Using Synthetic Controls to Evaluate an International Strategic Positioning Program in Uruguay: Feasibility, Data Requirements, and Methodological Aspects[1]

Alberto Abadie
John F. Kennedy School of Government
Harvard University
Cambridge, MA
USA

1. Introduction

In 2011, Uruguay initiated an economic reform ("Programa de Posicionamiento Estrategico Internacional") with the goal to promote foreign direct investment and facilitate foreign trade. This document discusses the feasibility, data requirements, and methodological aspects of an evaluation of the effect of this intervention on foreign direct investment (FDI) flows in Uruguay. For this purpose, this document describes the use of synthetic control methods, which have been successfully implemented in similar settings to evaluate the effects of large scale interventions on single regions or countries, and discusses their applicability to the evaluation of the effects of the 2011 Uruguayan economic reform.

The rest of the document is organized as follows. Section 2 provides an introduction to the synthetic control methodology and related methods. Section 3 discusses some the formal aspects of the synthetic control methodology that are of particular interest for empirical applications. Section 4 and 5 discuss contextual and data requirements for synthetic control empirical studies. We examine the validity of these requirements in the context of the evaluation of the effect of the 2011 Uruguayan economic reform and discuss potential ways to adapt the research design when the requirements do not hold in practice. This section provides also a discussion of the available data for an evaluation of the 2011 Uruguayan economic reform, along with data sources. Section 6 discusses previous studies that have applied the synthetic control method in settings similar to the Uruguayan economic reform. Section 7 discusses computational aspects. Section 8 concludes.

[1] This study was commissioned by the Trade and Integration Sector of the Inter-American Development

2. Synthetic Controls and Related Methods

Synthetic control methods (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010) aim to estimate the effects of aggregate interventions, that is, interventions that are implemented at an aggregate level affecting a small number of large units (such as a cities, regions, or countries), on some aggregate outcome of interest. Synthetic control methods have been applied to study a variety of problems: What was the effect of California's large scale tobacco control program on cigarette consumption in California? What was the economic cost of the German reunification? Do Kyoto Protocol's emission targets reduce greenhouse gas emissions? What is the effect of economic liberalization on per-capita GDP?[2] Traditional regression analysis techniques require large samples and many observed instances of the event or intervention of interest, and as a result they are ill-suited to estimate the effects of infrequent events, like large policy interventions. Economists have approached the estimation of the effects of large scale but infrequent interventions using time-series analysis and comparative case studies. Single-unit time-series analysis is an effective tool to study the short-time effects of policy interventions, in cases when we expect short-time effects to be of a substantial magnitude.[3] However, the use of single-unit time series techniques to estimate medium and long-run effects of policy intervention is complicated by the presence of shocks to the outcome of interest that occur after the intervention. Comparative cases studies are based on the idea that the effect of an intervention on some variables of interest can be inferred from the comparison of the evolution of the variables of interest between the unit exposed to the event or intervention of interest and a group of units that are similar to the exposed unit but that were not affected by the event/intervention.

Comparative case studies have long been applied to the evaluation of large-scale events or aggregate interventions. For example, to estimate the effects of the massive arrival of Cuban expatriates to Miami during the 1980 Mariel Boatlift on native unemployment in Miami, Card (1990) compares the evolution of native unemployment in Miami at the time of the boatlift to the average evolution of native unemployment in four other cities in the US. Similarly, Card and Krueger (1994) uses Pennsylvania as a comparison to estimate the effects of an increase in the New Jersey minimum wage on employment in fast food restaurants in New Jersey. A drawback of comparative case studies of this type is that the selection of the comparison units is not formalized and often relies on vague statements of affinity between the unaffected units

[2] See Abadie and Diamond, and Hainmueller (2010, 2012), Almer and Winkler (2012), and Billmeier and Nannicini (in press).

[3] The literature on "interrupted time-series" is particularly relevant in the context of policy evaluation. See, for example, Cook and Campbell (1979) which discusses the limitations of this methodology if interventions are gradual rather than abrupt and/or if the causal effect of an intervention is delayed in time. Interrupted time-series methods are closely related to regression-discontinuity design techniques (see, e.g., Lee and Lemieux, 2010).

and the set of comparison units. Moreover, when the units of observation are a small number of aggregate entities, like countries or regions, no single unit alone may provide a good comparison for the unit affected by the intervention.

The synthetic control method is based on the idea that a combination of unaffected units often provides a more appropriate comparison than any single unaffected unit alone. The synthetic control methodology seeks to formalize the selection of the comparison units using a data driven procedure. As we will discuss later, this formalization also opens the door to precise quantitative inference in comparative case studies.

