FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

VÍTOR AUGUSTO POSSEBOM

THEORETICAL AND EMPIRICAL ESSAYS ON MICROECONOMETRICS

São Paulo
2016

VÍTOR AUGUSTO POSSEBOM

THEORETICAL AND EMPIRICAL ESSAYS ON MICROECONOMETRICS

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas como requisito para obtenção do título de Mestre em Economia de Empresas

Campo de Conhecimento:
Microeconomia – Modelos Econométricos

Orientador: Prof. Dr. Sergio Pinheiro Firpo
Co-orientador: Profa. Dra. Cristine Campos de Xavier Pinto

São Paulo
2016

Possebom, Vítor Augusto.
Theoretical and Empirical Essays on Microeconometrics / Vítor Augusto Possebom. - 2016.
60 f.

Orientadores: Sergio Pinheiro Firpo, Cristine Campos de Xavier Pinto
Dissertação (mestrado) - Escola de Economia de São Paulo.

1. Microconomia - Modelos econométricos. 2. Testes de hipóteses estatísticas. 3. Portos e zonas francas - Manaus (AM). 4. Análise de painel. I. Firpo, Sergio Pinheiro. II. Pinto, Cristine Campos de Xavier. III. Dissertação (mestrado) - Escola de Economia de São Paulo. IV. Título.

CDU 330.101.542

VÍTOR AUGUSTO POSSEBOM

THEORETICAL AND EMPIRICAL ESSAYS ON MICROECONOMETRICS

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas como requisito para obtenção do título de Mestre em Economia de Empresas.

Campo de Conhecimento:
Microeconomia – Modelos Econométricos

Data de Aprovação:
___/___/_____

Banca examinadora:

_____
Prof. Dr. Sergio Pinheiro Firpo (Orientador)
Insper

_____
Profa. Dra. Cristine Campos de Xavier Pinto
FGV-EESP

_____
Prof. Dr. Bruno Ferman
FGV-EESP

_____
Prof. Dr. Ricardo Paes de Barros
Insper

To my parents.

Watching a coast as it slips by the ship is like thinking about an enigma. There it is before you, smiling, frowning, inviting, grand, mean, insipid, or savage, and always mute with an air of whispering, "Come and find out".

Joseph Conrad, Heart of Darkness.

AGRADECIMENTOS

Sou grato aos professores, Sergio Firpo, meu orientador, e Cristine Pinto, minha co-orientadora, por tudo que eles me ensinaram não apenas durante o processo de escrever a minha dissertação de mestrado, mas também durante os meus estudos na graduação e na pós-graduação. Eles certamente contribuíram imensamente para minha escolha de carreira e para o economista e econometrista que eu sou hoje.

Eu também sou grato aos professores Bruno Ferman e Ricardo Paes de Barros por serem dois dos primeiros leitores de alguns dos artigos que, talvez, marquem o começo da minha carreira acadêmica. Paes de Barros influenciou mais de uma geração de economistas brasileiros e sou grato por tê-lo encontrado como meu professor na EESP-FGV. Bruno Ferman é muito entusiasmado sobre Economia e contribuiu profundamente para meu conhecimento apesar de nós termos começado a trabalho junto apenas há alguns meses.

Mayra Lora, Enlinson Mattos e Paulo Picchetti também me ensinaram Estatística e Econometria no começo da minha carreira como economista, tendo um papel central na minha decisão de me tornar um pesquisador.

Eu agradeço todos os professores na Escola de Economia de São Paulo por tudo que eles me ensinaram desde 2009. Em particular, Ana Cristina Braga Martes, Vladimir Ponczek, Paulo Furquim, Braz Camargo, André Portela e Paulo Tenani influenciaram minhas decisões profissionais em algum momento. Bernardo Guimarães, que sempre é ávido por economia, também foi um personagem marcante na minha vida profissional, particularmente por ter escrito um interessante livro junto com Carlos Eduardo Gonçalves.

Yoshiaki Nakano sempre foi o *quarterback* da Escola de Economia de São Paulo. Sou grato a ele por ter construído um projeto tão incrível, sempre dando suporte aos professores e alunos. Ele também me convenceu a tomar a decisão certa quando eu era um adolescente de 18 anos.

Eu também sou grato a meus professores do ensino básico. Sem eles, não teria sido capaz de aprender tanto durante a minha vida.

Pessoalmente, estarei sempre em dívida com meus pais, Rute Augusto and Francisco Possebom. Eles sempre me oferecerão tudo com amor e me criaram em um ambiente intelectualmente estimulante. (Todos sabem que Desenvolvimento na Primeira Infância importa. MUITO.) Meu caráter pessoal e meus primeiros passos acadêmicos foram inteiramente influenciados por eles.

Lucky é também um membro da minha família e sempre me fez companhia, mesmo quando eu estava explicando a ele o Primeiro e o Segundo Teoremas do Bem-Estar. Ele me olhava tão intensamente durante nossas aulas que eu acho que ele sabe mais Microeconomia do que eu saberei em algum momento.

Durante o meu mestrado, eu fiz novos amigos e fiquei mais próximo dos antigos. Sou grato a eles pelo suporte acadêmico e, principalmente, emocional. Eles certamente provaram a importância de *peer effects*. Particularmente, Daniel de Lima, Flávio Riva, João Garcia e Marcela Mello conquistaram um lugar especial no meu coração. Nós, agora, nos mudamos para diferentes partes do globo, mas espero que sempre estejamos ao lado um dos outros.

Os amigos da minha graduação tem me dado apoio há muito tempo. Nunca esquecerei todos os momentos felizes que tive com eles (e ainda tenho). Um agradecimento muito especial é devido a Natasha Reis e Verena Paiva por todas as nossas extremamente longas e dramáticas conversas. Vocês me mantiveram caminhando durante os últimos dois anos. Sempre desejarei o melhor para vocês.

Meus amigos mais antigos do Bandeirantes estiveram lá apesar da distância física e das minhas frequentes ausências nos raros eventos que reuniam todos em São Paulo. Vocês sempre me ofereceram uma rota de fuga de problemas de matemática para risadas. Vocês se tornaram meu quarto andar.

Também agradeço à FAPESP por financiar minha pesquisa por meio fa bolsa número 2014/23731-3 e à CAPES por fornecer suporte financeiro durante o começo do meu mestrado. Como é fácil provar, ganhar o pão de cada dia é de extrema importância.

Ao fim, sou muito grato à Escola de Economia de São Paulo da Fundação Getulio Vargas por providenciar a infraestrutura necessária e por criar um ambiante academicamente desafiador durante a minha graduação e pós-graduação. *This is EESParta!*

ACKNOWLEDGMENTS

I am grateful to professors Sergio Firpo, my supervisor, and Cristine Pinto, my co-supervisor, for everything they have taught me not only during the process of writing my Master's Thesis, but also during my undergraduate and graduate studies. They surely have contributed immensely to my choice of career and to the economist and econometrician that I am now.

I am also thankful to professors Bruno Ferman and Ricardo Paes de Barros for being two of the first readers of some of the articles that hopefully will mark the start of my academic career. Paes de Barros have greatly influenced more than one generation of Brazilian economists and I am glad to have met him as my professor at EESP-FGV. Bruno Ferman is very enthusiastic about Economics and deeply contributed to my knowledge about it even though we started to work together just a couple of months ago.

Mayra Lora, Enlinson Mattos and Paulo Picchetti have also taught me Statistics and Econometrics at the beginning of my career as an economist, playing a central role in my decision to become a researcher too.

I thank all the professors at Sao Paulo School of Economics for everything they have taught me since 2009. In particular, Ana Cristina Braga Martes, Vladimir Ponczek, Paulo Furquim, Braz Camargo, André Portela and Paulo Tenani influenced my career decisions at some point. Bernardo Guimarães, who is always avid about Economics, was also an important character in my career life, particularly because of an interesting book he wrote with Carlos Eduardo Gonçalves.

Yoshiaki Nakano was the quarterback of São Paulo School of Economics. I am grateful to him for building such an awesome project, always supporting its faculty members and students. He also convinced me to take the right decision when I was a 18 years-old teenager.

I am also thankful to my elementary, middle and high school teachers. Without them, I would not be able to learn so much during my life.

Personally, I am forever in debt to my parents, Rute Augusto and Francisco Possebom. They have provided me everything with love and raised me in an intellectually stimulating environment. (Everyone knows that Early Childhood Development matters. A LOT.) My personal character and my first academic steps were entirely influenced by them.

Lucky is also a member of my family and have always made me company, even when I was explaining to him the First and Second Welfare Theorems. He looked to me so intensely during our lecture sections that I guess he may know more Microeconomics than I will ever know.

During my Master's Studies, I have made new friends and become closer to old ones. I am grateful to them for their academic and, mainly, emotional support. They have surely proved

me the importance of peer effects. Particularly, Daniel de Lima, Flávio Riva, João Garcia and Marcela Mello conquered a special place in my heart. We now move to different parts of the globe, but I hope we will always be by each other side.

The friends of my undergraduate studies have supported me for a long time now. I will never forget all the happy moments I had with them (and still have). A very special thanks goes to Natasha Reis and Verena Paiva for all our extremely long and dramatic conversations. You kept me walking during the last two years. I will always wish you all the best.

My oldest friends from Bandeirantes were there despite their physical distance or my frequent absences during the rare events when everyone was in São Paulo. You have always provided me a scape route from mathematical problems to laughs. You have become my Floor n. 4.

I am also thankful to FAPESP for funding my research through grant number 2014/23731-3 and to CAPES for supporting me financially during the beginning of my Master's studies. As it is easy to prove, earning's one's daily bread is of paramount importance.

Finally, I am grateful to São Paulo School of Economics and Fundação Getulio Vargas for providing the necessary infrastructure and creating an academic challenging environment during my undergraduate and Master's studies. This is EESParta!

ABSTRACT

This Master Thesis consists of one theoretical article and one empirical article on the field of Microeconometrics.

The first chapter[1], called *Synthetic Control Estimator: A Generalized Inference Procedure and Confidence Sets*, contributes to the literature about inference techniques of the Synthetic Control Method. This methodology was proposed to answer questions involving counterfactuals when only one treated unit and a few control units are observed. Although this method was applied in many empirical works, the formal theory behind its inference procedure is still an open question. In order to fulfill this lacuna, we make clear the sufficient hypotheses that guarantee the adequacy of Fisher's Exact Hypothesis Testing Procedure for panel data, allowing us to test any *sharp null hypothesis* and, consequently, to propose a new way to estimate Confidence Sets for the Synthetic Control Estimator by inverting a test statistic, the first confidence set when we have access only to finite sample, aggregate level data whose cross-sectional dimension may be larger than its time dimension. Moreover, we analyze the size and the power of the proposed test with a Monte Carlo experiment and find that test statistics that use the synthetic control method outperforms test statistics commonly used in the evaluation literature. We also extend our framework for the cases when we observe more than one outcome of interest (simultaneous hypothesis testing) or more than one treated unit (pooled intervention effect) and when heteroskedasticity is present.

The second chapter, called *Free Economic Area of Manaus: An Impact Evaluation using the Synthetic Control Method*, is an empirical article. We apply the synthetic control method for Brazilian city-level data during the 20th Century in order to evaluate the economic impact of the Free Economic Area of Manaus (FEAM). We find that this enterprise zone had positive significant effects on Real GDP per capita and Services Total Production per capita, but it also had negative significant effects on Agriculture Total Production per capita. Our results suggest that this subsidy policy achieve its goal of promoting regional economic growth, even though it may have provoked mis-allocation of resources among economic sectors.

Key-words: Synthetic Control Estimator, Hypothesis Testing, Confidence Sets, Free Economic Area of Manaus, Enterprise Zones.

[1] We also thank useful suggestions by Marinho Bertanha, Gabriel Cepaluni, Brigham Frandsen, Dalia Ghanem, Ricardo Masini, Marcela Mello, Áureo de Paula, Cristine Pinto, Edson Severnini and seminar participants at São Paulo School of Economics, the California Econometrics Conference 2015 and the 37th Brazilian Meeting of Econometrics.

RESUMO

Esta dissertação de mestrado consiste em um artigo teórico e um artigo empírico no campo da Microeconometria.

O primeiro capítulo contribui para a literatura sobre técnica de inferência do método de controle sintético. Essa metodologia foi proposta para responder a questões envolvendo contrafactuais quando apenas uma unidade tratada e poucas unidades controle são observadas. Apesar de esse método ter sido aplicado em muitos trabalhos empíricos, a teoria formal por trás de seu procedimento de inferência ainda é uma questão em aberto. Para preencher essa lacuna, nós deixamos claras hipóteses suficientes que garantem a validade do Procedimento Exato de Teste de Hipótese de Fisher para dados em painel, permitindo que nós testássemos qualquer hipótese nula do tipo *sharp* e, consequentemente, que nós propuséssemos uma nova forma de estimar conjuntos de confiança para o Estimador de Controle Sintético por meio da inversão de uma estatística de teste, o primeiro conjunto de confiança quando temos acesso apenas a dados agregados cuja dimensão de *cross-section* pode ser maior que a dimensão temporal. Ademais, nós analisamos o tamanho e o poder do teste proposto por meio de um experimento de Monte Carlo e encontramos que estatísticas de teste que usam o método de controle sintético apresentam uma performance superior àquela apresentada pelas estatísticas de teste comumente analisadas na literatura de avaliação de impacto. Nós também estendemos nosso procedimento para abarcar os casos em que observamos mais de uma variável de interesse (teste simultâneo de hipótese) ou mais de uma unidade tratada (efeito agregado da intervenção) e quando heterocedasticidade está presente.

O segundo capítulo é um artigo empírico. Nós aplicamos o método de controle sintético a dados municipais brasileiros durante o século 20 com o intuito de avaliar o impacto econômico da Zona Franca de Manaus (ZFM). Nós encontramos que essa zona de empreendimento teve efeitos positivos significantes sobre o PIB Real per capita e sobre a Produção Total per capita do setor de Serviços, mas também teve um efeito negativo e significante sobre a Produção total per capita do setor Agrícola. Nossos resultados sugerem que essa política de subsídio alcançou seu objetivo de promover crescimento econômico regional, apesar de possivelmente ter provocado falhas de alocação de recursos entre setores econômicos.

Palavras-chaves: Estimador de Controle Sintético, Teste de Hipótese, Conjuntos de Confiança, Zona Franca de Manaus, Zonas de Empreendimentos.

List of Figures

List of Tables

TABLE OF CONTENTS

1 Synthetic Control Estimator: A Generalized Inference Procedure and Confidence Sets

1.1 Introduction

The Synthetic Control Method was proposed by Abadie e Gardeazabal (2003), Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015) to address counterfactual questions involving only one treated unit and a few control units. Intuitively, this method constructs a weighted average of control units that is as similar as possible to the treated unit regarding the pre-treatment outcome variable and covariates. For this reason, this weighted average of control units is known as the synthetic control. Although the empirical literature applying the Synthetic Control Method is vast[1], this tool's theoretical foundation is still under development.

Our first contribution to this literature is to formalize the current existing inference procedure proposed by Abadie, Diamond e Hainmueller (2010). Adapting the framework described by Imbens e Rubin (2015) to a panel data context, we clearly state hypotheses that guarantee the validity of Fisher's Exact Hypothesis Testing Procedure, a method that compares an observed test statistic to its empirical distribution in order to verify whether there is enough evidence to reject the null hypothesis. Particularly, our framework allow us to test not only the null hypothesis of no effect whatsoever, but also any kind of *sharp null hypothesis*, generalizing the current existing inference procedure. The possibility of testing any *sharp null hypothesis* is relevant in order to approximate the intervention effect function by simpler functions that can be used to predict its future behavior. Most importantly, being able to test more flexible null hypothesis is fundamental to compare the costs and benefits of a policy. For example, one can interpret the intervention effect as the policy's benefit and test whether it is different than its costs. It also enables the empirical researcher to test theories related to the analyzed phenomenon, particularly the ones that predict some specific kind of intervention effect.

Based on our generalization of the current existing inference procedure, we propose a novel way to estimate Confidence Sets for the Synthetic Control Estimator by inverting a test statistic. We modify the method proposed by Imbens e Rubin (2015) to estimate confidence

[1] This tool was applied to an extremely diverse set of topics, including, for instance, issues related to terrorism, civil wars and political risk (Abadie e Gardeazabal (2003), Bove, Elia e Smith (2014), Li (2012), Montalvo (2011), Yu e Wang (2013)), natural resources and disasters (Barone e Mocetti (2014), Cavallo et al. (2013), Coffman e Noy (2011), DuPont e Noy (2012), Mideksa (2013), Sills et al. (2015), Smith (2015)), international finance (Jinjarak, Noy e Zheng (2013), Sanso-Navarro (2011)), education and research policy (Belot e Vandenberghe (2014), Chan et al. (2014), Hinrichs (2012)), health policy (Bauhoff (2014), Kreif et al. (2015)), economic and trade liberalization (Billmeier e Nannicini (2013), Gathani, Santini e Stoelinga (2013), Hosny (2012)), political reforms (Billmeier e Nannicini (2009), Carrasco, Mello e Duarte (2014), Dhungana (2011) Ribeiro, Stein e Kang (2013)), labor (Bohn, Lofstrom e Raphael (2014), Calderon (2014)), taxation (Kleven, Landais e Saez (2013), Souza (2014)), crime (Pinotti (2012b), Pinotti (2012a), Saunders et al. (2014)), social connections (Acemoglu et al. (2013)), and local development (Ando (2015), Gobillon e Magnac (2016), Kirkpatrick e Bennear (2014), Liu (2015), Severnini (2014)).

intervals based on Fisher's Exact Hypothesis Testing Procedure in order to apply it to a panel data framework, using test statistics generated by the Synthetic Control Method. To the best of our knowledge, this is the first work to propose Confidence Sets for the Synthetic Control Estimator when we observe aggregate level data for only one treated unit and a few control units (i.e., small finite samples) in a context whose cross-section dimension may be larger than its time dimension. With our confidence sets, a researcher can quickly show, by using a graph, not only the significance of the estimated intervention effect, but also the precision of this point-estimate. This plot summarizes a large amount of information that is important to measure the strength of qualitative conclusions achieved after an econometric analysis.

Since this generalized inference method and the associated confidence sets can use many different test statistics, we verify, by a Monte Carlo experiment, the size and the power of five test statistics when they are used in this inference procedure. We choose them based on our review of the empirical literature that applies the Synthetic Control Method. More specifically, we compare test statistics that use the Synthetic Control Estimator to test statistics that use simpler methods (e.g.: difference in means and a permuted differences-in-differences test that are commonly used in the evaluation literature) and to the asymptotic inference procedure for the difference-in-differences estimator proposed by Conley e Taber (2011). We find that a inference procedure based on a test statistic that uses the Synthetic Control Method performs much better than the ones that do not use this method when we compare their size and power.

We also extend our framework to cover hypothesis testing and confidence set estimation for a pooled effect among few treated units, as a formalization and a generalization of the test proposed by Cavallo et al. (2013), and to simultaneously test null hypotheses for different outcome variables. This last extension, that also expands the framework described by Anderson (2008) to a panel data context, is important, for example, to evaluate political reforms (Billmeier e Nannicini (2009), Billmeier e Nannicini (2013), Carrasco, Mello e Duarte (2014), Jinjarak, Noy e Zheng (2013), Sanso-Navarro (2011)) that generally affect multiple outcomes variables, such as income levels and investment. Moreover, we can also interpret each post-intervention time period as a different outcome variable, allowing us to investigate the timing of an intervention effect — a relevant possibility when the empirical researcher aims to uncover short and long term effects. As one last extension, we make some brief comments about cases in which heteroskedasticity is a concern. We stress that choosing a test statistic that is robust to this issue — e.g., the t-test or a modified version of the *RMSPE* test statistic — allows us to apply our generalized inference procedure and our confidence sets to empirical problems that present heteroskedasticity.

