Policy Analysis using Synthetic Controls in Continuous-Time

Alexis Bellot [1 2] **Mihaela van der Schaar** [1 2 3]

Abstract

Counterfactual estimation using synthetic controls is one of the most successful recent methodological developments in causal inference. Despite its popularity, the current description only considers time series aligned across units and synthetic controls expressed as linear combinations of observed control units. We propose a continuous-time alternative that models the latent counterfactual path explicitly using the formalism of controlled differential equations. This model is directly applicable to the general setting of irregularly-aligned multivariate time series and may be optimized in rich function spaces – thereby improving on some limitations of existing approaches.

1. Introduction

Counterfactual estimation poses the question of what would have been the outcome if a different treatment, policy or intervention[1] had been applied. To answer this question one often seeks a control group of *comparable* units e.g., individuals, patients, or states, to approximate a target unit's outcome trajectory had a different treatment been applied.

We focus on the case where a *single* target unit adopts the treatment of interest at a particular point in time, and then remains exposed to this treatment at all times afterwards. Both pre-treatment and post-treatment outcomes are assumed to be available. We ask whether we can infer the counterfactual trajectory over time had the unit not been exposed to the treatment using a population of control units never exposed to the treatment.

The synthetic control method (Abadie et al., 2010; Abadie & Gardeazabal, 2003) is one of the most important recent innovations in causal inference to solve this problem. It recognizes that a weighted combination of control units (instead of the standard procedure of seeking a single control or average in a neighbourhood of controls) often provides a more informative comparison for treatment effect estimation and then formalizes the selection of the comparison units using a data driven procedure.

Synthetic controls have become widely popular in the fields of policy analysis due to their simplicity and transparency. They have been used to assess the effect of tax hikes on the consumption of cigarettes (Abadie et al., 2010; Abadie, 2019), of drug programs on drug use and crime (Robbins et al., 2017), of immigration policies (Borjas, 2017; Bohn et al., 2014), of minimum wages (Allegretto et al., 2017), of terrorism on economic growth (Abadie & Gardeazabal, 2003), of large political changes and events (Hope, 2016), and also frequently in the biomedical sciences for estimating public health interventions (Pieters et al., 2016; Bouttell et al., 2018).

1.1. The synthetic control method

The typical setting considers n units $Y_i = (Y_{i,t_1}, \dots, Y_{i,t_m}) \in \mathbb{R}^m$, $i = 1, \dots, n$, each observed on the same grid of time points $t_1, \dots, t_m$. By convention, and without loss of generality, one unit ($i = 1$) receives the treatment or intervention at time $T \in (t_0, t_m)$ while the rest act as the control group. Let $Y^0_{i,t}$ be the potential outcome for unit i at time t in a hypothetical world where the intervention did not occur, and analogously let $Y^1_{i,t}$ be the corresponding potential outcome assuming the intervention did occur. Both are functions evaluated at time t.

The observed outcome for unit i at time t, denoted by $Y_{i,t}$, therefore satisfies:

$$\begin{aligned} Y_{1,t} &= Y^0_{1,t} + (Y^1_{1,t} - Y^0_{1,t})D_{1,t}, \\ Y_{i,t} &= Y^0_{i,t}, \qquad i = 2, \dots, n, \end{aligned} \tag{1}$$

where $D_{1,t}$ is a binary indicator of whether unit 1 is treated at time t, taken to be 0 at all times before treatment at time T and 1 at all time after time T. $\tau_1(t) = Y^1_1(t) - Y^0_1(t)$ is the causal effect of intervention on unit 1 at time t.

Synthetic control methods suppose that there exist weights $w_2, \dots, w_n$ such that $Y^0_{1,t}$ can be written as a weighted average of observed control outcomes: $Y^0_{1,t} = \sum_{i=2}^n w_i Y_{i,t}$, for $t \in [t_0, T)$ before treatment assignment. And then use this approximation to compute the causal effect of the

[*]Equal contribution [1]University of Cambridge, UK [2]Alan Turing Institute, UK [3]University of California Los Angeles, USA. Correspondence to: Alexis Bellot <alexis.b11@hotmail.com>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

[1]We use treatment, policy and intervention interchangeably.

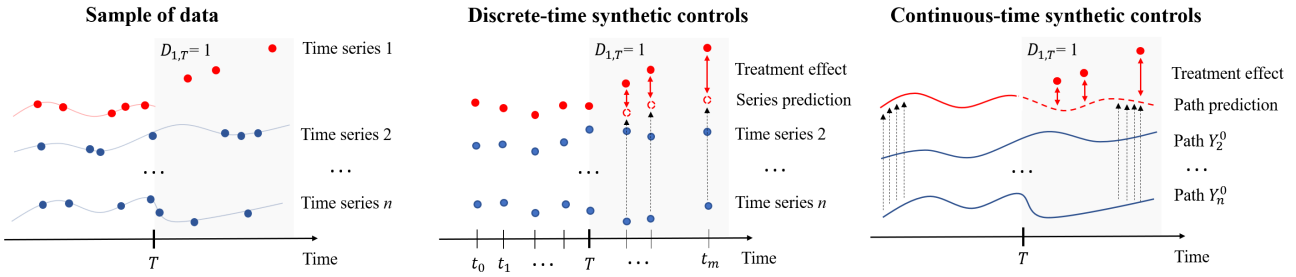


Figure 1. **Left panel**: Some data process is observed at misaligned observation times. The problem is to approximate the counterfactual trajectory of Time series 1 after time T. **Middle panel**: Previous work typically requires aligned observation times and synthetic controls $\sum_{i=2}^{n} w_i Y_{i,t}$ are defined to depend discretely and linearly on control observations. **Right panel**: In contrast, the proposed continuous-time synthetic control $Y_{1,T} + \int_T^t f(Y_{1,s}^0)\, d\mathbf{Y}_s^0$ depends continuously on control paths over time and naturally accommodates for misaligned observations and complex dynamics.

intervention:

$$\hat{\tau}_1(t) = Y_{1,t} - \sum_{i=2}^{n} w_i Y_{i,t}, \tag{2}$$

for every $t > T$.

