# Bayesian Predictive Synthetic Control Methods

**Akira Fukuda** [1] **Masahiro Kato** [2] **Kenichiro McAlinn** [3] **Kosaku Takanashi** [4]

## Abstract

We study Bayesian model synthesis for synthetic control methods (SCMs). SCMs have garnered significant attention as an indispensable tool for comparative case studies. The fundamental concept underlying SCMs involves the prediction of counterfactual outcomes for a treated unit by a weighted summation of observed outcomes from untreated units. In this study, we reinterpret the untreated outcomes as predictors for the treated outcomes and employ Bayesian predictive synthesis (BPS) to synthesize these forecasts. We refer to our novel approach as Bayesian Predictive SCM (BPSCM). The BPSCM represents a comprehensive, and foundational framework encompassing diverse statistical models, including dynamic linear models and mixture models, and generalizes SCMs significantly. Moreover, our proposal possesses the capability to synthesize a range of predictive models utilizing covariates, such as random forests. From a statistical decision-making perspective, our method can be interpreted as a Bayesian approach aimed at minimizing regrets in the prediction of counterfactual outcomes. Additionally, Bayesian approaches can effectively address challenges encountered in frequentist SCMs, such as statistical inference with finite sample sizes, time-varying parameters, and model misspecification. Through the utilization of simulation examples and empirical analysis, we substantiate the robustness of our proposed BPSCM.

## 1. Introduction

Synthetic control methods (SCMs, Abadie & Gardeazabal, 2003; Abadie et al., 2010) have garnered attention as essential tools for causal inference in comparative case studies across a range of fields, including economics, statistics, and machine learning.

SCMs are applicable in scenarios involving multiple units, where some units are subject to a policy intervention while others are not. Our interest lies in estimating the causal effect of the treated units by using respective outcomes, measured in each population before and after the policy intervention, potentially across numerous periods. Control units are utilized to account for unobserved trends in the outcome over time that are not associated with the policy intervention effect. The principle behind SCMs is that an appropriately weighted average of available control units, termed the *synthetic control (SC) unit*, often offers a more suitable comparison than a solitary control group (Abadie, 2021). This method has been implemented in various settings, such as in the analysis of tobacco control programs, economic effects of terrorism and German reunification, and the decriminalization of indoor prostitution. For further details, see Abadie (2021) and Cunningham (2021).

While original SCMs have been proposed as frequentist methods, Kim et al. (2020) points out that the original SCMs have at least three limitations: (1) restrictive constraints, (2) no inference theory, and (3) lack of an explicit mechanism for the "large $N$, small $T$" problem. The authors argue that those problems can be avoided by proposing Bayesian SCMs (BSCMs) [1]. Another issue is model misspecification because existing guarantees for synthetic control are typically derived under a linear factor model or a vector autoregressive model of the outcomes (Abadie et al., 2010; Ben-Michael et al., 2019; 2021; Ferman & Pinto, 2021; Viviano & Bradic, 2019). While the guarantees formally hold under these outcome models, there is a wide sense of optimism that the synthetic control method is robust to these modeling assumptions. For instance, (Ben-Michael et al., 2019) write, "Outcome modeling can also be sensitive to model mis-specification, such as selecting an incorrect number of factors in a factor model. Finally, [...] synthetic control can be appropriate under multiple data generating processes (e.g., both the autoregressive model and the linear factor model) so that it is not necessary for the applied researcher to take a strong stand on which is

[1]Takushoku University [2]The University of Tokyo [3]Temple University [4]Riken. Correspondence to: Masahiro Kato <mkato.csecon@gmail.com>.

*ICML (International Conference on Machine Learning) Workshop on Counterfactuals in Minds and Machines*, Honolulu, Hawaii, USA. 

[1]We refer to all SCMs based on Bayesian modeling as BSCMs, including our proposed methods.

correct." Abadie & i Bastida (2022) write, "Synthetic controls are intuitive, transparent, and produce reliable estimates for a variety of data generating processes." Shi et al. (2022) investigates implicit assumptions in the SC model of Abadie (2002). Based on this finding, Nazaret et al. (2023) proposes a SCMs that are robust against model misspecification. In BSCMs, Klinenberg (2022) proposes using time-varying coefficients to capture dynamics of outcomes. Finally, we point out an issue on incorporation of covariates.

This study proposes a Bayesian Predictive SCM (BPSCM) as a generalization of BSCMs, and develop a special case with time-varying coefficients. We propose our SCMs by applying Bayesian predictive synthesis (BPS, McAlinn & West, 2019; McAlinn et al., 2020). BPS is a general, coherent framework that incorporates information from different sources and synthesize them to improve predictions a target variable. In this study, by regarding outcomes of untreated units as information, we predict counterfactual outcome of the treated unit. By extending this basic formulation, we further propose incorporating predictors constructed based on covariates using various predictive models, such as random forests and neural networks. Under the framework of BPSCMs, we can posit various Bayesian models. Among them, we introduce dynamic linear latent models and mixture models from McAlinn & West (2019) and Johnson & West (2016), respectively. We confirm the soundness of the proposed methods by using simulation and empirical studies.

The BPSCMs offer three primary advantages. Firstly, akin to other BSCMs, BPSCMs benefit from the advantages of Bayesian modeling. For instance, BPSCMs can conduct Bayesian inference using posterior distributions even with a small sample size. Secondly, our method can be justified as Bayesian regret minimization in online learning. Thirdly, even under model misspecification, BPSCMs can either perform minimax predictions or appropriately utilize prior information to address the issue. Lastly, BPSCMs can flexibly employ covariate information by transforming it into outcome predictions. For example, we can use various machine learning methods to predict outcomes based on covariates and subsequently incorporate these predictions using BPS.

Among recent studies on frequentist SCMs, our work has close connections to Viviano & Bradic (2019) and Chen (2022). Viviano & Bradic (2019) proposes an ensemble scheme to aggregate predictions from multiple predictive models, which can include synthetic control, interactive fixed effect models, and random forests. Utilizing results from the online learning literature, the ensemble scheme in Viviano & Bradic (2019) possesses the no-regret property, making the ensemble predictions competitive against the prediction of any fixed predictive model in the ensemble. Under sampling processes that yield good performance for some predictive model in the ensemble, Viviano & Bradic (2019) derives performance guarantees for the ensemble learner. On the other hand, Chen (2022) examines SCMs directly in the worst-case setting, connecting guarantees in this setting to assurances on statistical risk in a design-based framework.

