The Annals of Applied Statistics
2022, Vol. 16, No. 1, 414–435
https://doi.org/10.1214/21-AOAS1498

ESTIMATING THE EFFECTIVENESS OF PERMANENT PRICE REDUCTIONS FOR COMPETING PRODUCTS USING MULTIVARIATE BAYESIAN STRUCTURAL TIME SERIES MODELS

BY FIAMMETTA MENCHETTI[1,a] AND IAVOR BOJINOV[2,b]

[1]*DiSIA, Universitá di Firenze,* [a]*fiammetta.menchetti@unifi.it*
[2]*Technology and Operations Management, Harvard Business School,* [b]*ibojinov@hbs.edu*

The Florence branch of an Italian supermarket chain recently implemented a strategy that permanently lowered the price of numerous store brands in several product categories. To quantify the impact of such a policy change, researchers often use synthetic control methods for estimating causal effects when a subset of units receive a single persistent treatment and the rest are unaffected by the change. In our applications, however, competitor brands not assigned to treatment are likely impacted by the intervention because of substitution effects; more broadly, this type of interference occurs whenever the treatment assignment of one unit affects the outcome of another. This paper extends the synthetic control methods to accommodate partial interference, allowing interference within predefined groups but not between them. Focusing on a class of causal estimands that capture the effect both on the treated and control units, we develop a multivariate Bayesian structural time series model for generating synthetic controls that would have occurred in the absence of an intervention, enabling us to estimate our novel effects. In a simulation study we explore our Bayesian procedures' empirical properties and show that it achieves good frequentists coverage, even when the model is misspecified. We use our new methodology to make causal statements about the impact on sales of the affected store brands and their direct competitors. Our proposed approach is implemented in the CausalMBSTS R package.

1. Introduction. On October 4, 2018, the Florence branch of a large Italian supermarket chain permanently lowered the price of 707 store brands in several product categories. In the past the firm had regularly used temporary promotions (discounting products for a brief period); however, the new permanent price reduction represented a significant strategic shift in its business model. The firm hypothesized that the lower price would expand its customer base, increasing sales and, ultimately, revenue. To evaluate the success of the new strategy, we model the permanent price reduction as a single persistent intervention. In this paper we focus on the cookies product category and estimate the causal effect of permanently reducing the price of 10 store branded cookies on daily sales.

A popular approach for obtaining estimates of causal effects from panel data with a single intervention is to use synthetic control methods (e.g., Abadie, Diamond and Hainmueller (2010), Abadie, Diamond and Hainmueller (2015), Abadie and Gardeazabal (2003), Brodersen et al. (2015)). Unlike traditional difference-in-difference methods, synthetic controls provide a more flexible framework as they directly impute the unobserved outcome for treated time series by combining data from multiple control series that were not directly impacted by the treatment but are, nevertheless, correlated with the counterfactual outcome (O'Neill et al. (2016)). In our supermarket study, daily wine sales could be a suitable control series because changes in the price of cookies are unlikely to affect wine sales; instead, the

Received October 2020; revised May 2021.
Key words and phrases. Causal inference, partial interference, synthetic controls, Bayesian structural time Series.

control series captures temporal trends that are useful in modeling how the sales of cookies would have evolved in the absence of a price reduction. More broadly, synthetic control methods have been successfully applied to evaluate the effectiveness of policy changes in healthcare (Kreif et al. (2016), Papadogeorgou et al. (2018), Viviano and Bradic (2019)), economics (Abadie, Diamond and Hainmueller (2015), Ben-Michael, Feller and Rothstein (2018), Billmeier and Nannicini (2013), Dube and Zipperer (2015), Gobillon and Magnac (2016)), marketing and online advertising (Brodersen et al. (2015), Li (2020)), among others.

Typically, synthetic control methods assume that there is no interference between experimental units; that is, the assignment any unit receives has no bearing on the outcome of any other unit (Cox (1958)). However, there are many applications where this assumption is violated (e.g., Hudgens and Halloran (2008), Tchetgen Tchetgen and VanderWeele (2012), and Basse, Feller and Toulis (2019)). In our study, for each store brand cookie the firm identified a competitor brand that is a direct substitute, differing primarily in the brand name. As traditional economic theory suggests, if two goods are substitutes, lowering the price of one will impact the sales of the other (Nicholson and Snyder (2012)); therefore, any price changes to the store brand will impact the sales of the direct competitor, and vice versa, violating the no-interference assumption. Beyond the direct competitor it is reasonable to assume that the price reduction will have negligible effects on the sales of other products.

More broadly, the setting where units interfere within predefined groups without interfering across these groups is known as partial interference (Sobel (2006)) and has been extensively studied for cross-sectional data (e.g., Rosenbaum (2007), Hudgens and Halloran (2008), and Forastiere, Airoldi and Mealli (2021)). In panel settings, like the supermarket study, partial interference has received relatively less attention, partly because of the added complications induced by the temporal component. In practice, authors often sidestep the issue by aggregating units that are likely to interfere with each other, generating a single-treated time series that now satisfies the no-interference assumption (Bojinov, Chen and Liu (2020)). One obvious downside of this approach is the inherent loss of information and a decreased ability to detect heterogeneous treatment effects.