The time series defined by $\sum_{i=2}^{n} w_i Y_{i,t}$ is the **synthetic control**.

It is called synthetic control because it is constructed such as to be representative of the treated unit ($i = 1$) had the treated unit not received treatment, and can be justified by assuming an underlying linear data generating mechanism for the data (with or without observed or unobserved confounders) (Abadie et al., 2010).

A large and varied set of strategies for the estimation of $(w_2, \ldots, w_n)$ have been proposed. We review some of these in section 4.

1.2. Limitations from a dynamical systems perspective

We seek to improve upon two limitations of discrete-time synthetic controls.

1. In reality, the time series $(Y_{i,t_1}, \ldots, Y_{i,t_m})$ is often assumed to be sequence of observations from an underlying *continuous* process. Synthetic controls may then be interpreted as a discrete approximation of the latent counterfactual path of the treated unit. However, this approximation typically breaks down if units are misaligned in time or irregularly sampled. An issue that may be solved only imperfectly by discarding information or interpolating the data.

2. Moreover, for complex problems, the *linearity* of control combinations may be restrictive. And, for general dynamical systems the assumption that the outcome correlations $(w_2, \ldots, w_n)$ are *static and invariant over time* is not plausible. For instance, the correlation between control and treated units may change over time if driven by weakly coupled dynamical systems. Examples of such systems arise in social sciences (Ranganathan et al., 2014), biology (Heltberg et al., 2019), and have been studied in the context of synthetic controls in (Ding & Toulis, 2020). For extrapolation over time to be consistent these dynamics should be captured.

1.3. Contributions

In this paper we take a different approach to synthetic control estimation rooted in the theory of dynamical systems.

We propose to model the synthetic control as the solution to a *controlled differential equation* (Lyons et al., 2007),

$$dY_{1,t}^0 = f\left(dY_{2,t}, \ldots, dY_{n,t}, Y_{1,t}^0\right), \quad Y_{1,t_0}^0 = y_{1,t_0}, \tag{3}$$

where y_{1,t_0} is an initial value and f is a latent vector field that is learnt from data and serves to combine control paths to approximate the counterfactual dynamics of the treated unit. In this context, at each time t, $Y_{1,t}^0$ describes the counterfactual state of the treated unit and it evolves as a function of its present state and the infinitesimal variation of control trajectories $Y_{2,t}, \ldots, Y_{n,t}$.

By integrating both sides, we construct a **continuous-time synthetic control** that is driven by a combination of the latent paths of control units. We thus retain the interpretation of equation (2) with the exception that we model a combination of the *latent infinitesimal variation* of control outcomes instead of the explicit *discrete-time observations* of control outcomes.

This model has three key features.

1. It is capable of processing irregularly aligned data and may be evaluated at any point over time.

2. f may be modelled in rich function spaces that may capture non-linearities and varying dependencies between units over time.

3. It may be trained efficiently with existing adjoint back-propagation algorithms and is easy to implement[2] – thereby offering a practical, fully non-parametric, and continuous-time alternative to existing synthetic control methods.

2. Problem formulation

This section extends the formulation of synthetic controls to a more general time series setting, illustrated in Figure 1.

Suppose that each latent path $Y_i : [t_0, t_m] \to \mathbb{R}^q$ is partially-observed through m irregular time series samples, $\{(t_0, Y_{i,t_0}), (t_1, Y_{i,t_1}), \ldots, (t_m, Y_{i,t_m})\}$, with each $t_j \in \mathbb{R}$ the timestamp of the observation $Y_{i,t_j} \in \mathbb{R}^q$, and $t_0 < \cdots < t_m$. q refers to the dimensionality of the outcome. To avoid notation clutter, the time subscript refers to function evaluation and the case where each i-th observation sequence has its own m_i irregular time stamps $t_{i,0}, \ldots, t_{i,m_i}$ will be described later.

Without loss of generality, only the counterfactual path of the first unit $Y^0_{1,t}$ after treatment assignment at time $t > T$ is of interest. As in equation (1), Y^0_1 is partially observed through discrete observations in the data up to time T and Y^1_1 is partially observed after time T. All other units are not administered treatment and act as control paths. Let $\mathbf{Y}^0 = (Y^0_2, \ldots, Y^0_n) : [t_0, t_m] \to \mathbb{R}^{(n-1)\times q}$ be the $(n-1) \times q$ dimensional path that includes all $n-1$ control paths.

We make the assumption that there exists a continuous function $f : \mathbb{R}^q \to \mathbb{R}^{q\times(n-1)}$ such that the counterfactual path of the treated unit $Y^0_1 : [t_0, t_m] \to \mathbb{R}^q$ is defined as the solution to the following controlled differential equation (CDE),

$$Y^0_{1,t} = Y^0_{1,t_0} + \int_{t_0}^{t} f(Y^0_{1,s})\, d\mathbf{Y}^0_s, \quad t \in (t_0, t_m], \tag{4}$$

where the integral is a Riemann–Stieltjes integral and "$f(Y^0_{1,s})\, d\mathbf{Y}^0_s$" is understood as matrix-vector multiplication (Kidger et al., 2020). (4) is the integral of (3) where the vector field f in (3) is taken to act linearly on $d\mathbf{Y}^0_s$. We say that Y^0_1 is controlled or driven by $\mathbf{Y}^0$, hence the name controlled differential equations.