## 2. Problem Setting

We consider the following setting of SCMs, as considered in Abadie & Gardeazabal (2003). Suppose that there are $J$ units, $j \in \mathcal{J} := \{0, 1, 2, \dots, J\}$ and a time series $t \in \mathcal{T} := \{1, \dots, T\}$ for $T \in \mathbb{N}$ such that $T > 2$.

Without loss of generality, we assume that the first unit ($j = 0$) is the treated unit, that is, the unit affected by the policy intervention of interest. A set of potential comparisons, $j \in \mathcal{J}^U = \mathcal{J} \backslash \{0\}$, is a collection of untreated units not affected by the intervention. We assume also that our data span $T$ periods and that the intervention occurs at $t = T_0$ such that $1 < T_0 < T$. The intervention is also referred to as intervention. Observations at a first period $t \in \mathcal{T}_0 := \{1, \dots, T_0\}$ are ones before the intervention. Let us also define a set of the post-intervention periods as $\mathcal{T}_1 := \mathcal{T} \backslash \mathcal{T}_0$ and denote the size by $T_1 := |\mathcal{T}_1| = T - T_0$.

### 2.1. Potential Outcomes

For each unit $j \in \mathcal{J}$ and period $t \in \mathcal{T}$, there is a potent outcome of interest, $y^I_{j,t}, y^N_{j,t} \in \mathbb{R}$, where $y^I_{j,t}$ denotes a potential outcome under the intervention and $y^N_{j,t}$ denotes a potential outcome without the intervention.

### 2.2. Observations

For each unit $j \in \mathcal{J}$, we can only observe one of the outcomes corresponding to actual intervention; that is, we observe $y_{j,t} \in \mathbb{R}$ such that

$$y_{0,t} = \begin{cases} y^N_{0,t} & \forall t \in \mathcal{T}_0 \\ y^I_{0,t} & \forall t \in \mathcal{T}_1 \end{cases}, \quad y_{j,t} = y^N_{j,t} \ \ \forall j \in \mathcal{J}^U, \ \forall t \in \mathcal{T}.$$

### 2.3. Causal Effects

We are interested in causal effects represented by the following treatment effects

$$\tau_{0,t} = y^I_{0,t} - y^N_{0,t}.$$

Because unit $j = 0$ is exposed to the intervention after period $T_0$, it follows that for $t > T_0$ we have $y_{0,t} = y^I_{0,t}$. It means that by definition, $y^N_{0,t}$ is a counterfactual outcome because of the intervention after $t = T_0$.

The goal of SCMs is to conduct inference for the treatment effects $\tau_{0,t}$ by predicting $y^N_{0,t}$ for $t > T_0$, which is a counterfactual outcome that would have been observed for the treated unit in the absence of the intervention. We usually use one unit or a small number of units that have similar characteristics as the treated unit at the time of the intervention but not exposed to the intervention. However, when the data consist of a few units, such as regions or countries, it is often difficult to find a single untreated unit that provides a suitable comparison for the unit treated by the policy intervention of interest.

In inference, we investigate the posterior distribution of $\tau_{0,t}$ or some statistics such as its expected value and quantiles. While expected values (average treatment effects; ATEs) are often used in standard SCMs (Abadie, 2002), researchers also focus on distributional treatment effects (Park et al., 2021; Kallus & Oprescu, 2022; Chikahara et al., 2022; Gunsilius, 2023; Chen, 2020, DTEs).

In this study, we employ Bayesian predictive synthesis (BPS) to model the distribution of $\tau_{0,t}$. In the following, we refer to SCMs with BPS as BPSCMs.

## 3. Expert Problems and Statistical Decision-Making

We consider a decision-maker who predicts the counterfactual outcome $y^N_{0,t}$ to estimate the treatment effects. Formally, a decision-maker sequentially predicts $y^N_{0,t}$ for a time series $t \in \mathcal{T}$. Then, the decision-maker estimate the treatment effect as $\widehat{\tau}_{0,t} = y_{0,t} - \widehat{y}^N_{0,t}$. Recall that at $t \in \mathcal{T}_0$, $\widehat{\tau}_{0,t} = y^N_{0,t} - \widehat{y}^N_{0,t}$, which should be zero; at $t \in \mathcal{T}$ The decision-maker's prediction $\hat{y}^N_{0,t}$ belongs to a decision space $\mathcal{D}$, which we assume to be a convex subset of a vector space. For instance, in prediction with experts, we take $\mathcal{D} = \mathcal{Y}$; in dynamic linear regression, the decision space $\mathcal{D}$ becomes the parameter space of the linear regression.

For predicting $y_t$, at each time $t$, the decision maker obtains a set of additional information $\mathcal{U}_t = \{y_{1,t}, y_{2,t}, \dots, y_{J,t}\}$.

### 3.1. Regret

Let us denote a prediction model as $M := \{m_{0,1}, m_{0,2}, \dots, m_{0.T}\} \in \mathcal{M} = \mathcal{D}^T$. First, we define the regret of a time-series prediction as

$$\mathrm{Regret}_T(M; (y_{0,t})_{t\in\mathcal{T}}) := \sum_{t=1}^{T} \ell(y_{0,t}, m_{0,t}) - \min_{\widetilde{y}_{0,1},\widetilde{y}_{0,2},\dots,\widetilde{y}_{0,T}\in\mathcal{D}^T} \sum_{t=1}^{T} \ell(y_{0,t}, \widetilde{y}_{0,t}),$$

where $\ell : \mathcal{Y} \times \mathcal{D} \to \mathbb{R}$ denotes a loss function of a prediction of $y_t$ using $\hat{d}_t$. Given an input sequence $\{y_{0,1}, \dots, y_{0,T}\} \in \mathcal{Y}^T$, this regret represents a deviation between the cumulative losses incurred by decision maker's model $M$ and the ideal case. Then, depending on the behavior of the inputs $y_{0,1}, \dots, y_{0,T}$, we consider two goals, minimax and Bayes regret minimizations, where the decision maker has different optimality.