3. Formal Aspects of the Synthetic Control Method

Suppose that we obtain data for a sample of J countries: $j = 1, 2, \dots, J$. Without loss of generality, we assume that the first country ($j = 1$) is the unit affected by the policy intervention of interest. The "donor pool", that is, the set of potential comparisons, $j = 2, \dots, J$ is a collection of countries not affected by the intervention. We assume also that our data span T periods and that the first T_0 periods are before the intervention. For each country, j, and time, t,we observe the outcome of interest, Y_{jt}. For each country, j, we observe also a set of k predictors of the outcome: $X_{1j}, \cdots, X_{kj}$ (which may include pre-intervention values of Y_{jt}). For the country affected by the intervention, $j = 1$, and a post-intervention period, $t > T_0$, we define the potential outcomes that would have been observed with and without the intervention, Y_{1t}^I and Y_{1t}^N, respectively.[4] Then, the effect of the intervention of interest for the affected country in period t (with $t > T_0$) is:

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N. \tag{1}$$

Because country "one" is exposed to the intervention after period T_0, it is clear that for $t > T_0$ we have $Y_{1t} = Y_{1t}^I$. Simply put, for the country affected by the intervention and a post-intervention period we observe the potential outcome under the intervention. The great policy evaluation challenge is to estimate Y_{1t}^N for $t > T_0$: how the outcome of interest would have evolved in the affected country *in the absence of the intervention*. This is a counterfactual outcome, as the affected country was, by definition, exposed to the intervention of interest after $t = T_0$. As equation (1) makes clear, given that Y_{1t}^I is observed, the problem of estimating the effect of a policy intervention is equivalent to the problem of estimating Y_{1t}^N.[5]

[4] These are the "potential responses" of Rubin's Model for Causal Inference (see, e.g., Rubin, 1974, Holland, 1986).

[5] Notice that equation (1) allows the effect of the intervention to change in time. This is crucial because intervention effects may not be instantaneous and may accumulate or dissipate as time after the intervention passes.

Comparative case studies aim to reproduce Y_{1t}^N, that is the value of the outcome variable that would have been observed for the affected unit in the absence of the intervention, using one unaffected unit or a small number of unaffected units that had similar characteristics as the affected unit at the time of the intervention. As discussed above, when the data consist of a few aggregate entities, such as countries, it is often difficult to find a single unaffected country that provides a suitable comparison for the country affected by the policy intervention of interest. The synthetic control method is based on the observation that a combination of units in the donor pool may resemble the characteristics of the affected unit substantially better than any unaffected unit alone. A synthetic control is defined as a weighted average of the units in the donor pool. Formally, a synthetic control can be represented by a set of weights, $w = \{w_2, \dots, w_J\}$, attached to the countries in the donor pool. Given a set of weights, w, the synthetic control estimators of Y_{1t}^N and τ_{1t} are, respectively:

$$\hat{Y}_{1t}^N = w_2 Y_{2t} + \cdots + w_J Y_{Jt},$$

and,

$$\hat{\tau}_{1t} = Y_{1t} - \hat{Y}_{1t}^N. \qquad (2)$$

To avoid extrapolation, the weights are restricted to be non-negative and to sum to one, so synthetic controls are weighted averages of the units in the donor pool.[6,7] As an example, a synthetic control that assigns equal weights, $w_j = 1/J$, to each of the units in the control group results in the following estimator for τ_{1t}:

$$\hat{\tau}_{1t} = Y_{1t} - \frac{1}{J}\left(Y_{2t} + \cdots + Y_{Jt}\right).$$

In this case, the synthetic control is the simple average of all the units in the donor pool. If, however, a single unit, z, in the donor pool is used as a comparison, then $w_z = 1$, $w_j = 0$ for $j \neq z$, and:

$$\hat{\tau}_{1t} = Y_{1t} - Y_{zt}.$$

[6] The requirement that country weights should be nonnegative and no greater than one can be relaxed at the cost of allowing extrapolation. Abadie, Diamond, and Hainmueller (2012) prove that in the context of estimating the effect of a policy intervention there is a regression estimator that can be represented as a synthetic control with country weights that are unrestricted except for that the sum of the country weights is equal to one. By not restricting country weights to be nonnegative and no greater than one, regression allows extrapolation.

[7] Notice also that restricting country weights to sum to one may be warranted only if the dependent variable is rescaled, so it is not affected by country size (for example, using GDP per capita instead of country GDP).

Expressing the comparison unit as a synthetic control motivates the question of how the weights, $w_2, \dots, w_J$, should be chosen. Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) propose to choose $w_2, \dots, w_J$ so that the resulting synthetic control best resembles the pre-intervention characteristics of the affected unit. That is, given a set of non-negative weights, $v_1, \dots, v_k$, Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) propose to chose the synthetic control , $w^* = \{w_2{}^*, \dots, w_J{}^*\}$ that minimizes:

$$v_1\left(X_{11} - w_2 X_{12} - \cdots - w_J X_{1J}\right)^2 + \cdots + v_k\left(X_{k1} - w_2 X_{k2} - \cdots - w_J X_{kJ}\right)^2. \tag{3}$$

The weights $v_1, \dots, v_k$ reflect the relative importance of the synthetic control reproducing the values of the predictors $X_{11}, \cdots, X_{k1}$. For a given set of weights, $v_1, \dots, v_k$, minimizing equation (3) can be easily accomplished using constrained quadratic optimization.[8] Of course, a question remains about how to chose $v_1, \dots, v_k$, the weights that represent the relative importance of reproducing the values, $X_{11}, \cdots, X_{k1}$, of each of the outcome predictors. We next describe four methods to chose $v_1, \dots, v_k$, starting with the simplest and in increasing order of sophistication.