At the end, we apply our generalized inference procedure, its associated new confidence sets and its extension to the case of simultaneous hypothesis testing to evaluate the statistical significance of the economic impact of ETA's terrorism estimated by Abadie e Gardeazabal (2003). With this empirical exercise, we illustrate how our proposed confidence set summarizes a large amount of information in a simple graph. We conclude that it is not possible to draw any

conclusion about the size or the sign of the impact of ETA's terrorism on Basque Country's GDP per-capita.

Literature Review

Regarding the inference of the Synthetic Control Method, other authors have surely made important previous contributions. Abadie, Diamond e Hainmueller (2010)[2] are the first authors to propose a inference procedure that consists in estimating p-values through permutation tests and Abadie, Diamond e Hainmueller (2015) suggest a different test statistic for the same procedure. However, they do not make clear the sufficient hypotheses that guarantee that their proposed p-values are valid. One recent advancement in this direction is Ando e Sävje (2013), who discuss the importance of the *Identical and Independent Distribution* hypothesis for the inference procedure[3] and propose two new test statistics that have adequate size and more power when applied to the above mentioned hypothesis test than the ones proposed by Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015).

Bauhoff (2014), Calderon (2014) and Severnini (2014) propose a way to apply the Synthetic Control Estimator to many treated and control units that is similar to a matching estimator for panel data, but none of them discusses its statistical properties in detail. Following a similar but more formal approach, Wong (2015) extends the synthetic control estimator to a cross-sectional setting where individual-level data is available and derives its asymptotic distribution when the number of observed individuals goes to infinity. Wong (2015) also explores the synthetic control estimator when panel data (or repeated cross-sections) are available in two levels: an aggregate level (regions), where treatment is assigned, and an individual level, where outcomes are observed. In this framework, he derives the asymptotic distribution of the synthetic control estimator when the number of individuals in each region goes to infinity. Finally, Cavallo et al. (2013) and Dube e Zipperer (2013) develop different ways to apply the Synthetic Control Estimator when there are more than one treated unit and propose tests[4] that are similar to the ones proposed by Abadie, Diamond e Hainmueller (2010), although they do not address the statistical properties of their inference procedures either.

Gobillon e Magnac (2016), also working on a context with more than one treated unit, propose a way to compute confidence intervals for their synthetic control estimator based on bootstrapping the point estimate. Although the authors have not clearly stated the assumptions behind their inference procedure either, it requires a large number of treated and control regions in order to be valid and focus exclusively on the time average of the post-intervention effect. Our

[2] They also discuss the asymptotic unbiasedness of their method. Kaul et al. (2015) deepen this topic by arguing that using all pre-intervention outcomes as economic predictors might provoke bias by forcing the synthetic control estimator to ignore all other predictor covariates.

[3] Our hypotheses are different than the ones advocated by Ando e Sävje (2013). In particular, we do not assume that units are independent and identically distributed.

[4] Acemoglu et al. (2013) follows a procedure similar to the one proposed by Cavallo et al. (2013). However, the former is less computationally demanding than the latter.

approach differs from theirs in two ways: it is valid in small samples and allow the construction of confidence sets for the post-intervention effect as a function of time. Consequently, while their inference procedure allows the empirical researcher to test only constant in time intervention effects, our generalized inference procedure allows the empirical researcher to test any function of time as the intervention effect.

Moreover, Carvalho, Mansini e Medeiros (2015) propose the Artificial Counterfactual Estimator (ArCo), that is similar in purpose to the Synthetic Control Estimator, and derive its asymptotic distribution when the time dimension is large (long panel data sets). However, many of the problems to which the Synthetic Control Method is applied present a cross-section dimension larger than their time dimension, making it impossible to apply the ArCo to them. Wong (2015) also conducts an asymptotic analysis when the pre-intervention period goes to infinity.

Finally, our approach is similar to the way Conley e Taber (2011) estimate confidence intervals for the difference-in-differences estimator in the sense that we also construct confidence sets by inverting a test statistic. However, we differ from them in many aspects. Firstly, while they make a contribution to the difference-in-differences framework, our contribution is inserted in the Synthetic Control literature. Secondly, they assume a functional form for the potential outcomes — imposing that the treatment effect is constant in time — and an arbitrarily large number of control units, while we assume a fixed and (possibly) small number of control units and make no assumptions concerning the potential outcome functional form — i.e., treatment effects can vary in time.

This paper is divided as follows: section 2 explains the Synthetic Control Method as proposed by Abadie e Gardeazabal (2003), Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015), and formalizes and generalizes its inference procedure; section 3 proposes a way to estimate Confidence Sets for the Synthetic Control Estimator; section 4 analyzes size and power of different tests statistics employed in this hypothesis test through a Monte Carlo experiment; section 5 develops possible extensions to our framework; section 6 applies our proposed inference procedure to the data set about the Basque Country made available by Abadie, Diamond e Hainmueller (2011) and section 7 concludes. Finally, in the appendices, we didactically explain how to compute the test statistics described in section 2, expand the results of our Monte Carlo Experiment and offer a guide to empirical researchers who wish to employ the synthetic control method in their studies.

1.2 Synthetic Control Method

This section is organized in three subsections. The first one presents the Synthetic Control Estimator, while the second one explains its inference procedure based on permutation tests. The ideas and notation that are used in the next two subsections are mostly based on Abadie,

Diamond e Hainmueller (2010). Finally, in the third subsection, we clearly state the hypotheses that guarantees that the current existing inference procedure is valid, generalizing it to test any *sharp null hypothesis* using any test statistic.

1.2.1 Synthetic Control Estimator

Suppose that we observe data for $(J+1) \in \mathbb{N}$ regions[5] during $T \in \mathbb{N}$ time periods. Additionally, assume that there is an intervention[6] that affects only region 1[7] from period T_0+1 to period T uninterruptedly[8], where $T_0 \in (1,T) \cap \mathbb{N}$. Let the scalar $Y_{j,t}^N$ be the potential outcome that would be observed for region j in period t if there were no intervention for $j \in \{1,...,J+1\}$ and $t \in \{1,...,T\}$. Let the scalar $Y_{j,t}^I$ be the potential outcome that would be observed for region j in period t if region j faced the intervention at period t. Define

$$\alpha_{j,t} := Y_{j,t}^I - Y_{j,t}^N \tag{1.1}$$

as the intervention effect (or gap) for region j in period t and $D_{j,t}$ as a dummy variable that assumes value 1 if region j faces the intervention in period t and value 0 otherwise. With this notation, we have that the observed outcome for unit j in period t is given by

$$Y_{j,t} := Y_{j,t}^N + \alpha_{j,t} D_{j,t}.$$

Since only the first region faces the intervention from period T_0+1 to T, we have that:

$$D_{j,t} := \begin{cases} 1 & \text{if } j=1 \text{ and } t > T_0, \\ 0 & \text{otherwise.} \end{cases}$$

We aim to estimate $(\alpha_{1,T_0+1},...,\alpha_{1,T})$. Since $Y_{1,t}^I$ is observable for $t > T_0$, equation (2.1) guarantees that we only need to estimate $Y_{1,t}^N$ to accomplish this goal.

---

[5] We use the word "region"instead of more generic terms, such as "unit", because most synthetic control applications analyze data that are aggregated at the state or country level. We use the term *donor pool* to designated the entire group of $(J+1)$ observed regions.

[6] Although the treatment effect literature commonly uses the more generic expression "treated unit", we adopt the expression "the region that faced an intervention"because it is more common in the comparative politics literature, an area where the synthetic control method is largely applied.

[7] In subsection 1.5.2, we extend this framework to include the case when multiple units face the same or a similar intervention.

[8] Two famous examples of interventions that affect uninterruptedly a region are Proposition 99 — an Tobacco Control Legislation in California — and the German Reunification, that were studied by Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015). If the intervention is interrupted (e.g.: ETA's Terrorism in the Basque Country studied by Abadie e Gardeazabal (2003)), we just have to interpret our treatment differently. Instead of defining the treatment as "region 1 faces an intervention", we define treatment as "region 1 have been exposed to an event that potentially has long term consequences". For example, instead of defining our treatment as "the Basque Country faces constant bombings perpetrated by ETA", we define our treatment as "the Basque Country suffered some bombings perpetrated by ETA".

Let $\mathbf{Y_j} := [Y_{j,1}...Y_{j,T_0}]'$ be the vector of observed outcomes for region $j \in \{1, ..., J+1\}$ in the pre-intervention period and $\mathbf{X_j}$ a $(K \times 1)$-vector of predictors of $\mathbf{Y_j}$.[9] Let $\mathbf{Y_0} = [\mathbf{Y_2}...\mathbf{Y_{J+1}}]$ be a $(T_0 \times J)$-matrix and $\mathbf{X_0} = [\mathbf{X_2}...\mathbf{X_{J+1}}]$ be a $(K \times J)$-matrix.

Since we want to make region 1's synthetic control as similar as possible to the actual region 1, the Synthetic Control Estimator of $Y_{1,t}^N$ is given, for each $t \in \{1, ..., T\}$, by

$$\widehat{Y}_{1,t}^N := \sum_{j=2}^{J+1} \widehat{w}_j Y_{j,t}, \tag{1.2}$$

where $\widehat{\mathbf{W}} = [\widehat{w}_2...\widehat{w}_{J+1}]' := \widehat{\mathbf{W}}(\widehat{\mathbf{V}}) \in \mathbb{R}^J$ is given by the solution to a nested minimization problem:

$$\widehat{\mathbf{W}}(\mathbf{V}) := \arg\min_{\mathbf{W} \in \mathcal{W}} (\mathbf{X_1} - \mathbf{X_0}\mathbf{W})'\mathbf{V}(\mathbf{X_1} - \mathbf{X_0}\mathbf{W}) \tag{1.3}$$

where $\mathcal{W} := \left\{\mathbf{W} = [w_2...w_{J+1}]' \in \mathbb{R}^J : w_j \geq 0 \text{ for each } j \in \{2, ..., J+1\} \text{ and } \sum_{j=2}^{J+1} w_j = 1\right\}$ and $\mathbf{V}$ is a diagonal positive semidefinite matrix of dimension $(K \times K)$ whose trace equals one. Moreover,

$$\widehat{\mathbf{V}} := \arg\min_{\mathbf{V} \in \mathcal{V}} (\mathbf{Y_1} - \mathbf{Y_0}\widehat{\mathbf{W}}(\mathbf{V}))'(\mathbf{Y_1} - \mathbf{Y_0}\widehat{\mathbf{W}}(\mathbf{V})) \tag{1.4}$$

where $\mathcal{V}$ is the set of diagonal positive semidefinite matrix of dimension $(K \times K)$ whose trace equals one.

Intuitively, $\widehat{\mathbf{W}}$ is a weighting vector that measures the relative importance of each region in the synthetic control of region 1 and $\widehat{\mathbf{V}}$ measures the relative importance of each one of the K predictors. Consequently, this technique makes the synthetic control of region 1 as similar as possible with the actual region 1 considering the K predictors and the pre-intervention values of the outcome variable when we choose the Euclidean metric (or a reweighed version of it) to evaluate the distance between the observed variables for region 1 and the values predicted by the Synthetic Control Method.[10]

[9] Some lines of matrix $\mathbf{X_j}$ can be linear combinations of the variables in $\mathbf{Y_j}$.

[10] Abadie e Gardeazabal (2003), Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015) propose two other ways to choose $\widehat{\mathbf{V}}$. The first and most simple one is to use subjective and previous knowledge about the relative importance of each predictor. Since one of the advantages of the Synthetic Control Method is to make the choice of comparison groups in comparative case studies more objective, this method of choosing $\mathbf{V}$ is discouraged by those authors. Another choice method for $\widehat{\mathbf{V}}$ is to divide the pre-intervention period in two sub-periods: one training period and one validation period. While data from the training period are used to solve problem (2.2), data for the validation period are used to solve problem (2.3). Intuitively, this technique of cross-validation chooses matrix $\widehat{\mathbf{W}}(\widehat{\mathbf{V}})$ to minimize the out-of-sample prediction errors, an advantage when compared to the method described in the main text. However, the cost of this improvement is the need of a longer pre-intervention period. Moreover, the Stata command made available by those authors also allows the researcher to use a regression-based method in order to compute matrix $\widehat{\mathbf{V}}$. It basically regress matrix $\mathbf{Y_1}$ on $\mathbf{X_1}$ and imposes $v_k = |\beta_k| / \left(\sum_{k'=1}^{K} |\beta_{k'}|\right)$, where v_k is the k-th diagonal element of matrix $\mathbf{V}$ and β_k is the k-th coefficient of the regression of $\mathbf{Y_1}$ on $\mathbf{X_1}$. The choice method that we have chosen to present in the

Finally, we define the Synthetic Control Estimator of $\alpha_{1,t}$ (or the estimated gap) as

$$\widehat{\alpha}_{1,t} := Y_{1,t} - \widehat{Y}^N_{1,t} \tag{1.5}$$

for each $t \in \{1, ..., T\}$.

1.2.2 Hypothesis Testing

Abadie, Diamond e Hainmueller (2010) propose a inference procedure that

> examines whether or not the estimated effect of the actual intervention is large relative to the distribution of the effects estimated for the regions not exposed to the intervention. This is informative inference if under the hypothesis of no intervention effect the estimated effect of the intervention is not expected to be abnormal relative to the distribution of the placebo effects. (p. 497)

In order to do that, they run a permutation test, i.e., they permute which region is assumed to be treated and estimate, for each $j \in \{2, ..., J+1\}$ and $t \in \{1, ..., T\}$, $\widehat{\alpha}_{j,t}$ as described in subsection 1.2.1. Then, they compare the entire vector $\widehat{\boldsymbol{\alpha}}_1 = [\widehat{\alpha}_{1,T_0+1} ... \widehat{\alpha}_{1,T}]'$ with the empirical distribution of $\widehat{\boldsymbol{\alpha}}_j = [\widehat{\alpha}_{j,T_0+1} ... \widehat{\alpha}_{j,T}]'$ estimated through the permutation test. If the vector of estimated effects for region 1 is very different (i.e., large in absolute values), they reject the null hypothesis of no effect.

Abadie, Diamond e Hainmueller (2015) note a problem with this approach: $|\widehat{\alpha}_{1,t}|$ can be abnormally large when compared to the empirical distribution of $|\widehat{\alpha}_{j,t}|$ for some $t \in \{T_0 + 1, ..., T\}$, but not for other time periods. In this case, it is not clear at all whether one should reject the null hypothesis of no effect or not. In order to solve this problem, they recommend to use the empirical distribution of

$$RMSPE_j := \frac{\sum_{t=T_0+1}^{T} \left(Y_{j,t} - \widehat{Y^N_{j,t}}\right)^2 / (T - T_0)}{\sum_{t=1}^{T_0} \left(Y_{j,t} - \widehat{Y^N_{j,t}}\right)^2 / T_0}$$

where the acronym RMSPE stands for *ratio of the mean squared prediction errors*. Moreover, they propose to calculate a p-value

$$p := \frac{\sum_{j=1}^{J+1} \mathbb{1}\left[RMSPE_j \geq RMSPE_1\right]}{J+1}, \tag{1.6}$$

where $\mathbb{1}[\diamond]$ is the indicator function of event $\diamond$, and reject the null hypothesis of no effect if p is less than some pre-specified significance level, such as the traditional value of 0.1.

Although this RMSPE test statistic solve the problem generated by the time dimension of the Synthetic Control Estimator, Abadie, Diamond e Hainmueller (2015) does not state the

main text is the most used one in the empirical literature.

sufficient conditions that guarantee the validity of this procedure[11] nor discuss this test's size and power. We address the former issue in the next subsection and the latter in section 1.4.

1.2.3 Formalizing and Generalizing the Inference Procedure

In this section, we follow Imbens e Rubin (2015), adapting their framework to a panel data context. We want to formalize and generalize the inference procedure described in subsection 1.2.2. The first hypothesis that we make is the *stable unit treatment value assumption* (SUTVA):

Assumption 1. *The potential outcome vectors* $\mathbf{Y}_j^I := \left[Y_{j,1}^I ... Y_{j,T}^I\right]'$ *and* $\mathbf{Y}_j^N := \left[Y_{j,1}^N ... Y_{j,T}^N\right]'$ *for each region* $j \in \{1, ..., J+1\}$ *do not vary based on whether other regions face the intervention or not (i.e., no spill-over effects in space) and, for each region, there are no different forms or versions of intervention (i.e., single dose treatment), which lead to different potential outcomes (IMBENS; RUBIN, 2015, p. 19).*

The second assumption concerns the treatment assignment:

Assumption 2. *The choice of which unit will be treated (i.e., which region is our region 1) is random* conditional on the choice of the donor pool.[12]

Although assumption 2 seems strong, it holds true for many empirical applications of the Synthetic Control Estimator. For example, Barone e Mocetti (2014), Cavallo et al. (2013), Coffman e Noy (2011) and DuPont e Noy (2012) evaluate the economic effect of large scale natural disasters, such as earthquakes, hurricanes or volcano eruptions. Although the regions in the globe that frequently faces these disasters are not random, the specific region among them that will be hit by a natural disaster and the timing of this phenomenon is fully random.[13] Moreover, Pinotti (2012b) and Pinotti (2012a) evaluate the economic and political cost of organized crime in Italy exploring the increase in Mafia activities after two large earthquakes. Two other examples of the plausibility of assumption 2 are Smith (2015), who argues that the discovery of large natural resources reserves is *as-if-random*, and Liu (2015), who argues that the location of land-grand universities in the 19th century is *as-if-random* too.[14]

[11] Particularly, it is not clear at all what they mean by "the hypothesis of no intervention effect"(ABADIE; DIAMOND; HAINMUELLER, 2010, p. 497). Is it a null average effect? Or a null median effect? Or even a null effect for all units in all time periods? Moreover, asking what these authors mean by "the estimated effect of the intervention is not expected to be abnormal"(ABADIE; DIAMOND; HAINMUELLER, 2010, p. 497) is also a valid question.

[12] This assumption is our precise definition of "not expected to be abnormal"in footnote 11.

[13] In this example, the donor pool contains all countries that frequently faces natural disasters. Conditional on being in the donor pools, being treated (i.e., being hit by a natural disaster in the analyzed time window) is random.

[14] Even in randomized control trials, the synthetic control method may be more interesting than traditional statistical methods when there are only a few treated units — an issue that may emerge due to budget constraints. As we show in section 1.4, test statistics that use the synthetic control estimator are more powerful than the ones that do not use it.

We also stress that the possibility of choosing the donor pool based on observable covariates implies that assumption 2 can be interpreted as imposing only random treatment assignment *conditional on observables*, a standard condition in the evaluation literature also known as *ignorability* or *unconfoundness*.

The third assumption is related to how we interpret the potential outcomes:

Assumption 3. *The potential outcomes* $\mathbf{Y}_j^I := \left[Y_{j,1}^I ... Y_{j,T}^I\right]'$ *and* $\mathbf{Y}_j^N := \left[Y_{j,1}^N ... Y_{j,T}^N\right]'$ *for each region* $j \in \{1, ..., J+1\}$ *are fixed but* a priori *unknown quantities.*[15]

Implicitly, we assume that we observe the *realization* of a random variable for the *entire population of interest* instead of a random sample of a larger superpopulation.[16]

We note that assumptions 2 and 3 implies that the units in the donor pool are *exchangeable*. In reality, *exchangeability* is the weakest assumption that guarantees the valid of our formal and generalized inference procedure, because it is simply based in a permutation test. However, we believe that, although stronger, assumptions 2 and 3 makes interpretation easier. In particular, assumption 2 justifies one of the robustness checks described in appendix 1.8.3.

