral trade volumes. First, in order to avoid confounding the effects of bilateral trade on the openness and total international trade of a country, we control for the degree of commercial openness of each country, which is a country's total annual exports and imports as a percentage of its GDP. This variable has typically been included in previous studies of military spending (see Bové and Nisticò 2014; Kollias et al. 2018; George, Hou and Sandler 2019, among others). Theoretically, in line with Rosh (1988), openness facilitates access to arms markets and finance for arms procurement; however, empirical studies usually find that trade has a peace-promoting influence diminishing the likelihood of conflicts and, with it, defense spending efforts (Seitz, Tarasov and Zakharenko 2015; Huang and Trosby 2011; Kollias et al. 2018).

Having controlled for the degree of openness, our first main variable is the bilateral trade conducted between each country and the United States, measured as the trade weight – including both imports and exports – that each country has within the U.S. GDP, that is, how much the trade of each country represents in U.S. GDP (in billions), per annum (*Trade_Weight*). The volume of trade is obtained from the Direction of Trade Statistics of the International Monetary Fund -IMF- (DOT-IMF). According to the theoretical model, our prediction is that the higher this share, the lower the military spending efforts of the U.S.'s commercial partner will be. By way of an alternative, we also consider the U.S. trading surplus with each country as a percentage of U.S. GDP (in billions) (*Trade_Surplus*). Testing the coefficients associated with these variables separately should validate, or otherwise, the main predictions of our model.

Among the rest of the covariates, we include four main groups of determinants of military spending. First, we include those related to the demographic and economic size of the country. Thus, we employ the total annual population in millions (*Pop*) and the square of this population (*Pop^2*) in order to account for any possible non-linear relationships. The values for this variable are drawn from the World Bank's (WB) World Development Indicators. The size of the economy is captured by the country's annual GDP, expressed at 2017 prices in PPP International U.S. dollars (RGDP), and calculated as the product of its GDP per capita and total population. The variable is obtained from the IMF's World Economic Outlook. Its square is also considered as a covariate (*RGDP^2*). Demographic and economic variables are typically found to be statistically significant drivers of military spending (Dunne et al., 2003; Collier and Hoeffler 2007; Nikolaidou 2007; Pamp and Thurner 2017; Hou 2018; Kollias et al. 2018; George et al. 2019; Droff and Malizard 2022).

The second group of covariates constitutes a country's political attributes. A diversity of political and regime-type variables has been employed in past studies (Dunne and Perlo-Freeman 2003; Albalate, Bel and Elias 2012; Yesilyurt and Elhorst 2017; Kollias et al. 2018; Hou 2018). In our model, the first of these variables is the *Political rights* (PR) scores taken from the Freedom in the World Survey (Freedom House database). The variable takes a value between 1 and 7, where 1 represents maximum freedom and 7 the minimum. Freedom and democracy are usually found to be negatively correlated with military spending. The second variable here is *System*, which indicates whether the political system is Presidential (0), Assembly-elected Presidential (1) or Parliamentary (2). This variable is taken from Scartascini, Cruz and Keefer (2021) and is available at the Database of Political Institutions (DPI2020) maintained by the Inter-American Development Bank (IADB)

Third, we account for variables related to current and past military conflicts as drivers of current military spending. These variables are binary and take a value of one if the country is currently engaged in a civil war/violent internal conflict (*Cwar*) or was previously engaged in an international war (*Pwar*). A civil war is considered as a conflict involving a minimum of 25 battle-related deaths per year and per dyad between the government of a State and at least one opposition group without intervention from other States, while a past conflict is considered as having occurred if the country was involved in a war between 1987 and 1992. Information on these two variables is drawn from the PROP Conflict Recurrence Database. A third variable on potential conflicts or threats is included by considering the *Emulation* variable, as defined in Collier and Hoeffler (2007) and Albalate et al. (2012). For each country, we take its neighboring countries (shared borders) and compute the share of aggregate military spending in terms of its own aggregate GDP. This is similar to the Security Web variable employed in earlier studies (Dunne and Perlo-Freeman 2003; Kollias et al. 2018). The expectation is that having armed neighbors is an indication of a potential conflict and, as such, countries adapt their military spending to that of their neighbors (Yesilyurt and Elhorst 2017). Yet, the literature reports mixed findings on its contribution (see, for example, Collier and Hoeffler 2007; Albalate, Bel and Elias 2012; Pamp and Thurner 2017; for non-significant effects).

The last group of covariates is those related to military alliances. Previous research has shown the importance of accounting for international defense agreements in estimating countries' military spending (Digiuseppe and Poast 2018). The expectation is that when a country belongs to a military alliance it tends to have, ceteris paribus, higher military spending (Albalate, Bel and Elias 2012; Droff and Malizard 2022), but the empirical evidence has been mixed in determining whether arms and alliances are substitutes or complementary (Digiuseppe and Poast 2018). There are two principal military alliances in the world for the period considered in this study. To account for the participation of countries in these alliances, we created the variable *NATO* and the variable *CSTA*, to account for participation in either the North Atlantic Treaty Organization or the Collective Security Treaty Alliance. These are binary variables taking a value of 1 if the country belongs to either alliance in a given year and 0 otherwise. The main actors in NATO and the CSTA are the U.S. and Russia, respectively, and, as such we are considering the alliances made by the world's two largest armies in the period examined. Finally, if a country has nuclear weapon capabilities, we also expect them to spend more on defense. Here, a dummy variable with a value of 1 is attributed to a country acknowledged internationally as having nuclear weapons, and 0 otherwise. Table 5.2 displays the descriptive statistics of the variables employed in our empirical exercise and their sources.

Table 5.2. Descriptive statistics and source

Variable	Mean	Std. Dev.	Min	Max	Source
% Mspending	2.08	2.02	0	34.37	SIPRI Database
Trade_weigth	0.11	0.39	0.00	4.06	IMF
Trade_surplus	0.01	0.07	-1.34	0.77	IMF
Openness	0.73	2.03	0.00	71.88	IMF
Pop (millions)	34.27	133.73	0.01	1,411	World Bank

RGDP (109 Int. USD 2017)	403.15	1223	0.032	22,935	IMF
PR	3.45	2.18	1	7	Freedom House (FH-FWS)
System	0.75	0.93	0	2	Inter-American Development Bank (IADB)
Emulation	0.02	0.02	0	0.21	SIRPI & IMF
Cwar	0.11	0.32	0	1	PROP Conflict Recurrence Database
Pwar	0.02	0.13	0	1	PROP Conflict Recurrence Database
NATO	0.12	0.32	0	1	NATO
CSTA	*0.04*	0.19	0	1	CSTO
Nuclear_weapon	*0.03*	0.18	0	1	Federation of American Scientists
Year	2008	7.21	1996	2020	

We adopt different regression methods to evaluate our hypotheses. First, we run a cross-sectional regression with an OLS estimator and robust-to-heteroskedasticity standard errors for the year 2019. We chose 2019 as it is the most recent year for which largely complete data are available prior to the Covid-19 shock of 2020. Note this analysis only considers cross-sectional variation in one year and, as such, it presents only a limited picture of a point in time. As Dunne and Perlo-Freeman (2003) stress, focusing on cross-sectional analyses limits our understanding of dynamic processes at work within countries.

Second, we estimate an unbalanced pooled data model, considering all observations for the period 1996–2020, with time-fixed effects (yearly specific dummies) to exploit both time and cross-sectional dimensions for a long period of time (25 years) in all observations available. This model also employs an OLS estimator to estimate coefficients.

Third, we consider the panel data structure of our data, accounting for unobserved heterogeneity and autocorrelation with panel-corrected standard errors (PCSEs). Panel data are recommended after implementing the Breusch-Pagan test, which rejected the null hypothesis of no panel effects (p-value=0.000). Autocorrelation is found in our time series according to the Wooldridge autocorrelation test for panel data.

Finally, we employ a dynamic panel data model to account for inertia in military spending and to address potential endogeneity and/or reverse causality concerns. The PCSEs and the dynamic panel data models have been typically employed in the literature on military demand and spending (see for example Dunne and Perlo-Freeman 2003; Pamp and Thurner 2017; and Hou 2018 for dynamic models on military spending). In line with previous studies, we specify the following linear model for the cross-sectional regression:

$$\begin{aligned}\%MSpending_{it} &= \alpha + \beta_1\, Trade\_Weight_{it} + \beta_2\, Openness_{it} \\ &+ \beta_3\, Pop_{it} + \beta_4 Pop_{it}^2 + \beta_5\, RGDP_{it} + \beta_6\, RGDP_{it}^2 \\ &+ \beta_7\, PR_{it} + \beta_8\, System_{it} + \beta_9\, Emulation_{it} \\ &+ \beta_{10}\, Cwar_{it} + \beta_{11}\, Pwar_{it} + \beta_{12}\, NATO_{it} \\ &+ \beta_{13}\, CSTA_{it} + \beta_{14}\, Nuclear\_Weapon_{it} + y_t \\ &+ u_{it}\end{aligned} \qquad (5.1)$$

Where i and t refer to the country (i) and year (t) of the observation, respectively, y_t denotes the presence of year-specific fixed effects – which in some models must be substituted by a common time trend- and u_{it} is the error term. Note that we also estimate equation (5.1) by replacing $\beta_1\, Trade\_Weight_{it}$ with $\beta_1\, Trade\_Surplus_{it}$ (5.1.b)