1. In some instances, the weights $v_1, \dots, v_k$ may reflect subjective measures of the relative importance of each of the variables, $X_{1j}, \cdots, X_{kj}$, explaining the outcome of interest, or their values may be calibrated by inspecting how different choices of $v_1, \dots, v_k$ affect the discrepancies between the characteristics of the unit affected by the intervention and the resulting synthetic controls: $(X_{11} - w_2 X_{12} - \cdots - w_J X_{1J})$, …, $(X_{k1} - w_2 X_{k2} - \cdots - w_J X_{kJ})$.
2. The weights $v_1, \dots, v_k$ can also be determined in a first step exploratory analysis using regression to determine the relative predictive power of the variables $X_{1j}, \cdots, X_{kj}$.
3. Another way to select $v_1, \dots, v_k$ is to choose the set of values for $v_1, \dots, v_k$ that produce the best "fit" in terms of how closely the synthetic control tracks the trajectory of the outcome variable of the affected unit during the pre-intervention period. In order words, the idea is chosing $v_1, \dots, v_k$ so that the resulting synthetic controls minimizes the size of the prediction error, $Y_{1t} - \hat{Y}_{1t}^N$, during some set of pre-intervention periods (typically the entire pre-intervention period). This can be implemented by solving a nested optimization problem where v is chosen so that w minimizes the mean square prediction error (defined below) over a pre-specified set of pre-intervention periods.
4. The preceding method to choose $v_1, \dots, v_k$ maximizes in-sample fit. An alternative way to choose $v_1, \dots, v_k$ is to maximize out-of-sample fit via cross-validation. The

[8] The role of the constraints is to restrict the resulting weights $w_2^*, \dots, w_J^*$ to be positive and sum to one.

ideas behind cross-validation selection of $v_1, \dots, v_k$ are described next.[9] The goal of the synthetic control is to approximate the trajectory that would have been observed for Y_{1t} and $t > T_0$ in the absence of the intervention. For that purpose, the synthetic control method selects a set of weights $w = \{w_2, \dots, w_J\}$ in such way that the resulting synthetic control resembles the affected unit before the intervention along the values of the variables $X_{11}, \cdots, X_{k1}$. The question of choosing $v = \{v_1, \dots, v_k\}$ boils down to assessing the relative importance of each of $X_{11}, \cdots, X_{k1}$ as a predictor of Y_{1t}^N. That is, v_h reflects the relative importance of approximating the value of X_{h1}, for $1 \leq h \leq k$. Y_{1t}^N is not observed during the post-intervention period. As a result, we cannot directly evaluate the relative importance of fitting each predictor to approximate Y_{1t}^N in the post-intervention period. However, because Y_{1t}^N is observed for the pre-intervention periods $t = 1, 2, \dots, T_0$, it is possible to use pre-intervention data to assess the predictive power on Y_{1t}^N of the variables $X_{1j}, \cdots, X_{kj}$. This can be accomplished, for example, following the steps described next:

(1) Divide the pre-intervention periods into a *training* period $t = 1, \dots, t_0$ and a *validation* period, $t = t_0 + 1, \dots, T_0$.

(2) Now, using data from the training period only, each potential choice of $v = \{v_1, \dots, v_k\}$ produces a synthetic control, $w(v)$, which can be determined minimizing equation (3), subject to the restriction that the weights $w(v)$ are positive and sum to one. The mean square prediction error (MSPE) of this synthetic control with respect to Y_{1t}^N in the validation period is:

$$\left(Y_{1t_0+1} - w_2(v)Y_{2t_0+1} - \cdots - w_J(v)Y_{Jt_0+1}\right)^2 + \\ \cdots + \left(Y_{1T_0} - w_2(v)Y_{2T_0} - \cdots - w_J(v)Y_{JT_0}\right)^2. \quad (4)$$

(3) Minimize the mean square prediction error in the previous equation with respect to v.

(4) Use the resulting v and data on the predictors for the last t_0 periods before in the intervention, $t = T_0 - t_0 + 1, \dots, T_0$, to calculate w^*.

Choosing v by by minimizing the MSPE over the a set of pre-treatment periods or by cross-validation involves a nested optimization problem. Each choice of the predictor weights v implies a choice of the country weights $w(v)$, obtained from minimizing equation (3), which in turn implies a value for the MSPE. Solving this optimization problem presents some challenges, which are discussed below (see section 7).

[9] This procedure was proposed in Abadie, Diamond, and Hainmueller (2012).