Definition 1 *A continuous-time synthetic control, approximating the counterfactual path $Y^0_{1,t}$ is defined as*

$$Y^0_{1,T} + \int_{T}^{t} f(Y^0_{1,s})\, d\mathbf{Y}^0_s, \qquad t \in (T, t_m],$$

and may be interpreted as a non-linear continuous-time extension to the linear discrete-time synthetic control $\sum_{i=2}^{n} w_i Y^0_{i,t}$ of (Abadie et al., 2010).

Similarly, the causal effect at time $t > T$ of an intervention administered at time T can be estimated through:

$$\hat{\tau}_{1,t} = Y^1_{1,t} - Y^0_{1,T} - \int_{T}^{t} f(Y^0_{1,s})\, d\mathbf{Y}^0_s, \quad t \in (T, t_m],$$

The first term $Y^1_{1,t}$ is observed for $t \in (T, t_m]$ while the integral term is learned by optimizing f such as to approximate the (observed, through irregular samples) $Y^0_{1,t}$ for $t \in (t_0, T)$ before intervention at time T.

This strategy may be justified i.e., the treatment effect estimator is unbiased, in the case that f is linear, starting with an underlying linear data generating dynamical system, similarly to (Abadie, 2019). We show this in the Appendix. In the non-linear case, an estimator of f will typically be biased due to multiple different local minima.

2.1. Remarks

We make the following remarks as a comparison to the original synthetic control methodology.

- **Continuous-time.** The proposed formalism explicitly models observed sequences as processes evolving continuously in time. It therefore uses the full information of path observations and time intervals between observations and can be evaluated at any point in time, as illustrated in Figure 1.

- **Regularity of dynamics.** The implicit assumption is that the dynamics of the system are regular enough: the dynamics before time T can be extrapolated to the dynamics after time T. This is in contrast with the invariance in correlations at *all* times that the vector of weights in equation (2) specifies. In weakly coupled dynamical systems we know this assumption to break down with important consequences for the validity of the extrapolation of synthetic controls, as shown in (Ding & Toulis, 2020).

- **Latent state.** In realistic scenarios, it is often the case that observations are a function of an underlying *latent* state whose dynamics follow a differential equation e.g., a country's economy may evolve according to a differential equation although in practice only economics indicators are observed.

  Accordingly, one may define a latent state $g(Y^0_1) =: z_1 : [t_0, t_m] \to \mathbb{R}^h$ of the counterfactual path Y^0_1 as the solution to equation (4), with h the dimension of the latent state. The synthetic control is then the projection of this latent state onto the space of observations, just as indicators are a projection of the latent state of a country's economy. This formalism is described in section 3.

[2] Our implementation will be made available upon acceptance.

- **Transparency.** Synthetic controls are desirable also because of their accessible and transparent interpretation. Non-linearities inevitably trade-off some transparency for greater flexibility but we will see that we may regularize the solution space to promote sparsity in the control paths $\mathbf{Y}^0$ that influence the product "$f(Y^0_{1,s})\ d\mathbf{Y}^0_s$". As a result, only few controls drive the counterfactual path of interest, which may be inspected by the user for interpretation.

- **Misaligned observations.** Since only discrete observations are usually available, each path may be approximated in practice with natural cubic splines. One benefit is that this allows for irregular time stamps between units as each path may be interpolated independently.

3. Neural Continuous Synthetic Controls

In this section, we propose to parameterize f in equation (4) as a neural network with constraints on the sparsity of control path contributions – an instance of Neural CDEs (Kidger et al., 2020).

Neural CDEs are a family of continuous-time models that explicitly define the latent vector field f_θ by a neural network parameterized by θ, and allow for the dynamics to be modulated by the values of an auxiliary path over time. It generalizes the popular Neural ODE formulation of (Chen et al., 2018), whose dynamics in contrast are fully specified by an initial state, and may be implemented similarly by augmenting the vector field to vary as a function of $\mathbf{Y}^0_t$ (the set of control paths) as well as $Y_{1,t}$ (the treated path of interest).

Let $\tilde{Y}^0_i : [t_0, t_m] \to \mathbb{R}^q$ be the natural cubic spline with knots at $t_0, \dots, t_m$ (or more generally at $t_{i,0}, \dots, t_{i,n_i}$) such that $\tilde{Y}^0_{i,t_j} = (Y^0_{i,t_j}, t_j)$, for $j = 0, \dots, m$. As we observe only a discretization of the underlying process, $\tilde{Y}^0_i$ is an approximation for which derivatives may be easily computed.

Let $g_\eta : \mathbb{R}^q \to \mathbb{R}^l$ be a neural network that embeds the observations into a l-dimensional latent state $z_t := g_\eta(Y^0_{1,t})$. Let $f_\theta : \mathbb{R}^l \to \mathbb{R}^{(n-1)\times l}$ be a neural network parameterizing the latent vector field and let and $h_\nu : \mathbb{R}^l \to \mathbb{R}^q$ be a neural network that defines the observation mechanism projecting the latent state into the observation space to recover an estimate of the counterfactual path $\hat{y}_{1,t} := h_\nu(z_t)$.