The specific estimating equation for panel data dynamic models is:

$$\begin{aligned}\%MSpending_{it} &= \alpha + \%MSpending_{i,t-1} + \%MSpending_{i,t-2} \\ &+ \%MSpending_{i,t-3} + \beta_1\, Trade\_Weight_{it} \\ &+ \beta_2\, Openness_{it} + \beta_3\, Pop_{it} + \beta_4 Pop_{it}^2 \\ &+ \beta_5\, RGDP_{it} + \beta_6\, RGDP_{it}^2 + \beta_7\, PR_{it} \\ &+ \beta_8\, System_{it} + \beta_9\, Emulation_{it} + \beta_{10}\, Cwar_{it} \\ &+ \beta_{11}\, Pwar_{it} + \beta_{12}\, NATO_{it} + \beta_{13}\, CSTA_{it} \\ &+ \beta_{14}\, Nuclear\_Weapon_{it} + year_t + s_i + u_{it}\end{aligned} \qquad (5.2)$$

where i and t refer to the country (i) and year (t) of the observation, respectively, $year_t$ denotes the presence of a common time trend, s_i is the country fixed effect and u_{it} is the error term. As above, we also estimate equation (5.2) by replacing $\beta_1\, Trade\_Weight_{it}$ with $\beta_1\, Trade\_Surplus_{it}$ (5.2.b)

5.4. Results

The first approach to testing our main hypothesis, estimating a cross-sectional regression (OLS) for the year 2019 (that is, prior to the Covid-19 shock in 2020 and the Russian Invasion of Ukraine in 2022), indicates that the coefficients associated with both the *Trade_Weight* and the *Trade_Surplus* variables are both negative and statistically significant. This means, very much in line with expectations, that the greater the commercial importance of a country in the U.S. economy, the less this country dedicates to defense spending. The same outcomes are obtained when we run a pooled data model, considering time effects with yearly specific binary variables. Table 5.3 shows our main results for both the cross-sectional and pooled data models for the *Trade-Weight* and *Trade_Surplus* variables. In addition to bilateral trade variables, most of the other covariates are found to be relevant drivers of military spending in our pooled data model estimates.

Our demographic and economic variables, used to capture the size of the economy, are statistically significant and present non-linear relationships. Indeed, a U-shaped function is associated with the population variable, while the reverse is the case for real GDP. Our political variables are also relevant according to the estimates conducted, with both political rights and emulation being statistically significant: the former outcome suggests that regimes with weaker political rights spend more on their military; the latter indicates that countries tend to increase spending in line with the spending of their neighbors. Our pooled models also point to the importance of conflicts, with the civil war variable presenting a positive sign in pooled models (3) and (4). In contrast, a past conflict presents a negative sign, suggesting that while there might have been a period of high expenditure associated with that war, spending has subsequently fallen. Interestingly, NATO allies are positively associated with military spending in pooled models, while CSTA allies exhibit the contrary relationship. Finally, countries with a nuclear capability spend more on the military than their counterparts.

However, these models also suggest a counterintuitive relationship between trade openness and military spending. Contrary to the main findings reported in the literature, the coefficient associated with openness is statistically significant in some models; however, it presents a positive sign. This is certainly a source of concern and one that points to potential problems with these simple methods.

Table 5.3. Estimates for the cross-sectional and Pooled data models.

	OLS (2019) (1)	OLS (2019) (2)	Pooled (1996-2020) (3)	Pooled (1996-2020) (4)
Trade_Weigth	-0.5537***	-	-0.5398***	-
	(0.1500)		(0.0419)	
Trade_Surplus	-	-2.74e-07***	-	-2.20e-07***
		(7.74e-10)		(1.93e-10)
Openness	-0.1364	-0.1565	0.2602***	0.2394***
	(0.2742)	(0.2698)	(0.0798)	(0.0789)
Pop	-0.0087**	-0.0087**	-0.0104***	-0.0104***
	(0.0038)	(0.0038)	(0.0009)	(0.0009)
Pop^2	5.68e-06**	5.49e-06**	6.19e-6***	6.36e-06***
	(2.39e-06)	(2.39e-06)	(6.40e-07)	(6.36e-07)
Rgdp	0.3978*	0.3957*	0.5076***	0.48905***
	(0.2252)	(0.2340)	(0.0562)	(0.0571)
Rgdp^2	-0.0191**	-0.0133*	-0.0232***	-0.0186***
	(0.0078)	(0.0069)	(0.0029)	(0.0025)
Pr	0.2709***	0.2704***	0.2792***	0.2868***
	(0.0740)	(0.0728)	(0.0225)	(0.02269)
System	0.0002	0.0002	0.0002	-0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.00028)
Emulation	53.34***	52.79***	40.70***	40.185***
	(10.931)	(10.84)	(3.376)	(3.3178)
Cwar	0.1258	0.1181	0.7675***	0.7693
	(0.4567)	(0.4505)	(0.1175)	(0.1183)
Pwar	-0.8151	-0.8093	-0.3139**	-0.3469**
	(0.6200)	(0.6628)	(0.1700)	(0.1683)
Nato	0.34778	0.3135	0.1919***	0.1571***
	(0.2375)	(0.2354)	(0.0513)	(0.0511)
Csta	-0.6435	-0.6425	-0.9562***	-0.9536***

	(0.7742)	(0.7705)	(0.1230)	(0.1234)
Nuclear_Weapon	1.6244**	1.4766**	1.5937***	1.481***
	(0.6363)	(0.7007)	(0.1035)	(0.1131)
time fe	no	no	yes	yes
N	138	138	3,390	3,390
R2	0.44	0.45	0.34	0.34
F-Test	9.36***	-	29.60***	-

Notes: ***, **, and * denote, respectively, significance at the 1%, 5%, and 10% levels. Robust Standard errors in parentheses.

Although these models seem to corroborate our theoretical predictions, they do not always account correctly for potential features of our data that might bias the estimated coefficients and affect standard errors. First, (1) the pooled data model does not account for unobserved heterogeneity, and a panel data approach is needed, as confirmed by the Breusch-Pagan test. Second, (2) the Wooldridge test for autocorrelation (Wooldridge 2002) confirms that serial correlation is a source of concern in our panel, by rejecting the null hypothesis of no first-order autocorrelation (F= 7.46, p-value= 0.0071). Thus, we estimate a Prais–Winsten (AR1) regression with correlated PCSEs. Third, (3) static panel data models do not account for the dynamic nature of military spending and its inertia and, moreover, they are unable to address any endogeneity concerns that may emerge if there are reasons to believe that it is military spending that shapes bilateral trade with the U.S. and not vice versa. Although we did not find any strong arguments or mechanisms that might result in such endogeneity, we did find clear evidence of inertia.

For all these reasons, our preferred model for testing our prediction is that of dynamic panel data.[13] This model implements the Arellano–Bond estimator (Arellano and Bond 1991), which uses moment conditions where lags of the dependent variable and first differences of the exogenous variables are instruments for the first-differenced equation. The moment conditions of these generalized method of moments (GMM) estimators are valid only if there is no serial correlation in the idiosyncratic errors. Because the first difference of white noise is necessarily autocorrelated, we focus our attention on second and higher autocorrelation. Based on the Arellano–Bond autocorrelation test result, we reject higher-order autocorrelation in our

[13] Lagged dependent variables become statistically insignificant from the second lag on. More lags do not change our main results.

model (Table A.5.2, in the appendix). Table 5.4 reports our main results after running the aforementioned panel data models.

The results obtained from the PCSE and dynamic panel data model confirm our hypotheses and model predictions, to the effect that the greater the commercial relevance of a country with respect to U.S. GDP, the lower its efforts in terms of military expenditure. The same result is obtained for the weight of the U.S. trade surplus in its bilateral trading activity with other countries: here, the higher the trade surplus achieved by the U.S. with a given country, the lower the military spending of that country. Therefore, our empirical estimates point in the direction predicted by our models. This suggests a substitution effect between military spending and trade, with a country able to protect its commercial partners. Undoubtedly, all countries need safety and protection, and our results confirm that this need can be satisfied by means of their own military spending efforts or by their becoming a trade ally of the U.S. The greater the benefits the U.S. reaps from bilateral trade, the lower the effort its commercial partners must incur to feel protected from external threats. Note that we detected, as expected, a positive effect of lagged military burden for a one-year lag, confirming previous reports to the same effect in the literature (Dunne and Perlo-Freeman 2003; Pamp and Thurner 2017; Pamp, Dendorfer and Thurner 2018; Hou, 2018).

Finally, the panel models in Table 5.4 confirm the consistent role across models played by political rights (*PR*), *Emulation* and the common time trend (decreasing trend). Our economic and demographic variables are only statistically significant in the PCSE models, while trade openness appears to be statistically significant but now presenting the expected negative sign only in the dynamic models.

Table 5.4. Estimates for Panel Data models.