Abadie, Diamond, and Hainmueller (2011) discuss a mode of inference for the synthetic control framework that is based on the comparison between the intervention effect estimated for the affected unit and the distribution of "placebo" intervention effects estimated for the units in the donor pool. They deem an estimate significant when the estimate is of large magnitude relative to the distribution of placebo effects obtained for units that were not affected by the intervention.[10] It is important to note that the availability of a well-defined procedure to select the comparison unit, like the one provided by the synthetic control method, makes the estimation of the effects of placebo interventions feasible. The reason is that, without a formal procedure to chose comparison units, it would be difficult to re-apply the same estimation procedure used for the affected unit to the units in the donor pool. In this sense, the formalization of the choice of the comparison unit provided by the synthetic control method opens the door to precise quantitative inference in the context of comparative case studies.

4. Contextual Requirements

This section discusses "contextual requirements", that is, the conditions under which synthetic controls are appropriate tools for policy evaluation, as well as appropriate ways to modify the analysis when these conditions do not hold perfectly. It is important, however, to point out that most of the requirements listed in this section pertain not only to synthetic control methods but also to any other type of comparative case study research design. We also discuss how to assess the requirements outlined here for the case of the evaluation of the 2011 Uruguayan economic reform.

1. Size of the effect and volatility of the outcome: As previously discussed, the goal of comparative case studies is to estimate the effect of a policy intervention on the unit (e.g., country) exposed to the intervention. That is, comparative case studies typically estimate the effect of an intervention for a single unit. In particular, in the case of the 2011 Uruguayan economic reform the goal is to estimate the effect of the reform on the flows of FDI to Uruguay following the reform. The nature of this exercise, which focuses on a single unit, indicates that small effects will be indistinguishable from random shocks to the outcome of the affected country, especially if the outcome variable of interest is highly volatile.[11] As a result, the impact of "small" interventions with effects of a magnitude similar to the volatility of the outcome are difficult to

[10] See Abadie, Diamond, and Hainmueller (2011) for additional detail.

[11] In studies that seek to estimate the average effect of an intervention that is observed in a large number of instances, volatility is reduced by averaging. In contrast, comparative case studies typically focus on the effect of a single event or intervention.

detect. Even if the effect of an intervention is large, it may be difficult to detect if the volatility of the outcome is also large. In cases where substantial volatility is present in the outcome of interest it is advisable to remove it (via time-series filtering) in both the exposed unit as well as in the units in the donor pool before applying synthetic control techniques. Foreign direct investment in Uruguay was substantially volatile before the economic reform.[12] Notice, however, that the challenge posed by volatility comes only from the fraction of this volatility that is generated by Uruguay-intrinsic factors. FDI volatility generated by common factors affecting other countries can be differentiated-out by choosing an appropriate synthetic control (that is, a synthetic control composed by countries that resemble Uruguay in the factors that determine FDI inflows).

2. <u>Availability of a comparison group</u>: The very nature of comparative case studies implies that inference based on these methods will be faulty in the absence of a suitable comparison group. First and foremost, in order to have countries available for the donor pool, it is important that not all countries adopt interventions similar to the one under investigation during the period of the study. Countries that adopt an intervention that is similar to the one adopted by the unit of interest should not be included in the donor pool because they are affected by the intervention, very much like the unit of interest.[13] It is also important to eliminate from the donor pool any country that may have suffered large idiosyncratic shocks to the outcome of interest during the study period, if it is judged that such shocks would not have affected the outcome of the unit of interest in the absence of the intervention. Moreover, it is important to restrict the donor pool to units with characteristics that are similar to the affected unit. The reason is that, while the restrictions placed on the weights $w = \{w_2, \dots, w_J\}$ do not allow extrapolation, interpolation biases may still be important if the synthetic control matches the characteristics of the affected unit by averaging away large discrepancies between the characteristics of the affected unit and the characteristics of the units in the synthetic control. As noticed in Abadie, Diamond, and Hainmueller (2010), controlling the discrepancies between the characteristics on the unit affected by the intervention and the individual units that compose the synthetic control can be accomplished by adding to the objective function in equation (3) a set of penalty terms that depend on the discrepancies between the characteristics of the affected unit and the characteristics of the individual units included in the synthetic control. Unfortunately, there is currently no

[12] See López and Niembro (2012), Figure 2.
[13] As an example, in their study of the effect of California's tobacco control legislation, Abadie, Diamond, and Hainmueller (2010) discard from the donor pool several states that adopted large-scale tobacco programs during the sample period of the study.

ready-to-use computer code to calculate synthetic control weights using this type of penalty terms. In many cases, however, it is possible to directly restrict the set of units in the donor pool so that they have characteristics similar to the country affected by the policy intervention of interest.[14] In the case of the 2011 Uruguayan economic reform, it is natural to restrict the donor pool to other Latin American countries, provided that: (i) these countries did not adopt large FDI promotion policies or where affected by other events or interventions that had substantial effects on the FDI inflows, (ii) they were not substantially affected by the Uruguayan economic reform (see point 4, below), and (iii) there is a combination of them that closely approximated the pre-2011 values of FDI predictors for Uruguay.