### 3.2. Minimax Regret Minimization

In minimax regret minimization, the decision maker aims to minimize the worst case regret; that is, the decision maker follows a prediction model defined as

$$\hat{M}_{\mathrm{minimax}} = \arg\min_{M\subset\mathcal{M}} \max_{\{y_{0,1},\dots y_{0,T}\}\subset\mathcal{Y}^T} \mathrm{Regret}_T(M; (y_{0,t})_{t\in\mathcal{T}}).$$

The minimax regret rule views the prediction protocol as the repeated game between "forecaster", who makes guesses $m_{0,t}$, and "environment", who chooses the expert advice and sets the true outcomes $y_t$. For a decision maker's prediction, an outcome $y_{0,t}$ is determined by an adversary, who tries to make the deviation from $m_{0,t}$ large.

### 3.3. Bayesian Regret Minimization

In Bayes regret minimization, an outcome $y_t$ is generated from the posterior distribution conditioned on $\mathcal{X}_{t-1}$ and the past observations. For $t \in \mathcal{T}$ and $j = 0$, let $p^N(y_{0,t}|y_{0,1}, \dots, y_{0,t-1})$ be a posterior distribution of outcomes without intervention. When $t \in \mathcal{T}_1$, $p^N(y_{0,t}|y_{0,1}, \dots, y_{0,t-1})$ a posterior distribution of a counterfactual outcome. Then, the decision maker's optimal prediction model is defined as

$$\hat{M}_{\mathrm{Bayes}} := \arg\min_{M\subset\mathcal{M}} \int \mathrm{Regret}_T(M; (y_{0,t})_{t\in\mathcal{T}}) \prod_{t=2}^{T} p^N(y_{0,t} \mid y_{0,1})\pi(y_{0,1})\mathrm{d}y_{0,1}\cdots\mathrm{d}y_{0,T},$$

where $\pi$ is the prior of $y_{0,1}$.

We refer to a decision maker minimizing the minimax regret as a *minimax decision-maker* and one minimizing the Bayes regret a *Bayes decision-maker*. Chen (2022) proposes SCMs of a minimax decision-maker. In this study, we propose SCMs of a Bayes decision-maker. In particular, we follow a formulation of BPS and refer to our proposed SCMs as BPSCMs. We describe our proposed BPSCMs in the following sections.

## 4. BPSCMs

As discussed above, we consider a Bayesian decision maker, who predicts $y^N_{0,t}$. The Bayesian decision maker has some opinion for $y^N_{0,t}$, quantified by a subjective prior density $\pi(y)$. Based on such a decision-maker, we propose a novel Bayesian approach for SCMs, based on BPS.

### 4.1. BPS

**Predictive distributions** For the untreated units $j \in \mathcal{J}^U$ and $t \in \mathcal{T}$, let $p_{j,t}(y)$ be the probability density of $y_{j,t}$. For the treated unit $j = 0$ and for each $t \in \mathcal{T}_1$, let $p^N_{j,t}(y)$ be the counterfactual probability density of $y$ without intervention. We denote the set of the $y_{j,t}$'s density $p_{j,t}(y_{j,t})$ as $\mathcal{P}_t = \{p_{1,t}(y_{1,t}), p_{2,t}(y_{2,t}), \dots, p_{J,t}(y_{J,t})\}$. The corresponding outcome $y_t$ is revealed after the decision maker decides $\hat{y}_{0,t}$ based on $\mathcal{U}_t$ and the past observations obtained until the time $t$.

**Synthesis function.** First, we define a conditional density function of $y_{0,t}$ given $\left\{y^N_{j,t}\right\}_{j\in\mathcal{J}^U}$. Because it can be interpreted as a synthesis of $\left\{y^N_{j,t}\right\}_{j\in\mathcal{J}^U}$ for predicting $y_{0,t}$, we refer to the conditional density function as the synthesis function, denoted as

$$\alpha\left(y|\left\{y^N_{j,t}\right\}_{j\in\mathcal{J}^U}, \Phi_t\right).$$

where $\Phi_t$ represents the time-varying parameters defining the synthesis function, $\alpha\big(y^N_{0,t}|\left\{y^N_{j,t}\right\}_{j\in\mathcal{J}^U}, \Phi_t\big)$.

**Posterior distribution.** Then, the Bayesian decision maker predicts $y^N_{0,t}$ using its implied posterior predictive distribution, $p(y^N_{0,t}|\mathcal{P})$ as

$$\begin{aligned} &p^N\left(y|\left\{y^N_{0,t}\right\}_{t\in[1:t]}, \Phi_t, \{\mathcal{P}_s\}_{s=1}^t\right) \\ &:= \int \alpha\left(y|\left\{y^N_{j,t}\right\}_{j\in\mathcal{J}^U}, \Phi_t\right) \prod_{j\in\mathcal{J}^U} p_{j,t}(y^N_{j,t})\mathrm{d}y^N_{j,t}, \end{aligned} \tag{1}$$

where $[1:t] := \{1, 2, \dots, t\}$.

Here, $\alpha\big(y^N_{0,t}|\left\{y^N_{j,t}\right\}_{j\in\mathcal{J}^U}, \Phi_t\big)$ determines how the predictive distributions are synthesized and it includes a variety of existing combination methods, such as standard and advanced model averaging methods (e.g. Hoeting et al., 1999; Geweke & Amisano, 2011; Aastveit et al., 2018), as special cases.

**Regret minimization.** Once we obtain a posterior distribution, we compute the regret minimizer as

$$\begin{aligned} \{\widehat{y}_{0,t}\}_{t\in\mathcal{T}} := \mathop{\arg\min}_{\{m_{0,t}\}_{t\in\mathcal{T}}\subset\mathcal{D}^{\mathcal{T}}} &\int \mathrm{Regret}_T(\{m_{0,t}\}_{t\in\mathcal{T}}; (y_{0,t})_{t\in\mathcal{T}}) \\ &\prod_{t=2}^T p^N(y_{0,t} \mid y_{0,1})\pi(y_{0,1})\mathrm{d}y_{0,1}\cdots\mathrm{d}y_{0,T}. \end{aligned}$$

Consider a case where $\ell$ is a squared loss; that is, $\ell(a,b) := (a-b)^2$. In this case, choosing the posterior mean of $y_{0,t}$ as $\widehat{y}_{0,t}$ minimizes the Bayesian regret.