We generalize our problem formulation to assume that a *latent* path z_t (instead of the actual observed $Y^0_{1,t}$) can be expressed as the solution to a controlled differential equation of the form

$$z_t = z_{t_0} + \int_{t_0}^{t} f(z_s)\ d\mathbf{Y}^0_s, \quad t \in (t_0, t_m]. \tag{5}$$

3.1. Interpretability via sparse contributions

Arguably one of the reasons for the success of synthetic controls is their natural interpretation as a weighted, sparse combination of control paths that can be inspected by the user.

While non-linearities inevitably make the resulting fit more complex and less interpretable, one may enforce sparsity by explicitly including a weighted diagonal matrix that restricts the contribution of control paths. This extension defines the latent counterfactual state as a solution to

$$z_t = z_{t_0} + \int_{t_0}^{t} f(z_s)\ \mathbf{W} d\mathbf{Y}^0_s, \quad t \in (t_0, t_m], \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{(n-1)\times(n-1)}$ is a time-independent diagonal matrix of trainable parameters that are optimized subject to an l_1 penalty on its values to encourage sparsity. It is clear then that Y^0_1 is independent of Y^0_{i+1} if $[\mathbf{W}]_{ii} = 0$, for $i = 1, \dots, n-1$. In fact, we have the following proposition, which implies that this constraint precisely identifies the set of CDEs that are independent of Y^0_i:

Proposition 1. *Consider the class of CDEs $\mathcal{C}$ defined by (5) that are independent of Y^0_i and the class of CDEs $\mathcal{C}_0$ defined by (6) such that the i-th diagonal entry of $\mathbf{W}$ is zero. Then $\mathcal{C} = \mathcal{C}_0$.*

Proof. Given in the Appendix.

This proposition provides a rigorous way to enforce that a Neural CDE approximation depends only on a few control paths, and implies also that there is no loss of expressivity or approximating power in the estimation of CDEs that are independent of some control paths using this parameterization.

3.2. Algorithm

For each estimate of f_θ and g_η the forward latent trajectory in time that these functions define through (6) can be computed using any numerical ODE solver:

$$\hat{z}_{t_1}, \dots, \hat{z}_{t_m} = \texttt{ODESolve}(f_\theta, \hat{z}_{t_0}, (t_1, \mathbf{Y}^0_{t_1}), \dots, (t_m, \mathbf{Y}^0_{t_m})). \tag{7}$$

The goodness of f_θ, g_η, and h_ν, is then quantified by a loss function $\mathcal{L} : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ that compares the reconstructed signal $\hat{y}_{1,t} := h_\nu(\hat{z}_t)$ with the observed trajectory values $y_{1,t}$ for $t < T$ (before treatment assignment). Gradients with respect to θ may be computed with adjoint sensitivities treating the ODE solver as a black-box and outputting the predicted state of the system $\hat{z}(t)$ at multiple times, as described by (Chen et al., 2018; Kidger et al., 2020).

The optimization problem is defined as

$$\arg\min_{\theta,\eta,\nu,\mathbf{W}} \quad \mathcal{R}(f_\theta, g_\eta, h_\nu, \mathbf{W}), \tag{8}$$

$$\mathcal{R}(f_\theta, g_\eta, h_\nu, \mathbf{W}) := \sum_{i:t_i<T} \mathcal{L}\left(y_{1,t_i}, \hat{y}_{1,t_i}\right) + \lambda \sum_{j=1}^{n-1} |\mathbf{W}_{jj}|,$$

where $\lambda > 0$ is a hyperparameter and $|\cdot|$ denotes the absolute value.

We call this algorithm for estimating the counterfactual trajectory of a treated unit the Neural Continuous Synthetic Control method (**NC-SC**).

4. Related work

In light of the remarks and methods presented above, in this section we review the methodological body of work that has extended the original proposal of Abadie et al. (Abadie et al., 2010), we review the literature on treatment effect estimation with structural models and we review other methods for differential equation modelling.

Synthetic controls. A particular choice for estimating weights depending on data structure and model assumptions is often the distinguishing feature of recent synthetic control methods. For instance, (Doudchenko & Imbens, 2016) propose to use negative weights and intercept terms, (Amjad et al., 2018; Chernozhukov et al., 2017) propose regularization terms to promote sparsity and robustness, (Ding & Toulis, 2020) propose time-varying weights to model changing correlations between variables and (Athey et al., 2018) interpret counterfactual estimation as a matrix completion problem regularized with matrix norms. However, whilst often improving goodness of fit, matching in discrete time remains difficult with irregularly aligned data i.e., unit observations not aligned in time.

Structural models. The synthetic control literature is often discussed in contrast to structural time series methods, that explicitly fit the trajectory of counterfactual outcomes using lagged outcomes and covariates. These include the g-computation formula, marginal structural models (Robins et al., 2000; Cole & Hernán, 2008) and several flexible extensions using neural networks and Gaussian processes including (Bica et al., 2020; Soleimani et al., 2017; Schulam & Saria, 2017). One contrast is that structural methods *balance* distributions between treated and control units (relying on regularity of the treated unit trajectory over time to extrapolate counterfactual estimates) while synthetic controls *match* treated units to control units (relying on regularity across units to extrapolate counterfactual estimates).

There is an important contrast also in the data requirements of these two approaches. For accurate extrapolation structural models require a large number of control paths and many covariates to approximate the underlying causal structure, while synthetic controls require a large number of path observations. Many applications where synthetic controls have proven successful (mostly with $20-40$ control paths, complex dynamics and hidden variables) are inherently not amenable to structural modelling.