3. No anticipation: As in any research design that exploits time variation in the outcome variable to estimate the effect of an intervention, synthetic control estimators may be biased if forward looking economic agents react in advance of the policy intervention under investigation or if certain components of the intervention were put in place in advance of the formal implementation/enactment of the intervention. In the case of an evaluation of the Uruguayan economic reform it is necessary to investigate if FDI started flowing into the country in anticipation of the policy reform and/or the economic reform was preceded by similarly aimed policies that could have increased FDI in Uruguay. If there are signs of anticipation, it is advisable to backdate the intervention in the data set to a period before any anticipation effect can be expected, so the full extent of the effect of the economic reform can be estimated.[15]

4. No interference: The issue of no interference is closely related to the issue of availability of a comparison group in point 2 above. In some instances, an intervention may have spillover effects on units that are not directly affected by it. For example, if the 2011 economic reform in Uruguay was successful attracting FDI to Uruguay, other countries in the region may have received lower FDI flows.[16] If spillover effects are

[14] For example, in Abadie, Diamond, and Hainmueller (2012)'s study of the effect of the 1990 German reunification on economic growth in the former West Germany, the donor pool is restricted to a set of OECD countries.

[15] Notice that backdating the intervention in the data does not mechanically bias the estimator of the effect of the intervention even if some periods before the intervention are mistakenly recoded as post-intervention periods. The reason is that, as shown in equations (1) and (2), the synthetic control estimator does not restrict the time variation in the effect of the intervention. Therefore, periods barely affected by the intervention may show small or zero effects, while subsequent periods may produce a large estimated effect. This is in contrast with much of the practice using panel data model, where in many instances the effect of an intervention is restricted to be constant across post-intervention periods.

[16] Also, related to point 2, other countries in the region may have reacted to the Uruguayan reform by instituting their own FDI promotion policies. Notice that although this may arguably be conceived of as a

judged to be substantial, it may be advisable to keep the countries indirectly affected by the Uruguayan economic reform out of the donor pool. Notice that there is a potential tension between the issues discussed here and those discussed in point 2 above ("Availability of a Comparison Group"). On the one hand, it is advisable to select for the donor pool countries that are affected by the same regional economic shocks as the country where the intervention happens. However, if spillover effects are substantial and affect countries in close geographical proximity, those countries may provide a biased estimate of the counterfactual outcome without intervention for the country affected by the intervention. In those cases, if countries affected by spillover effects are included in the synthetic control, the researcher should be aware of the potential direction of the bias of the resulting estimator. For example, if the 2011 Uruguayan economic reform increased FDI flows to Uruguay and, as a result, reduced the FDI flows to other countries in the region, and those countries contribute to the synthetic control, then we would expect the synthetic control estimator to be upwards biased (that, is to overestimate the effect of the economic reform on FDI in Uruguay). Notice, however, that such *bias exists only if the economic reform did have an effect on FDI in Uruguay.* In this case, if the policy has an effect, the bias exacerbates the estimated effect of the policy. As a result, this type of bias helps us detecting the existence of an effect, although it exaggerates its estimated size.

5. Convex hull condition: Synthetic control estimates are predicated on the idea that a combination of unaffected units can approximate the pre-intervention characteristics of the affected unit. Once the synthetic control is constructed it should be checked that the differences in the characteristics of the affected unit and the synthetic control are small, that is:

$$X_{11} - w_2 X_{12} - \cdots - w_J X_{1J} \approx 0, \cdots, X_{k1} - w_2 X_{k2} - \cdots - w_J X_{kJ} \approx 0.$$

In mathematical parlance, we need that $(X_{11},\ X_{21}, \ldots, X_{k1})$ falls close to the convex hull of the set of points $\{(X_{12},\ X_{22}, \ldots, X_{k2}), \ldots, (X_{1J},\ X_{2J}, \ldots, X_{kJ})\}$.[17] If the unit affected by the intervention of interest is "extreme" in the value of a particular variable, such value may not be closely approximated by a synthetic control.[18] The fact

part of the overall intervention effect, countries that react to the Uruguayan reform should be eliminated to the donor pool because they are also affected by the intervention.

[17] The convex hull of the set $\{(X_{12},\ X_{22}, \ldots, X_{k2}), \ldots, (X_{1J},\ X_{2J}, \ldots, X_{kJ})\}$ is the set of all convex combinations of the points in the set. It is identical to the set of all possible weighted averages of the points in the set. Figure 1 in the Appendix shows a set of points in $\mathbb{R}^2$ and their convex hull region.