	PCSE (AR1) (5)	PCSE (AR1) (6)	PCSE (AR1) (7)	PCSE (AR1) (8)	Dynamic (9)	Dynamic (10)
%MSpending (-1)	-	-	-	-	0.5313***	0.5320***
					(0.1097)	(0.1097)
%MSpending (-2)	-	-	-	-	0.0479	0.04780
					(0.06815)	(0.0680)
%MSpending (-3)	-	-	-	-	-0.0586	-0.0592
					(0.0733)	(0.0733)
Trade_Weight	**-0.0063*** **	**-**	**-0.0062*** **	**-**	**-0.5127** **	**-**
	(0.0007)		**(0.0007)**		**(0.2618)**	
Trade_Surplus	**-**	**-1.82e-9*** **	**-**	**-1.77e-9*** **	**-**	**-1.61e-07* **
		(3.32e-12)		**(3.31e-12)**		**(8.28e-10)**
Openness	0.0004	0.0004	0.0011	0.0010	-0.5439**	-0.5272**
	(0.0016)	(0.0016)	(0.0017)	(0.0017)	(0.2352)	(0.2272)
Pop	0.0016***	-0.0001***	-0.0001***	0.0001***	-0.0030	-0.0041
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0064)	(0.0067)
Pop^2	7.45e-08***	7.59e-08***	7.38e-08***	7.37e-08***	3.27e-07	9.29e-07
	(1.06e-08)	(1.03e-08)	(1.00e-08)	(1.01e-08)	(2.78e-06)	(2.78e-06)
Rgdp	.0036***	0.00306***	0.0043***	0.0037***	0.3527	0.3243
	(.0008)	(0.0008)	(0.0007)	(0.0007)	(0.2752)	(0.2751)
Rgdp^2	-0.0001***	-0.0001***	-0.0002***	-0.0001***	-0.0103	-0.0078
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0074)	(0.0069)
Pr	0.0028***	0.0028***	0.0027***	0.0027***	0.0765**	0.0765**
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0318)	(0.0313)

System	1.89e-06	1.91e-06	2.38e-06	2.40e-06	-0.0003	-0.0003
	(2.39e-06)	(2.36e-06)	(2.39e-06)	(2.36e-06)	(0.00045)	(0.0004)
Emulation	0.2525	0.2496	0.2427***	0.2401***	11.673**	11.602**
	(0.0304)	(0.0303)	(0.0310)	(0.0309)	(5.621)	(5.5795)
Cwar	0.0008	0.0009	0.00078	0.0008	-0.0621	-0.0598
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.1596)	0.1599
Pwar	-0.0013	-0.0001	-0.0006	0.0003	-	-
	(0.0039)	(0.0030)	(0.0034)	(0.0035)		
Nato	-0.0002	-0.00001	-0.0009	-0.0006	0.1251	0.1257
	(0.0008)	(0.0009)	(0.00089)	(0.0009)	(0.2038)	(0.2043)
Csta	-0.0064**	-0.0063**	-0.0073***	-0.0070***	0.3330	0.3007
		(0.0025)	(0.0024)	(0.0020)	(0.7149)	(0.6869)
Nuclear_Weapon	0.0176***	0.0186***	0.0183***	0.0186***	-0.8721	-0.7675
	(0.0023)	(0.0031)	(0.0028)	(0.0032)	(0.9170)	(0.8516)
Time Trend	-0.0003***	-0.0003***	-	-	-0.0149**	-0.0144**
	(0.0000)	(0.0001)			(0.0072)	(0.0072)
Time Fe	no	no	yes	yes	no	no
N	3,390	3,390	3,390	3,390	3,224	3,153
R2	0.44	0.43	0.44	0.44	-	-
F-Test	454.69***	398,19***	-	-	-	-
Wald Chi2	-	-	689.40***	625.64***	703.45***	732.10***

Notes: ***, **, and * denote, respectively, significance at the 1%, 5%, and 10% levels. Robust standard errors in parentheses. Pwar is excluded due to collinearity in dynamic models (9) and (10). Dependent's variable lags become statistically insignificant since the second lag. More lags do not change our main results.

5.5. Discussion / Conclusion

Throughout history, warfare has played a fundamental role in the evolution of human societies. Indeed, until the 20th century and the expansion of the welfare state, military spending was the main item in the government budget. In Ancient Rome, military spending represented more than two-thirds of total government expenditure (McLaughlin 2014), while during the Middle and Modern Ages, it continued to represent more than 50 per cent of government spending in most western countries (Kennedy 1987). For instance, in France, war expenditure accounted for some 57 per cent of total spending in 1683 and for around 52 per cent in 1714 (Bonney 1999). In some regions of the Dutch state, it is reported to have climbed above 90 per cent during the Franco-Dutch war (Hart 1999). In the course of history, military alliances, of many different kinds, have been of vital importance, often founded on the provision of protection or guarantees of safety in return for payments and services. As the balance of power in recent centuries has become increasingly organized around trade relations, the relationship between the trade and military alliances forged by sovereign states has become a central issue in both international relations and scholarly analyses.

Seminal studies by Gowa (1989, 1994) and Gowa and Mansfield (1993, 1994) have laid the groundwork for the analysis of the effect of military alliances on trade, and studies by other scholars have followed in their wake (e.g., Mansfield and Bronson 1997; Morrow, Siverson and Tabares 1998; Long 2003; Long and Leeds 2006). Generally speaking, alliances are found to have a positive influence on trade between partners, especially when certain characteristics (e.g., parallel trade agreements, defense-pact, etc.) are fulfilled. Since alliance leaders tend to incur greater military spending, the question has been raised as to whether these additional costs (Alley and Fuhrmann 2021) are offset by their benefits (Egel et al. 2016, Eichengreen, Mehl and Chiţu 2019; Poast 2022). However, much less attention has been paid to what – if anything – minor partners in these alliances might obtain, beyond protection from third-party aggression.

In this paper, we built a theoretical model to understand how defense and trade alliances affect military spending. By uniting forces, countries reduce the risk of war and the need to invest in the military in two ways: First, by increasing the opportunity cost of war, they make themselves more valuable to each other; and, second, by intensifying their trade relations with military leaders, they ensure their protection. The more valuable non-leaders can make themselves to the leader, the greater the protection the leader will be prepared to provide in case of military aggression against the non-leader, and the higher the cost for a non-allied country to initiate military conflict with the non-leader.

Our results show that trade relations with an alliance leader are a highly significant driver of military spending for minor partners. For each percentage point of U.S. GDP attributable to trade between a certain country and the U.S., the military spending of that country falls by 0.5 percentage points. Moreover, when the trade balance is especially favorable to the U.S., this effect is even greater.

While most of the extant literature on military alliances and trade has focused on the effects of such pacts on commerce and examined their benefits and costs to the alliance leader, it has been assumed that almost the sole benefit for minor partners is the increased protection they obtain. This study has taken the analysis of the outcomes of military alliances for minor partners further. As such, our main findings are that increasing the weight of trade with the alliance leader and, in particular, bilateral trade surplus in favor of alliance leader – which can be interpreted as a payment for protection – brings minor partners an additional benefit insofar as they are able to reduce their direct defense spending. Thus, our study provides a more comprehensive analysis of the benefits and costs of trade partnerships for military protection.

One of the most significant impacts that the Russian invasion of Ukraine seems likely to have in the near future is an increase in global military spending, but especially that of European NATO members as they seek to reduce their almost absolute reliance on U.S. military power when faced with a military threat. Our analysis suggests that these developments will result in a decrease in the weight of the bilateral trade of minor partners with the U.S. and an improvement in their bilateral trade balance with this major power. These developments constitute, we believe, a very interesting avenue for future research.

5.6. Appendix:

Table A.5.1. List of countries that could not be included in the empirical analysis

Andorra	Estonia	Maldives	Seychelles
Antigua y Barbuda	Grenada	Micronesia	Sierra Leone
Aruba	Haiti	Monaco	South Sudan
Bahamas	Holy See	Nauru	The Comoros
Barbados	Hong Kong	North Korea	Marshall Islands
Bhutan	Kiribati	Palestine	Tonga
Channel Islands	Kosovo	Papua New Guinea	Turk Cyprus
Djibouti	Latvia	Saint Kitts and Nevis	Tuvalu
Dominica	Lesotho	Saint Vincent	Yemen, North
Dominican Republic	Liechtenstein	Santa Lucia	Yemen South
Equatorial Guinea	Macao	Serbia	

Table A.5.2. Arellano-Bond test for zero autocorrelation in first-difference errors.

Order	z	Prob>z
1	-2.7254	0.0064
2	0.7998	0.4238
3	-0.8594	0.3901

Chapter 6: Conclusions and Policy Implications

This dissertation presents the results of five studies regarding the role of institutions in public policy implementation and economic outcomes. All five studies contribute to the growing body of public policy evaluation literature by providing theoretical frameworks and empirical evidence to better understand the government's decision-making and its impact on economic outcomes, and methodological novelties to evaluate public policies.

The first study introduces a methodological improvement to evaluate causal effects under quasi-experimental designs. By decoupling covariate relevance and counterfactual weights in the synthetic control method, we show that the procedure offers higher accuracy and stability while ensuring the economic meaningfulness of covariates used in building the counterfactual. We evaluate the economic effect of the government formation deadlock in Spain-2016 with the new methodology and find that economic growth was not harmed by the deadlock. This result has a clear policy implication. While there are strong reasons to favor government formation, starting with the fact that it is in the essence of democracy that political parties form governments, pressure to facilitate government formation in order to prevent economic harm, especially if it comes from the winning party or economical lobbies that could benefit from government formation, could be hiding self-interest. Hence, it could be better for the public interest that minor political parties or opposition parties do not facilitate government formation until their requests (and those of the citizens they represent) are met.