Differential equation modelling. Recently, differential equation models for irregular time series data are increasingly commonplace in the machine learning literature. Of note are Neural Ordinary Differential Equations (ODEs) (Chen et al., 2018), several extensions that modulate the trajectory of interest with incoming data (Rubanova et al., 2019; De Brouwer et al., 2019), and many other proposals that extend the design of vector fields (Dupont et al., 2019; Zhang et al., 2020; Chen et al., 2020), improve optimization performance (Li et al., 2020) and incorporate processes driven by stochastic noise (Tzen & Raginsky, 2019). In the context of causality, (De Brouwer et al., 2021; Bellot et al., 2021) study causal discovery in continuous-time using Neural ODEs and (Soleimani et al., 2017) model counterfactuals in continuous-time using specialized pharmacokinetic data generating systems for applications in medicine. This paper in contrast proposes the first continuous-time formalism for synthetic control estimation.

Control in reinforcement learning. The "continuous control" terminology is also found in the reinforcement learning literature to designate physical control tasks with continuous (real valued) action spaces (Lillicrap et al., 2015). This is different and not to be confused with the causality definition where the term refers to absence of treatment.

5. Experiments

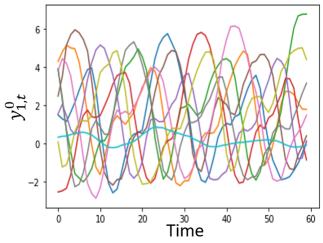
In this section we experiment with synthetic data from Lorenz's chaotic dynamical system and discuss 2 studies that have received attention in the public policy literature.

Evaluation metric. In all experiments, we report mean and standard deviations of the *control* estimation error

$$\frac{1}{|\mathcal{T}|} \sum_{t\in\mathcal{T}} ||y^0_{1,t} - \hat{y}^0_{1,t}||_2^2, \tag{9}$$

over 10 model runs, where for real data $\mathcal{T} = \{t_i : t_i < T, i = 2, \dots, n\}$ is the pre-intervention observation times (where untreated data is observed). For synthetic data we use all observation times $\mathcal{T} = \{t_1, \dots, t_m\}$ for evaluation (as we are free to generate any amount of data).

Evaluation methods. Comparisons are made with the original approach of (Abadie et al., 2010) with weights constrained to be non-negative and summing to 1 (**SC**), with an extension that instead matches pre-treatment outcomes in a reproducing kernel Hilbert space using an instance of kernel mean matching (Gretton et al., 2009) (**KMM-SC**),



(a) Sample control paths.

	Aligned	30% **Dropped**	50% **Dropped**	70% **Dropped**
SC	.0521 (.003)	.0541 (.003)	.0545 (.003)	.0552 (.003)
KMM-SC	.0531 (.003)	.0564 (.004)	.0566 (.005)	.0589 (.005)
R-SC	.0510 (.002)	.0533 (.004)	.0560 (.004)	.0555 (.004)
MC-SC	.0552 (.003)	.0570 (.003)	.0595 (.002)	.0637 (.003)
NC-SC (ours)	.0484 (.003)	.0488 (.003)	.0496 (.003)	.0490 (.003)

(b) Counterfactual estimation performance (lower better) on aligned and mis-aligned observations.

Figure 2. Experiments on Lorenz's model.

with robust synthetic controls (**R-SC**) that use penalized weight estimation as in (Doudchenko & Imbens, 2016), and with the matrix completion (**MC-SC**) approach of (Athey et al., 2018).

For experiments involving misaligned data, all methods except NC-SC require some form of prior interpolation and evaluation on a regular grid of time points. Here we use cubic spline interpolations with knots at observation times and smoothing parameter chosen visually for good fit.

Precise experimental details including neural network architectures, optimization software, implementation details and data sources may be found in the Appendix.

5.1. Lorenz's chaotic model

We begin by demonstrating the efficacy of NC-SC on **irregularly aligned time series** from Lorenz's model for chaotic dynamical systems (Lorenz, 1996).

The dynamics in a d-dimensional Lorenz model are

$$\frac{d}{dt}x_i(t) = (x_{i+1}(t) - x_{i-2}(t)) \cdot x_{i-1}(t) - x_i(t) + F,$$

for $i = 1, \ldots, d$, where $x_{-1}(t) := x_{d-1}(t)$, $x_0(t) := x_p(t)$, $x_{d+1}(t) := x_1(t)$ and F is a treatment variable that has the effect of changing the level of non-linearity and chaos in the series. We take $F = 5$ (mild chaotic behaviour) as the baseline control behaviour and $F = 10$ to define the dynamics of the treated regime. The initial state of each variable is sampled from a standard Gaussian distribution and d is set to 10.

Experiment design. For simplicity, only the counterfactual trajectory of the *first* dimension of the system is of interest, $y^0_{1,t} := x_1(t)$, while control trajectories are each similarly defined but with *different* random initializations of Lorenz's model. That is, $y^0_{2,t} := x_1(t)$ with some random initialization, $y^0_{3,t} := x_1(t)$ with some different random initialization and so on (this is equivalent to having units with different features in static models). We set the number of control paths to 20.

The problem is to construct a synthetic control for the treated unit had $F = 5$ for all t, given that we observed 200 time observations ($t < 200$) with $F = 5$ before treatment assignment at time $T = 200$.

Sequences of observations from these paths are observed in two configurations.

1. *Regularly aligned* with a fixed grid of observation times.
2. *Irregularly aligned* by removing randomly 30%, 50% and 70% of the aligned data, independently for each unit.