[18] For example, Abadie, Diamond, and Hainmueller (2012) find that because inflation levels were particularly low for West Germany before the reunification, the value of this variable cannot be closely reproduced by a synthetic control composed by other OECD countries.

that the value of a particular variable cannot be closely approximated by the synthetic control may not cause much concern as long as the approximation is good enough so that the synthetic control closely tracks the trajectory of the outcome variable for the unit affected by the intervention during the pre-intervention periods. In some cases, however, the unit affected by the intervention of interest may be extreme in the values of the outcome variable before the intervention. For example, if Uruguay had particularly low or particularly high levels of FDI before 2011 relative to the countries in the donor pool, then no weighted average of countries in the donor pool would be able to closely reproduce the FDI series for Uruguay before 2011. A potential way to proceed in those cases is to transform the outcome to time differences, ΔY_{it}=$Y_{it} - Y_{it-1}$, or growth rates, $100 \times \Delta Y_{it}/Y_{it-1}$. In some cases this strategy may be fruitful, because as evidenced in the difference-in-differences literature, there are instances when a comparison group can reproduce the changes in the outcome variable for the unit of interest even if the level of the outcome variable cannot be reproduced.

6. Time horizon: The effect of many interventions, especially those like the 2011 Uruguayan economic reform that intent to produce structural changes in an economy, may take time to arise or to be of enough magnitude to be quantitatively perceived. Given the complexity of the program, the uncertainty that international investors may entertain about the implications of the reform for their potential investments in Uruguay, and the irreversible nature of many investment decisions, it is likely that the full extent of the 2011 economic reform in Uruguay will not be realized for years. An obvious but unsatisfying approach to this problem is to wait until the effects economic reform runs its course. A more proactive approach is to use forward looking indicators of FDI (like business climate surveys).

5. Data Requirements

This section discusses specific data requirements to evaluate the 2011 Uruguayan economic reform.

1. Aggregate data on the outcome and predictors for Uruguay and a set of comparison units: From the previous discussion, it can be seen that the synthetic control method requires the availability of data on outcomes and predictors of the outcome for Uruguay and a set of comparison countries. Aggregate data on FDI and determinants of FDI are available in cross-country databases like the World Bank's World Development Indicators. It is important that the synthetic control approximates Uruguay in the values of the most important predictors of FDI. The literature on the determinants of FDI flows indentifies factors like geographical location, culture, GDP

per-capita, market size, infrastructure, investment treaties, corruption, political stability, and human capital endowments.[19] (Note: For a list of FDI determinants with sources see Abadie and Gardeazabal, 2008

2. Sufficient pre-intervention information: The credibility of a synthetic control estimator depends in great part on its ability to steadily track the trajectory of the outcome variable for the affected unit before the intervention. In this respect, Abadie, Diamond, and Hainmueller (2011) show that under certain conditions, the bias of the synthetic control estimator is bounded by a function that goes to zero as the number of pre-intervention periods during which the synthetic control closely tracks the trajectory of the outcome variable for the affected unit increases.[20] Therefore, when designing a synthetic control study it is of crucial importance to collect information on the affected unit and the donor pool for a large pre-treatment window.

3. Sufficient post-intervention information: This data requirement derives from the time horizon contextual requirement in section 4 (see point 6 in that section). The evaluation data must include outcome measures possibly affected by the intervention. This may be problematic if the effect of an intervention is expected to arise gradually over time and if no forward looking measures of the outcome are available.

6. Related Applications

Several previous studies have applied the synthetic control method to estimate the effects of country-level interventions, like the 2011 Uruguayan economic reform. In this section, we summarize some of these studies as they provide useful templates for the evaluation of the economic reform in Uruguay.

7. Software and Computational Aspects

Abadie, Diamond, and Hainmueller (2010) provide companion software to estimate synthetic controls in R, Matlab, and Stata. In this section, we briefly review computation of synthetic controls using Stata.[21] Estimating synthetic controls in Stata requires installation of the Synth

[19] See, e.g., Blonigen and Piger (2011), Javorcik and Wei (2009), Lim (2001), Stein and Daude (2007).

[20] The intuition is quite straightforward. With a large number of units in the donor pool and a small number of pre-intervention periods in the data, overfitting may cause that a synthetic control that reproduces the trajectory of the outcome variable for the affected unit during the pre-intervention periods is not a good predictor of the trajectory of Y_{1t} for $t > T_0$.

[21] See http://www.mit.edu/~jhainm/synthpage.html for additional information.

package, which can be downloaded from Jens Hainmueller's website.[22] We first recommend that Stata is updated before the installation of Synth. To update Stata type "update all" on the Stata command line. To download Synth, type:

```
net from "http://www.mit.edu/~jhainm/Synth"
net describe synth
net install synth, all replace force
```

Before using Synth, we need to specify the variable that identifies the units of observation, $j = 1, \dots, J$, and the variable that identifies the time periods $t = 1, \dots, T$. This information can be passed to Stata using the command:

tsset *panelvar timevar*

where *panelvar* identifies units of observation (e.g., countries) and *timevar* identifies time periods (e.g., years). Once this is done, we can start using the command synth. A simple example of usage of synth is:

synth *depvar predictorvars* , trunit(#) trperiod(T_0+1) figure

where

depvar	is the outcome variable, Y_{it}
predictorvars	is the set of predictors, $X_{1j}, \cdots, X_{kj}$
trunit(#)	specifies which is the affected unit
trperiod(#)	specifies which is the period of the intervention
figure	requests a figure with the results