To tackle this issue directly, we extend the synthetic control framework to partial interference setting by leveraging the extended potential outcomes that allow both spillovers across units and time (Bojinov, Rambachan and Shephard (2020), Bojinov and Shephard (2019), Robins (1986), Robins, Greenland and Hu (1999), VanderWeele (2010)). We then define new classes of causal effects that capture the impact of an intervention on both the unit that received it and the units within the same group. To perform inference, we derive the multivariate version of the popular Bayesian structural time-series model for causal inference, introduced in Brodersen et al. (2015). Like its univariate counterpart, our model allows for a great deal of flexibility, due to its ability to incorporate trends and seasonality effects. Indeed, our approach uses preintervention patterns (such as trend or seasonality) to forecast the counterfactual outcome in the absence of intervention and then updates these predictions, based on the posttreatment values of covariates linked to the outcome. To fit the model, we provide a Markov chain Monte Carlo algorithm and describe how to use the resulting draws to estimate our causal effects; all algorithms are implemented in the CausalMBSTS (Bojinov and Menchetti (2020)) R package. We then use a small simulation study to investigate the frequentist properties of our proposed approach and our ability to use posterior predictive checks (Rubin (1984)) to assess the model fit.

In the supermarket study our framework treats every store-competitor pair jointly, allowing us to model the group-specific interference directly. To determine the intervention's impact, we use our Bayesian structural time-series model to estimate the causal effect at various time horizons. The results show that the new strategy had a minor, short-term impact on store brands' sales; interestingly, we do not detect significant effects on the competitor brands. In

contrast, performing the equivalent aggregated analysis that ignores the interference incorrectly concludes that the price reduction positively affected sales.

Two papers consider our setup of partial-interference on panel or time series data, Cao and Dowd (2019) and Grossi et al. (2020). Cao and Dowd (2019) develop a model that requires that the impact of an intervention on one unit to the other is linear with an unknown parameter. Our paper imposes no such restriction, making it much more generally applicable. The formulation by Grossi et al. (2020) focuses on a context where only a single unit is intervened on, while the others are assigned to control. Since the treatment received by that unit may affect the outcomes of the other units, they rely on the partial interference assumption and identify different clusters such that the units belonging to different clusters do not interfere with each other. The inference is restricted to the group containing the treated unit, while the others form the "donor pool," used to construct synthetic controls by combining the donor outcomes. Our work presents a generalization of their specific context. Again, we study partial interference, but we allow for the existence of multiple treated units. By extending the univariate Bayesian structural time series model to the multivariate setting, we can also model the interference between units in the same cluster by explicitly modeling their dependence structure while transparently dealing with the surrounding uncertainty.

The paper is structured as follows. In Section 2 we present our causal framework, defining the treatment assignments, potential outcomes, causal effects and our underlying assumptions. In Section 3 we introduce the multivariate Bayesian structural time series model and explain how to apply it to our setting. In Section 4, we detail a simulation study that tracks our approach's performance. In Section 5 we provide the details of our supermarket study analysis. The final section presents our concluding remarks. In Appendices A and B we then provide the proof of the relations presented in this paper as well as additional results. The appendices, the data set and the source code to replicate the results of the empirical analysis are included in the Supplementary Material (Menchetti and Bojinov (2022)).

2. Causal framework. In this section we outline our framework for estimating the causal effect of an intervention in a panel setting with partial interference among statistical units. Throughout, we illustrate key concepts and definitions by leveraging our analyses of the Italian supermarket chain's new price policy. In our empirical example the statistical units are grouped into pairs, and so we begin by introducing the notation for a bivariate outcome variable; we then provide extensions to general group sizes. We conclude the section by defining our causal effects.