Results. Performance is computed on a held-out segment of the data (extrapolating the counterfactual path of the treated unit over $t \in (200, 400)$). Performance results, as well as an illustration of control paths is given in Figure 2. Continuous-time synthetic controls outperform every other model considered and furthermore have relatively stable performance with irregular data while other methods exhibit a decrease in performance which we hypothesize is due to worsening imputation performance.

	Smoking	**Eurozone**
SC	.0248 (.00)	.0339 (.00)
KMM-SC	.0221 (.00)	.0321 (.00)
R-SC	.0002 (.00)	.0230 (.00)
MC-SC	.0005 (.00)	.0299 (.00)
NC-SC (ours)	.0001 (.00)	.0003 (.00)

Table 1. Counterfactual estimation performance (lower better).

5.2. The Eurozone and current account deficits

Next, we consider an experiment that further highlights the need for **non-linear combinations** of control paths to accurately approximate the control path of the treated unit.

Experiment design. The problem is to evaluate the impact of Eurozone membership on the path of current account deficits, thought to have considerably aggravated the recovery after the 2007-2008 financial crisis.

By the end of 2009, Europe was at the beginning of a multi-year sovereign debt crisis, in which several Eurozone mem-

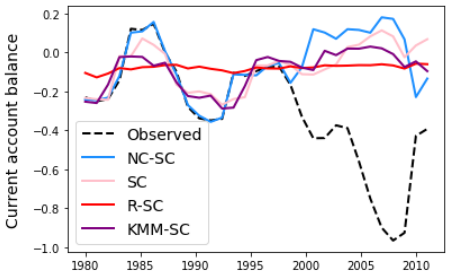


Figure 3. Counterfactual current account predictions for Spain $\hat{Y}^0_{1,t}$ over time. The euro was made effective in 1998.

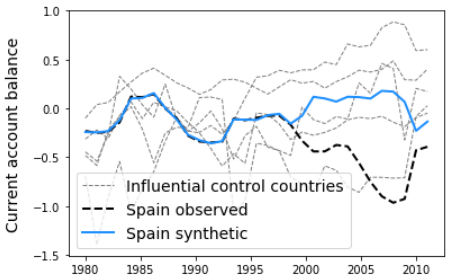


Figure 4. Counterfactual current account balance predictions for Spain $\hat{Y}^0_{1,t}$ over time and contrast with the trajectory of the most influential control countries. The influential control countries are Chile, Hungary, Japan, Mexico and Sweden.

bers were unable to repay their government debt or to bail out over-indebted banks. The eurozone crisis is thought to have been caused in part by a sudden stop of foreign capital investments into countries that had substantial deficits, fueled by low borrowing costs as a consequence, arguably, of Eurozone membership (Frieden & Walter, 2017). One may ask then whether there is any evidence for this claim, whether or not a country's current account deficits would have been different had it not join the Eurozone.

We focus on one country, Spain. The data consists of yearly current account figures from 1980 to 2010 for Spain as well as 15 other countries outside the Eurozone, as collected by David Hope in (Hope, 2016). The pre-treatment period ranges from 1980 to 1998 when the Eurozone was made effective (as illustrated in Figure 3).

Results. Performance results are given in Table 1. These demonstrate that NC-SC can substantially improve performance in this case, as also illustrated in the model fit to the observed control trajectory in Figure 3.

Figure 4 shows which other countries were influential in determining the synthetic control for Spain, inferred by inspecting the non-zero entries of $\mathbf{W}$. NC-SC uses combinations of current account balance figures from Chile, Hungary, Japan, Mexico and Sweden, in contrast to those of Great Britain, Israel, Mexico and Sweden used by the original synthetic control method (Abadie et al., 2010). Interestingly, as a result, the projection given by NC-SC gives a slightly different interpretation estimating a positive current account balance had Spain not adopted the Euro in contrast to a zero current account balance given by (Abadie et al., 2010).

5.3. Smoking control in California

Next, we consider one of the most popular benchmarks for synthetic control estimation, namely the evaluation of the effect of the influential 1988 anti-smoking legislation in California on cigarette sales.

Experiment design. At that time, California lead a wave of anti-smoking legislation, known as Proposition 99, that served as a model for policy interventions in other states later on and arguably reduced the prevalence of smoking. The problem is to assess its effect in comparison to California's cigarette sales had the legislation not been passed.

We follow the experiment by (Abadie et al., 2010) and use annual state-level panel data for the period 1970-2000, giving us 19 years of pre-intervention cigarette sales data.

Results. Performance comparisons are given in Table 1 and the corresponding fit and treatment effect (as the difference between the counterfactual estimation and observed trajectory) is illustrated in Figure 5. Continuous-time synthetic controls, as well as baseline methods match almost exactly the pre-treatment trajectory of the treated unit and all counterfactual projections point towards an important treatment effect. The California anti-smoking legislation was responsible for part of the lowering of cigarette sales.

We show in addition the contribution of each state to the synthetic control in Figure 6, inferred by inspecting the non-zero entries of $\mathbf{W}$. In this case there is little contrast with existing baselines; most methods use Nevada, Utah, Montana, Colorado and Connecticut as the most influential states for the construction of synthetic controls which serves to confirm the estimates of NC-SC.

6. Discussion

We conclude with some additional remarks and clarifications that may be of practical importance.