**Remark** (Stationarity). *The representation of* (1) *permits that the parameters are time-variant. The model also does not need a complete specification of the joint distribution,* $p(y^N_{0,t}, \mathcal{P}t)$*, and does not impose restrictions on the functional form of the synthesis function,* $\alpha\big(y^N_{0,t}|f, \Phi_t\big)$*. This flexibility allows Bayes decision-makers to specify how they want the information to be synthesized. For instance, if the focus is on predicting a univariate time series, a dynamic linear model can be specified (McAlinn & West, 2019). Alternatively, if the aim is to predict a multivariate time series, a dynamic seemingly unrelated regression can be specified (McAlinn et al., 2020). It is important to note that* (1) *is only a valid posterior if it satisfies the consistency condition. This condition stipulates that, prior to observing* $\mathcal{P}_t$*, Bayes decision-makers must specify their own prior predictive,* $p(y^N_{0,t})$*, as well as their prior expectation of the model forecasts,* $\mathrm{E}[\prod_{j\in\mathcal{J}^U} h_j(y_{j,t})]$*. Then* $p(y^N_{0,t}) = \int \alpha\big(y^N_{0,t}|f, \Phi_t\big) \prod_{j\in\mathcal{J}^U} p_{j,t}\big(y_{j,t}\big)\mathrm{d}y_{j,t}$ *must hold, meaning that the two priors that Bayes decision-makers specify must be consistent with each other.*

In the following parts, we explain how we derive our formulation and demonstrate examles.

### 4.2. Example 1: Dynamic Latent Factor Linear Regression Models

Our specification in (1) is general and allows various regression model by specifying the function $\alpha(\cdot)$. For example, following the context of the studies of SCMs, we specify the counterfactual outcome model as a dynamic linear model (DLM) defined as

$$\alpha(y^N_{0,t}|\mathcal{P}_t, \Phi_t) = \phi\left(y^N_{0,t}; w_{0,t} + \sum_{j=1}^J w_{j,t}y_{j,t}, \nu_t\right),$$

where $\phi(\cdot; a, b^2)$ denotes the univariate normal density with mean $a$ and variance $b^2$, and $\Phi_t = (w_{0,t}, w_{1,t}, \dots, w_{J,t}, \nu_t)$ represents the time-varying parameters defining the synthesis pdf parameters. Here, $w_{j,t}$ defines the time-varying coefficient for the $j$th model and $w_j$ is assumed to be a smooth function of $s$.

The specification above defines the following latent factor time-varying coefficient model for $y^N_{0,t}$:

$$\begin{aligned} y^N_{0,t} &= w_{0,t} + \sum_{j\in\mathcal{J}^U} w_{j,t}y^N_{j,t} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu_t), \\ w_{j,t} &= w_{j,t-1} + \eta_{j,t}, \quad \eta_{j,t} \sim N(0, \nu_t\boldsymbol{W}_t), \end{aligned} \tag{2}$$

where $\nu_t$ is a time-varying variance, and $\nu_t\boldsymbol{W}_t$ is a time-varying variance matrix that can be decomposed into a scalar $\nu_t$ and a $(J\times J)$-matrix $\boldsymbol{W}_t$. The model in (2) is quite similar to the time-varying coefficient model (Gelfand et al., 2003), but the difference is that the latent factor $x_{tj}$ in (2) is a random variable rather than fixed variable, as in the standard varying coefficient model.

Here, $w_{1,t}, \ldots, w_{J,t}$ evolves in time according to a linear/normal random walk with innovations variance matrix $\nu_t \boldsymbol{W}_t$ at time $t$, and $\nu_t$ is the residual variance in predicting $y_{0,t}^N$ based on past information and the set of agent forecast distributions. The residuals $\nu_t$ and evolution $\eta_{1,s}, \ldots, \eta_{J,s}$ are independent over time and mutually independent for all $t$ and $s$.

The DLM specification is completed using standard discount factor methods (West & Harrison, 1997; Prado & West, 2010). Here, the time-varying intercept and agent coefficients $(w_{0,t}, w_{1,t}, \ldots, w_{J,t})$ follow the random walk evolution of (2) where $\boldsymbol{W}_t$ is defined via a standard, single discount factor specification (West & Harrison 1997, Section 6.3; Prado & West 2010, Sect 4.3), using a state evolution discount factor $\beta \in (0, 1]$. Besides, the residual variance $\varepsilon_t$ follows a standard beta-gamma random walk volatility model (West & Harrison 1997, Section 10.8; Prado & West 2010, Section 4.3), with $\varepsilon_t = \varepsilon_{t-1}\delta/\gamma_t$ for some discount factor $\delta \in (0, 1]$ and where $\gamma_t$ are beta distributed innovations, independent over time and independent of $\nu_s$ and $\eta_{1,r}, \ldots, \eta_{J,r}$ for all $t, s, r$. Given choices of discount factors underlying these two components, and a (conjugate normal/inverse-gamma) prior for $(w_{0,0}, w_{1,0}, \ldots, w_{J,0}, \nu_0)$ at $t = 0$, the model is specified.

### 4.3. Example 2: Mixture Models

We consider a mixture of point masses with a base prediction $\pi_0(y)$ using agent state-dependent probabilities .

$$\alpha \left( y | \left\{ y_{j,t}^N \right\}_{j \in \mathcal{J}^U}, \Phi_t \right) :=$$
$$\alpha_0(\{y_{j,t}^N\}_{j\in\mathcal{J}^U})\pi_0(y) + \sum_{j\in\mathcal{J}^U} q_j \alpha_j(\{y_{j,t}^N\}_{j\in\mathcal{J}^U}) p_{j,t}(y).$$

Extending earlier work on expert opinion analysis (Genest & Schervish, 1985; West & Crosse, 1992; West, 1992), (McAlinn & West, 2019) proposes the Bayesian predictive synthesis framework, which provides a general and coherent way for Bayesian updating, given multiple predictive distributions. Following the study, we specify the Bayesian posterior counterfactual outcome model as

With this specific form of calibration function, it follows from (1) that

$$p^N \left( y | \left\{ \mathcal{P}_s \right\}_{s=1}^t \right) =$$
$$q_0(y | \{\mathcal{P}_s\}_{s=1}^t)\pi_0(y) + \sum_{j\in\mathcal{J}^U} q_j(y | \{\mathcal{P}_s\}_{s=1}^t) p_j(y).$$

where the probability weights $q_j(H) = \int a_j(x) p_{0,t}(y) \prod_j h_j(x_j) \mathrm{d}x$ (for $j \in \mathcal{J}$) depend on $\mathcal{P}$ and naturally act to recalibrate the contributions from each hj$(\cdot)$ modulo the specified forms of the calibration weights aj(x). Some specific examples connect with existing methods as well as to new and practically relevant features.