Finally, our null hypothesis is given by a *sharp null hypothesis*:

$$H_0 : Y_{j,t}^I = Y_{j,t}^N + f_j(t) \text{ for each region } j \in \{1, ..., J+1\} \text{ and time period } t \in \{1, ..., T\}, \tag{1.7}$$

where $f_j : \{1, ..., T\} \to \mathbb{R}$ is a function of time that is specific to each region j.

Observe that a *sharp null hypothesis* allows us to know all potential outcomes for each region regardless of its treatment assignment. Note also that the *exact null hypothesis*

$$H_0 : Y_{j,t}^I = Y_{j,t}^N \text{ for each region } j \in \{1, ..., J+1\} \text{ and time period } t \in \{1, ..., T\}, \tag{1.8}$$

is a particular case of the *sharp null hypothesis* (1.7) and can be interpreted as an hypothesis of no intervention effect whatsoever. We underscore that equation (1.8) is our precise definition of "no intervention effect"in footnote 11.[17] We also note that, under assumptions 1-3 and the null hypothesis (1.8), the p-value in equation (1.6) is valid and known as *Fisher's Exact p-Value*, after Fisher (1971). In this sense, our inference procedure with the *sharp null hypothesis* is a generalization of the inference procedure proposed by Abadie, Diamond e Hainmueller (2015).

Although the *sharp null hypothesis* (1.7) is theoretically interesting due to its generality, we almost never have a meaningful null hypothesis that is precise enough to specify individual

[15] As a consequence of this assumption, all the randomness of our problem come from the treatment assignment.

[16] See Imbens e Rubin (2015) for details regarding this interpretation.

[17] Observe that the *exact null hypothesis* (1.8) is stronger than assuming that the *typical* (mean or median) effect across regions is zero.

intervention effects for each observed region. For this reason, we can simply assume a simpler *sharp null hypothesis*[18]:

$$H_0 : Y^I_{j,t} = Y^N_{j,t} + f(t) \text{ for each region } j \in \{1, ..., J+1\} \text{ and time period } t \in \{1, ..., T\}, \tag{1.9}$$

where $f : \{1, ..., T\} \to \mathbb{R}$.

After formally stating conditions that guarantee the validity of the inference procedure proposed by Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015), we generalize it to other test statistics and to any *sharp null hypothesis*. We, again, follow Imbens e Rubin (2015).

We define a test statistic θ_f as a known positive real-valued function $\theta_f(\iota, \tau, \mathbf{Y}, \mathbf{X}, f)$ of:

1. the vector $\iota := [\iota_1 ... \iota_{J+1}]' \in \mathbb{R}^{J+1}$ of treatment assignment, where $\iota_j = 1$ if region j faces the intervention at some moment in time and zero otherwise;
2. $\tau := [\tau_1 ... \tau_T]' \in \mathbb{R}^T$, where $\tau_t = 1$ if $t > T_0$ and zero otherwise;
3. the matrix
$$\mathbf{Y} := \begin{bmatrix} Y^I_{1,1}\iota_1\tau_1 + Y^N_{1,1}(1 - \iota_1\tau_1) & ... & Y^I_{1,T}\iota_1\tau_T + Y^N_{1,T}(1 - \iota_1\tau_T) \\ & \ddots & \\ Y^I_{J+1,1}\iota_{J+1}\tau_1 + Y^N_{J+1,1}(1 - \iota_{J+1}\tau_1) & ... & Y^I_{J+1,T}\iota_{J+1}\tau_T + Y^N_{J+1,T}(1 - \iota_{J+1}\tau_T) \end{bmatrix}$$
of observed outcomes;
4. the matrix $\mathbf{X} := [\mathbf{X}_1 \ \mathbf{X}_0]$ of predictor variables;
5. the intervention effect function $f : \{1, ..., T\} \to \mathbb{R}$ given by the *sharp null hypothesis* (1.9).

The observed test statistic is given by $\theta^{obs}_f := \theta(e_1, \tau, \mathbf{Y}, \mathbf{X}, f)$ and, under assumptions 1-3 and the *sharp null hypothesis* (1.9), we can estimate the entire empirical distribution of θ_f by permuting which region faces the intervention, i.e., by estimating $\theta_f(e_j, \tau, \mathbf{Y}, \mathbf{X}, f)$ for each $j \in \{1, ..., J+1\}$, where e_j is the j-th canonical vector of $\mathbb{R}^{J+1}$.[19] We reject the *sharp null hypothesis* (1.9) if

$$p_{\theta_f} := \frac{\sum_{j=1}^{J+1} \mathbb{1}\left[\theta(e_j, \tau, \mathbf{Y}, \mathbf{X}, f) \geq \theta^{obs}_f\right]}{J+1} \leq \gamma \tag{1.10}$$

[18] We stress that the *exact null hypothesis* is still a particular case of the simpler *sharp null hypothesis* (1.9).

[19] For a step-by-step guide on how to compute the test statistic and its empirical distribution, see appendix 1.8.1.

γ is some pre-specified significance level[20] Note that rejecting the null hypothesis implies that there is some region with a non-zero effect for some time period. Moreover, observe that *RMSPE* and any linear combination of the absolute estimated synthetic control gaps are test statistics according to this definition. Consequently, the hypothesis tests proposed by Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015) are inserted in this framework.[21]

Regarding the choice of function f, there are many interesting options for a empirical researcher. For example, after estimating the intervention effect function $(\widehat{\alpha}_{1,1}, ..., \widehat{\alpha}_{1,T_0+1}, ..., \widehat{\alpha}_{1,T})$, the researcher may want to fit a linear, a quadratic or a exponential function to the estimated points associated with the post-intervention period. He or she can then test whether this fitted function is rejected or not according to our inference procedure. This possibility is useful in order to predict, in a very simple way, the future behavior of the intervention effect function.

Another and possibly the most interesting option for function f is related to cost-benefit analysis. If the intervention cost and its benefit are in the same unit of measurement, function f can be the intervention cost as a function of time and our inference procedure allows the researcher to test whether the intervention effect is different than its costs.

Moreover, function f can be chosen in order to test a theory that predicts a specific form for the intervention effect. For example, imagine that a researcher is interested in analyzing the economic impact of natural disasters (Barone e Mocetti (2014), Cavallo et al. (2013), Coffman e Noy (2011), DuPont e Noy (2012)). Theory predicts three different possible intervention effects in this case: (i) GDP initially increases due to the aid effect and, then, decreases back to its potential level; (ii) GDP initially decreases due to the destruction effect and, then, increases back to its potential level; and (iii) GDP decreases permanently due to a reduction in its potential level. The researcher can choose a inverted U-shaped function f_i, a U-shaped function f_{ii} and a decreasing function f_{iii} and apply our inference procedure to each one of those three *sharp null hypotheses* in order to test which theoretical prediction is not rejected by the data.

1.3 Confidence Sets for the Synthetic Control Estimator

Following the inference procedure described at the end of subsection 1.2.3, we can test many different types of *sharp null hypothesis*. Consequently, we can invert the test statistic to estimate confidence sets for the treatment effect function. Formally, under assumptions 1-3, we

[20] Yates (1984) stresses that γ should be chosen carefully and always clearly reported since the discreteness of data (the number of regions is always a finite, usually small, natural number) may preclude the choice of the usual significance levels of 10% or 5%.

[21] In section 1.4, we analyze five different test statistics that were previously proposed in the synthetic control literature in order to select the ones that have power against an alternative hypothesis similar to $H_a: Y^I_{1,t} = Y^N_{1,t} + c_t$ for all time periods $t \in \{T_0 + 1, ..., T\}$, where $c_t \in \mathbb{R}$.

can construct a γ-confidence set in the space $\mathbb{R}^{\{1,...,T\}}$ as

$$CS_{\gamma,\theta} := \left\{ f \in \mathbb{R}^{\{1,...,T\}} : p_{\theta_f} > \gamma \right\}. \tag{1.11}$$

Note that it is easy to interpret $CS_{\gamma,\theta}$: it contains all intervention effect functions whose associated *sharp null hypotheses* are not rejected by the inference procedure described in subsection 1.2.3.

However, although theoretically possible to define such a general confidence set, null hypothesis (1.9) might be too general for practical reasons since the space $\mathbb{R}^{\{1,...,T\}}$ is too large to be informative and estimating such a confidence set would be computationally infeasible. For these reasons, we propose to assume the following null hypothesis:

$$H_0 : Y^I_{j,t} = Y^N_{j,t} + c \times \mathbb{1}(t \geq T_0 + 1) \tag{1.12}$$

for each region $j \in \{1, ..., J+1\}$ and time period $t \in \{1, ..., T\}$, where $c \in \mathbb{R}$. Intuitively, we assume that there is a constant (in space and in time) intervention effect. Note that we can apply the inference procedure described in subsection 1.2.3 to any $c \in \mathbb{R}$, estimating the empirical distribution of θ_c. Under assumptions 1-3, we can then construct a γ-confidence interval for the constant intervention effect as

$$CI_{\gamma,\theta} := \{c \in \mathbb{R} : p_{\theta_c} > \gamma\} \subseteq CS_{\gamma,\theta} \tag{1.13}$$

where $\gamma \in (0, 1) \subset \mathbb{R}$. It is easy to interpret $CI_{\gamma,\theta}$: it contains all constant in time intervention effects whose associated *sharp null hypotheses* are not rejected by the inference procedure described in subsection 1.2.3.

We can easily extend (1.12) and (1.13) to a linear in time intervention effect (with intercept equal to zero). Assume

$$H_0 : Y^I_{j,t} = Y^N_{j,t} + \widetilde{c} \times (t - T_0) \times \mathbb{1}(t \geq T_0 + 1) \tag{1.14}$$

for each region $j \in \{1, ..., J+1\}$ and time period $t \in \{1, ..., T\}$, where $\widetilde{c} \in \mathbb{R}$. Intuitively, we assume that there is a constant in space, but linear in time intervention effect (with intercept equal to zero). Note that we can apply the inference procedure described in subsection 1.2.3 to any $\widetilde{c} \in \mathbb{R}$, estimating the empirical distribution of $\theta_{\widetilde{c}}$. Under assumptions 1-3, we can then construct a γ-confidence set for the linear intervention effect as

$$\widetilde{CS}_{\gamma,\theta} := \left\{ \begin{array}{cc} f \in \mathbb{R}^{\{1,...,T\}} : & f(t) = \widetilde{c} \times (t - T_0) \times \mathbb{1}(t \geq T_0 + 1) \\ & \text{and } p_{\theta_{\widetilde{c}}} > \gamma \end{array} \right\} \subseteq CS_{\gamma,\theta} \tag{1.15}$$

where $\gamma \in (0, 1) \subset \mathbb{R}$. It is also easy to interpret $\widetilde{CR}_{\gamma,\theta}$: it contains all linear in time intervention effects (with intercept equal to zero) whose associated *sharp null hypotheses* are not rejected by the inference procedure described in subsection 1.2.3.

We also note that extending our confidence intervals to two-parameter functions (e.g.: quadratic, exponential and logarithmic functions) is theoretically straightforward as equation (1.11) makes clear. However, since we believe that computationally estimating such confidence sets would be extremely time consuming for the practitioner, we opted for restricting our main examples to one-parameter functions (equations (1.13) and (1.15)).

Finally, we highlight that confidence sets (1.13) and (1.15) summarizes a large amount of relevant information since they not only show the statistical significance of the estimated intervention effect, but also provide a measure of the precision of the point-estimate, indicating the strength of qualitative conclusions. Section 1.6 exemplifies the communication efficacy of this graphical device.

At the authors' webpage (<https://goo.gl/4Jvd2W>), there is a R function that implements confidence sets (1.13) and (1.15) using the *RMSPE* as a test statistic.

1.4 Analyzing Size and Power

In this section, we analyze the size and the power of five different test statistics when they are applied to the inference procedure described in subsection 1.2.3.[22] In order to do that, we assume seven different intervention effects, simulate 5,000 data sets for each intervention effect through a Monte Carlo experiment and, for each data set, we test, at the 10% significance level, the *exact null hypothesis* (equation (1.8)), following the mentioned inference procedure and using each test statistic. Firstly, we explain how we generated our data sets. Then, we describe our five test statistics. Finally, at the end of this section, we present and discuss the results of our Monte Carlo experiment.

The first step in our Monte Carlo experiment is to decide the values of the parameters: $J + 1$ (number of regions), T (number of time periods), T_0 (number of pre-intervention time periods) and K (number of predictors). In our review of the empirical literature, we found that typical values of these parameters are, approximately, $T = 25$, $T_0 = 15$ and $K = 10$ (nine control variables and the pre-intervention average of the outcome variable). We also set $J + 1 = 20$ (one treated region and nineteen control regions). Our data generating process follows equation (5) of Abadie, Diamond e Hainmueller (2010) and is different from the one used by Ando e Sävje (2013):

$$\begin{aligned} Y_{j,t+1}^N &= \delta_t Y_{j,t}^N + \boldsymbol{\beta}_{t+1}\mathbf{Z}_{j,t+1} + u_{j,t+1} \\ \mathbf{Z}_{j,t+1} &= \kappa_t Y_{j,t}^N + \boldsymbol{\pi}_t \mathbf{Z}_{j,t} + \mathbf{v}_{j,t+1} \end{aligned} \tag{1.16}$$

[22] In appendix 2, we discuss the size and the power of other thirteen test statistics.

for each $j \in \{1, ..., J+1\}$ and $t \in \{0, ..., T-1\}$, where $\mathbf{Z}_{j,t+1}$ is a $(K-1) \times 1$-dimension vector of control variables[23]. The scalar $u_{j,t+1}$ and each element of the $(K-1) \times 1$-dimension vector $\mathbf{v}_{j,t+1}$ are independent random draws from a standard normal distribution. The scalars δ_t and κ_t and each element of $\boldsymbol{\beta}_{t+1}$ and $\boldsymbol{\pi}_t$ are independent random draws from a uniform distribution with lower bound equal to -1 and upper bound equal to +1. We make $\mathbf{Z}_{j,0} = \mathbf{v}_{j,0}$ and $Y_{j,0}^N = \boldsymbol{\beta}_0 \mathbf{Z}_{j,0} + u_{j,0}$. Finally, the potential outcome when region 1 faces the intervention in period $t \in \{1, ..., T\}$ is given by

$$Y_{1,t}^I = Y_{1,t}^N + \lambda \times sd(Y_{1,t}^N | t \leq T_0) \times (t - T_0) \times \mathbb{1}\left[t \geq T_0 + 1\right], \tag{1.17}$$

where $\lambda \in \{0, 0.05, 0.1, 0.25, 0.5, 1.0, 2.0\}$ is the intervention effect and $sd(\clubsuit|\diamond)$ is the standard deviation of variable $\clubsuit$ conditional on event $\diamond$. Hence, our alternative hypothesis is that there is a linear intervention effect only for region 1, implying that our Monte Carlo experiment investigates what are the most powerful test statistics against this alternative hypothesis[24].

Note that, in each one of the 35,000 Monte Carlo repetitions, we create an entire population of regions. Hence, after realizing the values of the potential outcome variables, we can interpret them as fixed but *a priori* unknown quantities in accordance to assumption 3.[25]

Now that we have explained our data generating process for our 35,000 Monte Carlo repetitions (5,000 repetitions for each different intervention effect λ), we describe the five different test statistics that we use to analyze the size and the power of the inference procedure described in subsection 1.2.3:

- $\theta^1 := mean\left(\left|\widehat{\alpha}_{\widetilde{j},t}\right| | t \geq T_0 + 1\right)$ is implicitly suggested by Abadie, Diamond e Hainmueller (2010).
- $\theta^2 := RMSPE_{\widetilde{j}}$ is used by Abadie, Diamond e Hainmueller (2015).
- θ^3 is the absolute value of the statistic of a t-test that compares the estimated average post-intervention effect against zero. More precisely,

$$\theta^3 := \left| \frac{\overline{\alpha}_{post} / (T - T_0)}{\widehat{\sigma} / \sqrt{T - T_0}} \right|$$

where $\overline{\alpha}_{post} := \dfrac{\left(\sum_{t=T_0+1}^{T} \widehat{\alpha}_{\widetilde{j},t}\right)}{(T - T_0)} =: \theta^1$ and $\widehat{\sigma} := \dfrac{\left(\sum_{t=T_0+1}^{T} \left(\widehat{\alpha}_{\widetilde{j},t} - \overline{\alpha}_{post}\right)^2\right)}{(T - T_0)}$. This test statistic is used by Mideksa (2013).

[23] $\mathbf{X}_j$ is a vector that contains the pre-intervention averages of the control variables and the outcome variable.

[24] In a previous version of this text, that circulated under the title *Synthetic Control Estimator: A Walkthrough with Confidence Intervals*, we used a constant in time intervention effect. The results of that smaller Monte Carlo experiment were similar to the ones presented below.

[25] If we treat our hypothesis test as conditional on the realized outcome variable, assumption 3 holds automatically.

- $\theta^4 := mean\left(Y_{\widetilde{j},t} | t \geq T_0 + 1\right) - \frac{\sum_{t=T_0+1}^{T} \sum_{j \neq \widetilde{j}} Y_{j,t}}{(T - T_0) \times J}$ is a simple difference in means between the treated region and the control regions for the realized outcome variable during the post-intervention period. This test statistic is suggested by Imbens e Rubin (2015).

- θ^5 is the coefficient of the interaction term in a differences-in-differences model. More precisely, we estimate the model

$$Y_{j,t} = \eta_1 \times \mathbb{1}\left[j = \widetilde{j}\right] + \eta_2 \times \mathbb{1}\left[j = \widetilde{j}\right] \times \mathbb{1}\left[t \geq T_0 + 1\right] + Z_{j,t} \times \boldsymbol{\zeta} + \xi_j + \mu_t + \varepsilon_{j,t},$$

  where ξ_j and μ_t are, respectively, region and time fixed effects, and we make $\widehat{\theta^5} = \widehat{\eta}_2$.

where $\widetilde{j}$ is the region that is assumed to face the intervention in each permutation, $mean(\clubsuit|\diamond)$ is the mean of variable $\clubsuit$ conditional on event $\diamond$. We construct the empirical distribution of each test statistic for each Monte Carlo repetition and test the null hypothesis at the 10% significance level. In practice, we reject the null hypothesis if the observed test statistic is one of the two largest values of the empirical distribution of the test statistic.

Note that, although test statistic θ^4 and θ^5 do not use the synthetic control method, they are included in our Monte Carlo Experiment for being commonly used in the literature about permutation tests. Since the synthetic control estimator is a time-consuming and computer-demanding methodology, it is important to analyze whether it outperforms much simpler methods that are commonly used in the evaluation literature and that are also adequate in our framework (assumption 1-3). For this same reason, we also report rejection rates for the differences-in-differences inference procedure proposed by Conley e Taber (2011) (CT)[26].

Table 1 shows the results of our Monte Carlo Experiment. Each cell presents the rejection rate of the permutation test described in subsection 1.2.3 that uses the test statistic in each row or the rejection rate of the test proposed by Conley e Taber (2011) when the true intervention effect is given by the value mentioned in the column's heading. Consequently, while column (1) presents tests' sizes[27], the columns (2)-(7) present their power.

Analyzing column (1), we note that the five permutation tests of our Monte Carlo Experiment (θ^1-θ^5) present the correct nominal size as expected by the decision rule of Fisher's Exact Inference Procedure. The most interesting result in this column is the conservativeness of the inference procedure proposed by Conley e Taber (2011) (CT), that under-rejects the null hypothesis. This finding can be explained by the fact that, while our sample size is small

[26] We estimate the model $Y_{j,t} = \eta_1 \times \mathbb{1}\left[j = \widetilde{j}\right] + \eta_2 \times \mathbb{1}\left[j = \widetilde{j}\right] \times \mathbb{1}\left[t \geq T_0 + 1\right] + Z_{j,t} \times \boldsymbol{\zeta} + \xi_j + \mu_t + \varepsilon_{j,t}$, where ξ_j and μ_t are, respectively, region and time fixed effects, and test the null hypothesis $H_0 : \eta_2 = 0$ using the confidence intervals recommend by Conley e Taber (2011). Since their inference procedure uses only the control regions in order to estimate the test statistic distribution, the true nominal size of this test is 10.53%.