- **On the role of covariates.** Auxiliary covariates have not played a role in the development of continuous-time synthetic controls. While (Abadie et al., 2010) demonstrate the treatment effect to be asymptotically unbiased under a perfect match on both pre-treatment outcomes and relevant covariates (among other conditions) this is not strictly necessary as long as sufficient pre-treatment outcomes are observed (see Theorem 1 (Botosaru & Ferman, 2019)). The intuition behind this result is that it

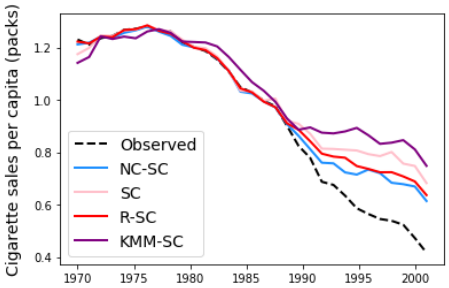


Figure 5. Counterfactual cigarette sales predictions for California $\hat{Y}^0_{1,t}$ over time. Anti-smoking legislation was introduced in 1988.

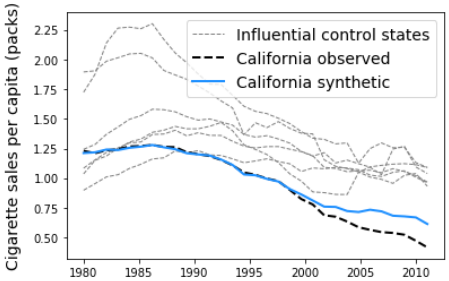


Figure 6. Counterfactual cigarette sales predictions for California $\hat{Y}^0_{1,t}$ over time and contrast with the trajectory of the most influential control states.

would not be possible to match on a large number of pre-treatment outcomes without matching on both observed and unobserved relevant covariates.

However, if desired, matching on time-variant or time-invariant covariates to define continuous-time synthetic controls with our formalism is possible and straightforward by using a data-dependent lasso regularization scheme on the matrix $\mathbf{W}$. Instead of penalizing all entries of $\mathbf{W}$ equally, one may adopt a relevance weighting approach $\sum_{i=1}^{n-1} \frac{1}{\hat{p}_i}|\mathbf{W}_{ii}|$, where $\hat{p}_i > 0$ measures the relevance of the i-th control covariates towards matching the covariates of the treated unit (higher values of $\hat{p}_i$ corresponding to more relevant control paths in this case).

For example, with n units of d-dimensional static covariates $\mathbf{X} \in \mathbb{R}^{n \times d}$ observed along the outcome paths of interest, $\hat{p} = (\hat{p}_2, \ldots, \hat{p}_n)$ may be defined by the linear projection of $\mathbf{X}_1$ onto $\mathbf{X}_{2:n}$, $\hat{p} = (\mathbf{X}_{2:n}^T \mathbf{X}_{2:n})^{-1} \mathbf{X}_{2:n}^T \mathbf{X}_1$.

- **Control population.** The similarity between the control population and the treated unit in the pre-treatment period and the posterior counterfactual trajectory underlies the validity of synthetic controls. The choice of control population is therefore important and there are two important requirements that must be met.

  A first requirement is that control units not be affected by the intervention or treatment of interest so that they faithfully describe the counterfactual trajectory of the treated unit. This assumption has to be justified and is not necessarily always plausible. For instance one may argue that public policy interventions have spillover effects e.g., cigarette sales of neighbouring control states to California being affected by anti-smoking legislation or the economy of control countries with strong commercial ties to Eurozone members influenced by the monetary union, in which case counterfactual estimates will be biased. In these particular two examples, this assumption however has been carefully justified (Abadie et al., 2010; Hope, 2016).

  A second requirement is that units, after the intervention, not be subject to large idiosyncratic shocks that would not have affected the treated unit in the absence of treatment as such control units would not remain representative of the counterfactual trajectory. This assumption relates to the regularity of the correlations between control and treated units over time, which holds if a common underlying causal model for the data can plausibly be assumed.

- **Data requirements.** The credibility of synthetic controls hinges on the accuracy of pre-treatment control path approximation. Therefore a sizeable number of pre-treatment observations should be available for valid extrapolations. This is perhaps in contrast with structural models that require observation of all covariates and a larger number of control paths for accurate extrapolation.

  We have omitted an explicit comparison with structural models because of this practical difference. If not all variables in the data generating mechanism are observed, as in our experiments (for instance, we do not have any information on the driving forces that determine the current account balance in the Eurozone experiment), it is not plausible to fit a structural equation model.

- **Uncertainty estimation**. As presented here, continuous synthetic controls do not explicitly quantify uncertainty in counterfactual estimation. Such extensions are feasible given that stochastic differential equations (SDEs) can be expressed as a controlled differential equation driven by a stochastic process, and given existing work on back-propagating through SDE solvers (Liu et al., 2019; Kong et al., 2020) that may be used with a neural vector field in analogy to Neural ODEs.

7. Conclusion

This paper demonstrates how synthetic control estimation (Abadie et al., 2010) may be extended to continuous-time using the mathematics of controlled differential equations (Lyons et al., 2007).

Our proposal for counterfactual estimation, called Neural Continuous Synthetic Controls, models explicitly the latent paths of the observed time series defining a synthetic control

as a combination of paths rather than as a combination of discrete observations. Neural Continuous Synthetic Controls are conceptually natural for modelling processes continuously unfold over time and accommodate for irregularly aligned data and more complex dynamics than previously analysed.

Acknowledgements

We thank the anonymous reviewers for valuable feedback. This work was supported by the Alan Turing Institute under the EPSRC grant EP/N510129/1, the ONR and the NSF grants number 1462245 and number 1533983.

References

Abadie, A. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 2019.

Abadie, A. and Gardeazabal, J. The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132, 2003.