### 4.4. Predictive Model Synthesis using Auxiliary Covariates

For each unit, $j$, we consider a situation where there are also $(K-1)$-dimensional covariates $x_{j,t} \in \mathbb{R}^{K-1}$ unaffected by the intervention. In the previous sections, we do not use covariates. There are several ways in incorporating covariates. In this study, we consider constructing predictors for $y_{j,t}^N$ by using $x_{j,t}$ and synthesize them with the BPS.

Formally, we extend the previous BPSCMs as follows. We consider making $L$ predictoros for each untreated unit. Let us define $L$ predictors for $y_{j,t}^N$ by $\{\hat{f}^l(x_{j,t})\}_{l=1}^L$. Then, we consider there are $K = J + JL$ agents who have opinions (predictions) for $y_{0,t}^N$, where opinions of agents are

$$\boldsymbol{z}_t = \Big\{ y_{1,t}^N, \ldots, y_{J,t}^N, \hat{f}^1(x_{1,t}), \ldots, \hat{f}^L(x_{1,t}),$$
$$\hat{f}^1(x_{2,t}), \ldots, \hat{f}^L(x_{J-1,t}), \hat{f}^1(x_{J,t}), \ldots, \hat{f}^L(x_{J,t}) \Big\}.$$

Then, we conduct the BPSC method as if there are $J + JL$ untreated units that can be used for SCs; that is,

$$p^N \left( y | \Phi_t, \left\{ y_{0,t}^N \right\}_{t\in[1:t]}, \mathcal{P}_t \right)$$
$$= \int \alpha \left( y | \boldsymbol{z}_t, \Phi_t \right) \prod_{j\in\{1,2,\ldots,(1+L)J\}} p_{j,t}(z_{j,t}) \mathrm{d}z_{j,t},$$

where $z_{j,t}$ is an $j$-th element of $\boldsymbol{z}_t$.

Although we can incorporate covariates in such a way, for simplicity, we explain BPSCM with $J$ outcomes without loss of generality. The following arguments can be held even if we include covariates. In experiments of Section 7, we synthesize predictive models using linear regression and random forests.

### 4.5. Model Misspecification

The BPS can deal with model misspecification. In Example 1, dynamic latent factor linear regression models, we can show that even under model misspecification, we can obtain a minimax optimal predictor for $y_{j,t}^N$ in the sense that the KL divergence between the "true" distribution of $y_{0,t}$ and our posterior distribution of $y_{0,t}$. For the details, see Takanashi & McAlinn (2021).

## 5. BPSCMs with Dynamic Latent Factor Linear Regression Models

Example 1 in Section 4.2 employs standard DLMs to define synthesis function $\alpha_t(y_t | y_{1,t}^N, \ldots, y_{J,t}^N, \Phi_t)$ because they are flexible to incorporate time-varying structure. In this section, we investigate the detailed properties.

### 5.1. Bayesian Analysis

In Section 4.2, (2) defines a dynamic latent factor model, where $\{y_{1,t}^N, \ldots, y_{J,t}^N\}$ are latent variables. At time $t-1$, the set of densities of outcomes available; then from the BPSC formulation, each $y_{j,t}^N$ is a latent draw from $h_{t,j}(y)$. Note that the latent factor generating process has the $y_{j,t}^N$ drawn independently from their $h_{t,j}(y)$ and externally to the BPSC model. Then, from (2), we have

$$p^N\left(y_{1,t}^N, \ldots, y_{J,t}^N | \{y_{0,s}^N\}_{s\in[1:t-1]}, \Phi_t, \{\mathcal{P}_s\}_{s=1}^t\right) \\ := p^N\left(y_{1,t}^N, \ldots, y_{J,t}^N | \{\mathcal{P}_s\}_{s=1}^t\right) = \prod_{j\in\mathcal{J}^U} p_{j,t}\left(y_{j,t}^N\right).$$

At any current time $t$, the Bayesian decision maker has available the history of the BPS analysis to that point, including the now historical information $\{\{y_t^N\}_{t\in[1:t-1]}, \{\mathcal{P}_s\}_{s=1}^t\}$. Over times $t \in \mathcal{T}_0$ the BPS analysis will have involved inferences on both the latent agent states $\{y_{j,t}^N\}$ as well as the dynamic BPS model parameters $\{\Phi_t\}_{t\in\mathcal{T}}$. Importantly, inferences on the former provide insights into the nature of dependencies among the agents, as well as individual agent forecast characteristics. The former addresses key and topical issues of overlap and redundancies among groups of forecasting models or individuals, as well as information sharing and potential herding behavior within groups of forecasters. The "output" of full posterior summaries for the $\{y_{j,t}^N\}_{j\in\mathcal{J}^U}$ series is thus a key and important feature of BPS.

For posterior analysis, the holistic view is that a Bayes decision-maker is interested in computing the posterior for the full set of past latent agent states (latent factors) and dynamic parameters $\{\{y_t^N\}_{t\in[1:t-1]}, \{\mathcal{P}_s\}_{s=1}^t\}$, rather than restricting attention to forward filtering to update posteriors for current values $\{\{y_t^N\}_{t\in[1:t-1]}, \{\mathcal{P}_s\}_{s=1}^t\}$; the latter is of course implied by the former. This analysis is enabled by Markov chain Monte Carlo (MCMC) methods, and then forecasting from time $t$ onward follows by theoretical and simulation-based extrapolation of the model; both aspects involve novelties in the BPS framework but are otherwise straightforward extensions of traditional methods in Bayesian time series (West & Harrison 1997, Chap 15; Prado & West 2010).