The symbol # represents the numerical identifier of the affected unit. In this example, the weights $v = \{v_1, \dots, v_k\}$ are calculated using regression analysis by default.[23] Each predictor in *predictorvars* can be evaluated at different time periods. Suppose that gdppc measures annual GDP per-capita, and that *timevar* measures years. If we want the set of predictors to include the values of GDP per-capita in 2001 and 2010 this can be done including gdppc(2001) and gdppc(2010) in *predictorvars*. Similarly, if we want that the set of predictor includes the average GDP per-capita during the period 2001 to 2010, this can be accomplished including

[22] *Ibid.*

[23] For this purpose, synth runs a set of regressions of the outcome in the post-intervention periods on the predictors using only the donor pool as a sample. Then, the weights $v_1, \dots, v_k$ reflect the relative magnitude of the coefficients on the variables $X_{1j}, \cdots, X_{kj}$ in those regressions.

gdppc(2001(1)2010) in *predictorvars*. Also, the synth command has the option xperiod(numlist), which allows users to specify the time periods over which the values of all other predictors are averaged. If the user does not specify a time period over which to average the value of the predictors using one of these methods, then the default is to average them over the entire pre-intervention period.

If the user wants to provide values for $v = \{v_1, \ldots, v_k\}$ this can be accomplished with the option customV

synth *depvar predictorvars* , trunit(#) trperiod(T_0+1) customV(numlist)

where numlist is a Stata numerical list that contains $(v_1, \ldots, v_k)$.

As discussed in point 3 of section 3, a different way to chose $v = \{v_1, \ldots, v_k\}$ is to minimize the MSPE over some set of pre-intervention periods, which results in a nested optimization problem. This can be implemented by Synth in Stata using the following options:

synth *depvar predictorvars* , trunit(#) trperiod() mspeperiod(numlist) nested allopt figure

where

nested requests nested optimization procedure (see point 3 in section 3)
mspeperiod(numlist) set of periods over which the MSPE is minimized
allopt starts the optimization procedure from three different initial values

If mspeperiod is not used with nested, then the MSPE is minimized over the entire pre-intervention period. The option allopt can only be used along with nested. This option starts the algorithm at three different initial values to try to potentially improve over local optima.

Finally, as discussed in section 3, a more sophisticated approach to choose $v = \{v_1, \ldots, v_k\}$ is to use cross-validation (see point 4 of section 3). Cross-validation can be implemented as follows. First, we must calculate the set of weights $v = \{v_1, \ldots, v_k\}$ that minimize the MSPE over the validation period. With that purpose we run:

synth *depvar predictorvars* , trunit(#) trperiod(t_0+1) mspeperiod(*validationperiod*) nested allopt

In this step, the variables in *predictorvars* are evaluated in the training period, $1, \dots, t_0$. Next, we store the resulting $v = \{v_1, \dots, v_k\}$ in a Stata numerical list. In the current version of the software, this requires extracting the main diagonal of the V_matrix produced by the synth command and transforming it into a Stata numerical list. This can be done using the following code snippet:

```
mat d = vecdiag(e(V_matrix))
local B = ""
forval i = 1/`=colsof(d)' {
        local B `B'  `=d[1,`i']'
}
```

Next, calculate the synthetic control using the weights in B and data from the validation period:

synth *depvar predictorvars*, trunit(#) trperiod(T_0+1) customV(`B')

In this step, the variables in *predictorvars* are evaluated in the last t_0 periods before the intervention.

The Stata distribution of Synth includes also the file smoking.dta, which contains the data analyzed in Abadie, Diamond, and Hainmueller (2010) to evaluate the impact of the 1988 California tobacco control program. Those data are provided with the code so that new users of the code have the opportunity to practice and gain familiarity with Synth. The Synth online help, which can be accessed by typing

```
help synth
```

on the Stata command line provides examples of usage applied to the smoking.dta dataset.

8. Conclusions

Synthetic control methods have been applied to the estimation of the effect of aggregate interventions in a variety of contexts. This document provides a set of conditions (or "requirements") that increase the viability and credibility of synthetic controls estimation methods. We divide the requirements into "contextual requirements", which depend on the context in which the evaluation takes place, and "data requirements", which refer to data availability issues. We describe these requirements in the context of the 2011 Uruguayan economic reform and discuss potential ways to adapt the analysis if these requirements are not met. This document provides the results of a review of the literature on the determinants of FDI along with data sources. It also discusses previous applications of the synthetic control method in related contexts. Finally, we discuss computation of synthetic controls in Stata.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California Tobacco Control Program," *Journal of the American Statistical Association, 105*, 493-505.