Abadie, A., Diamond, A., and Hainmueller, J. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

Allegretto, S., Dube, A., Reich, M., and Zipperer, B. Credible research designs for minimum wage studies: A response to neumark, salas, and wascher. *ILR Review*, 70 (3):559–592, 2017.

Amjad, M., Shah, D., and Shen, D. Robust synthetic control. *The Journal of Machine Learning Research*, 19(1):802–852, 2018.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. Matrix completion methods for causal panel data models. Technical report, National Bureau of Economic Research, 2018.

Bellot, A., Branson, K., and van der Schaar, M. Consistency of mechanistic causal discovery in continuous-time using neural odes. *arXiv preprint arXiv:2105.02522*, 2021.

Bica, I., Alaa, A. M., Jordon, J., and van der Schaar, M. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. *arXiv preprint arXiv:2002.04083*, 2020.

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

Borjas, G. J. The wage impact of the marielitos: A reappraisal. *ILR Review*, 70(5):1077–1110, 2017.

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

Bouttell, J., Craig, P., Lewsey, J., Robinson, M., and Popham, F. Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Community Health*, 72(8):673–678, 2018.

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. In *Advances in neural information processing systems*, pp. 6571–6583, 2018.

Chen, R. T., Amos, B., and Nickel, M. Learning neural event functions for ordinary differential equations. *arXiv preprint arXiv:2011.03902*, 2020.

Chernozhukov, V., Wuthrich, K., and Zhu, Y. An exact and robust conformal inference method for counterfactual and synthetic controls. *arXiv preprint arXiv:1712.09089*, 2017.

Cole, S. R. and Hernán, M. A. Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, 168(6):656–664, 2008.

De Brouwer, E., Simm, J., Arany, A., and Moreau, Y. Gru-ode-bayes: Continuous modeling of sporadically-observed time series. In *Advances in Neural Information Processing Systems*, pp. 7379–7390, 2019.

De Brouwer, E., Arany, A., Simm, J., and Moreau, Y. Latent convergent cross mapping. 2021.

Ding, Y. and Toulis, P. Dynamical systems theory for causal inference with application to synthetic control methods. In *International Conference on Artificial Intelligence and Statistics*, pp. 1888–1898. PMLR, 2020.

Doudchenko, N. and Imbens, G. W. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research, 2016.

Dupont, E., Doucet, A., and Teh, Y. W. Augmented neural odes. In *Advances in Neural Information Processing Systems*, pp. 3140–3150, 2019.

Frieden, J. and Walter, S. Understanding the political economy of the eurozone crisis. *Annual Review of Political Science*, 20:371–390, 2017.

Gretton, A., Smola, A., Huang, J., Schmittfull, M., Borgwardt, K., and Schölkopf, B. Covariate shift by kernel mean matching. *Dataset shift in machine learning*, 3(4): 5, 2009.

Heltberg, M. L., Krishna, S., and Jensen, M. H. On chaotic dynamics in transcription factors and the associated effects in differential gene regulation. *Nature communications*, 10(1):1–10, 2019.

Hope, D. Estimating the effect of the emu on current account balances: A synthetic control approach. *European Journal of Political Economy*, 44:20–40, 2016.

Kidger, P., Morrill, J., Foster, J., and Lyons, T. Neural controlled differential equations for irregular time series. *arXiv preprint arXiv:2005.08926*, 2020.

Kong, L., Sun, J., and Zhang, C. Sde-net: Equipping deep neural networks with uncertainty estimates. *arXiv preprint arXiv:2008.10546*, 2020.

Li, X., Wong, T.-K. L., Chen, R. T., and Duvenaud, D. K. Scalable gradients and variational inference for stochastic differential equations. In *Symposium on Advances in Approximate Bayesian Inference*, pp. 1–28, 2020.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Liu, X., Si, S., Cao, Q., Kumar, S., and Hsieh, C.-J. Neural sde: Stabilizing neural ode networks with stochastic noise. *arXiv preprint arXiv:1906.02355*, 2019.

Lorenz, E. N. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.

Lyons, T. J., Caruana, M., and Lévy, T. *Differential equations driven by rough paths*. Springer, 2007.

Pieters, H., Curzi, D., Olper, A., and Swinnen, J. Effect of democratic reforms on child mortality: a synthetic control analysis. *The Lancet Global Health*, 4(9):e627–e632, 2016.

Ranganathan, S., Spaiser, V., Mann, R. P., and Sumpter, D. J. Bayesian dynamical systems modelling in the social sciences. *PloS one*, 9(1):e86468, 2014.

Robbins, M. W., Saunders, J., and Kilmer, B. A framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention. *Journal of the American Statistical Association*, 112(517):109–126, 2017.

Robins, J. M., Hernan, M. A., and Brumback, B. Marginal structural models and causal inference in epidemiology, 2000.

Rubanova, Y., Chen, R. T., and Duvenaud, D. Latent odes for irregularly-sampled time series. *arXiv preprint arXiv:1907.03907*, 2019.

Schulam, P. and Saria, S. Reliable decision support using counterfactual models. In *Advances in Neural Information Processing Systems*, pp. 1697–1708, 2017.

Soleimani, H., Subbaswamy, A., and Saria, S. Treatment-response models for counterfactual reasoning with continuous-time, continuous-valued interventions. *arXiv preprint arXiv:1704.02038*, 2017.

Tzen, B. and Raginsky, M. Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. *arXiv preprint arXiv:1905.09883*, 2019.

Zhang, H., Gao, X., Unterman, J., and Arodz, T. Approximation capabilities of neural odes and invertible residual networks. In *International Conference on Machine Learning*, pp. 11086–11095. PMLR, 2020.