[27] Note that one possible measure of the coverage rate of our confidence set is one minus the rejection rates presented in column (1).

Tabela 1 – Monte Carlo Experiment's Rejection Rates

	Intervention Effect						
Test Statistic	(1) $\lambda = .0$	(2) $\lambda = .05$	(3) $\lambda = .1$	(4) $\lambda = .25$	(5) $\lambda = .5$	(6) $\lambda = 1.0$	(7) $\lambda = 2.0$
$\widehat{\theta}^1$	0.10	0.19	0.23	0.35	0.45	0.59	0.69
$\widehat{\theta}^2$	0.10	0.30	0.37	0.48	0.56	0.70	0.77
$\widehat{\theta}^3$	0.10	0.62	0.71	0.79	0.88	0.93	0.95
$\widehat{\theta}^4$	0.10	0.20	0.27	0.37	0.46	0.57	0.65
$\widehat{\theta}^5$	0.10	0.19	0.23	0.37	0.45	0.60	0.70
CT	0.06	0.15	0.24	0.36	0.38	0.60	0.64

Source: Authors' own elaboration. *Notes:* Each cell presents the rejection rate of the test associated to each row when the true intervention effect is given by the value λ in the columns' headings. Consequently, while column (1) presents tests' sizes, the columns (2)-(7) present their power. $\widehat{\theta}^1$-$\widehat{\theta}^3$ are associated to permutation tests that uses the Synthetic Control Estimator. $\widehat{\theta}^4$-$\widehat{\theta}^5$ are associated to permutation tests that are frequently used in the evaluation literature. CT is associated with the asymptotic inference procedure proposed by Conley e Taber (2011).

$(J + 1 = 20)$, their inference procedure is an asymptotic test based on the number of control regions going to infinity.

Analyzing the other columns, we note that the test statistic *RMSPE*, proposed by Abadie, Diamond e Hainmueller (2015) (θ^2), is uniformly more powerful than the simple test statistics (θ^4, θ^5) that are commonly used in the evaluation literature. This result suggests that, in a context where we observe only one treated unit, we should use the synthetic control estimator even if the treatment were randomly assigned. We also stress that the hypothesis test based on the statistic *RMSPE* (θ^2) outperforms the test proposed by Conley e Taber (2011) (CT) in terms of power, suggesting that, in a context with few control regions, we should use the synthetic control estimator instead of a differences-in-differences model.

We also underscore that the most powerful test statistic is the t-test, θ^3. This result makes clear the gains of power when the researcher chooses to use the synthetic control estimator instead of a simpler method, such as the difference in means (θ^4) or the permuted differences-in-differences test (θ^5). We also note that the large power of the t-test have been previously observed in contexts that are different from ours: Lehmann (1959) looks to a simple test of mean differences, Ibragimov e Muller (2010) analyzes a two-sample test of mean differences where samples' variances are different from each other, and Young (2015) focus on a linear regression coefficient.

Finally, we note that the simple average of the absolute post-intervention treatment effect (θ^1), despite using the synthetic control method, is as powerful as the simple test statistics that are commonly used in the evaluation literature (θ^4, θ^5). Consequently, we do not recommend to use it to conduct inference, because it is as time-consuming to estimate as the more powerful test statistics that uses the synthetic control method, (θ^2 and, specially, θ^3). We avoid making any

stronger suggestion about which test statistic the empirical researcher should use, because, as (EUDEY; KERR; TRUMBO, 2010, p. 14) makes clear, this choice is data dependent since the empirical researcher's goal is to match the test statistic to the meaning of the data. For example, if outliers are extremely important, θ^2 may be a better option than θ^3 even though the latter is more powerful than the former.

In appendix 1.8.2, we expand the results of this section to other test statistics.

1.5 Extensions to the Inference Procedure

1.5.1 Simultaneously Testing Hypotheses about Multiple Outcomes

Imbens e Rubin (2015) states that the validity of the procedure described in subsection 1.2.3 depends on a prior (i.e., before seeing the data) commitment to a test statistic. Moreover, Anderson (2008) shows that simultaneously testing hypotheses about a large number of outcomes can be dangerous, leading to an increase in the number of false rejections.[28] Consequently, applying the inference procedure described in subsection 1.2.3 to simultaneously test hypotheses about multiple outcomes can be misleading, because there is no clear way to choose a test statistic when there are many outcome variables and because our test's true size may be smaller than its nominal value in this context. After adapting the *familywise error rate control methodology* suggested by Anderson (2008) to our framework, we propose one way to test any *sharp null hypothesis* for a large number of outcome variables, preserving the correct test size for each variable of interest.

First, we modify the framework described in section 1.2, assuming that there are $M \in \mathbb{N}$ observed outcome variables — $\mathbf{Y}^1, ..., \mathbf{Y}^M$ — with their associated potential outcomes. We change assumptions 1-3 to:

Assumption 4. *The potential outcome vectors* $\mathbf{Y}_j^{m,I} := \left[Y_{j,1}^{m,I} ... Y_{j,T}^{m,I}\right]'$ *and* $\mathbf{Y}_j^{m,N} := \left[Y_{j,1}^{m,N} ... Y_{j,T}^{m,N}\right]'$ *for each region* $j \in \{1, ..., J+1\}$ *and each outcome variable* $m \in \{1, ..., M\}$ *do not vary based on whether other regions face the intervention or not (i.e., no spill-over effects in space) and, for each region, there are no different forms or versions of intervention (i.e., single dose treatment), which lead to different potential outcomes.*

Assumption 5. *The choice of which unit will be treated (i.e., which region is our region 1) is random* conditional on the choice of the donor pool.[29]

[28] List, Shaikh e Xu (2016) argues that false rejections can harm the economy since vast public and private resources can be misguided if agents base decisions on false discoveries. They also point that multiple hypothesis testing is a especially pernicious influence on false positives.

[29] Again, we stress that the possibility of choosing the donor pool based on observable covariates implies that assumption 5 can be interpreted as imposing only *ignorability*, a standard condition in the evaluation literature.

Assumption 6. *The potential outcomes* $\mathbf{Y}_j^{m,I} := \left[Y_{j,1}^{m,I} ... Y_{j,T}^{m,I}\right]'$ *and* $\mathbf{Y}_j^{m,N} := \left[Y_{j,1}^{m,N} ... Y_{j,T}^{m,N}\right]'$ *for each region* $j \in \{1, ..., J+1\}$ *and each outcome variable* $m \in \{1, ..., M\}$ *are fixed but* a priori *unknown quantities.*

Now, our null hypothesis is slightly more complex than the one described in equation (1.9):

$$H_0 : Y_{j,t}^{m,I} = Y_{j,t}^{m,N} + f_m(t) \tag{1.18}$$

for each region $j \in \{1, ..., J+1\}$, each time period $t \in \{1, ..., T\}$ and each outcome variable $m \in \{1, ..., M\}$, where $f_m : \{1, ..., T\} \to \mathbb{R}$ is a function of time that is specific to each outcome m. Note that we could index each function f_m by region j, but we opt not to do so because we almost never have a meaningful null hypothesis that is precise enough to specify individual intervention effects. Observe also that it is important to allow for different functions for each outcome variable because the outcome variables may have different units of measurement or different scales.

Under assumptions 4-6 and the null hypothesis (1.18), we can, for each $m \in \{1, ..., M\}$, calculate an observed test statistic, $\theta_{f_m}^{obs} = \theta^m(e_1, \tau, \mathbf{Y}^m, \mathbf{X}, f_m)$, and their associated observed p-value,

$$p_{\theta_{f_m}}^{obs} := \frac{\sum_{j=1}^{J+1} \mathbb{1}\left[\theta^m(e_j, \tau, \mathbf{Y}, \mathbf{X}, f_m) \geq \theta_{f_m}^{obs}\right]}{J+1}$$

where we choose the order of the index m to guarantee that $p_{\theta_{f_1}}^{obs} < p_{\theta_{f_2}}^{obs} < ... < p_{\theta_{f_M}}^{obs}$.

Since this p-value is itself a test statistic, we can estimate, for each outcome $m \in \{1, ..., M\}$, its empirical distribution by computing $p_{\theta_{f_m}}^{\widetilde{j}} := \frac{\sum_{j=1}^{J+1} \mathbb{1}\left[\theta^m(e_j, \tau, \mathbf{Y}, \mathbf{X}, f_m) \geq \theta^{m,\widetilde{j}}\right]}{J+1}$ for each region $\widetilde{j} \in \{1, ..., J+1\}$, where $\theta^{m,\widetilde{j}} := \theta^m\left(e_{\widetilde{j}}, \tau, \mathbf{Y}^m, \mathbf{X}, f_m\right)$. Our next step is to calculate $p_{\theta_{f_m},*}^{\widetilde{j}} := \min\left\{p_{\theta_{f_m}}^{\widetilde{j}}, p_{\theta_{f_{m+1}}}^{\widetilde{j}}, ..., p_{\theta_{f_M}}^{\widetilde{j}}\right\}$ for each $m \in \{1, ..., M\}$ and each $\widetilde{j} \in \{1, ..., J+1\}$. Then, we estimate $p_{\theta_{f_m}^{obs}}^{fwer*} := \frac{\sum_{j=1}^{J+1} \mathbb{1}\left[p_{\theta_{f_m},*}^{j} \leq p_{\theta_{f_m}}^{obs}\right]}{J+1}$ for each $m \in \{1, ..., M\}$. We enforce monotonicity one last time by computing $p_{\theta_{f_m}^{obs}}^{fwer} := \min\left\{p_{\theta_{f_m}^{obs}}^{fwer*}, p_{\theta_{f_{m+1}}^{obs}}^{fwer*}, ..., p_{\theta_{f_M}^{obs}}^{fwer*}\right\}$ for each $m \in \{1, ..., M\}$. Finally, for each outcome variable $m \in \{1, ..., M\}$, we reject the *sharp null hypothesis* (1.18) if $p_{\theta_{f_m}^{obs}}^{fwer} \leq \gamma$, where γ is a pre-specified significance level.

It is important to observe that rejecting it for some outcome variable $m \in \{1, ..., M\}$ implies that there is some region whose intervention effect differs from $f_m(t)$ for some time period $t \in \{1, ..., T\}$ for that specific outcome variable.

We also note that, when we observe only one outcome variable of interest as in section 1.2, we can reinterpret it as case with multiple outcome variables where each post-intervention time period is seem as a different outcome variable. With this interpretation, the inference procedure described in subsection 1.2.3 is still valid and is similar in flavor with the *summary index test* proposed by Anderson (2008), because we summarized the entire time information in a single test statistic. Since Anderson (2008) argues that the *summary index test*[30] has more power than the *familywise error rate control* approach, we recommend that the empirical researcher uses the inference procedure described in subsection 1.2.3 if he or she is interested in knowing whether there is an intervention effect or not, but is not interested in the timing of this effect. If the empirical researcher is interested in the timing of this effect, he or she should interpret each post-intervention time period as a different outcome variable and apply the inference procedure described in this subsection. Both approaches deliver valid statistical inference in small samples.

1.5.2 Hypothesis Testing and Confidence Sets with Multiple Treated Units

Cavallo et al. (2013) extend the Synthetic Control Method developed by Abadie e Gardeazabal (2003) and Abadie, Diamond e Hainmueller (2010) to the case when we observe multiple treated units. Firstly, we explain their innovation and, then, we clearly state the hypotheses that guarantee the validity of their method since they have not done it either. We also generalize their inference procedure in order to test any kind of *sharp null hypothesis* and, then, propose a way to estimate confidence sets for the pooled intervention effect.

Assume that there are $G \in \mathbb{N}$ similar interventions that we are interested in analyzing simultaneously. For each intervention $g \in \{1, ..., G\}$, there are $J^g + 1$ observed regions and we denote the region that faces the intervention as the first one, 1^g. Following the procedure described in subsection 1.2.1, we define the Synthetic Control Estimator of $\alpha_{1^g,t}$ as

$$\widehat{\alpha}_{1^g,t} := Y_{1^g,t} - \widehat{Y}^N_{1^g,t} \tag{1.19}$$

for each $t \in \{1, ..., T\}$ and each intervention $g \in \{1, ..., G\}$. The pooled intervention effect according to the Synthetic Control Estimator is given by:

$$\overline{\widehat{\alpha}}_{1,t} := \frac{\sum_{g=1}^{G} \widehat{\alpha}_{1^g,t}}{G} \tag{1.20}$$

for each $t \in \{1, ..., T\}$.

[30] The *summary index test* can also be adapted to our framework of multiple outcomes and be applied in place of the procedure described in this subsection. In order to do that, the researcher must aggregate all the information contained in test statistics $\theta^1, ..., \theta^M$ in a single index test statistic $\widetilde{\theta}$ and use $\widetilde{\theta}$ as the test statistic for the inference procedure described in subsection 1.2.3. In this case, a rejection of the null hypothesis implies that there is some region whose intervention effect differs from $f_m(t)$ for some time period $t \in \{1, ..., T\}$ for some specific outcome variable $m \in \{1, ..., M\}$.

In order to run hypothesis testing for each time period $t \in \{T_0+1, ..., T\}$, Cavallo et al. (2013) suggests the following procedure:

1. For each intervention $g \in \{1, ..., G\}$, permute which region is assumed to be treated and estimate, for each control region $j^g \in \{2, ..., J^g+1\}$, $\widehat{\alpha}_{j^g,t}$ as described in subsection 1.2.1.
2. Estimate a placebo pooled intervention effect as $\overline{\widehat{\alpha}}_{q,t} := \dfrac{\sum_{g=1}^{G} \widehat{\alpha}_{\widetilde{j^g},t}}{G}$, where $q \in \mathbb{N}$ indexes placebo estimations and $\widetilde{j^g} \in \{1, ..., J^g+1\}$ for each intervention $g \in \{1, ..., G\}$. Note that there are $Q := \prod_{g=1}^{G} (J^g+1)$ possible placebo pooled intervention effects.
3. Compute $p_{CGNP,t} := \dfrac{\sum_{q=1}^{Q} \mathbb{1}\left[|\overline{\widehat{\alpha}}_{q,t}| \geq |\overline{\widehat{\alpha}}_{1,t}|\right]}{Q}$ for each $t \in \{T_0+1, ..., T\}$.
4. Reject the null hypothesis if $p_{CGNP,t}$ is less than some pre-specified significance level.

We want to formalize and generalize this inference procedure. Moreover, differently from Cavallo et al. (2013), we summarize the entire time information in a single test statistic in order to avoid over-rejecting the null hypothesis as pointed out by Anderson (2008)[31]. We need to assume:

Assumption 7. *The potential outcome vectors* $\mathbf{Y}^I_{j^g} := \left[Y^I_{j^g,1} ... Y^I_{j^g,T}\right]'$ *and* $\mathbf{Y}^N_{j^g} := \left[Y^N_{j^g,1} ... Y^N_{j^g,T}\right]'$ *for each intervention* $g \in \{1, ..., G\}$ *and each region* $j^g \in \{1, ..., J^g+1\}$ *do not vary based on whether other regions face the intervention or not (i.e., no spill-over effects in space) and, for each region, there are no different forms or versions of intervention (i.e., single dose treatment), which lead to different potential outcomes.*

Assumption 8. *The choice of which unit will be treated in each intervention (i.e., which region is our region* 1^g *for each* $g \in \{1, ..., G\}$*) is random* conditional on the choice of the donor pool of each region $g \in \{1, ..., G\}$.[32]

Assumption 9. *The potential outcomes* $\mathbf{Y}^I_{j^g} := \left[Y^I_{j^g,1} ... Y^I_{j^g,T}\right]'$ *and* $\mathbf{Y}^N_{j^g} := \left[Y^N_{j^g,1} ... Y^N_{j^g,T}\right]'$ *for each intervention* $g \in \{1, ..., G\}$ *and each region* $j^g \in \{1, ..., J^g+1\}$ *are fixed but* a priori *unknown quantities.*

Finally, our *sharp null hypothesis* is now given by:

$$H_0 : Y^I_{j^g,t} = Y^N_{j^g,t} + f(t) \tag{1.21}$$

[31] For more information about over-rejecting the null hypothesis, see the articles mentioned in subsection 1.5.1.

[32] Once more, we stress that the possibility of choosing the donors pool based on observable covariates implies that assumption 8 can be interpreted as imposing only *ignorability*, a standard condition in the evaluation literature. It is very easy to see that when we interpret each intervention $g \in \{1, ..., G\}$ as a subpopulation defined by discrete predictor variables or subcategories based on continuous predictor variables.

for each intervention $g \in \{1, ..., G\}$, each region $j^g \in \{1, ..., J^g + 1\}$ and time period $t \in \{1, ..., T\}$, where $f : \{1, ..., T\} \to \mathbb{R}$. Note that we could index the function f by intervention g and region j^g, but we opt not to do so because we almost never have a meaningful null hypothesis that is precise enough to specify individual intervention effects for each observed region.

Now, we define a test statistic $\theta_{pld,f}$ as a known positive real-valued function $\theta_{pld}((\iota^g, \tau^g, \mathbf{Y}^g, \mathbf{X}^g)_{g=1}^G, .$ that summarizes the entire information of the post-intervention period.

The observed test statistic is given by $\theta_{pld,f}^{obs} := \theta_{pld}((e_{1^g}, \tau^g, \mathbf{Y}^g, \mathbf{X}^g)_{g=1}^G, f)$, where e_{j^g} is the j^g-th vector of the canonical base of $\mathbb{R}^{J^g+1}$. Under assumptions 7-9 and the *sharp null hypothesis* (1.21), we can estimate the entire empirical distribution of $\theta_{pld,f}$ by estimating $\theta_q := \theta_{pld}((e_{j^g}, \tau^g, \mathbf{Y}^g, \mathbf{X}^g)_{g=1}^G, f)$ for each possible placebo pooled intervention effect[33] $q \in \{1, ..., Q\}$. We, then, reject the null hypothesis (equation (1.21)) if

$$p_{\theta_{pld,f}} := \frac{\sum_{q=1}^{Q} \mathbb{1}\left[\theta_q \geq \theta_{pld,f}^{obs}\right]}{Q} \leq \gamma$$

where γ is some pre-specified significance level. Note that rejecting the null hypothesis implies that there is some intervention with some region whose intervention effect differs from $f(t)$ for some time period $t \in \{1, ..., T\}$.

Now, we extend our confidence sets to the pooled intervention effect. Under assumptions 7-9, we can then construct a γ-confidence set for the pooled intervention effect as

$$CS_{\gamma,\theta_{pld}} := \left\{ f \in \mathbb{R}^{\{1,...,T\}} : p_{\theta_{pld,f}} > \gamma \right\}. \tag{1.22}$$

Note that it is easy to interpret $CS_{\gamma,\theta_{pld}}$: it contains all pooled intervention effect functions whose associated *sharp null hypotheses* are not rejected by the inference procedure described in this subsection.

However, although theoretically possible to define such a general confidence set, null hypothesis (1.21) might be too general for practical reasons since the space $\mathbb{R}^{\{1,...,T\}}$ is too large to be informative and estimating such a confidence set would be computationally infeasible. For these reasons, we propose to assume the following null hypothesis:

$$H_0 : Y_{j^g,t}^I = Y_{j^g,t}^N + c \times \mathbb{1}(t \geq T_0 + 1) \tag{1.23}$$

for each intervention $g \in \{1, ..., G\}$, each region $j^g \in \{1, ..., J^g + 1\}$ and time period $t \in \{1, ..., T\}$, where $c \in \mathbb{R}$, and estimate the empirical distribution of $\theta_{pld,c}$ following the procedure

[33] Note that each possible placebo pooled intervention effect q is just a possible combination of $(e_{j^1}, ..., e_{j^G})$.

described in this subsection. Under assumptions 7-9, we can then construct a γ-confidence interval for the constant pooled intervention effect as

$$CI_{\gamma,\theta_{pld}} := \left\{c \in \mathbb{R} : p_{\theta_{pld,c}} > \gamma\right\} \subseteq CS_{\gamma,\theta_{pld}} \tag{1.24}$$

where $\gamma \in (0,1) \subset \mathbb{R}$. It is easy to interpret $CI_{\gamma,\theta_{pld}}$: it contains all constant in time pooled intervention effects whose associated *sharp null hypotheses* are not rejected by the inference procedure described in this subsection.