In Chapter 3 we developed a new theoretical model to understand government decision-making in a highly complex context with incomplete information, as it was the COVID-19 outbreak. The model suggests that in such a context governments not only rely on technical criteria related to the crisis (in this case, healthcare capacity), but also balance economic and political costs, and could be influenced by cognitive and emotional biases due to the incompleteness of information. We empirically test the hypothesis derived from the model and find evidence of the effects of expected economic and political costs in delaying a response that governments would have anticipated had they had complete information. We also find evidence that federal governments reacted in a more agile way, but, in the case of the US, at the expense of lower coordination. Republican-lead states reacted later due to the pressure for President Trump to avoid state lockdowns. Two important policy implications emerge from this analysis. On the one hand, the COVID-19 pandemic generated frequent demands to increase health expenditure. While such expenditure may improve health-system performance on a regular day-to-day basis, as long as the additional capacity meets positive social cost-benefit requirements, it will not provide full insurance for managing future pandemics, as strong healthcare-system capacity can induce governments to make riskier decisions, particularly under incomplete information. On the other hand, multi-level governance in federal countries fostered institutional competition and allowed to implement and compare different policies to deal with the COVID-19 outbreak. In such a context, where the risk of adopting non-optimal policies was extremely high due to the unprecedentedness of the crisis, implementing multiple policies constituted an overall advantage. Decision-makers could learn from actions taken by competitors and could be forced to modify their decisions due to competitive pressure. However, the concurrency of different policies also leads to performance inequality, as shown in the case of the US.

The study presented in Chapter 4 showed that geographic and natural factors have a large influence on population density in Europe and that deviating from nature's outcomes has a significant negative effect on economic growth and regional convergence. The result has a clear policy implication. Societies that choose to exploit the opportunities of the best locations, according to the natural endowment, rather than promoting a different distribution of the population across regions by means of institutional intervention, achieve better economic performance.

In Chapter 5, we presented a theoretical model to understand the effect of military expenditure on trade and defense alliances with military leaders. The model suggests non-leader countries have a clear incentive to build a trade alliance as beneficial as possible for the military leader in the exchange for protection, and, as a consequence, reduce their military expenditure. We empirically tested the model and found for each percentage point in US GDP in trade between a certain country and the US (military leader), the military spending of the country reduces by 0.5 percentage points. This result helps us understand why the US has sustained (and still does= a trade deficit for almost 50 years, and why it could be beneficial for its allies to keep it. Indeed, in terms of effects on trade, Egel et. al. (2016) estimates that a 50% reduction in U.S. security commitments abroad could reduce U.S. bilateral trade in goods and services annually by 18 percent (excluding trade with Canada and Mexico). Being a military leader and providing protection to their allies, the US could be able to get more advantageous trade alliances that let them reduce the cost of financing budget and current account deficits. In return, allies can reduce their military expenditure.

The studies presented in this dissertation also touched on some practical challenges and limitations of impact evaluations. The most relevant of them is the fact that they are quasi-experimental designs. We do not have two Spains, one with government formation deadlock and the other with no deadlock, to be sure that the impact is properly estimated. There are no multiple Europes with different population distributions and no other crisis as COVID-19. Hence, although we provided theoretical models (with a very simple hypothesis) to understand empirical results, and we use different alternatives among the most advanced empirical techniques in these contexts, our findings need to be taken as pieces of evidence that build on top of a growing literature on public policy analysis and the role of institutions, rather than conclusive or timeless knowledge. Future research will face the same difficulty but will take advantage of the studies presented in here.

Bibliography

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.

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 America Statistical Association, 105 (490): 493-505.

Abadie, A., Diamond, A., Hainmueller, J. Synth: (2011) An R Package for Synthetic Control Methods in Comparative Case Studies. Journal of Statistical Software, 42(13):1-17.

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 L'Hour, J. (2019). A penalized synthetic control estimator for disaggregated data, Working paper, Massachusetts Institute of Technology, Mimeo. Available on-line at http://economics.mit.edu/files/18642

Abadie A (2020) Using synthetic controls: feasibility, data requirements, and methodological aspects. JEcon Lit 59:391–425

Acemoglu, D., Johnson, S., and Robinson, J. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review 91(5), 1369–1401.

------, Johnson, S., and Robinson, J. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. Quarterly Journal of Economics 117(4), 1231-1294

------, and Johnson, S. (2005). Unbundling institutions. Journal of Political Economy 113(5), 949-995.

------, Johnson, S., Kermani, A., Kwak, J. and Mitton, T. (2016) The value of connections in turbulent times: Evidence from the United States. Journal of the American Statistical Association, 107(498), 833-843

Aisen, A., and Veiga, FJ. (2013). How does political instability affect economic growth? European Journal of Political Economy 29, 151-167

Akerloff, George, and Robert J. Shiller. 2009 Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism. Princeton (NJ): Princeton University Press.

Akin,O, García-Montalvo, J. García-Villar, Peydró, J.L, and Raya, J.M. (2014). The real estate and credit bubble: evidence from Spain. SERIEs: Journal of the Spanish Economic Association, 5: 223-243.

Albalate, D., Bel, G., and Elias, F. (2012). Institutional Determinants of military spending. Journal of Comparative Economics 40(2): 279-290.

Albalate, D., Bel, G, and Fageda, X. (2012). Beyond the efficiency-equity dilemma: Centralization as a determinant of government investment in infrastructure. Papers in Regional Science 91(3), 599-615.

Albalate D, Bel G, and Mazaira-Font F (2020) Ensuring Stability, accuracy and meaningfulness in synthetic control methods: the regularized SHAP-distance method. IREA-B, Working Paper 2020-05, posted 26 April 2020

Albalate, D. and Bel, G. (2020). Do government formation deadlocks really damage economic growth? Evidence from history's longest period of government formation impasse. Governance. 33 (1): 155-171.

Alesina, A., Ozler, S., Roubini, N. and Swagel, P. (1999). Political instability and economic growth. Journal of Economic Growth 1(2), 189-211.

Alley, J. and Fuhrmann, M. (2021). Budget Breaker? The Financial Cost of US Military Alliances. Security Studies 30(5): 661-690.

Álvarez-Mon, M. et al. (2021). Predictive Model and Risk Factors for Case Fatality of COVID-19. Journal of Personalized Medicine. 11(1): 36.

Anderson, Roy M., Heesterbeek, H., Klinkenberg,D., and Hollingsworth, T. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet 395(10228): 391-394.

Angelopoulos, K. and Economides, G. (2008). Fiscal policy, rent seeking, and growth under electoral uncertainty: theory and evidence from the OECD. Canadian Journal of Economics 41(4), 1375-1405

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies, 58(2): 277-297.

Aristotle (IV c. B.C.). Politics (Book 3).

Arnopoulos, P. (2001). Between Isolationism and Interventionism: The Exopolitics of Plato and Aristotle: Sovereignty vs Prosperity. Skepsis: A Journal for Philosophy and Interdisciplinary Research 12.

Arora, S., and Barak, B. (2007) Computational Complexity: A Modern Approach. Cambridge University Press, NP, Cambridge

Athey, S., and Imbens,G. W. (2017) . The State of Applied Econometrics: Causality and Policy Evaluation. Journal of Economic Perspectives, 31 (2): 3-32.

Baden, L., and Hana, M., et al. (2021). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. New England Journal of Medicine 384: 403-416.

Baekkeskov, E., and Rubin, O.(2014). Why Pandemic Response is Unique: Powerful Experts and Hands-Off Political Leaders. Disaster Prevention and Management 23(1): 81–93.

Baekkeskov, E. (2016). Same threat, different responses: experts steering politicians and stakeholders in 2009 H1N1 vaccination policy-making. Public Administration 94(2): 299-315.

Ballesteros, L., and Kunreuther, H. (2018). Organizational decision making under uncertainty shocks. NBER WP 24924. National Bureau of Economic Research.

Barnes, T.D., and Beaulieu, E. (2019). Women Politicians, Institutions, and Perceptions of Corruption. Comparative Political Studies 52(1): 134–167.

Barro, R. J. (1998). Determinants of Economic Growth: A Cross-Country Empirical Study. Cambridge, MA: MIT Press.

------, and Sala-I-Martin, X. (1995). Economic Growth Theory. Boston, MA: Mc-Graw Hill.

Bechtel, M., and Hainmueller, J. (2011). How lasting is voter gratitude? An analysis of the short- and long-term electoral returns to beneficial policy. American Journal of Political Science 55(4): 851-867.

Beckers Hospital Review (2021). https://www.beckershospitalreview.com/rankings-and-ratings/states-ranked-by-hospital-beds-per-1-000-population-2.html (retrieved 28 October 2021).

Beeson, P. E., Dejong, D. N., and Troesken, W. (2001). Population Growth in US Counties, 1840-1990. Regional Science & Urban Economics 31, 669-699.

Bel, G. (2011). Infrastructure and nation building: The regulation and financing of network transportation infrastructures in Spain (1720-2010). Business History 53(5), 688-705.

------ (2012). Infrastructure and the Political Economy of Nation Building in Spain, 1720–2010. Eastbourne: Sussex Academic Press.

------ (2020). Gestió de crisis i la síndrome Churchill. ARA, 25 March 2020. (https://www.ara.cat/opinio/germa-bel-gestio-crisis-sindrome-churchill_0_2423157816.html)

------., Gasulla, O., and Mazaira-Font, F. (2021). The effect of health and economic costs on governments' policy responses to COVID-19 crisis under incomplete information. Public Administration Review 81(6): 1131-1146.

------, and Esteve, M. (2022). Resilient Managed Competition during Pandemics: Lessons from the Italian experience. Health Economics, Policy and Law 17(2): 220-223

Benton, J.E. (2020). Challenges to Federalism and intergovernmental relations and takeaways amid the COVID-19 experience. American Review of Public Administration 50(6-7): 536–542.