### 5.2. Posterior Computations via MCMC

The dynamic latent factor model of (2) leads to a two-component block Gibbs sampler for sets of the latent agent states $\{y_t^N\}_{t\in[1:t-1]}$ and DLM dynamic parameters $\Phi_t$. These are iteratively resimulated from two conditional posteriors noted below, with obvious initialization based on agent states drawn independently from priors $h_*(*)$.

First, conditional on values of agent states, the next MCMC step draws new parameters from $p(\{\Phi_t\}_{s=1}^t | \{y_t^N\}_{t\in[1:t-1]})$. By design, this is a discount-based dynamic linear regression model, and sampling uses the standard forward filtering, backward sampling (FFBS) algorithm (e.g. Frühwirth-Schnatter 1994; West & Harrison 1997, Sect 15.2; Prado & West 2010, Sect 4.5).

Second, conditional on values of dynamic parameters, the MCMC draws new agent states from $p^N(\{y_{j,t}^N\}_{j\in\mathcal{J}^U} | \{y_{0,s}^N\}_{s=1}^t, \Phi_t, \{\mathcal{P}_s\}_{s=1}^t)$. It is immediate that the $\{y_{j,t}^N\}_{j\in\mathcal{J}^U}$ are conditionally independent over time $t$ in this conditional distribution, with time $t$ conditionals

$$p^N(y_{1,t}^N, \ldots, y_{J,t}^N | \Phi_t, y_{0,t}^N, \{\mathcal{P}_s\}_{s=1}^t) \\ \propto N\left(y_{0,t}^N | w_{0,t} + \sum_{j\in\mathcal{J}^U} w_{t,j} y_{j,t}^N, \nu_t\right) \prod_{j\in\mathcal{J}^U} p_{j,t}(x_{tj}).$$

In cases when each of the agent forecast densities is normal, this yields a multivariate normal for $\{y_{j,t}^N\}_{j\in\mathcal{J}^U}$ that is trivially sampled. In other cases, this will involve either a Metropolis-Hastings simulator or an augmentation method. A central, practically relevant case is when agent forecasts are T distributions; each $h_{tj}(\cdot)$ can then be represented as a scale mixture of normals, and augmenting the posterior MCMC to include the implicit underlying latent scale factors generates conditional normals for each $\{y_{j,t}^N\}_{j\in\mathcal{J}^U}$ coupled with conditional inverse gammas for those scales. This is again a standard MCMC approach and much used in Bayesian time series, in particular (e.g. Frühwirth-Schnatter 1994; West & Harrison 1997, Chap 15).

Full technical details of the MCMC computations, and additional discussion are given in the supplementary of McAlinn & West (2016).

### 5.3. Posterior Computation

At any current time $t$, the Bayes decision-maker has available the history of the BPS analysis to that point, including the now historical information $\{(y_{0,1}, \ldots, y_{0,t}), (\mathcal{P}_1, \ldots, \mathcal{P}_t)\}$. Over times $1, 2, \ldots, t$, the BPS analysis will have involved inferences on both the latent agent states $(y_{j,1})_{j\in\mathcal{J}^U}, (y_{j,2})_{j\in\mathcal{J}^U}, \ldots, (y_{j,t})_{j\in\mathcal{J}^U}$ as well as the dynamic BPS model parameters $\Phi_1, \ldots, \Phi_t$. Importantly, inferences on the former provide insights into the nature of dependencies among the agents, as well as individual agent forecast characteristics. The former addresses key and topical issues of overlap and redundancies among groups of forecasting models or individuals, as well as information sharing and potential herding behavior within groups of forecasters. The "output" of full posterior summaries for the $\boldsymbol{x}_t$ series is thus a key and important feature of BPS. For posterior analysis, the holistic view is that D is interested in computing the posterior for the full set of past latent agent states (latent factors) and dynamic parameters $\{((y_{j,1})_{j\in\mathcal{J}^U},\ldots,(y_{j,t})_{j\in\mathcal{J}^U}),(\Phi_1,\ldots,\Phi_t)\}$, rather than restricting attention to forward filtering to update posteriors for current values $\{((y_{j,1})_{j\in\mathcal{J}^U},\ldots,(y_{j,t})_{j\in\mathcal{J}^U}),(\Phi_1,\ldots,\Phi_t)\}$; the latter is of course implied by the former.

### 5.4. Inference

By using observations until $t = T_0$ we predict a counterfactual outcome $y^N_{0,t}$ for $t \in \mathcal{T}_1$ from the BPSCM, as follows:

1. For each sampled $\Phi_t$ from the posterior MCMC above, draw $\nu_{t+1}$ from its discount volatility evolution model, and then $\{w_{0,t+1}, w_{1,t+1},\ldots,w_{J,t+1}\}$ conditional on $\{w_{0,t}, w_{1,t},\ldots,w_{J,t}\},\nu_{t+1}$ from the evolution model (2)– this gives a draw $\Phi_{t+1} = \{w_{0,t}, w_{1,t},\ldots,w_{J,t},\nu_{t+1}\}$ from $p^N(\Phi_{t+1}|\{y^N_{0,s}\}^t_{s=1},\{\mathcal{P}_s\}^t_{s=1})$;
2. Draw $y^N_{1,t+1},\ldots,y^N_{J,t+1}$ via independent sampling of the $p^N_{1,t+1}(y),\ldots,p^N_{J,t+1}(y)$;
3. Draw $y^N_{0,t+1}$ from the conditional normal of (2) given these sampled parameters and agent states. Repeating this generates a random sample from the 1-step ahead forecast distribution for time $t+1$.

## 6. Related Work

This section provides related work.

### 6.1. Frequentist SCMs

Comparative case studies aim to reproduce $y^N_{0,t}$; 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 have similar characteristics as the affected unit at the time of the intervention. When the data consist of a few aggregate entities, such as regions or countries, it is often difficult to find a single unaffected unit that provides a suitable comparison for the unit affected by the policy intervention of interest.

In the SCM, we define a synthetic control unit as a weighted average of control units to approximate the characteristics of counterfactual outcome unit for the treated unit under no treatment. Formally, a synthetic control can be represented by a $(J \times 1)$ vector of weights, $\boldsymbol{w} = (w_0, w_1,\ldots,w_J)$. Given a set of weights, $\boldsymbol{w}$, the synthetic control estimators of $y^N_{0,t}$ and $\alpha_t$ are $\hat{y}^N_{0,t} = \sum_{j\in\mathcal{J}^U} w_j y_{j,t}$ and $\hat{\tau}_{0,t} = y_{0,t} - \hat{y}^N_{0,t}$.