We can easily extend (1.23) and (1.24) to a linear in time pooled intervention effect. Assume

$$H_0 : Y^I_{j^g,t} = Y^N_{j^g,t} + \widetilde{c} \times (t - T_0) \times \mathbb{1}(t \geq T_0 + 1) \tag{1.25}$$

for each intervention $g \in \{1, ..., G\}$, each region $j^g \in \{1, ..., J^g + 1\}$ and time period $t \in \{1, ..., T\}$, where $\widetilde{c} \in \mathbb{R}$. Note that we can apply the inference procedure described above to any $\widetilde{c} \in \mathbb{R}$, estimating the empirical distribution of $\theta_{pld,\widetilde{c}}$. Under assumptions 7-9, we can then construct a γ-confidence set for the linear pooled intervention effect as

$$\begin{aligned}\widetilde{CS}_{\gamma,\theta_{pld}} &:= \left\{ f \in \mathbb{R}^{\{1,...,T\}} : \begin{array}{r} f(t) = \widetilde{c} \times (t - T_0) \times \mathbb{1}(t \geq T_0 + 1) \\ \text{and } p_{\theta_{pld,\widetilde{c}}} > \gamma \end{array} \right\} \\ &\subseteq CS_{\gamma,\theta_{pld}}\end{aligned} \tag{1.26}$$

where $\gamma \in (0,1) \subset \mathbb{R}$. It is also easy to interpret $\widetilde{CR}_{\gamma,\theta_{pld}}$: it contains all linear in time pooled intervention effects whose associated *sharp null hypotheses* are not rejected by the inference procedure described in this subsection.

Finally, if the researcher wants to analyze each intervention $g \in \{1, ..., G\}$ separately in order to investigate heterogeneous effects, he or she can apply our framework for multiple outcomes (see subsection 1.5.1) instead of implementing the pooled analysis describe in this subsection. The more detailed analysis based on the multiple outcomes framework has the cost of losing statistical power since the framework described in this subsection is based on the *summary index test* while the procedure explained in subsection 1.5.1 is based on the *familywise error rate*.[34]

1.5.3 Hypothesis Testing and Confidence Sets under Heteroskedasticity

Assumption 3 and its equivalent assumptions 6 and 9 under different contexts may be considered too strong for empirical applications since they implicitly impose homoskedasticity

[34] Anderson (2008) offers a detailed discussion about the differences between inference procedures based on the *summary index test* or on the *familywise error rate*.

among different regions. In particular, Ferman e Pinto (2016) stress that a common source of heteroskedasticity in empirical contexts is that regions with larger population sizes present smaller variances in their potential outcome values.

Fortunately, assumption 3 is stronger than needed. In order to have a valid inference procedure, we only need exchangeability between the control units and the treated unit. If we want to allow for heteroskedasticity, we can use test statistics that are robust to heterogeneity in the variance of the potential outcome values. For example, Hahn, Konietschke e Salmaso (2013) and Pauly, Brunner e Konietschke (2015) conclude that permutation tests using a t-test statistic are robust to heteroskedasticity. Moreover, Ferman e Pinto (2016) propose a modified RMSPE test statistic that is also robust to heteroskedasticity when not all pre-intervention outcome values are used to construct the synthetic control region since, in this case, its denominator is a measure of each region's variance. More generally, Canay, Romano e Shaikh (2015) argues that permutation tests are valid under heteroskedasticity if the approximate symmetry assumption holds. They also show that this assumption implies that the test statistic distribution does not depend on which unit is treated or not and that it holds for the t-statistic.[35]

Finally, we stress that this section is another example of the importance of a careful choice of test statistic as pointed out by Eudey, Kerr e Trumbo (2010). If the data present heteroskedasticity, the empirical researcher should use a test statistic that is robust to this issue.

1.6 Evaluating the Statistical Significance of the Economic Impact of ETA's Terrorism

In this section, we aim to illustrate that our inference procedure can cast new light on empirical studies that use the Synthetic Control Method. Not only we can test more flexible null hypotheses, but also we can summarize important information in a simple and effective graph. In order to achieve this goal, we use economic data for Spanish provinces made available by Abadie e Gardeazabal (2003). When they estimated the economic impact of ETA's terrorism on the Basque Country's economy, they did not discuss the statistical significance of their results because the first inference procedure for the synthetic control method would only be proposed by Abadie, Diamond e Hainmueller (2010). In order to fill this lacuna, we implement three empirical exercises:

1. We evaluate the statistical significance of the economic impact of ETA's terrorism and whether this effect can be reasonably approximated by a quadratic function using the *RMSPE* test statistic and the inference procedure described in section 1.2.3.

2. We estimate the $88.\overline{9}\%$-confidence set that contains all linear in time intervention effects (with intercept equal to zero) whose associated *sharp null hypotheses* are not rejected by our inference procedure (see equation (1.15)) when we use the *RMSPE* test statistic.

[35] We stress that, in a panel data framework, the t-test is valid if the post-intervention period is long enough and if there is no serial correlation.

3. We analyze the timing of the economic impact of ETA's terrorism using the procedure described in subsection 1.5.1.

The data set used by Abadie e Gardeazabal (2003) is available for download using the software *R*. We observe, as our outcome variable, annual real GDP per-capita in thousands of 1986 USD from 1955 to 1997 and, as covariates, biannual sector shares as a percentage of total production for agriculture, forestry and fishing, energy and water, industry, construction and engineering, marketable services and nonmarketable services from 1961 to 1969; annual shares of the working age population that was illiterate, that completed at most primary education and that completed at least secondary education from 1964 to 1969; the population density in 1969; and annual gross total investment as a proportion of GDP from 1964 to 1969. All those variables are observed at the province level and there are seventeen provinces, including the Basque Country $(J + 1 = 17)$. For historical details and descriptive statistics about this data set, see Abadie e Gardeazabal (2003) and Abadie, Diamond e Hainmueller (2011).

ETA's terrorism acts gained strength and relevance during the 70s. For this reason, our post intervention period goes from 1970 to 1997 $(T_0 = 1969)$. In order to estimate the Synthetic Control Unit, we plug, in equation (2.2), the averages of our covariates and the average of our outcome variable from 1960 to 1969. Moreover, we use data from 1960 to 1969 in equation (2.3).

When we estimate the intervention effect for the Basque Country and the placebo effect for all the other Spanish provinces, we find that the estimated intervention effect does not look abnormally large when compared to the estimated placebo effects as subfigure 1a shows.

This intuitive perception is confirmed by our formal inference procedure (see subsection 1.2.3) when we use the *RMSPE* test statistic. More specifically, we have that $p_{RMSPE} = 0.41$, implying that we can not reject the null hypothesis of no effect whatsoever.

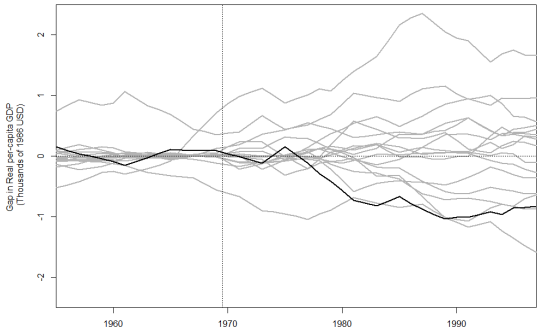
We also test whether the estimated intervention effect can be reasonably approximated by a quadratic function. In order to do that, we fit a second order polynomial to the estimated intervention effect by applying a ordinary least square estimator only in the post-intervention period. Subfigure 1b shows this fitted quadratic function. Applying our formal inference procedure and using the *RMSPE* test statistic, we do not reject the null hypothesis that the true intervention effect follows this quadratic function because $p_{RMSPE_{quadratic}} = 0.65$.

We also estimate a $88.\overline{9}\%$-Confidence Set[36] for a Linear in Time Intervention Effect whose intercept is equal to zero following equation (1.15) and using the *RMSPE* test statistic. This Confidence Set is represented in figure 2. This graph not only quickly shows that we can not reject the null hypothesis of no effect whatsoever (because the confidence set contains the

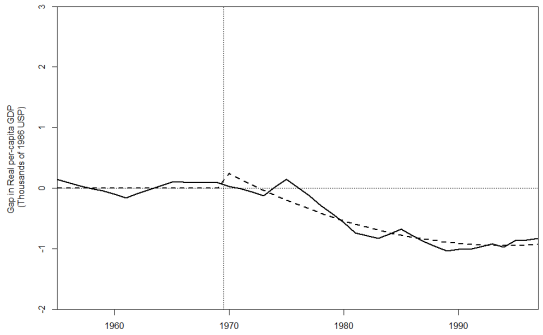
[36] Since we need at least 20 regions in order to estimate a 90%-Confidence Set, we use the possible confidence level that it is closest to 90%. Intuitively, we only reject the null hypothesis that generates one of the two largest values of the empirical distribution of the test statistic.

Figura 1 – Estimated Effects using the Synthetic Control Method

(a) Intervention and Placebo Effects



(b) Quadratic Intervention Effect Function



Note: While the gray lines show the estimated placebo effect for each Spanish province, the black lines show the estimated impact of ETA's terrorism on the Basque Country's economy and the dashed line shows the quadratic function that best approximates this effect.

linear function whose slope is equal to zero), but also shows that the economic impact of ETA's terrorism is not precisely estimated, precluding even conclusions about its true sign.[37] Due to its ability to summarize a large amount of information, our preferred confidence set (equation (1.15)) is useful to the empirical researcher even being only a subset of the general confidence set (equation (1.11)).

Figura 2 – $88.\overline{9}\%$-Confidence Set for Linear in Time Intervention Effects

Note: The black line shows the estimated impact of ETA's terrorism on the Basque Country's economy while the gray area shows the $88.\overline{9}\%$-Confidence Set for Linear in Time Intervention Effects (with intercept equal to zero) that were constructed using the *RMSPE* test statistic.

Differently from what we do in the last paragraphs, we can treat each year as a different outcome variable and apply the inference procedure described in subsection 1.5.1. This interpretation allow us to analyze the timing of the economic impact of ETA's terrorism, which may be

[37] Note that, if our estimated confidence set intersected only a small part of the positive quadrant, we could argue that the analyzed intervention effect is likely to be negative.

significant for some time periods even though we have not reject the null hypothesis of no effect whatsoever when we pooled together all the years using the *RMSPE* test statistic. We use the squared value of the estimated intervention effect for each year of the post-intervention period as a test statistic. Using the notation of subsection 1.5.1, we have that

$$\theta_{f_m}^{obs} = \theta^m(e_1, \tau, \mathbf{Y}^m, \mathbf{X}, f_m) = (\widehat{\alpha}_{1,m})^2 \,,$$

where $m \in \{1970, ..., 1997\}$ is a year of the post-intervention period.

Applying the procedure described in subsection 1.5.1, we find p-values between 0.42 and 0.88 for all years. Clearly, we can not reject the null hypothesis that ETA's terrorism has no economic impact whatsoever.

As a consequence of all our empirical exercises, we conclude that terrorists acts in the Basque Country had no statistically significant economic consequence even though the point estimate found by Abadie e Gardeazabal (2003) suggests a strong negative effect. We stress that we analyzed only the impact on GDP per-capita, ignoring possible other macroeconomic and microeconomic costs and, most importantly, social and human costs incurred by the Basque people.

1.7 Conclusion

In this article, we contribute to the theoretical literature about the Synthetic Control Method. First, we clearly state the assumptions that guarantee the validity of the inference procedure proposed by Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015) and vastly used in the empirical literature. As our main contribution, we, then, generalize this inference procedure to test any kind of *sharp null hypothesis*, allowing us to propose a new way to estimate confidence sets for the Synthetic Control Estimator by inverting a test statistic. Basically, our confidence sets contain any function of time — particularly, the constant and linear ones — whose associated *sharp null hypothesis* is not rejected by the mentioned inference procedure. To the best of our knowledge, that is the first way to estimate confidence sets for the Synthetic Control Estimator in a context with only aggregate level data whose cross-section dimension may be larger than its time dimension. We also extend our framework to the cases when there are more than one observed outcome variable or more than one treated unit. Finally, we make some brief comments about applying our generalized inference procedure when heteroskedasticity is a concern. When this issue is present, the empirical researcher should use a test statistic that is robust to this problem, such as the t-test statistic or the *RMSPE*.

The possibility to test any *sharp null hypothesis* is important to predict the future behavior of the intervention effect, to compare the costs and the benefits of a policy and to test theories that predict some specific kind of intervention effect. Constructing confidence sets is useful to summarize a large amount of information in a single graph, illustrating the statistical significance

of the intervention effect and the precision of a point-estimate. Moreover, simultaneous hypothesis testing can be used to analyze the timing of an intervention effect. Consequently, those tools not only allows the empirical researcher to be more flexible about his or her null hypothesis, but also help him or her to convey a message in a more effective way.

Since our inference procedure works for any test statistic, we analyze, using a Monte Carlo experiment, the size and power of five different test statistics that are applied to hypothesis testing in the empirical literature about the Synthetic Control Method. In this simulation, we compare test statistics that use the Synthetic Control Method to simpler test statistics that are commonly used in the evaluation literature (e.g.: difference in means and the coefficient associated with the interaction term in a differences-in-differences model) and to an asymptotic inference procedure proposed by Conley e Taber (2011). We find that test statistics that use the Synthetic Control Method perform much better than its competitors when there is only one region that faces the intervention.

As an application of our generalized inference procedure, its associated new confidence sets and its extension to the case of simultaneous hypothesis testing, we evaluate the statistical significance of the economic impact of ETA's terrorism estimated by Abadie e Gardeazabal (2003). This application clearly demonstrates the amount of information summarized by our proposed confidence sets, whose graphs quickly show not only the significance of the estimated intervention effect, but also the precision of this estimate. We stress that knowing the precision of a point-estimate is an important measure of the strength of qualitative conclusions. Finally, our empirical exercises clearly do not reject the null hypothesis of no effect whatsoever.

1.8 Appendices

1.8.1 Computing the Observed Test Statistic and its Empirical Distribution

In this appendix, we didactically explain how to compute the observed test statistic $\theta_f^{obs} := \theta(e_1, \tau, \mathbf{Y}, \mathbf{X}, f)$ and the empirical distribution of the test statistic.

First of all, we compute the observed test statistic θ_f^{obs}:

1. Simply solve the nested minimization problem given by equations (2.2) and (2.3) using the observed matrices $\mathbf{Y_1}$, $\mathbf{Y_0}$, $\mathbf{X_1}$ and $\mathbf{X_0}$.
2. Save the estimated intervention effect vector $\widehat{\boldsymbol{\alpha}}_{\mathbf{1}} := [\widehat{\alpha}_{1,1}, ..., \widehat{\alpha}_{1,T_0+1}, ..., \widehat{\alpha}_{1,T}]'$.
3. Save a vector containing the differences between the estimated intervention effect and the hypothesized intervention effect function given by the *sharp null hypothesis*, i.e., $\widetilde{\boldsymbol{\alpha}}_{\mathbf{1}} := [\widehat{\alpha}_{1,1} - f(1), ..., \widehat{\alpha}_{1,T_0+1} - f(T_0+1), ..., \widehat{\alpha}_{1,T} - f(T)]'$.
4. Aggregate the information contained in vector $\widetilde{\boldsymbol{\alpha}}_{\mathbf{1}}$ using some positive function. This function is your test statistic. For example, if we chose the test statistic *RMSPE*, we

compute

$$\theta_f^{obs} = RMSPE_f^1 = \frac{\sum_{t=T_0+1}^{T} (\widehat{\alpha}_{1,t} - f(t))^2 / (T - T_0)}{\sum_{t=1}^{T_0} (\widehat{\alpha}_{1,t} - f(t))^2 / T_0}$$

or, if you choose the first test statistic of our Monte Carlo Experiment (see section 1.4), we compute

$$\theta_f^{obs} = \frac{\sum_{t=T_0+1}^{T} |\widehat{\alpha}_{1,t} - f(t)|}{T - T_0}.$$

Now, we start our permutation test by assuming that region $\widetilde{j}$, where $\widetilde{j} \in \{2, ..., J+1\}$ is treated and estimate $\theta_f^{\widetilde{j}} := \theta(e_{\widetilde{j}}, \tau, \mathbf{Y}, \mathbf{X}, f)$:

1. Compute the counterfactual outcomes for regions 1 and $\widetilde{j}$ using the *sharp null hypothesis* (1.9).

2. Substitute those hypothesized counterfactual outcomes for the realized outcomes in matrices $\mathbf{Y_1}$ and $\mathbf{Y_0}$, saving the new matrices $\widetilde{\mathbf{Y}}_\mathbf{1}$ and $\widetilde{\mathbf{Y}}_\mathbf{0}$.

3. Change matrices $\mathbf{X_1}$ and $\mathbf{X_0}$ accordingly and save them as $\widetilde{\mathbf{X}}_\mathbf{1}$ and $\widetilde{\mathbf{X}}_\mathbf{0}$.

4. Solve the nested minimization problem given by equations (2.2) and (2.3) using the hypothesized matrices $\widetilde{\mathbf{Y}}_\mathbf{1}$, $\widetilde{\mathbf{Y}}_\mathbf{0}$, $\widetilde{\mathbf{X}}_\mathbf{1}$ and $\widetilde{\mathbf{X}}_\mathbf{0}$.

5. Save the estimated intervention effect vector $\widehat{\boldsymbol{\alpha}}_{\widetilde{\boldsymbol{j}}} := \left[\widehat{\alpha}_{\widetilde{j},1}, ..., \widehat{\alpha}_{\widetilde{j},T_0+1}, ..., \widehat{\alpha}_{\widetilde{j},T}\right]'$.

6. Save a vector containing the differences between the estimated intervention effect and the hypothesized intervention effect function given by the *sharp null hypothesis*, i.e., $\widetilde{\boldsymbol{\alpha}}_{\widetilde{\boldsymbol{j}}} := \left[\widehat{\alpha}_{\widetilde{j},1} - f(1), ..., \widehat{\alpha}_{\widetilde{j},T_0+1} - f(T_0+1), ..., \widehat{\alpha}_{\widetilde{j},T} - f(T)\right]'$.

7. Aggregate the information contained in vector $\widetilde{\boldsymbol{\alpha}}_{\widetilde{\boldsymbol{j}}}$ using some positive function. This function is your test statistic. For example, if we chose the test statistic *RMSPE*, we compute

$$\theta_f^{\widetilde{j}} = RMSPE_f^{\widetilde{j}} = \frac{\sum_{t=T_0+1}^{T} \left(\widehat{\alpha}_{\widetilde{j},t} - f(t)\right)^2 / (T - T_0)}{\sum_{t=1}^{T_0} \left(\widehat{\alpha}_{\widetilde{j},t} - f(t)\right)^2 / T_0}$$

or, if you choose the first test statistic of our Monte Carlo Experiment (see section 1.4), we compute

$$\theta_f^{\widetilde{j}} = \frac{\sum_{t=T_0+1}^{T} \left|\widehat{\alpha}_{\widetilde{j},t} - f(t)\right|}{T - T_0}.$$

Repeating this process for each $\widetilde{j} \in \{2, ..., J+1\}$, we compute the entire empirical distribution of the test statistic θ_f, given by $\left(\theta_f^{obs}, \theta_f^2, ..., \theta_f^{J+1}\right)$, and can estimate Fisher's Exact p-Value as in equation (1.10).