Bertoméu-Sánchez, S. and Estache, A. (2017). Unbundling political and economic rationality: A non-parametric approach tested on transport infrastructure in Spain. Transport Policy 59, 181-195.

Besley, T., and Coate, S. (1997). An Economic Model of Representative Democracy. The Quarterly Journal of Economics 112(1): 85–114.

Bhattacharyya, S. (2009). Unbundled Institutions, Human Capital and Growth. Journal of Comparative Economics 37(1), 106–210.

Billmeier, A., and Nannicini, T. (2013). Assessing economic liberalization episodes: a synthetic control approach. The Review of Economics and Statistics, 95 (3): 983–1001.

Birkland, T.A., Taylor, K., Crow, D.A., and DeLeo, R. (2021). Governing in a Polarized Era: Federalism and the Response of U.S. State and Federal Governments to the COVID-19 Pandemic. Publius 51(4): 650–672.

Black, D., and Henderson, V. (1998). Urban evolution in the USA. Working Paper No. 98-21, Brown University.

Bohn, S., Lofstrom, M., and Raphael, S. (2014). Did the 2007 legal Arizona workers act reduce the State's unauthorized immigrant population? The Review of Economics and Statistics, 96(2):258-269.

Boin, A., Hart, P., Stern, E., and Sundelius, B. (2005). The Politics of Crisis Management: Public Leadership under Pressure. Cambridge: Cambridge U. Press.

Boin, A. (2019). The Transboundary Crisis: Why we are unprepared and the road ahead. Journal of Contingencies and Crisis Management 27(1): 94-99.

Bonney, R., eds. (1999). "France, 1494-1815", in The Rise of the Fiscal State in Europe c. 1200-1815, ed. Bonney, Richard, 123–176. Oxford: Oxford University Press

Botero, G. (1588). The magnificence and greatness of cities. Translated by R. Peterson, 1606. Reprinted 1979. Amsterdam: Theatrum Orbis Terrarum; Norwood, NJ: Walter J. Johnson.

Botosaru, I., and Bruno, F. (2019) On the role of covariates in the synthetic control method. Econ J 22(2):117–130

Bouckaert, G., and Brans, M. (2012). Governing without Government: Lessons from Belgium's Caretaker Government. Governance 25(2), 173-176.

Bove, V. and Nisticò, R. (2014). Coups d'état and defense spending: a counterfactual analysis. Public Choice 161(3/4): 321–344.

Bowling C.J., J.M. Fisk, and J.C. Morris (2020). Seeking Patterns in Chaos: Transactional Federalism in the Trump Administration's Response to the COVID-19 Pandemic. The American Review of Public Administration 50(6-7): 512-518.

Brenner, R. (1990). "Feudalism." In Marxian Economics (The New Palgrave), eds. Eatwell, John, Milgate, Murray, and Newman, Peter, 170-185. London: Palgrave Macmillan.

Bueno de Mesquita, B., Alastair, S., Siverson, R., and D. Morrow, J. (2003). The Logic of Political Survival. Cambridge (MA): MIT Press.

Bulman-Pozen, J. (2016). Executive Federalism Comes to America. Virginia Law Review 102(4): 953-2030.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software 80(1), 1-28.

Cairncross, F. (2001). The Death of Distance 2.0. London: Texere Publishing Limited.

Cameron, D. (2021) The relative performance of federal and non-federal countries during the pandemic In R. Chattopadhyay et. al. (eds), Federalism and the Response to COVID-19. A Comparative Analysis (pp. 262-276). London: Routledge.

Card, D. (2009). Immigration and Inequality. American Economic Review, 99 (2), 1-21.

Cartone, A., Postiglione, P., and Hewings, G. F. D. (2021). Does economic convergence hold? A spatial quantile analysis on European regions. Economic Modelling 95, 408-417.

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013) Catastrophic natural disasters and Economic Growth. The Review of Economics and Statistics, 95(5): 1549-1561.

Cénat, J. M., Darly, R., K. Kokou-Kpolou, C., Mukunzi, J., and Rousseau, C. (2020). Social inequalities and collateral damages of the COVID-19 pandemic: when basic needs challenge mental health care. International Journal of Public Health, 65: 717–718.

Chattopadhyay, R., and Knüpling, F. (2021). "Comparative summary". In R. Chattopadhyay et. al. (eds), Federalism and the Response to COVID-19. A Comparative Analysis (pp. 277-307). London: Routledge.

Chattopadhyay, R., Knüpling, F., Chebenova, D., Whittington, L., and Gonzalez, P. eds. (2021). Federalism and the Response to COVID-19. A Comparative Analysis. London: Routledge.

Chirinko, R., and Harper Jr, E. (1993). Buckle up or slow down? New estimates of offsetting behavior and their implications for automobile safety regulation. Journal of Policy Analysis and Management 12(2): 270-296.

Christensen, T., Lægreid, P., and Rykkja, L. (2016). Organizing for Crisis Management: Building Governance Capacity and Legitimacy. Public Administration Review 76(6): 887-897.

Cigler, B.A. (2021). Fighting COVID-19 in the United States with Federalism and Other Constitutional and Statutory Authority. Publius 51(4): 673–692.

Clemens, M. A. (2011). Economics and Emigration: Trillion-Dollar Bills on the Sidewalk? Journal of Economic Perspectives 25 (3), 83-106.

Collier, P., and Hoeffler, A. (2007). Unintended Consequences: Does Aid Promote Arms Races?. Oxford Bulletin of Economics and Statistics 69(1): 1-27.

Comfort, L. K. (2007). Crisis management in hindsight: Cognition, communication, coordination, and control. Public Administration Review 67(s1): 189–197.

------, Waugh, W., and Cigler. B. (2012). Emergency management research and practice in public administration: Emergency, evolution, expansion, and future directions. Public Administration Review 72(4): 539-547.

Congleton, R.D. (2021). Federalism and pandemic policies: variety as the spice of life. Public Choice forthcoming. https://doi.org/10.1007/s11127-021-00915-9

Connor, P. (2020). More than nine-in-ten people worldwide live in countries with travel restrictions amid COVID-19. Washington DC: Pew Research Center.

Copeland, D. C. (1996). Economic Interdependence and War: A Theory of Trade Expectations. International Security 20(4): 5-41.

Cortes, P. (2008). The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data. Journal of Political Economy, 116(3), 381-422.

Cörvers, F., and Mayhew, K. (2021). Regional inequalities: causes and cures, Oxford Review of Economic Policy 37(1), 1–16.

Cutler, D.M, and Summers, L. (2020). The COVID-19 Pandemic and the $16 Trillion Virus. JAMA 324(15):1495–1496.

Dal Bó, E., Finan, F., Folke, O., Persson, T. and Rickne. J. (2020). Social Outsiders but Political Insiders: Sweden's Radical Right. Working papers. University of California, Berkeley.

Deslatte, A., Hatch, M., and Stokan, E. (2020). How Can Local Governments Address Pandemic Inequities? Public Administration Review 80(5): 827-831.

Desson, Z., Lambertz, L., Peters, J., Falkenbach, M. and Kauer, L. (2020). Europe's Covid-19 outliers: German, Austrian and Swiss policy responses during the early stages of the 2020 pandemic. Health Policy and Technology 9(4): 405-418.

Diamond, J. (1997). Guns, Germs, and Steel. New York: Norton.

Dietz, T., Ostrom, E., and Stern, P., (2003). The Struggle to Govern the Commons. Science, 302 (5652): 1907-1912.

Digiuseppe, M., and Poast, P. (2018). Arms versus Democratic Allies. British Journal of Political Science 48(4): 981-1003.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. NBER Working Papers, 22791.

Dougherty, S., C. Vammalle, P. De Biase and K. Forman (2020). COVID-19 and fiscal relations across levels of government. OECD Tackling Coronavirus. 2-22 April. http://oe.cd/il/COVID-FF.

Downey, D. C., and Myers. W. (2020). Federalism, Intergovernmental Relationships, and Emergency Response: A Comparison of Australia and the United States. American Review of Public Administration 50(6/7): 526-535

Droff, J. and Malizard, J. (2022). Determinants of Defense Spending: The Role of Strategic Factors in France. Defence and Peace Economics 33(8): 938-955,

Dror, Y. (1994). Statecraft as Prudent Risk Taking: The Case of the Middle East Peace Process. Journal of Contingencies and Crisis Management 2(3):126-135.

Dunne J. P., and Perlo-Freeman, S. (2003). The demand for military expenditure in developing countries: a dynamic panel analysis. Defence and Peace Economics 14(6): 461–474.

Duranton, G., and Puga, D. (2020). The economics of urban density. Journal of Economic Perspectives 34(3), 3-26.

Easterly, W.R. (2001). The elusive quest for growth: economists' adventures and misadventures in the topics. Cambridge MA: MIT Press.

------, and Levine, R. (2003). Tropics, Germs, and Crops: How Endowments Influence Economic Development. Journal of Monetary Economics 50(1), 3–39.

Egel, D., Grissom, A.R., Godges, A.R., Kavanagh, J., and Shatz, H. J. (2016). Estimating the Value of Overseas Security Commitments, RAND Corporation, RR-518-AF, 2016. https://www.rand.org/pubs/research_reports/RR518.html (retrieved 31 January 2023)

Eichengreen, B., Mehl, A., and Chiţu, L. (2019). Mars or Mercury? The geopolitics of international currency choice. Economic Policy 34(98): 315–363.