Abadie, A., Diamond, A., and Hainmueller, J. (2012), "Comparative Politics and the Synthetic Control Method," working paper (available at: http://www.hks.harvard.edu/fs/aabadie/comparative.pdf).

Abadie, A., and Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review,* 93(1), 113-132.

Abadie, A., and Gardeazabal, J. (2008), "Terrorism and the World Economy," *European Economic Review,* 52, 1-27.

Almer, C., and Winkler, R. (2012), "The Effect of Kyoto Emission Targets on Domestic CO2 Emissions: A Synthetic Control Approach," Working Paper 12-02, Department of Economics, University of Bern.

Billmeier, A., and Nannicini, T. (in press), "Assessing Economic Liberalization Episodes: A Synthetic Control Approach," *Review of Economics and Statistics,* forthcoming.

Blonigen, B. A., and Piger, J. (2011), "Determinants of Foreign Direct Investment," NBER working paper, 16704.

Card, D. (1990), "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, 44, 245–257.

Card, D., and Krueger, A. B. (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84, 772–793.

Coffman, M., and Noy, I. (2011), "Hurricane Iniki: Measuring the Long-Term Economic Impact of a Natural Disaster Using Synthetic Control," *Environment and Development Economics*, 17, 187–205.

Cook, T. D., and Campbell, D. T. (1979), *Quasi-Experimentation. Design & Analysis Issues for Field Settings.* Boston: Houghton Mifflin.

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945-960.

Javorcik, B. S., and Wei, S.-J. (2009), "Corruption and cross-border investment in emerging markets: Firm-level evidence," *Journal of International Money and Finance*, 28, 605-624.

Nannicini, T., and Billmeier, A. (2011) "Economies in Transition: How Important Is Trade Openness for Growth?," *Oxford Bulletin of Economics and Statistics*, 73(3), 287-314.

Lee, D. S., and Lemieux, T. (2010), “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 48, 281-355.

Lee, W-S. (2010), “Comparative Case Studies of the Effects of Inflation Targeting in Emerging Economies,” *Oxford Economic Papers*, 63, 375–397

Lim, E.-G. (2001), “Determinants of, and the Relation Between, Foreign Direct Investment and Growth: A Summary of the Recent Literature,” IMF working paper, 01/175.

López, A., and Niembro, A. (2012), “Los Mecanismos para la Promoción de Inversiones: Análisis, Evidencia Internacional y Sugerencias para Uruguay,” IDB.

Rubin, D. B. (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Educational Psychology,* 66, 688-701.

Sanso-Navarro, M. (2011), “The Effects on American Foreign Direct Investment in the United Kingdom from Not Adopting the Euro,” *Journal of Common Market Studies*, 49(2), 463–483.

Stein, E., and Daude, C. (2007), “Longitude Matters: Time Zones and the Location of Foreign Direct Investment,” *Journal of International Economics*, 71, 96-112.

Appendix

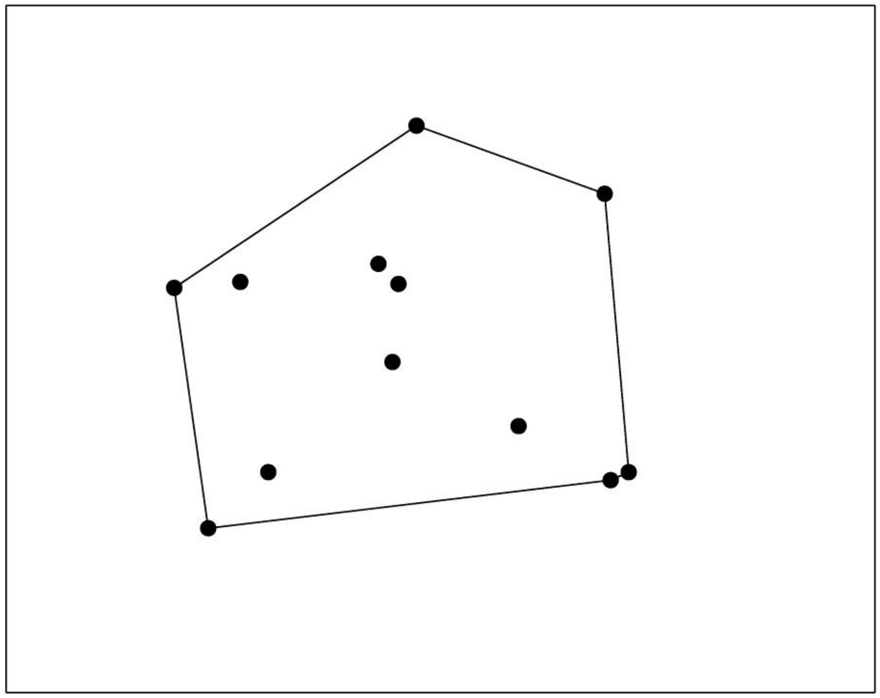

Figure 1: A Set of Points and Its Convex Hull