1.8.2 Monte Carlo Experiment's Complete Set of Results

In this appendix, we report our Monte Carlo Experiment's rejection rates for thirteen test statistics in addition to the ones shown in section 1. The new test statistics are:

- $\theta^6 := \left|mean\left(\widehat{\alpha}_{\widetilde{j},t}|t \geq T_0+1\right)\right|$ is suggested by Mideksa (2013) and used by Ando (2014).
- $\theta^7 := mean\left(\widehat{\alpha}^2_{\widetilde{j},t}|t \geq T_0+1\right)$.
- $\theta^8 := \left|median\left(\widehat{\alpha}_{\widetilde{j},t}|t \geq T_0+1\right)\right|$ is suggested by Sanso-Navarro (2011).
- $\theta^9 := median\left(\left|\widehat{\alpha}_{\widetilde{j},t}\right| |t \geq T_0+1\right)$.
- $\theta^{10} := median\left(\widehat{\alpha}^2_{\widetilde{j},t}|t \geq T_0+1\right)$.
- $\theta^{11} := min\left(\left|\widehat{\alpha}_{\widetilde{j},t}\right| |t \geq T_0+1\right)$.
- θ^{12} is the absolute value of the first (AS 1) test statistic proposed by Ando e Sävje (2013). Intuitively, it is a rescaled post-intervention time average of the estimated intervention effects.[38]
- θ^{13} is the statistic of the Kolmogorov-Smirnov Test that compares the vector of estimated post-intervention effects against a vector of zeros. This test statistic is suggested by Imbens e Rubin (2015).
- θ^{14} is the Rank statistic for region $\widetilde{j}$ using the post-intervention time average of the observed outcome. This test statistic is suggested by Imbens e Rubin (2015).
- θ^{15} is the Rank statistic for region $\widetilde{j}$ using the post-intervention time median of the observed outcome.
- θ^{16} is the Rank statistic for region $\widetilde{j}$ using the post-intervention time minimum of the observed outcome.
- θ^{17} is the Rank statistic for region $\widetilde{j}$ using the post-intervention time maximum of the observed outcome.

where $\widetilde{j}$ is the region that is assumed to face the intervention in each permutation, $mean(\clubsuit|\diamond)$ and $median(\clubsuit|\diamond)$ are, respectively, the mean and the median of variable $\clubsuit$ conditional on event $\diamond$.

[38] For details on how to calculate θ^{12}, we recommend reading Ando e Sävje (2013). These authors also propose a second test statistic (AS 2). Since it is very computationally demanding and presents similar size and power to AS 1 in Ando e Sävje (2013)'s Monte Carlo Experiment, we decided to test only AS 1 in our simulation.

Note that test statistics θ^{14}, θ^{15}, θ^{16} and θ^{17} do not use the synthetic control method and are commonly used in the literature about permutation tests. We also report rejection rates for the differences-in-differences estimator (DID)[39].

Table 2 shows the results of our Monte Carlo Experiment for all our analyzed tests. Each cell presents the rejection rate of the permutation test described in subsection 1.2.3 that uses the test statistic in each row, or the rejection rates of the tests recommend by Bertrand, Duflo e Mullainathan (2004) or proposed by Conley e Taber (2011) when the true intervention effect is given by the value mentioned in the column's heading. Consequently, while column (1) presents tests' sizes, columns (2)-(7) present their power.

Looking at the size of the tests associated with the new test statistics, there are only two interesting findings that, although not surprising according to the previous literature, are worth noting. On one hand, the Kolmogorov-Smirnov Test Statistic (θ^{13}) is associated with a conservative test due to ties in its empirical distribution. On the other hand, the rejection rate associated to a differences-in-differences model that uses standard errors clusterized at the region level is much higher than the nominal size of 10%. This last result is explained by the small number of cluster $(J + 1 = 20)$, as already pointed out by Bertrand, Duflo e Mullainathan (2004), Cameron, Gelbach e Miller (2008) and Conley e Taber (2011).

Analyzing the power of the tests associated with the new test statistics, we observe that all the test statistics that are frequently used in the evaluation literature (θ^4-θ^5 and θ^{14}-θ^{17}) are dominated by the test statistics in the main text that uses the synthetic control method (θ^2-θ^3). We also stress that the permuted t-test that uses the synthetic control method and the numerator of the t-test (θ^3 and θ^6, respectively) are the most powerful tests. Those results suggests, again, that, in a context where we observe only one treated unit, we should use the synthetic control estimator even if the treatment were randomly assigned.

Comparing the test statistics *AS 1* proposed by Ando e Sävje (2013) (θ^{12}) with the *RMSPE* recommended by Abadie, Diamond e Hainmueller (2015), we find ambiguous results. While *AS 1* is more powerful to detect small intervention effects $(\lambda \in \{0.05, 0.1\})$, *RMSPE* is more powerful to detect intermediate and large intervention effects $(\lambda \in \{0.25, 0.5, 1.0, 2.0\})$. Consequently, choosing between the two test statistics may be influenced by the magnitude of the expected intervention effect, providing another example of the importance of a careful choice of test statistic as pointed out by Eudey, Kerr e Trumbo (2010).

Finally, we note that the simple average of the squared post-intervention treatment effect (θ^7), despite using the synthetic control method, is as powerful as the simple test statistics that are commonly used in the evaluation literature (θ^4, θ^5). We also observe that test statistics θ^8-θ^{11}, that uses the synthetic control method, are as powerful as (θ^{15}), the most powerful test statistic

[39] We estimate the model $Y_{j,t} = \eta_1 \times \mathbb{1}\left[j = \widetilde{j}\right] + \eta_2 \times \mathbb{1}\left[j = \widetilde{j}\right] \times \mathbb{1}\left[t \geq T_0 + 1\right] + Z_{j,t} \times \boldsymbol{\zeta} + \xi_j + \mu_t + \varepsilon_{j,t}$, where ξ_j and μ_t are, respectively, region and time fixed effects, and test the null hypothesis $H_0 : \eta_2 = 0$ using standard errors clusterized at the region level as recommend by Bertrand, Duflo e Mullainathan (2004).

Tabela 2 – Monte Carlo Experiment's Rejection Rates

	Intervention Effect						
Test Statistic	(1) $\boldsymbol{\lambda}$= .0	(2) $\boldsymbol{\lambda}$= .05	(3) $\boldsymbol{\lambda}$= .1	(4) $\boldsymbol{\lambda}$= .25	(5) $\boldsymbol{\lambda}$= .5	(6) $\boldsymbol{\lambda}$= 1.0	(7) $\boldsymbol{\lambda}$= 2.0
Test Statistics in the Main Text							
$\widehat{\theta}^{1}$	0.10	0.19	0.23	0.35	0.45	0.59	0.69
$\widehat{\theta}^{2}$	0.10	0.30	0.37	0.48	0.56	0.70	0.77
$\widehat{\theta}^{3}$	0.10	0.62	0.71	0.79	0.88	0.93	0.95
$\widehat{\theta}^{4}$	0.10	0.20	0.27	0.37	0.46	0.57	0.65
$\widehat{\theta}^{5}$	0.10	0.19	0.23	0.37	0.45	0.60	0.70
CT	0.06	0.15	0.24	0.36	0.38	0.60	0.64
Additional Test Statistics							
$\widehat{\theta}^{6}$	0.10	0.31	0.40	0.52	0.63	0.75	0.80
$\widehat{\theta}^{7}$	0.10	0.17	0.21	0.32	0.41	0.55	0.64
$\widehat{\theta}^{8}$	0.10	0.24	0.30	0.45	0.55	0.68	0.80
$\widehat{\theta}^{9}$	0.10	0.24	0.30	0.45	0.55	0.68	0.80
$\widehat{\theta}^{10}$	0.10	0.24	0.29	0.45	0.55	0.68	0.79
$\widehat{\theta}^{11}$	0.10	0.29	0.36	0.50	0.62	0.75	0.86
$\widehat{\theta}^{12}$	0.10	0.32	0.39	0.46	0.51	0.59	0.61
$\widehat{\theta}^{13}$	0.05	0.51	0.61	0.75	0.85	0.91	0.95
$\widehat{\theta}^{14}$	0.10	0.20	0.27	0.37	0.46	0.57	0.65
$\widehat{\theta}^{15}$	0.10	0.26	0.31	0.48	0.53	0.66	0.77
$\widehat{\theta}^{16}$	0.10	0.14	0.15	0.23	0.31	0.40	0.54
$\widehat{\theta}^{17}$	0.10	0.13	0.13	0.19	0.25	0.33	0.44
DID	0.61	0.65	0.67	0.73	0.77	0.82	0.86

Source: Authors' own elaboration. *Notes:* Each cell presents the rejection rate of the test associated to each row when the true intervention effect is given by the value $\boldsymbol{\lambda}$ in the columns' headings. Consequently, while column (1) presents tests' sizes, the columns (2)-(7) present their power. $\widehat{\theta}^{1}$-$\widehat{\theta}^{3}$ and $\widehat{\theta}^{6}$-$\widehat{\theta}^{13}$ are associated to permutation tests that uses the Synthetic Control Estimator. $\widehat{\theta}^{4}$-$\widehat{\theta}^{5}$ and $\widehat{\theta}^{14}$-$\widehat{\theta}^{17}$ are associated to permutation tests that are frequently used in the evaluation literature. CT is associated with the asymptotic inference procedure proposed by Conley e Taber (2011). DID is associated with the differences-in-differences model that uses standard errors clusterized at the region level.

that does not use the synthetic control estimator. Consequently, we do not recommend to use them θ^{7}-θ^{11} to conduct inference.

1.8.3 Synthetic Control Method: A Walkthrough

In our review of the empirical literature that applied the Synthetic Control Method, we have found, on one hand, some innovative articles that proposed new, intuitive and relevant robustness checks and, on the other hand, confusing articles that were impossible to replicate due to missing information about their data. In this short walkthrough guide, we aim to not only summarize the best practices found in the literature about the Synthetic Control Method, but also

point out all the information that an author must provide in order to make his or her research replicable[40].

The first thing that must be extremely clear in a article that applies the Synthetic Control Method is the choice of the donor pool. The author must state not only how many units are in the donor pool $(J + 1)$, but how they have been chosen. In view of assumption 2, understanding the method of choice behind the donor pool is fundamental in order to evaluate whether the treatment assignment is indeed random conditional on the choice of the donor pool. This recommendation is particularly important for articles that analyze more than one intervention with different donor pools since this information must be available for all the analyzed cases.

Related to the choice of the donor pool, Abadie, Diamond e Hainmueller (2015) propose two robustness checks. The first one, the leave-one-out test, consists in dropping, from the donor pool, one of the comparison regions that received a positive weight in the synthetic control unit and reestimating the treatment effect. It aims to verify the influence of a particular unit in the estimated result. The second robustness check consists in trying different donor pool sizes in order to verify the sensibility of the estimated results to the choice of the donor pool. This last test is also applied by other authors, e.g.: Ando (2015), Barone e Mocetti (2014), Kreif et al. (2015) and Mideksa (2013).

The size and duration of the pre-intervention period (T_0) and of the entire studied period (T) must also be clearly stated. Since the main identification result about the Synthetic Control Method (ABADIE; DIAMOND; HAINMUELLER, 2010) relies on the pre-intervention period going to infinity $(T_0 \to \infty)$, the value of T_0 is important to know in order to evaluate whether the Synthetic Control Estimates are close to the true counterfactual. Moreover, the author must inform the real date associated with T_0 and T in order to access causality. In a similar way to the differences-in-differences estimator, the Synthetic Control Method relies on a time difference in order to identify the intervention effect. If there are unobservable variables that affect the outcome of interest and change at T_0, it is impossible to know whether the estimated effect is due to the analyzed treatment or due to those unobservable variables. An example will clarify this problem. Imagine that a researcher wants to investigate the effect of winning FIFA World Cup on a country's GDP. Brazil won the FIFA World Cup in 1994 and went through a fast and effective stabilization plan (*Plano Real*). If we apply the synthetic control method using other Latin American Countries as comparison units, GDP per capita as the outcome of interest and 1994 as the beginning of the treatment period $(T_0 + 1 = 1994)$, we have no way to disentangle which part of the estimated effect is due to the World Cup or to the stabilization plan.

One indirect way to test for the presence of unobservable variables that harm the interpretation of the estimated effect is to run the in-time placebo test recommend by Abadie, Diamond e Hainmueller (2015). This procedure consists in assigning the beginning of the post-treatment period to a earlier period $t^* < T_0$ and look for a treatment effect before T_0. If there is one, there

[40] Chang e Li (2015) offers a deeper discuss about replicability in Economics.

is evidence that the estimated effect is not caused by the investigated treatment.

The components of matrix $\mathbf{X}$ must also be very clear. Not only the author must state which variables are included in this matrix (e.g.: investment rate, population size, pre-intervention outcome values), but also which time periods or linear combinations of time periods are included if the predictor variables are observed in more than one time period (e.g.: investment in T_0, $T_0 - 1$, $T_0 - 2$, $T_0 - 3$ and $T_0 - 4$; population size only in the last census before T_0 and the pre-intervention time average of the outcome variable). In particular, the choice of which linear combinations of pre-intervention outcome values to include in matrix $\mathbf{X}$ is important because it may affect the significance of the estimated intervention effect as pointed out by Ferman, Pinto e Possebom (2016). Consequently, an author must report results for different specifications of matrix $\mathbf{X}$ in order to address the robustness of his or her estimates.

The researcher must also be precise about which procedure he or she have used to estimate matrix $\widehat{\mathbf{V}}$. Since there are four different procedures (ad-hoc choice, nested minimization, cross-validated nested minimization, regression based method), the author could report results for some of them in order to access the robustness of his or her estimates. Since the cross-validated nested minimization requires a larger pre-intervention period, it may not be reasonable to report results for all the four methods. However, it is always possible to report results for the nested minimization, the regression based method and for the ad-hoc choice that imposes $\widehat{\mathbf{V}} = \mathbf{I}$.

Finally, Abadie, Diamond e Hainmueller (2015) proposes one last robustness check: the restricted synthetic control unit. In order to avoid over-fitting, this test force the synthetic control method to allocate positive weights only to a fixed number of control units, mimicking the n nearest neighbors matching estimator.

2 Free Trade Zone of Manaus: An Impact Evaluation using the Synthetic Control Method

2.1 Introduction

In this article, we aim to estimate the effect of the enterprise zone known as *Free Trade Zone of Manaus* (FTZM)[1] on the city's real GDP per capita, Agriculture Total Production per capita, Manufacture Total Production per capita and Services Total Production per capita. In order to achieve this goal, we estimate the synthetic control unit of Manaus, using only other cities in the Brazilian North Region as control units, to approximate the counterfactual that would have happened if the enterprise zone had not been created. We conduct inference as explained by Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015).

The effects of the FTZM on Manaus' economy are ambiguous. There is no evidence of impact on its Manufacture Total Production per capita, suggesting that this subsidy policy failed to achieve its main goal. There is a significant at the 10% level *negative* effect on Agriculture Total Production per capita, an evidence that this subsidy policy may have distorted incentives, artificially reducing the primary sector due to mis-allocated resources. There is also a significant at the 5% level positive impact on Services Total Production per capita, an indicator that the main impact of the FTZM was due to lower import tariffs than the ones imposed in the rest of the country, attracting Brazilians tourists willing to buy imported goods for lower prices. Finally, the impact on real GDP per capita was positive and significant at the 10% level, suggesting that this subsidy policy have been successful in developing the city's economy.

Our most important and direct contribution is an evaluation of a costly and historical investment decision in Brazil related to the import substitution industrialization policy. Since FTZM's costs are hard to measure (particularly, during the beginning of its implementation), we can not state whether significant benefits found by us surpass its costs or not.[2] Making this comparison even harder, a large amount of FTZM's costs are paid by other cities or states instead of being paid by Manaus.[3] This phenomenon allows us to tentatively conclude that FTZM is likely to pass a cost-benefit analysis when we consider only the city of Manaus, but we can not draw any conclusion about its efficacy when we consider the entire country.

Moreover, recent cost-benefit analysis of FTZM have focused on the 21st Century and adopted descriptive evaluation tools. Different from those works, we focus on the historical

[1] This enterprise zone was a set of subsidies to manufacture production in the city of Manaus, located in the Brazilian Amazon Region.

[2] **??**) points out that a real cost-benefit study of any *special economic zone*, such as FTZM, requires exhaustive data on numerous parameters that are extremely hard to collect rigorously. Castilho, Meneédez e Sztulman (2015) argues that this type of study is even more challenging to implement in the case of FTZM due to a complex and evolving system of incentives and taxes.

[3] See Miranda (2013).

evolution of FTZM (20th Century) and adopt causal inference tools, allowing us to draw conclusions about FTZM's causal effects. This methodology and the conclusions that we have reached about this subsidy policy's effects also allow us to make a small and indirect contribution to the literature about mis-allocation of resources, since FTZM may distort investment incentives among sectors and regions.

The literature about FTZM's benefits and costs is ambiguous. On one hand, Oliveira e Souza (2012), descriptively analyzing data for the beginning of the 21st Century, conclude that FTZM's fiscal costs are lower than its socio-economic benefits. Although they argue that FTZM has increased Manaus' inequality, they found that it has increased Manaus' GDP and Human Development Index. In particular, they stress that the reduced import tariffs for Manaus attracted a large number of Brazilian tourists willing to buy imported goods for lower prices, stimulating the services sector. On the other hand, Miranda (2013) also descriptively analyzes data for the beginning of the 21st Century, but reaches completely different conclusions. He argues that FTZM's impact on the regional development level is modest and that the jobs created by this subsidy policy are low quality positions. Consequently, those benefits do not surpass this policy's elevated fiscal costs, even considering that he focus only on federal fiscal costs since state and municipal fiscal costs are extremely hard to measure. Finally, he concludes that the maintenance of FTZM will impose permanent costs to the public sector.

Regarding the FTZM's main goal (i.e., to estimulate the industrial sector in the Amazon region), Sá e Machado (2012) find that the FTZM presents higher rates of added value than Brazil as a whole between 2006 and 2010. He achieves this results by descriptively analyzing data from 1996 to 2010. Another interesting economic effect provoked by the FTZM is studied by Castilho, Meneédez e Sztulman (2015). They apply micro-decomposition techniques and find that, while labor income was a major driver of poverty and inequality declines for the municipality of Manaus in the 2000-2010 decade, non-labor income was far more important in the rest of the state of Amazonas.

The international literature about other enterprise zones around the world is also ambiguous. On one hand, Ham et al. (2011) and Busso, Gregory e Kline (2013) present a positive perspective on this type of policy. Ham et al. (2011) analyze, in the United States of America (USA), the economic impact of State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Community programs. Using a differences-in-differences approach, they found positive and significant effects on the unemployment rate, the poverty rate, the fraction with wage and salary income and the employment level, concluding that the enterprise zones are effective regarding their goal of boosting local labor markets. Busso, Gregory e Kline (2013) also analyze the Federal Empowerment Zones program using a differences-in-differences estimator. They find that this policy substantially increased local employment rate and wages without increases in the local population or cost of living. Consequently, they conclude that this policy's efficiency costs

are relatively modest. On the other hand, Gobillon, Magnac e Selod (2012) studies a French enterprise zone policy and, using a combination of a hazard model, a matching estimator and a differences-in-differences method, they find only a small short-run effect on the rate at which unemployed workers find a job. Based on this result, they conclude that this enterprise zone policy is likely to be cost-ineffective. A similar finding for the same enterprise zone policy is reached by **??**) using an interactive effect model.

The structure of this essay is simple and straightforward. I explain the synthetic control estimator in the next section and briefly describe the institutions that govern FTZM and my data in the third section. In the fourth section, I describe my main results and a robustness check, while, in the last section, I conclude and discuss possible future work.