Eleazar, D. J. (1993). International and Comparative Federalism. Political Science and Politics 26(2): 190-195.

Ellison, G., and Glaeser, E. L. (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? American Economic Review 89(2), 311–316.

Ertur, C. and Koch, W. (2007). Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence. Journal of Applied Econometrics, 22, 1033-62

Faggian, A., Rajbhandari, I., and Dotzel, K. R. (2017). The interregional migration of human capital and its regional consequences: a review. Regional Studies 51(1), 128–143.

Farazmand, A. (2007). Learning from the Katrina Crisis: A Global and International Perspective with Implications for Future Crisis Management. Public Administration Review 67(s1), 149-159.

Fearon, James D. (1995). Rationalist Explanations for War. International Organization. 49(3): 379–414.

Ferguson, N., et al., (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, Imperial College COVID-19 Response Team, Report 9

Ferman B, Pinto C, Possebom V (2020) Cherry picking with synthetic controls. J Policy Anal Manag., 39(2):510–532

Ferro, L. (2020). Il virus in Lombardia. Storia di un disastro. Medium, 22 March 2020. https://medium.com/@luigiferro010/il-virus-in-lombardia-storia-di-un-disastro-552bfa51673b

Fingleton, B., and López-Bazo, E. (2006). Empirical Growth Models with Spatial Effects. Papers in Regional Science, 85, 177-98

Fingleton, B. (2003). European Regional Growth. Berlin: Springer

Fleming, F. J. (1991). Profits and visions: British capital and railway construction in Argentina, 1854–1886. In Railway Imperialism, ed. C. B. Davis, K. E. Wilburn, & R. E. Robinson, 71–84. New York, NY: Greenwood Press

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics 29(5), 1189-1232.

------ (2005). The World is Flat. New York, NY: Farrar, Straus and Giroux.

Furceri, D., Loungani, P., and Ostry. J.D. (2020). How Pandemics Leave the Poor Even Farther Behind. IMFBlog, May 11th, 2020 https://blogs.imf.org/2020/05/11/how-pandemics-leave-the-poor-even-farther-behind/

Gallup, J. L., and Sachs., J. D. (2001). The Economic Burden of Malaria. The American Journal of Tropical Medicine and Hygiene 64(1–2) (Suppl.), 85–96.

------, Sachs, J. D., and Mellinger, A. D. (1999). Geography and economic development. International Regional Science Review 22(2), 179–232.

George, J., Hou, D., and Sandler, T. (2019). Asia-Pacific Demand for Military Expenditure: Spatial Panel and SUR Estimates. Defence and Peace Economics, 30(4): 381–401.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. (2004). Do institutions cause growth? Journal of Economic Growth 9(3), 271-303.

------, and Shapiro, J. (2003). Urban Growth in the 1990s: Is city living back? Journal of Regional Science 43(1), 139-165.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. Review of Economics and Statistics, 98(3), 535–551.

Goelzhauser, G. & D.M. Konisky (2020). The State of American Federalism 2019–2020: Polarized and Punitive Intergovernmental Relations. Publius, 50(3): 311–343.

Gong, Y., and Janssen, M. (2012). From policy implementation to business process management: Principles for creating flexibility and agility. Government Information Quarterly 29(S1): S61-S71

Gowa, Joanne. 1989. Bipolarity, Multipolarity, and Free Trade. American Political Science Review 83(4): 1245-1256.

------, and Mansfield, E. D. (1993). Power Politics and International Trade. American Political Science Review 87(2): 408-420.

------,. (1994). Allies, Adversaries, and International Trade. Princeton (NJ): Princeton University Press.

------, and Mansfield, E. D. (2004). Alliances, Imperfect Markets, and Major-Power Trade.” International Organization 58(4): 775-805.

Graham, S. (1998). The End of Geography or the Explosion of Place? Conceptualizing Space, Place and Information Technology. Progress in Human Geography 22, 165–185.

Grasselli, G., Pesenti, A., and Cecconi, M. (2020). Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. JAMA. Journal of the American Medical Association 323(16):1545–1546.

Grossman, H. I. (2013). Choosing Between Peace and War, Annals of Economics and Finance 14(2), 765-783

Gunnigle, P., and McGuire, D. (2001). Why Ireland? A Qualitative Review of the Factors Influencing the Location of US Multinationals in Ireland with Particular Reference to Labour Issues. The Economic and Social Review 32(1), 43-67

Gupta, S., Nguyen, T.D., Lozano Rojas, F., Raman, S., Lee, B., Bento, A., Simon, K. and Wing, C. (2020). Tracking public and private responses to the covid-19 epidemic: evidence from state and local government actions. NBER WP 27027. National Bureau of Economic Research.

Hale, T. et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nature Human Behaviour 5, 529–538.

Hallas L., Thttps://www.bsg.ox.ac.uk/people/thomas-hale Hale, A. Hatibie, S. Majumdar, M. Pyarali, R. Koch and A. Wood (2020). Variation in US states' responses to COVID-19 BSG-WP-2020/034. Oxford University

Hansen, P. Jaumard, B. and Savard, G. (1992). New branch-and-bound rules for linear bilevel programming. SIAM Journal on Scientific and Statistical Computing, 13(5), 1194-1217.

Hart, M. (1999). "The United Provinces, 1579-1806", in The Rise of the Fiscal State in Europe c. 1200-1815, ed. Bonney, Richard, 309-325. Oxford: Oxford University Press.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning: Data mining, inference, and Prediction. Second Edition. Springer-Verlag. 763 pages. New York.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations CRC

Hegele, Y., and Schnabel, J. (2021). Federalism and the management of the COVID-19 crisis: centralisation, decentralisation and (non-) coordination. West European Politics 44(5-6): 1052-1076.

Hilton, J., and Keeling, M. (2020). Estimation of country-level basic reproductive ratios for novel Coronavirus (SARS-CoV-2/COVID-19) using synthetic contact matrices. PLoS Computational Biology 16(7): e1008031

Hooghe, M. (2012). Does Multi-Level Governance Reduce the Need for National Government? European Political Science 11(1), 90-95

Hou, D. (2018). The Determinants of Military Expenditure in Asia and Oceania, 1992–2016: A Dynamic Panel Analysis. Peace Economics, Peace Science, and Public Policy 24(3): 2018004.

Howlett, M., Capano, G., and Ramesh, M. (2018). Designing for robustness: surprise, agility and improvisation in policy design. Policy and Society 37(4): 405-421.

Irving, H YF. (2020). Fighting COVID-19 through government Initiatives and collaborative governance: The Taiwan experience. Public Administration Review 80(4): 665-670.

Sterling, H. and Throsby, D. (2011). Economic, political, and social determinants of peace. The Economics of Peace Security Journal 6(2): 5–14.

Ifantis, K. (2004). State Interests, External Dependency Trajectories and the European Union: Greece, in European Union Enlargement: A Comparative History, ed. J. Elvert & W. Kaiser, 70-92. London: Routledge.

Jacobs, N. (2021) Federalism, Polarization, and Policy Responsibility during COVID-19: Experimental and Observational Evidence from the United States. Publius 51(4): 693–719.

Janssen, M., and van der Voort, H. (2020). Agile and adaptive governance in crisis response: Lessons from the COVID-19 pandemic. International Journal of Information Management, 55: 102180.

Kahneman, D. (2011). Thinking, Fast and Slow. New York (NY): Farrar, Straus and Giroux.

Kapiriri, L., and Ross, A. (2020). The Politics of Disease Epidemics: a Comparative Analysis of the SARS, Zika, and Ebola Outbreaks. Global Social Welfare 7(1): 33-45.

Kaplan, R. D. (2012). The Revenge of Geography. What the Map Tells Us about Coming Conflicts and the Battle against Fate. New York, NY: Random House.

Gabriel G., Mrad, A., Bonetti, S., Manoli, G. and Parolari, A. (2020) Global convergence of COVID19 basic reproduction number and estimation from early-time SIR dynamics. PLoS ONE 15(9): e0239800

Klößner, K. Pfeifer, G., and Schieler, M. (2015) Synthetic control methods: never use all pre-intervention outcomes together with covariates. MPRA Paper 83790, University Library of Munich, Germany

------, Ashok K., and Pfeifer, G. (2018) Comparative politics and the synthetic control method revisited: a note on Abadie. Swiss J Econ Stat 154(1):651–667

Kautilya (III c. B.C.). Arthaśāstra (Book VII. Chapter II)

Kennedy, J., Sayers, A. and Alcantara, C. (2022). Does Federalism Prevent Democratic Accountability? Assigning Responsibility for Rates of COVID-19 Testing. Political Studies Review 20(1):158-165

Kennedy, P. (1987). The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000. New York (NY): Penguin Random House.

Kettl, D.F. (2020). States Divided: The Implications of American Federalism for COVID-19. Public Administration Review 80(4): 595-602.

Kincaid, J., and Leckrone, J.W. (2021). COVID-19 and American federalism: First-wave responses. In R. Chattopadhyay et al. (eds), Federalism and the Response to COVID-19. A Comparative Analysis (pp. 239-249). London: Routledge.

Kleinfeld, R. (2020). Do Authoritarian or Democratic Countries Handle Pandemics Better? Carnegie Endowment for International Peace. March 31, 2020 (https://carnegieendowment.org/2020/03/31/do-authoritarian-or-democratic-countries-handle-pandemics-better-pub-81404 downloaded July 31, 2020)

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the European football market. American Economic Review, 103(5), 1892–1924.