Because $\boldsymbol{w}$ is unknown, we select it using observations. Let us denote a selected weighted by $\hat{\boldsymbol{w}} = (\hat{w}_0,\hat{w}_1,\ldots,\hat{w}_J)^\top$. To avoid extrapolation, the weights are restricted to be nonnegative and to sum to one, so synthetic controls are weighted averages of the control units. Then, we can select $\hat{\boldsymbol{w}}$ as the solution to the following constrained minimization problem: $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}\in\Lambda}\sum_{t\in\mathcal{T}_0}\left(y_{0,t} - w_0 - \sum_{j\in\mathcal{J}^U} w_j y_{j,t}\right)^2$, where $w_0$ is an intercept, and $\Lambda$ represents constrains on $\hat{\boldsymbol{w}}$. A typical choice of the constraints is $\Lambda = \Big\{\boldsymbol{w}\in\mathbb{R}^{J+1} : w_0 = 0,\ w_j \geq 0\ \forall j\in\mathcal{J}^U \text{ and } \sum_{j\in\mathcal{J}^U} w_j = 1\Big\}$.

There are several variants and extensions in this basic formulation. Some include regularization term in weight estimation, such as Ridge (Ben-Michael et al., 2021) and ElastyNet (Doudchenko & Imbens, 2016). Ferman & Pinto (2021) analyzes conditions for pre-treatment fit.

### 6.2. Ensemble SCMs

Predictive model ensemble has been proposed by Viviano & Bradic (2019). Using results from the online learning literature (Hazan, 2016), Viviano & Bradic (2019) propose a ensemble SCM that has no-regret property.

### 6.3. Minimax SCMs

Chen (2022) also employs online convex learning framework for SCMs. In contrast to Viviano & Bradic (2019), they study SCMs directly in the worst-case analysis and show that SCMs themselves are no-regret online algorithms.

### 6.4. BSCMs

Bayesian framework also has intensively studied in the literature of SCMs. Kim et al. (2020) introduces BSCMs with shrinkage priors, such as Laplace and horse-shoe priors.

### 6.5. Time-Varying Coefficients

Klinenberg (2022) proposes time-varying coefficients BSMs with schrinkage priors as well as Kim et al. (2020). Such BSMs also have close relationship to causal impact methods proposed by Brodersen et al. (2015).

## 7. Empirical analysis

We conduct empirical studies discussed in Abadie & Gardeazabal (2003), Abadie et al. (2010), and Abadie et al.

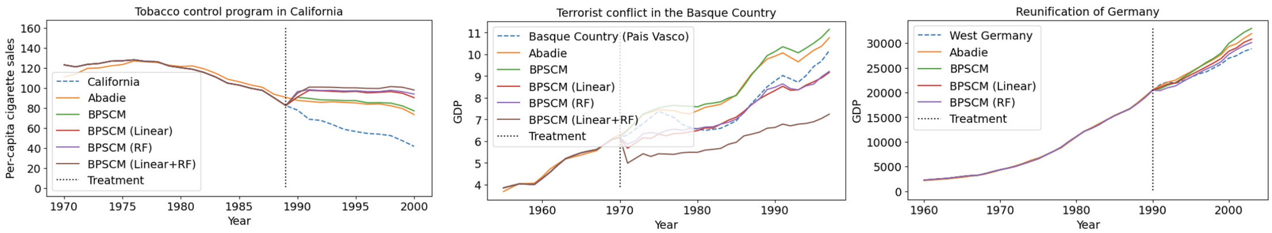


Figure 1: Counterfactual outcomes of the per-capita cigarette sales in California (left), the GDP in the Basque Country (center), and the GDP of Germany (right). The blue dashed lines represent the actual values for the treatment groups, and the solid orange lines are based on actual outcomes after treated, respectively. The green solid line represents the BPSCM only with un-treated units' outcomes, the red solid line represents the BPSCM with Linear, the purple solid line represents the BPSCM with RF, and the brown solid line represents the BPSCM with Linear and RF. Vertical dotted lines represent the timing of the treatment intervention.

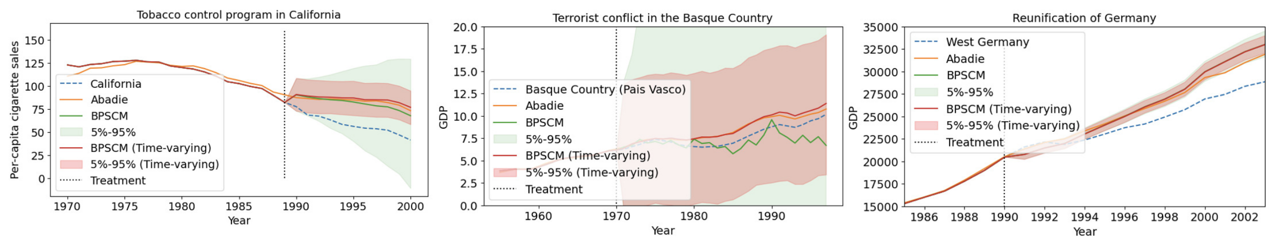


Figure 2: Credible intervals of the per-capita cigarette sales in California (left), the GDP in the Basque Country (center), and the GDP of Germany (right). The blue dashed lines represent the actual values for the treatment groups, and the solid orange lines are based on actual outcomes after treated, respectively. The green solid line represents the counterfactual outcome of BPSCM without parameter updating after $T_0$, and the red solid line represents the counterfactual outcome of BPSCM with parameter updating after $T_0$. The red intervals are credible intervals of BPSCM without parameter updating after $T_0$ while the green intervals are credible intervals of BPSCM with parameter updating after $T_0$.

(2015). Here, we briefly explain each example.

In our experiments, in addition to untreated units' outcomes $(y_{j,t})_{j\in\mathcal{J}^U}$, we employ predictors by using linear regression with least squares (Linear) and random forests (RF). In Figures 1 and 2, We refer to the BPSCM only with untreated units' outcomes as BPSCM, the BPSCM with Linear as BPSCM (Linear), the BPSCM with RF as BPSCM (RF), and the BPSCM with Linear and RF as BPSCM (Linear + RF).