2.2 Synthetic Control Estimator

Abadie, Diamond e Hainmueller (2010) propose the following model to estimate the impact of a treatment when only one unit is treated.

Suppose that we observe data for $(J+1) \in \mathbb{N}$ units during $T \in \mathbf{N}$ time periods. Additionally, assume that there is a treatment that affects only unit 1 from period T_0+1 to period T uninterruptedly, where $1 \leq T_0 < T$ is a natural number. Let $Y_{j,t}^N$ be the potential outcome that would be observed for unit j in period t if there were no treatment for $j \in \{1, ..., J+1\}$ and $t \in \{1, ..., T\}$. Let $Y_{j,t}^I$ be the potential outcome that would be observed for unit j in period t if unit j received the treatment from period T_0+1 to T. Define

$$\alpha_{j,t} = Y_{j,t}^I - Y_{j,t}^N \tag{2.1}$$

as the effect of the treatment for unit j in period t and D_{jt} as a dummy variable that assumes the value of 1 if unit j receives the treatment in period t and value 0 otherwise. With this notation, we have that the observed outcome for unit j in period t is given by

$$Y_{j,t} = Y_{j,t}^N + \alpha_{jt} D_{j,t}.$$

Since only the first unit receives the treatment from period T_0+1 to T, we have that:

$$D_{j,t} = \begin{cases} 1 & \text{if } j = 1 \text{ e } t > T_0, \\ 0 & \text{otherwise.} \end{cases}$$

We aim to estimate $(\alpha_{1,T_0+1}, ..., \alpha_{1,T})$. Since $Y_{1,t}^I$ is observable for $t > T_0$, equation (2.1) guarantees that we only need to estimate $Y_{1,t}^N$ to accomplish this goal.

Let $\mathbf{Y_1} = [Y_{1,1}, ..., Y_{1,T_0}]'$ be the vector of observed outcomes for unit 1 in the pre-treatment period and $\mathbf{X_1}$ a $(K \times 1)$-vector of predictors of $\mathbf{Y_1}$. Let $\mathbf{Y_0}$ be a $(T_0 \times J)$-matrix, whose $(j-1)$-th column is given by $\mathbf{Y_j} = [Y_{j,1}, ..., Y_{j,T_0}]'$ for each $j \in \{2, ..., J+1\}$, and $\mathbf{X_0}$ is

a $(K \times J)$-matrix that contains the values of the same K predictors for the J comparison units.[4] Define the weighting vector $\mathbf{W} = [w_2 \ldots w_{J+1}]'$ of $(J \times 1) - dimension$, where $w_j \geq 0$ for each $j \in \{2, ..., J+1\}$ and $\sum_{j=2}^{J+1} w_j = 1$. Intuitively, $\mathbf{W}$ measures the relative importance of each control unit in the synthetic control of unit 1. Moreover, define a positive semidefinite diagonal weighting matrix $\mathbf{V}$ of $(K \times K)$-dimension. Intuitively, $\mathbf{V}$ measures the relative importance of each one of the K predictors.

Since we want to make unit 1's synthetic control as similar as possible to the actual unit 1, we choose $\widehat{\mathbf{W}}(\mathbf{V})$ such that

$$\widehat{\mathbf{W}}(\mathbf{V}) := \arg\min_{\mathbf{W} \in \mathcal{W}} (\mathbf{X_1} - \mathbf{X_0}\mathbf{W})'\mathbf{V}(\mathbf{X_1} - \mathbf{X_0}\mathbf{W}) \tag{2.2}$$

where $\mathcal{W} = \{\mathbf{W} = [w_2 \ldots w_{J+1}]' \in \mathbb{R}^J : w_j \geq 0$ for each $j \in \{2, ..., J+1\}$ and $\sum_{j=2}^{J+1} w_j = 1\}$.

Our notation makes clear that the correspondence $\widehat{\mathbf{W}}(\mathbf{V})$ for problem (2.2) depends on $\mathbf{V}$. Abadie, Diamond e Hainmueller (2010) propose to use $\widehat{\mathbf{V}}$ such that

$$\widehat{\mathbf{V}} := \arg\min_{\mathbf{V} \in \mathcal{V}} (\mathbf{Y_1} - \mathbf{Y_0}\widehat{\mathbf{W}}(\mathbf{V}))'(\mathbf{Y_1} - \mathbf{Y_0}\widehat{\mathbf{W}}(\mathbf{V})) \tag{2.3}$$

where $\mathcal{V}$ is the set of positive semidefinite diagonal matrices of dimension $(K \times K)$. Intuitively, this technique makes the synthetic control of unit 1 as similar as possible with the actual unit 1 during the pre-treatment period when we choose the Euclidian metric to evaluate the distance between the observed outcomes for unit 1 and the values predicted by the synthetic control.

Chosen $\widehat{\mathbf{V}}$, the synthetic control weights of unit 1 are given by

$$\widehat{\mathbf{W}} := \widehat{\mathbf{W}}(\widehat{\mathbf{V}}) = [\widehat{w_2} \ldots \widehat{w_{J+1}}]'.$$

For each $t \in \{T_0 + 1, ..., T\}$, the estimator of $Y_{1,t}^N$ according to the synthetic control method is given by

$$\widehat{Y_{1,t}^N} = \sum_{j=2}^{J+1} \widehat{w_j} Y_{j,t}.$$

Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015) propose the following inference procedure[5].

In order to have an idea of how unlikely our estimator $\widehat{\alpha_{1,t}} = Y_{1,t} - \widehat{Y_{1,t}^N}$ for $t \geq T_0 + 1$ under the null of no effect is, Abadie, Diamond e Hainmueller (2010) propose to run placebo tests: assume that each control unit $j \in \{2, ..., J+1\}$ had received the treatment, estimate its synthetic counterfactual and compare the $\widehat{\alpha_{1,t}}$ with $\widehat{\alpha_{j,t}}$ for each $t \geq T_0 + 1$ and $j \in \{2, ..., J+1\}$. If $|\widehat{\alpha_{1,t}}|$ is abnormally large, there is some evidence to reject the null hypothesis.

[4] Some lines of matrix $\mathbf{X_1}$ and $\mathbf{X_0}$ can be linear combinations of the variables in $\mathbf{Y_1}$ e $\mathbf{Y_0}$.

[5] For a generalization of this methodology, see **??**).

One drawback of the last method is that $|\widehat{\alpha_{1,t}}|$ can be abnormally large for some time periods, but not for others. In order to handle this issue, Abadie, Diamond e Hainmueller (2015) propose a way to condensate information about all periods in order to conduct inference: estimate

$$RMSPE_j := \frac{\sum_{t=T_0+1}^{T} \left(Y_{j,t} - \widehat{Y_{j,t}^N}\right)^2 / (T - T_0)}{\sum_{t=1}^{T_0} \left(Y_{j,t} - \widehat{Y_{j,t}^N}\right)^2 / T_0} \tag{2.4}$$

for each $j \in \{1, ..., J+1\}$, compute

$$p := \frac{\sum_{j=1}^{J+1} \mathbb{1}\left[RMSPE_j \geq RMSPE_1\right]}{J+1}$$

and reject the null of no effect if p is less than some pre-specified significance level.

Methodologically, synthetic control estimator's main competitor is the differences-in-differences estimator. The latter option was previously applied to the estimation of enterprise zones' economic impact as our brief review of the literature makes clear. However, applying it to our dataset would be problematic, since its inference procedure requires a large number of treated and control units and our comparison group is relatively small.[6] Moreover, this methodology restricts the treatment effect to be constant during the post-treatment period, a very strong assumption in our context since our post-treatment period lasts for almost half a century.

Those two issues are successfully tackled by the synthetic control method. First of all, its inference procedure is completely appropriate for a small sample such as ours.[7] Moreover and most importantly, this method allows us to estimate a completely flexible treatment effect as a function of time.

2.3 Institutional and Data Description

The enterprise zone FTZM was created in February, 28th, 1967 as a subsidy policy to promote regional growth by giving incentives to the manufacture sector to substitute imports according to Decree Law n. 288. This policy reduced the tariffs to import inputs to produce industrialized goods only in the area close to the city of Manaus, aiming to increase the local production of final goods and decrease its imports in the entire country. Although the details of the law that regulated the ZFM changed through the years, FTZM's essence remained unchanged: a heavy subsidy policy to the manufacture sector in Manaus[8]. The service sector was indirectly subsidized through reduced import tariffs, because Manaus' retailers could sell imported goods for lower prices, attracting many Brazilian tourists as stressed by Oliveira e Souza (2012).

[6] Even if we applied the inference procedure developed by Conley e Taber (2011), we would still need a large number of control units.

[7] For more information about its inference procedure, see Firpo e Possebom (2016).

[8] See Oliveira e Souza (2012) and Miranda (2013) for details about the creation and implementation of FTZM. **??**) provides a critical historical overview about the FTZM.

Since FTZM is a subsidy policy to the Manufacture and the Services sectors in the city of Manaus, theory predicts that its GDP per capita level would increase in expense of other cities' income levels (mis-allocation among regions) and that its Manufacture Total Production per capita and Services Total Production per capita would increase in expense of other sectors' production (mis-allocation among sectors). Another possible explanation for an increase in Manaus' per capita GDP is that a stronger manufacture sector generates positive spill-overs due to its working dynamics, increasing total production of all economic sectors. We aim to evaluate those predictions by analyzing city-level data for economic variables during the 20th Century.

We collect data on GDP per capita, Agriculture, Manufacture and Services Total Production per capita[9], Agriculture, Manufacture, Services and Government Production as shares of GDP, and Population Density for 49 Minimum Comparable Areas (MCAs) in the Brazilian North Region for the years of 1920, 1939, 1949, 1959, 1970, 1975, 1980, 1985, 1996 and 1999.[10] I define the pre-treatment period as 1920-1959 and the treated region is the MCA 1920 #2097002[11], where the city of Manaus is currently located. I apply the methodology proposed by Abadie e Gardeazabal (2003), Abadie, Diamond e Hainmueller (2010) and Abadie, Diamond e Hainmueller (2015) to construct synthetic control units of Manaus for the variables GDP per capita, Agriculture Total Production per capita, Manufacture Total Production per capita and Services Total Production per capita, using their own pre-intervention mean values, sectoral GDP shares, government size and population density as predictors.

Table 3 reports the pre-treatment means of the outcome variables and the predictor variables for Manaus (column (1)) and our four Synthetic Versions of Manaus (columns (2)-(5)). It also reports the sample average (excluding Manaus) of those variables for each one of our samples (columns (6)-(9)). As expected, the synthetic units reproduce the values for the city of Manaus more precisely than a simple average.

Since Manaus is one of the least densely populated cities in the Brazilian North region, it is extremely hard for the synthetic control method to fit this predictor variable. Consequently, this methodology imposes a weight close to zero for this variable when constructing the synthetic units for all the outcome variables but Agriculture Total Production per capita.

We also note that Manaus presents the largest Manufacture Total Production per capita in the Brazilian North region, making very hard for the synthetic control method to fit this outcome variable. Despite this fact, our estimator reproduces closely the city of Manaus for this variable of interest even considering it assigns an weight of only 13.7% to it.

[9] The last four variables are in real values (Reais of 2000).

[10] Due to missing data, we have information about 46 MCAs for Agriculture Total Production per capita, 22 MCAs for Manufacture Total Production per capita and 40 MCAs for Services Total Production per capita.

[11] In order to make the text easier to read, I refer to this area simply as "Manaus", although it encompass an area that is larger than the current city of Manaus.

Tabela 3 – Descriptive Statistics

		Synthetic Manaus				Sample Average			
Variable	**Manaus** (1)	**GDP pc** (2)	**ATP pc** (3)	**MTP pc** (4)	**STP pc** (5)	**GDP pc** (6)	**ATP pc** (7)	**MTP pc** (8)	**STP pc** (9)
GDP pc (R$ of 2000)	1213.64	1213.49				470.96			
ATP pc (R$ of 2000)	193.56		191.01				248.43		
MTP pc (R$ of 2000)	349.35			279.71				63.87	
STP pc (R$ of 2000)	670.74				670.58				196.34
Population Density (Inhabitants per sq km)	0.60	32.96	0.73	36.57	24.02	3.76	3.49	5.46	4.35
Agriculture Share (%)	18.3	18.4	27.0	19.1	18.4	57.8	57.9	46.6	55.0
Manufacture Share (%)	22.6	21.3	16.6	22.5	22.7	9.6	9.5	11.1	8.9
Services Share (%)	59.0	61.3	56.9	58.3	59.1	35.6	35.3	42.7	37.3
Government Size (%)	9.5	9.9	9.8	10.1	9.7	6.7	6.4	7.5	7.1
Sample Size						49	46	22	40

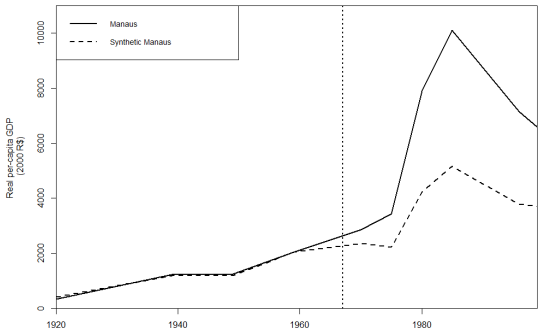
Notes: *XTP pc* and *X Share* stand for *Sector X Total Production per capita* and *Sector X Production as a share of GDP*, respectively. Since, for each outcome variable, we estimate a different Synthetic Unit using a different sample, we report descriptive statistics for each one of the four synthetic units and each one of the four samples. The different sample sizes are due to missing data.

2.4 Results

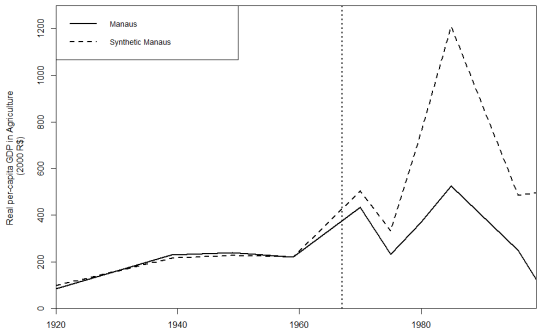
Figure 3 plots the time pattern of Manaus and its synthetic control versions for the variables real GDP per capita (subfigure 3a), Agriculture Total Production per capita (subfigure 3b), Manufacture Total Production per capita (subfigure 3c) and Services Total Production per capita (subfigure 3d). As those graphs show, FTZM seems to have a positive effect on real GDP per capital, Manufacture Total Production per capita and Services Total Production per capita, and a negative impact on Agriculture Total Production per capita. These results are supporting evidence of the theory about mis-allocation of resources as we discussed in the last section, contradicting the predictions of the positive spill-overs view. Moreover, as it is easy to see, those point estimates suggest that FTZM presents very large economic impacts.

Figura 3 – Estimated Effects using the Synthetic Control Method

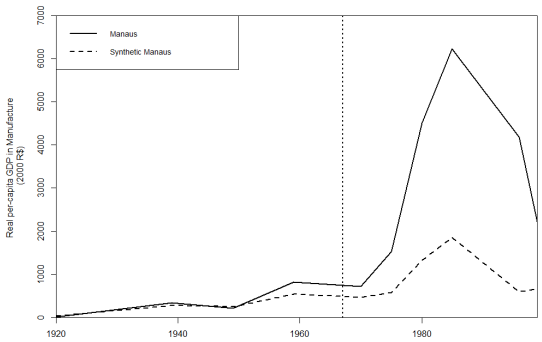
(a) Real GDP per capita



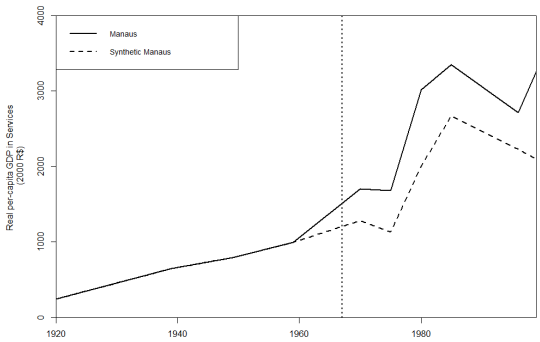
(b) Agriculture Total Production per capita



(c) Manufacture Total Production per capita



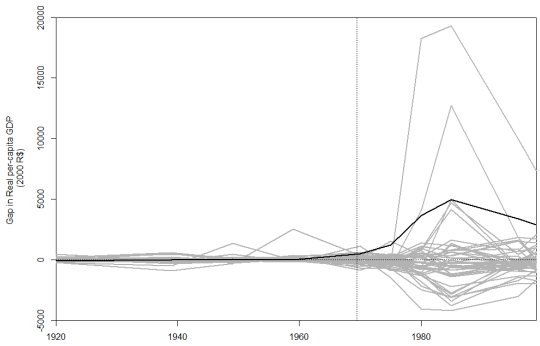
(d) Services Total Production per capita



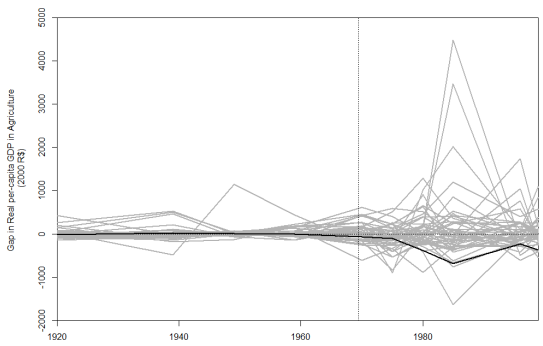
Now, we pay attention only to the black lines on subfigures 4a, 4b, 4c and 4d. They show FTZM's estimated economic impact, since they represent the differences between the black and the dotted lines on subfigures 3a, 3b, 3c and 3d. As one can see, FTZM's economic effect on GDP per capita and Agriculture and Manufacture Total Production per capita presents a parabolic form, while its impact on Services Total Production per capital increases at the beginning of the post-treatment period and, then, stabilizes. These results illustrates the importance of using the Synthetic Control Method instead of the differences-in-differences estimator, since the former allows us to estimate flexible treatment effects.

Figura 4 – Placebo Tests

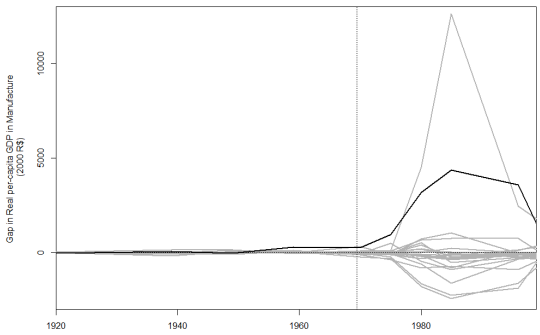
(a) Real GDP per capita



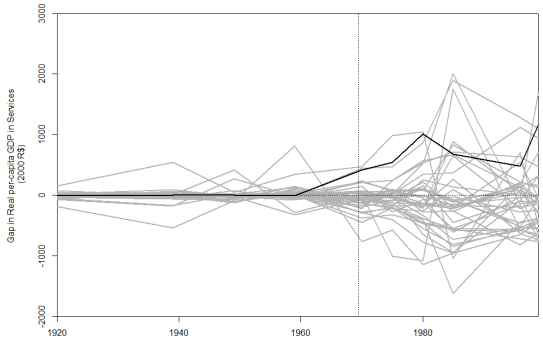
(b) Agriculture Total Production per capita



(c) Manufacture Total Production per capita



(d) Services Total Production per capita



Note: While the gray lines show the estimated placebo effect for each outcome variable and for each control city, the black lines show the estimated impact of FTZM on Manaus' economy for each outcome variable.