Knight, S. R. et. al. (2020). Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ 370: m3339

Kollias, C., Paleologou, S. M., Tzeremes, P., and Tzeremes, N.(2018). The demand for military spending in Latin American countries. Latin American Economic Review 27(1): 11.

Koo, J.R., Cook, A., Park, M., Sun, Y., Sun, H., Lim, J., Tam, C., and Dickens, B. (2020). Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet, Infectious Diseases 20(6): 678-688

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

Krugman, P. (1991). Increasing returns and economic geography. Journal of Political Economy 99(3), 483-499.

------ (1993). First nature, second nature, and metropolitan location. Journal of Regional Science 33(2), 129-144.

------ and Venables A. (1990). Integration and the competitiveness of peripheral industry. In Unity with diversity in the European economy: The Community's Southern Frontier, ed. C. Bliss & J. Braga De Macedo, 56-75. Cambridge: Cambridge University Press

La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. (2008). The economic consequences of legal origins. Journal of Economic Literature 46(2), 285-332.

Lai, A. Y-H. (2018). Agility amid uncertainties: evidence from 2009 A/H1N1 pandemics in Singapore and Taiwan. Policy and Society 37(4): 459-472.

LeSage, J. P., and Fischer, M. M. (2008). Spatial growth regressions: model specification, estimation and interpretation. Spatial Economic Analysis, 3(3), 275-304.

Levy Yeyati, E., Moscovich L., and Abuin, C. (2020). Leader over Policy? The Scope of Elite Influence on Policy Preferences. Political Communication 37(3): 398-422.

Lindsey, J.K. (2000). Applying generalized linear models. Springer.

Long, A. G. (2003). Defense Pacts and International Trade. Journal of Peace Research 40(5), 537–552.

------, and Leeds, B. A. (2006). Trading for Security: Military Alliances and Economic Agreements. Journal of Peace Research, 43(4): 433–451.

Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks (CA): Sage Publications.

López-Bazo, E., Vayá, E., and Artís, M. (2004). Regional Externalities and Growth: Evidence from European Regions. Journal of Regional Science, 44, 43-73

López-Santana, M., and Rocco, P. (2021). Fiscal Federalism and Economic Crises in the United States: Lessons from the COVID-19 Pandemic and Great Recession. Publius 51(3): 365–395

Lundberg, S. M., and Lee, S. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. NIPS proceedings, 30, 1-10.

Lundberg, S. M., and Lee, S. (2019). Consistent Individualized Feature Attribution for Tree Ensembles. Cornell University Working papers. Available at https://arxiv.org/abs/1802.03888v3.

Lundell, K. (2004). Determinants of Candidate Selection: The Degree of Centralization in Comparative Perspective. Party Politics 10(1), 25–47.

Machiavelli, N. (1519). Discourses on Livy. New York, NY: Oxford University Press, 1987.

Magone, J. M. (2011). The difficult transformation of state and public administration in Portugal. Europeanization and the persistence of neo-patrimonialism. Public Administration 89(3), 756-782.

Malhotra, N., Margalit, Y., and Mo, C. (2013). Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact. American Journal of Political Science, 57(2), 391-410.

Mallinson, D.J. (2020). Cooperation and Conflict in State and Local Innovation During COVID-19. American Review of Public Administration, 50(6–7): 543–550.

Malthus, T. R. (1798). An essay on the principle of population. London: J. Johnson.

Mankiw, N. G., Romer, D., Weil D. N. (1992). A Contribution to the Empirics of Economic Growth. Quarterly Journal of Economics 107(2), 407-437.

Mansfield, E. D., and Bronson, R. (1997). Alliances, Preferential Trading Arrangements, and International Trade. American Political Science Review, 91(1): 94-107.

Martin, P., Mayer, T., and Thoenig, M. (2008). Make Trade Not War? The Review of Economic Studies 75(3): 865–900.

Maxeiner, J. R. (2019). Restarting a Government of Laws in the United States of America. In A. Parise ed., 2019. Re- De- Co-dification? New Insights on the Codification of Private Law. Intersentia, Ius Commune Europaeum University, pp. 187-219. (available at https://ssrn.com/abstract=3261996)

McLaughlin, R. (2014). The Roman Empire and the Indian Ocean: The Ancient World Economy & the Kingdoms of Africa, Arabia & India. Barnsley: Pen & Sword Books.

Mecenas P., Bastos, R., Vallinoto, A., and Normando, D. (2020). Effects of temperature and humidity on the spread of COVID-19: A systematic review. PLoS ONE 15(9): e0238339.

Menifield, C., E., and Clark, C. (2020). Pandemic Planning in the United States: An Examination of COVID-19 Data. Public Administration Review forthcoming https://doi.org/10.1111/puar.13326

Milanovic, B. (2005). Worlds Apart: Global and International Inequality 1950-2000. Princeton, NJ: Princeton University Press.

Mitchener, K. J., McLean, I. W. (2003). The productivity of the US States since 1888. Journal of Economic Growth 8(1), 73-114.

Mitton, T. (2016). The wealth of subnations: Geography, institutions, and within-country development. Journal of Development Economics 118, 88-111.

Mizrahi, S., Vigoda-Gadot, S. and Cohen, N. (2021). How Well Do They Manage a Crisis? The Government's Effectiveness during the Covid-19 Pandemic. Public Administration Review forthcoming https://doi.org/10.1111/puar.13370

Montalvo, J.G. (2011). Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections. Review of Economics and Statistics, 93(4), 1146–1154.

Montesquieu, C. de S. (1748). The Spirit of the Laws. Paris.

Montserrat, J. et al. (2021). Impact of the Innate Inflammatory Response on ICU Admission and Death in Hospitalized Patients with COVID-19. Biomedicines. 11(9)

Moon, M. Jae. (2020). Fighting Against COVID-19 with Agile, Adaptive, and Transparent Government: Wicked Policy Problems and New Governance Challenges. Public Administration Review 80(4): 651-656

Morrow, James D., Siverson, Randolph M., and Tabares, Tressa E. (1998). The Political Determinants of International Trade: The Major Powers, 1907–1990. American Political Science Review 92(3): 649-661.

Moynihan, D. (2008). Learning under Uncertainty: Networks in Crisis Management. Public Administration Review 68(2): 350-365.

National Governors Association. https://www.nga.org/governors/ (retrieved 2 November 2021)

Neelon, B., F. Mutiso, N.T. Mueller, J.L. and Benjamin-Neelon, S-E. (2021). Associations between governor political affiliation and COVID-19 cases, deaths, and testing in the U.S. American Journal of Preventive Medicine 61(1): 115-119.

Nikolaidou, E. (2008). The demand for military expenditure: Evidence from the EU15 (1961–2005). Defence and Peace Economics 19(4): 273–292.

Nohrstedt D., and Weible, C. (2010). The Logic of Policy Change after Crisis: Proximity and Subsystem Interaction. Risk, Hazards & Crisis in Public Policy 1(2): 1–32.

Nolette, P., and Provost, C. (2018). Change and Continuity in the Role of State Attorneys General in the Obama and Trump Administrations. Publius 48(3): 469–494.

North, D. C. (1981). Structure and change in economic history. New York: W.W. Norton & Co

O'Brien, R. (1992). Global Financial Integration: The End of Geography. London: Pinter.

O'Sullivan, A. (2007). Urban Economics. McGraw-Hill/Irwin

Olson, M., and Zeckhauser, R. (1966). An Economic Theory of Alliances. The Review of Economics and Statistics 48(3): 266–279

Osborne, M.J., and Slivinski, A. (1996). A Model of Political Competition with Citizen-Candidates. The Quarterly Journal of Economics 111(1): 65–96.

Ottaviano, G., and Peri, G. (2012). Rethinking the effect of immigration on wages. Journal of the European Economic Association, 10(1), 152-197.

Palan, R. (2010). International Financial Centers: The British-Empire, City-States and Commercially Oriented Politics. Theoretical Inquiries in Law 11(1), 149-176.

Palivos, T. (2009). Welfare effects of illegal immigration, Journal of Population Economics 22(1), 131-144.

Pamp, O., and Thurner, P. W. (2017). Trading Arms and the Demand for Military Expenditures: Empirical Explorations Using New SIPRI-Data. Defence and Peace Economics 28(4), 457–472.

------, Dendorfer, F., and Thurner, P. W. (2018). Arm your friends and save on defense? The impact of arms exports on military expenditures. Public Choice 177: 165-187.

Partridge, M. D., Rickman, D. S., Ali, K., and Olfert, M. R. (2008). Employment Growth in the American Urban Hierarchy: Long Live Distance. B.E. Journal of Macroeconomics 8, 10

Peltzman, S. (1975). The Effects of Automobile Safety Regulation. Journal of Political Economy, 83(4): 677–726.

Percoco, M. (2015) Heterogeneity in the reaction of traffic flows to road pricing: a synthetic control approach applied to Milan. Transportation, 42, 1063-1079.

Peri, G. (2012). The Effect of Immigration on Productivity: Evidence from U.S. States, The Review of Economics and Statistics 94(1), 348-358.

Peterson, S., Hoffer, G., and Millner, E. (1995). Are Drivers of Air-Bag-Equipped Cars More Aggressive? A Test of the Offsetting Behavior Hypothesis. Journal of Law and Economics 38(2): 251-264.