**Terrorist conflict in the Basque Country.** Abadie & Gardeazabal (2003) investigates the effect of terrorist conflict in the Basque Country on gross domestic product (GDP). In this example, $J = 16$, $K = 14$, and $(t_0, T_0, T_1) = (1955, 1970, 1997)$.

**Tobacco control program in California.** Abadie et al. (2010) study the effect of a large tobacco control program adopted in California in 1988, which we explained in our introduction. In this example, $J = 20$, $K = 5$, and $(t_0, T_0, T_1) = (1970, 1989, 2000)$.

**Reunification of Germany.** Abadie et al. (2015) studies the effect of the reunification of Germany on GDP in 1990. Because the countries were dissimilar from any one country to make a comparison group, they apply SCMs to create a composite comparison group. In this example, $J = 8$, $K = 31$, and $(t_0, T_0, T_1) = (1960, 1990, 2003)$.

**Results.** For each dataset, we show the result of treatment effect estimation in Figure 1 and posterior credible intervals in Figure 2. In Figure 2, we show the 95% credible intervals of the treatment effects with and without parameter updating after $T_0$. Our proposed method shows good pre-treatment fits and show the seemingly valid credible intervals.

## 8. Conclusion

We proposed BPSCMs, which is a BSCMs with ensemble of predictive models. First, we showed a general framework and then discuss a specific implementation with focusing on the DLM. We confirmed the soundedness in empirical analysis with datasets used in existing studies.

## References


Aastveit, K. A., Ravazzolo, F., and Van Dijk, H. K. Combined density nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics*, 36(1):131–145, 2018. 4

Abadie, A. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, 97(457):284–292, 2002. 2, 3

Abadie, A. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425, June 2021. 1

Abadie, A. and Gardeazabal, J. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132, March 2003. 1, 2, 7, 8

Abadie, A. and i Bastida, J. V. Synthetic controls in action, 2022. 2

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. 1, 7, 8

Abadie, A., Diamond, A., and Hainmueller, J. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2015. 7, 8

Ben-Michael, E., Feller, A., and Rothstein, J. Synthetic controls with staggered adoption, 2019. 1

Ben-Michael, E., Feller, A., and Rothstein, J. The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803, 2021. 1, 7

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9: 247–274, 2015. 7

Chen, J. Synthetic control as online linear regression, 2022. 2, 3, 7

Chen, Y.-T. A distributional synthetic control method for policy evaluation. *Journal of Applied Econometrics*, 35 (5):505–525, 2020. 3

Chikahara, Y., Yamada, M., and Kashima, H. Feature selection for discovering distributional treatment effect modifiers. In *The 38th Conference on Uncertainty in Artificial Intelligence*, 2022. 3

Cunningham, S. *Causal Inference: The Mixtape*. Yale University Press, 2021. 1

Doudchenko, N. and Imbens, G. W. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Working Paper 22791, National Bureau of Economic Research, October 2016. 7

Ferman, B. and Pinto, C. Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221, 2021. 1, 7

Frühwirth-Schnatter, S. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994. doi: 10.1111/j.1467-9892.1994.tb00184.x. 6

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396, 2003. 4

Genest, C. and Schervish, M. J. Modelling expert judgements for Bayesian updating. *Annals of Statistics*, 13: 1198–1212, 1985. 5

Geweke, J. and Amisano, G. Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141, 2011. 4

Gunsilius, F. F. Distributional synthetic controls. *Econometrica*, 91(3):1105–1117, 2023. 3

Hazan, E. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2:157–325, 2016. doi: 10.1561/2400000013. 7

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. Bayesian model averaging: a tutorial. *Statistical science*, pp. 382–401, 1999. 4

Johnson, M. C. and West, M. Bayesian predictive synthesis for probabilistic forecast calibration and combination. Technical report, Department of Statistical Science, Duke University, 2016. 2

Kallus, N. and Oprescu, M. Robust and agnostic learning of conditional distributional treatment effects, 2022. 3

Kim, S., Lee, C., and Gupta, S. Bayesian synthetic control methods. *Journal of Marketing Research*, 57(5):831–852, 2020. 1, 7

Klinenberg, D. Synthetic control with time varying coefficients a state space approach with bayesian shrinkage. *Journal of Business & Economic Statistics*, 0 (0):1–12, 2022. doi: 10.1080/07350015.2022.2102025. URL `https://doi.org/10.1080/07350015.2022.2102025`. 2, 7

McAlinn, K. and West, M. Dynamic bayesian predictive synthesis in time series forecasting. *ArXiv*, 1601.07463, 2016. 6

McAlinn, K. and West, M. Dynamic bayesian predictive synthesis in time series forecasting. *Journal of econometrics*, 210(1):155–169, 2019. 2, 4, 5

McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. Multivariate bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association*, 115(531):1092–1110, 2020. 2, 4

Nazaret, A., Shi, C., and Blei, D. M. On the misspecification of linear assumptions in synthetic control, 2023. 2

Park, J., Shalit, U., Schölkopf, B., and Muandet, K. Conditional distributional treatment effect with kernel conditional mean embeddings and u-statistic regression. In *International Conference on Machine Learning*, volume 139, pp. 8401–8412, 2021. 3

Prado, R. and West, M. *Time Series: Modelling, Computation & Inference*. Chapman & Hall/CRC Press, 2010. 5, 6

Shi, C., Sridhar, D., Misra, V., and Blei, D. On the assumptions of synthetic control methods. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151, pp. 7163–7175, 2022. 2

Takanashi, K. and McAlinn, K. Predictions with dynamic bayesian predictive synthesis are exact minimax, 2021. 5

Viviano, D. and Bradic, J. Synthetic learner: model-free inference on treatments over time, 2019. URL `https://arxiv.org/abs/1904.01490`. 1, 2, 7

West, M. Modelling agent forecast distributions. *Journal of the Royal Statistical Society (Series B: Methodological)*, 54:553–567, 1992. 5

West, M. and Crosse, J. Modelling of probabilistic agent opinion. *Journal of the Royal Statistical Society (Series B: Methodological)*, 54:285–299, 1992. 5

West, M. and Harrison, P. J. *Bayesian Forecasting & Dynamic Models*. Springer Verlag, 2nd edition, 1997. 5, 6