At this moment, we have only presented point estimates and have not discussed the statistical significance of our results. In order to address this point, we follow Abadie, Diamond e Hainmueller (2010) and plot placebo tests for each analyzed variable (see figure 4). Looking at subfigures 4a, 4b, 4c and 4d, it seems that FTZM had a significant impact only on Manaus' real GDP per capita and Manufacture Total Production per capita. However, this impression is contradicted when we apply the formal inference procedure described by Abadie, Diamond e Hainmueller (2015) and use the RMSPE as a test statistic.[12]

We find that FTZM's economic effect presents a p-value of 6.12% on real GDP per capita, 6.52% on Agriculture Total Production per capita, 63.64% on Manufacture Total Production per capita and 2.50% on Services Total Production per capita. The lack of significance for the results regarding the manufacture sector is likely due to a larger pre-intervention mean squared prediction error and suggests that this policy failed to achieve its main goal. However, the three significant results suggest that the positive effects on the service sector were more than enough to compensate for the negative impacts on the agricultural sector, implying that FTZM had a

[12] Firpo e Possebom (2016) discuss this inference procedure, formalizing and generalizing it.

positive effect on Manaus' economy as a whole.

These findings allow us to conclude that FTZM is likely to pass a cost-benefit analysis when we consider only the city of Manaus. Although tentative, this conclusion is likely to be valid because most of FTZM's costs were paid by cities outside the Amazon region. However, we can not draw any conclusion about FTZM's efficacy when we consider the entire country without collecting data about its fiscal costs during the 20$^{\text{th}}$ Century.

As a cautionary tale, we stress that part of the FTZM's positive impacts on Manaus' Real GDP per capita may be due simply to mis-allocation of resources among cities in the North Region, i.e., investments that would have been made in the control cities were made in Manaus due to FTZM. If this is the case, FTZM's effect for the country as a whole may be negative.

Robustness Check

As a robustness check, we run a standard differences-in-differences regression

$$y_{it} = \theta D_{it} + \gamma T_i + \boldsymbol{x}_{it}\beta + \alpha_i + \delta_t + \epsilon_{it} \tag{2.5}$$

where i and t respectively index minimum comparable areas and time periods; y_{it} represents real GDP per capita or Sectoral Total Production per capita levels; D_{it} is a dummy variable that assumes the value of 1 only for Manaus during the post-treatment period; T_i is a dummy variable that assumes the value of 1 only for Manaus; $\boldsymbol{x}_{it}$ is a row vector of control variables that contains Agriculture, Manufacture, Services and Government Production as shares of GDP, and Population Density; α_i is a city fixed effect, δ_t is a time fixed effect and ϵ_{it} is a error term. Moreover, we estimate a 90%-confidence interval for the coefficient of interest, θ, using the procedure proposed by Conley e Taber (2011).

Table 4 reports the estimated results for model (2.5) using the inference method suggested by Conley e Taber (2011). Regarding the point estimates, the magnitudes of the coefficient of interest for each dependent variable are similar to the average estimated effects of the synthetic control method for each analyzed outcome, illustrating the robustness of our findings. We also note that the confidence intervals for the variables *real GDP per capita*, *Manufacture Total Production per capita* and *Services Total Production per capita* are extremely wide, indicating that our estimates are imprecise even if when they are statistically significant as it is the case for the coefficient associated with the dependent variable *Manufacture Total Production per capita*. The only coefficient that is precisely estimated is the one associated with the dependent variable *Agriculture Total Production per capita*. In this case, our point estimate is negative and statistically significant, in accordance with our Synthetic Control results.

Tabela 4 – Differences-in-Differences Results

Dependent Variable	**GDP pc**	**ATP pc**	**MTP pc**	**STP pc**
	(1)	(2)	(3)	(4)
Point Estimate $\left(\widehat{\theta}\right)$	2050	-532	1276	1066
Confidence Intervals	[-265, 3265]	[-809, -289]	[419, 1798]	[-106, 1606]
Sample Size	49	46	22	40

Notes: *XTP pc* stands for *Sector X Total Production per capita*. We report point estimates for the coefficient of interest, θ, in model (2.5) and its 90%-confidence interval based on the inference procedure proposed by Conley e Taber (2011). The different sample sizes are due to missing data.

2.5 Conclusion

Applying the Synthetic Control Method to city-level data for cities of the Brazilian North region during the 20$^{\text{th}}$ century, we evaluate the economic impact of the Free Trade Zone of Manaus. While FTZM's impact on Manaus' Manufacture Total Production per capita is non significant, its effect on the agriculture sector and the services sector are significant and present different signs — negative for the former and positive for the latter. At the end, FTZM's positive and significant effect on Manaus' real GDP per capita suggests that its positive impacts were larger than the negative ones, implying that this subsidy policy achieved its goal of promoting regional economic growth.

Although it is possible to think that FTZM would pass a cost-benefit analysis when we consider only the city of Manaus, we do not have enough information to draw any conclusion about the relative magnitude of its benefits and costs when we consider Brazil as a whole. This tentative conclusion is due to the fact that a large amount of FTZM's costs is paid not by Manaus, but by cities outside the Amazon region.

Moreover, our methodology does allow us to disentangle whether the positive impact on Manaus' real GDP per capita is due to positive spillovers that benefit all the city economic sectors and the entire country or due to mis-allocation of resources among cities that harms Brazil as a whole. Since the positive spill-overs view predicts a positive impact of the FTZM on Agriculture Total Production per capita, our results contradicts this theory. In reality, our estimated negative impact on the agriculture sector is in accordance with the mis-allocation literature.

References

ABADIE, A.; DIAMOND, A.; HAINMUELLER, J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statiscal Association*, v. 105, n. 490, p. 493–505, 2010.

ABADIE, A.; DIAMOND, A.; HAINMUELLER, J. Synth: An R Package for Synthetic Control Methods in Comparative Case Studies. *Journal of Statistical Software*, v. 42, n. 13, p. 1–17, 2011.

ABADIE, A.; DIAMOND, A.; HAINMUELLER, J. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, Wiley Online Library, v. 59, n. 2, p. 495–510, 2015.

ABADIE, A.; GARDEAZABAL, J. The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, v. 93, n. 1, p. 113–132, 2003.

ACEMOGLU, D. et al. The Value of Connections in Turbulent Times: Evidence from the United States. NBER Working Paper 19701. Available at: <http://www.nber.org/papers/w19701.pdf>. 2013.

ANDERSON, M. L. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool and Early Training Projects. *Journal of the American Statiscal Association*, v. 103, n. 484, p. 1481–1495, December 2008.

ANDO, M. Dreams of Urbanization: Quantitative Case Studies on the Local Impacts of Nuclear Power Facilities using the Synthetic Control Method. Working Paper, <https://75673a0a-a-62cb3a1a-s-sites.googlegroups.com/site/michihito7ando/Nuclear_20140623.pdf>. 2014.

ANDO, M. Dreams of Urbanization: Quantitative Case Studies on the Local Impacts of Nuclear Power Facilities using the Synthetic Control Method. *Journal of Urban Economics*, v. 85, p. 68–85, June 2015.

ANDO, M.; SäVJE, F. Hypothesis Testing with the Synthetic Control Method. Working Paper, <http://www.eea-esem.com/files/papers/eea-esem/2013/2549/scm.pdf>. 2013.

BARONE, G.; MOCETTI, S. Natural Disasters, Growth and Institutions: a Tale of Two Earthquakes. *Journal of Urban Economics*, p. 52–66, 2014.

BAUHOFF, S. The Effect of School Nutrition Policies on Dietary Intake and Overweight: a Synthetic Control Approach. *Economics and Human Biology*, p. 45–55, 2014.

BELOT, M.; VANDENBERGHE, V. Evaluating the Threat Effects of Grade Repetition: Exploiting the 2001 Reform by the French-Speaking Community of Belgium. *Education Economics*, v. 22, n. 1, p. 73–89, 2014.

BERTRAND, M.; DUFLO, E.; MULLAINATHAN, S. How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, v. 119, n. 1, p. 249–275, 2004.

BILLMEIER, A.; NANNICINI, T. Trade Openness and Growth: Pursuing Empirical Glasnost. *IMF Staff Papers*, v. 56, n. 3, p. 447–475, 2009.

BILLMEIER, A.; NANNICINI, T. Assessing Economic Liberalization Episodes: A Synthetic Control Approach. *The Review of Economics and Statistics*, v. 95, n. 3, p. 983–1001, 2013.

BOHN, S.; LOFSTROM, M.; RAPHAEL, S. Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population? *The Review of Economics and Statistics*, v. 96, n. 2, p. 258–269, 2014.

BOVE, V.; ELIA, L.; SMITH, R. P. The Relationship between Panel and Synthetic Control Estimators on the Effect of Civil War. Working Paper, <http://www.bbk.ac.uk/ems/research/BirkCAM/working-papers/BCAM1406.pdf>. 2014.

BUSSO, M.; GREGORY, J.; KLINE, P. Assessing the Incidence and Efficiency of a Proeminent Place Based Policy. *American Economic Review*, v. 103, n. 2, p. 897–947, 2013.

CALDERON, G. The Effects of Child Care Provision in Mexico. Working paper, <http://goo.gl/YSEs9B>. 2014.

CAMERON, A. C.; GELBACH, J. B.; MILLER, D. L. Bootstrap-based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, v. 90, n. 3, p. 414–427, 2008.

CANAY, I. A.; ROMANO, J. P.; SHAIKH, A. M. Randomization Tests under an Approximate Symmetry Assumption. Working Paper, <http://goo.gl/zCvu08>. 2015.

CARRASCO, V.; MELLO, J. M. P. de; DUARTE, I. A Década Perdida: 2003 – 2012. Texto para Discussão, <http://www.econ.puc-rio.br/uploads/adm/trabalhos/files/td626.pdf>. 2014.

CARVALHO, C. V.; MANSINI, R.; MEDEIROS, M. C. ArCo: An Artificial Counterfactual Approach for Aggregate Data. Working Paper. 2015.

CASTILHO, M.; MENEéDEZ, M.; SZTULMAN, A. Poverty and inequality dynamics in manaus: Legacy of a free trade zone? Working Paper n. 43. Available at <https://halshs.archives-ouvertes.fr/halshs-01245394>. 2015.

CAVALLO, E. et al. Catastrophic Natural Disasters and Economic Growth. *The Review of Economics and Statistics*, v. 95, n. 5, p. 1549–1561, 2013.

CHAN, H. F. et al. Academic Honors and Performance. *Labour Economics*, v. 31, p. 188–204, 2014.

CHANG, A.; LI, P. Is Economic Research Replicable? Sixty Published Papers from Thirteen Journal Say "Usually Not

. Finance and Economics Discussion Series 2015-083. Available at <http://dx.doi.org/10.17016/FEDS.2015.083>. 2015.

COFFMAN, M.; NOY, I. Hurricane Iniki: Measuring the Long-Term Economic Impact of Natural Disaster Using Synthetic Control. *Environment and Development Economics*, v. 17, p. 187–205, 2011.

CONLEY, T. G.; TABER, C. R. Inference with Difference-in-Differences with a Small Number of Policy Changes. *The Review of Economics and Statistics*, v. 93, n. 1, p. 113–125, 2011.

DHUNGANA, S. Identifying and Evaluating Large Scale Policy Interventions: What Questions Can We Answer? Available at: <https://openknowledge.worldbank.org/bitstream/handle/10986/3688/WPS5918.pdf?sequence=1>. 2011.

DUBE, A.; ZIPPERER, B. Pooled Synthetic Control Estimates for Recurring Treatments: An Application to Minimum Wage Case Studies. Available at: <http://www.irle.berkeley.edu/events/spring14/zipperer/dubezipperer_pooledsyntheticcontrol.pdf>. 2013.

DUPONT, W.; NOY, I. What Happened to Kobe? A Reassessment of the Impact of the 1995 Earthquake in Japan. Available at: <http://www.economics.hawaii.edu/research/workingpapers/WP_12-4.pdf>. 2012.

EUDEY, T. L.; KERR, J.; TRUMBO, B. Using R to Simulate Permutation Distributions for Some Elementary Experimental Designs. *Journal of Statistics Education*, v. 18, n. 1, 2010.

FERMAN, B.; PINTO, C. Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity. Working Paper. Available at <https://dl.dropboxusercontent.com/u/12654869/Ferman%20and%20Pinto%20-%20Inference%20in%20DID.pdf>. 2016.

FERMAN, B.; PINTO, C.; POSSEBOM, V. Specification-Search Possibilities with Synthetic Controls. 2016.

FIRPO, S.; POSSEBOM, V. Synthetic Control Estimator: A Generalized Inference Procedure and Confidence Sets. 2016.

FISHER, R. A. *The Design of Experiments*. 8th edition. ed. United States: Hafner Publishing Company, 1971.

GATHANI, S.; SANTINI, M.; STOELINGA, D. Innovative Techniques to Evaluate the Impacts of Private Sector Developments Reforms: An Application to Rwanda and 11 other Countries. Working Paper, <https://blogs.worldbank.org/impactevaluations/files/impactevaluations/methods_for_impact_evaluations_feb06-final.pdf>. 2013.

GOBILLON, L.; MAGNAC, T. Regional Policy Evaluation: Interative Fixed Effects and Synthetic Controls. *Review of Economics and Statistics*, 2016. Forthcoming.

GOBILLON, L.; MAGNAC, T.; SELOD, H. Do Unemployed Workers Benefit from Enterprise Zones? The French Experience. *Journal of Public Economics*, v. 96, p. 881–892, 2012.

HAHN, S.; KONIETSCHKE, F.; SALMASO, L. A Comparison of Efficient Permutation Tests for Unbalanced ANOVA in two by two Designs — and their Behavior under Heterocedasticity. Avaiable at: <arXiv:1309.7781>. 2013.

HAM, J. C. et al. Government Programs Can Improve Local Labor Markets: Evidence from State Enterprise Zones, Federal Emporwement Zones and Federal Enterprise Community. *Journal of Public Economics*, v. 95, p. 779–797, 2011.

HINRICHS, P. The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities. *Review of Economics and Statistics*, v. 94, n. 3, p. 712–722, March 2012.

HOSNY, A. S. Algeria's Trade with GAFTA Countries: A Synthetic Control Approach. *Transition Studies Review*, v. 19, p. 35–42, 2012.

IBRAGIMOV, R.; MULLER, U. K. t-Statistic Based Correlation and Heterogeneity Robust Inference. *Journal of Business & Economic Statistics*, v. 28, n. 4, p. 453–468, 2010.

IMBENS, G. W.; RUBIN, D. B. *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. 1st. ed. United Kingdom: Cambridge University Press, 2015.

JINJARAK, Y.; NOY, I.; ZHENG, H. Capital Controls in Brazil — Stemming a Tide with a Signal? *Journal of Banking & Finance*, v. 37, p. 2938–2952, 2013.

KAUL, A. et al. Synthetic Control Methods: Never Use All Pre-Intervention Outcomes as Economic Predictors. Working Paper. Available at: <http://www.oekonometrie.uni-saarland.de/papers/SCM_Predictors.pdf>. 2015.

KIRKPATRICK, A. J.; BENNEAR, L. S. Promoting Clean Enery Investment: an Empirical Analysis of Property Assessed Clean Energy. *Journal of Environmental Economics and Management*, v. 68, p. 357–375, 2014.

KLEVEN, H. J.; LANDAIS, C.; SAEZ, E. Taxation and International Migration of Superstars: Evidence from European Football Market. *American Economic Review*, v. 103, n. 5, p. 1892–1924, 2013.

KREIF, N. et al. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. *Health Economics*, 2015.

LEHMANN, E. *Testing Statistical Hypotheses*. New York: John Wiley & Sons, 1959.

LI, Q. Economics Consequences of Civil Wars in the Post-World War II Period. *The Macrotheme Review*, v. 1, n. 1, p. 50–60, 2012.

LIST, J.; SHAIKH, A. M.; XU, Y. Multiple Hypothesis Testing in Experimental Economics. NBER Working Paper 21875. Available at <http://www.nber.org/papers/w21875>. 2016.

LIU, S. Spillovers from Universities: Evidence from the Land-Grant Program. *Journal of Urban Economics*, v. 87, p. 25–41, 2015.

MIDEKSA, T. K. The Economic Impact of Natural Resources. *Journal of Environmental Economics and Management*, v. 65, p. 277–289, 2013.

MIRANDA, R. N. de. Zona Franca de Manaus: Desafios e Vulnerabilidades. Textos para Discussão 126. Available at: <http://goo.gl/Hj9NH0>. 2013.

MONTALVO, J. G. Voting after the Bombings: A Natural Experiment on the Effect of Terrorist Attacks on Democratic Elections. *Review of Economics and Statistics*, v. 93, n. 4, p. 1146–1154, 2011.

OLIVEIRA, J. L. da C.; SOUZA, R. B. de Lucena de. Um Estudo sobre a Renúncia Tributária dos Entes Federativos e os Benefícios Gerados pela Zona Franca de Manaus. *Revista Estudos do CEPE*, Porto Alegre, n. 36, p. 38–60, 2012.

PAULY, M.; BRUNNER, E.; KONIETSCHKE, F. Asymptotic Permutation Tests in General Factorial Designs. *Journal of the Royal Statistical Society. Series B*, v. 77, n. 2, p. 461–473, 2015.

PINOTTI, P. Organized Crime, Violence and the Quality of Politicians: Evidence from Southern Italy. Available at: <http://dx.doi.org/10.2139/ssrn.2144121>. 2012.

PINOTTI, P. The Economic Costs of Organized Crime: Evidence from Southern Italy. Temi di Discussione (Working Papers), <http://www.bancaditalia.it/pubblicazioni/temi-discussione/2012/2012-0868/en_tema_868.pdf>. 2012.

RIBEIRO, F.; STEIN, G.; KANG, T. The Cuban Experiment: Measuring the Role of the 1959 Revolution on Economic Performance using Synthetic Control. Available at: <http://economics.ca/2013/papers/SG0030-1.pdf>. 2013.

Sá, M. T. de V.; MACHADO, J. A. da C. Polo industrial de manaus (pim): A medida de seu valor adicionado comparado ao resto do brasil. Available at <http://goo.gl/2LwSjF>. 2012.

SANSO-NAVARRO, M. The effects on American Foreign Direct Investment in the United Kingdom from Not Adopting the Euro. *Journal of Common Markets Studies*, v. 49, n. 2, p. 463–483, 2011.

SAUNDERS, J. et al. A Synthetic Control Approach to Evaluating Place-Based Crime Interventions. *Journal of Quantitative Criminology*, 2014.

SEVERNINI, E. R. The Power of Hydroelectric Dams: Agglomeration Spillovers. IZA Discussion Paper, No. 8082, <http://ftp.iza.org/dp8082.pdf>. 2014.

SILLS, E. O. et al. Estimating the Impact of a Local Policy Innovation: The Synthetic Control Method Applied to Tropica Desforestation. *PLOS One*, 2015.

SMITH, B. The Resource Curse Exorcised: Evidence from a Panel of Countries. *Journal of Development Economics*, v. 116, p. 57–73, 2015.

SOUZA, F. F. A. de. *Tax Evasion and Inflation: Evidence from the Nota Fiscal Paulista Program*. Dissertação (Mestrado) — Pontifícia Universidade Católica, March 2014. Available at <http://www.dbd.puc-rio.br/pergamum/tesesabertas/1212327_2014_completo.pdf>.

WONG, L. *Three Essays in Causal Inference*. Tese (Doutorado) — Stanford University, March 2015.

YATES, F. Tests of Significance for 2 x 2 Contingency Tables. *Journal of the Royal Statistical Society. Series A (General)*, v. 147, n. 3, p. p. 426–463, 1984.

YOUNG, A. Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Eperimental Results. Available at: <http://goo.gl/zR07Bn>. 2015.

YU, J.; WANG, C. Political Risk and Economic Development: A Case Study of China. *Eknomska Istrazianja - Economic Research*, v. 26, n. 2, p. 35–50, 2013.