Poast, P. (2002). No Free Lunches: What the United States Gains from Its Alliances. (In Cooley, Alexander, Nexon, Daniel H., Poast, Paul, Alley, Joshua, and Fuhrmann, Matthew, "Estimating Alliance Costs: An Exchange"). Security Studies 31(3): 510-532.

Pondorfer. A., Barsbai, T., and Schmidt, U. (2017). Gender Differences in Stereotypes of Risk Preferences: Experimental Evidence from a Matrilineal and a Patrilineal Society. Management Science 63(10): 3268-3284.

Postiglione, P., Andreano, M. S., and Benedetti, R. (2013). Using constrained optimization for the identification of convergence clubs. Computational Economics 42(2), 151-174.

Powell, R. (1999). In the Shadow of Power: States and Strategies in International Politics. Princeton (NJ): Princeton University Press.

Rajbhandari, I., Faggian, A., and Partridge, M. (2020). Migrants and boomtowns: Micro evidence from the US shale boom. GSSI Discussion Papers in Regional Science & Economic Geography. DP no. 2020-11. L'Aquila: Gran Sasso Science Institute

Rey, S. J., and Montouri, B. D. (1999). US regional income convergence: a spatial econometric perspective. Regional studies, 33(2), 143-156.

Rietveld, P., Vickerman, R. W. (2004). Transport in Regional Science: The "Death of Distance" is Premature. Papers in Regional Science 83, 229–248.

Ritchie, H. et al. (2021). Coronavirus Pandemic (COVID-19). Published online at OurWorldInData.org. 'https://ourworldindata.org/coronavirus (Retrieved 30 October 2021).

Robinson, Scott. E., Ripberger, J., Gupta, K., Ross, J., Fox, A., Jenkins-Smith, H., and Silva, C. (2020). The Relevance and Operations of Political Trust in the COVID-19 Pandemic. Public Administration Review forthcoming https://doi.org/10.1111/puar.13333

Rocco, P., Béland, D., and Waddan, A. (2020). Stuck in neutral? Federalism, policy instruments, and counter-cyclical responses to COVID-19 in the United States, Policy and Society, 39 (3), 458-477.

Rodríguez-Pose, A., Tselios, V., Winkler, D., and Farole, T. (2013). Geography and the Determinants of Firm Exports in Indonesia. World Development, 44, 225–240.

Rodrik, D., Subramanian, A., and Trebbi, F. (2004). Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. Journal of Economic Growth 9(2), 131–165.

Rosenthal, U., and Kouzmin, A. (1997). Crises and Crisis Management: Toward Comprehensive Government Decision Making. Journal of Public Administration Research and Theory 7(2): 277-304.

Rosés, J. R., and Wolf, N. (2021). Regional growth and inequality in the long-run: Europe, 1900–2015, Oxford Review of Economic Policy 37(1), 17–48

Rosh, Robert M. (1988). Third world militarization: security webs and the states they ensnare. Journal of Conflict Resolution 32(4): 671–698.

Rozell, M.J. and Wilcox, C. (2020). Federalism in a time of plague: how federal systems cope with pandemic. American Review of Public Administration 50(6-7): 519–525

Sachs, J., and Warner, A. M. (2001). Natural resources and economic development: the curse of natural resources. European Economic Review 45, 827–838.

Sala-I-Martin, X. (1997). I just run two million regressions. American Economic Review 87(2), 178-183

Scartascini, C., Cruz, C., Keefer, P. (2021). The Database of Political Institutions 2020 (DPI2020). Inter-American Development Bank https://publications.iadb.org/en/database-political-institutions-2020-dpi2020 (retrieved 15 November 2022)

Scavo, C., Kearney, R., and Kilroy, R.J. Jr. (2008). Challenges to Federalism: Homeland Security and Disaster Response. Publius 38(1): 81-110.

Scharff, D.P., K.J. Mathews, P. Jackson, J. Hoffsuemmer, E. Martin, D. Edwards (2010). More than Tuskegee: Understanding Mistrust about Research Participation. Journal of Health Care for the Poor and Underserved 21(3): 879-897.

Scher, B. (2020). Coronavirus vs. Governors: Ranking the Best and Worst State Leaders. Politico, 1 April 2020 (https://www.politico.com/news/magazine/2020/04/01/coronavirus-state-governors-best-worst-covid-19-159945 , retrieved 14 March 2022).

Schwartz, J., and Yen M-H. (2017). Toward a collaborative model of pandemic preparedness and response: Taiwan's changing approach to pandemics. Journal of Microbiology, Immunology and Infection 50(2): 125-132.

Schwartz, J. (2012). Compensating for the ‘Authoritarian Advantage’ in Crisis Response: A Comparative Case Study of SARS Pandemic Responses in China and Taiwan. Journal of Chinese Political Science 17:313–331.

Scott, V., Crawford-Browne, S., and Sanders, D. (2016). Critiquing the response to the Ebola epidemic through a Primary Health Care Approach. BMC Public Health 16: 410.

Seitz, M., Tarasov, A., and Zakharenko, R. (2015). Trade costs, conflicts, and defense spending. Journal of International Economics 95(2): 305–318.

Senese, P. D. (2005). Territory, Contiguity, and International Conflict: Assessing a New Joint Explanation. American Journal of Political Science 49(4), 769-779.

Shvetsova O., Zhirnov, A., Giannelli, F.R., Catalano, M.A., and Catalano, O. (2022). Governor's Party, Policies, and COVID-19 Outcomes: Further Evidence of an Effect. American Journal of Preventive Medicine 62(3): 433-437.

Starr, H. (2005). Territory, Proximity, and Spatiality: The Geography of International Conflict. International Studies Review 7(3), 387-406.

Stock, J. H. (2020). Data gaps and the policy response to the novel coronavirus. NBER Working Paper 26902. National Bureau of Economic Research.

Summary, R. M. (1989). A Political-Economic Model of U.S. Bilateral Trade. The Review of Economics and Statistics 71(1): 179-182.

Sun J., Wang F., Yin H., and Zhang B. (2019) Money talks: the environmental impact of China's green credit policy. J Policy Anal Manag 38(3):653–680

Sun Tzu (VI c. B.C.). The Art of War (Article XI).

Thies, M. F. (2001). Keeping Tabs on Partners: The Logic of Delegation in Coalition Governments. American Journal of Political Science 45(3): 580–598.

Tobin, M.J. (2022). Fiftieth Anniversary of Uncovering the Tuskegee Syphilis Study: The Story and Timeless Lessons. American Journal of Respiratory and Critical Care Medicine 205(10): 1145-1158.

Torslov, T., Wier, L., and Zucman, G. (2020). The missing profits of nations. NBER Working Paper No. 24701.

Toshkov, D., B. Carroll, and K. Yesilkagit (2021). Government capacity, societal trust or party preferences: what accounts for the variety of national policy responses to the COVID-19 pandemic in Europe? Journal of European Public Policy forthcoming DOI:10.1080/13501763.2021.1928270

US Census Bureau. https://www.census.gov/ (retrieved 28 October 2021)

US Department of Labor. https://www.bls.gov (retrieved 28 October 2021)

US Travel Association. https://www.ustravel.org (retrieved 28 October 2021)

Vallgårda, S. (2007). Problematizations and Path Dependency: HIV/AIDS Policies in Denmark and Sweden. Medical History 51(1): 99–112.

Van Dooren, W., and Noordegraaf, M. (2020). Staging Science: Authoritativeness and Fragility of Models and Measurement in the Covid-19 Crisis. Public Administration Review 80(4): 610-615.

Verity. R., et al. (2020). Estimates of the severity of COVID-19 disease. The Lancet Infectious Diseases, 20(6): 669-677.

Vicente, L., Savard, G. and Judice, J. (1994) Descent approaches for quadratic bilevel programming. Journal of Optimization Theory and Applications, 81(2), 379-399.

Wallerstein, I. (1976). From Feudalism to Capitalism: Transition or Transitions?. Social Forces 55(2): 273–283.

Warner, M.E., and Zhang, X. (2021). Social Safety Nets and COVID-19 Stay Home Orders across US States: A Comparative Policy Analysis. Journal of Comparative Policy Analysis 23(2): 176-190.

Wehde, W., and Choi, J. (2021). Public Preferences for Disaster Federalism: Comparing Public Risk Management Preferences Across Levels of Government and Hazards. Public Administration Review forthcoming https://doi.org/10.1111/puar.13432

Western, B., and Jackman, S. (1994). Bayesian Inference for Comparative Research. American Political Science Review 88(2): 412-423.

Wong, David W.S., and Yun, L. (2020). Spreading of COVID-19: Density matters. PLoS ONE 15(12): e0242398

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

World Health Organization. 2020b. Clinical management of severe acute respiratory infection when novel coronavirus (nCoV) infection is suspected: interim guidance. January 12, 2020.

World Health Organization. 2020a. Risk communication and community engagement readiness and initial response for novel coronaviruses (nCoV): interim guidance. 10 January 2020.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. Political Analysis, 25(1), 57-76.

Yan, B., Zhang, X., Wu, L., Zhu, H, and Chen, B. (2020). Why Do Countries Respond Differently to COVID-19? A Comparative Study of Sweden, China, France, and Japan. American Review of Public Administration 50(6/7): 762-769

Yesilyurt, M. E., and Elhorst, P. (2017). Impacts of neighboring countries on military expenditures: A dynamic spatial panel approach. Journal of Peace Research 54(6): 777–790.