2.1. *Notation.* Throughout, we use a superscript s to denote the store brand and c the competitor brand. At time $t \in \{1, \dots, T\}$ and for each pair $j \in \{1, \dots, J\}$, let $\mathrm{W}_{j,t}^{(s)} \in \mathcal{W}$ be the treatment assignment for the store brand, $\mathrm{W}_{j,t}^{(c)} \in \mathcal{W}$ be treatment assignment for the competitor brand and $\mathbf{W}_{j,t} = (\mathrm{W}_{j,t}^{(s)}, \mathrm{W}_{j,t}^{(c)}) \in \mathcal{W}^2$ the pair assignment. We mostly focus on the binary treatment case, where $\mathcal{W} = \{0, 1\}$; following convention, we refer to "1" as treatment and "0" as control. In our supermarket study each pair is assigned to one of four possible treatments: no permanent price reduction $\mathbf{W}_{j,t} = (0, 0)$, both receive a permanent price reduction $\mathbf{W}_{j,t} = (1, 1)$, store brand receives a permanent price reduction only $\mathbf{W}_{j,t} = (1, 0)$ or competitor brand receives a permanent reduction only $\mathbf{W}_{j,t} = (0, 1)$. We then define the assignment path for each pair as the matrix $\mathbf{W}_{j,1:T} = (\mathbf{W}_{j,1}, \dots, \mathbf{W}_{j,T})' \in \mathcal{W}^{2\times T}$ and the assignment panel that captures the assignments of all units throughout the study as $\mathbf{W}_{1:J,1:T} = (\mathbf{W}'_{1,1:T}, \dots, \mathbf{W}'_{J,1:T}) \in \mathcal{W}^{2J\times T}$. We will use this vector and matrix notation for other variables but will sometimes drop the subscript if the dimensions are obvious from the context. Realizations of random variables will be denoted by their lower case; for example, $\mathbf{w}_{j,t}$ will denote a sample from $\mathbf{W}_{j,t}$.

In the panel set up, the pairs can change their assignment at any point in time, but, to keep our notation less cumbersome, we only focus on the case when there is a single persistent policy change, as was the case in our supermarket study.

ASSUMPTION 1 (Single intervention). We say pair j received a single intervention, if there exists a $t_j^* \in \{1, \ldots, T\}$ such that, for all $t \leq t_j^*$, we have $\mathbf{W}_{j,t} = (0, 0)$ and for all $t, t' > t_j^*$ we have $\mathbf{W}_{j,t} = \mathbf{W}_{j,t'}$. If all pairs receive a single intervention, then we say the study is a single intervention panel study. For simplicity, we also assume that the intervention happen simultaneously, that is, $t_j^* = t_{j'}^* = t^*$.

We maintain Assumption 1 which allows us to drop the t subscript from the treatment assignment so that $\mathbf{W}_j = (\mathrm{W}_j^{(s)}, \mathrm{W}_j^{(c)}) \in \{0, 1\}^2$ for all $t > t^*$ and $\mathbf{W}_j = (0, 0)$ for $t \leq t^*$.

2.1.1. *Potential outcomes*. We now define the potential outcomes that describe what would be observed for a particular pair at a fixed point in time for a given assignment panel. Generally, the potential outcomes are a function of the full treatment panel (e.g., Bojinov, Rambachan and Shephard (2020)); however, restricting our attention to nonanticipating potential outcomes[1] and Assumption 1 somewhat simplify the setup.

Assuming the intervention occurred at time $t^* + 1$, for each pair $j \in \{1, \ldots, J\}$ at time $t \in \{1, \ldots, t^*\}$, we observe an outcome $\mathbf{Y}_{j,t} = (\mathrm{Y}_{j,t}^{(s)}, \mathrm{Y}_{j,t}^{(c)})$, where $\mathrm{Y}_{j,t}^{(s)}$ is the outcome of the store brand and $\mathrm{Y}_{j,t}^{(c)}$ is the outcome of the competitor brand. In our application the outcome of interest is the average hourly sales for each product.

For $t > t^*$, generally, the outcomes depend on the treatment assignment matrix, $\mathbf{Y}_{j,t}(\mathbf{w}_{1:J}) = (\mathrm{Y}_{j,t}^{(s)}(\mathbf{w}_{1:J}), \mathrm{Y}_{j,t}^{(c)}(\mathbf{w}_{1:J}))$. In our empirical application the products within each pair are alike and only differ on their brand name and packaging; whereas, brands in different pairs differ on many characteristics (e.g., ingredients, flavor or weight). Therefore, we assume that a price reduction of one brand will impact its sales and the sales of its direct competitor. This assumption represents a model of consumer behavior in which customers' selection of the cookie type is not driven by price but rather by individual preferences; the choice within cookie type is then impacted by the price.[2] To connect the general setting to our empirical application, we assume that there is no interference across pairs.

ASSUMPTION 2 (Partial temporal no interference). For all $j \in \{1, \ldots, J\}$ and $t \in \{t^* + 1, \ldots, T\}$, we assume that, for any $\mathbf{w}_{1:J}, \mathbf{w}'_{1:J} \in \mathcal{W}^{2\times J}$ such that $\mathbf{w}_j = \mathbf{w}'_j$,

$$\mathbf{Y}_{j,t}(\mathbf{w}_{1:J}) = \mathbf{Y}_{j,t}(\mathbf{w}'_{1:J}).$$

This allows us to simplify out notation and write $\mathbf{Y}_{j,t}(\mathbf{w}_{1:J}) = \mathbf{Y}_{j,t}(\mathbf{w}_j)$.

In our application there are four potential outcome paths that can occur, corresponding to the four different assignments. For each store-competitor pair, we can combine the posttreatment outcomes to define four potential outcome paths or *potential outcome time series*,

$$\mathbf{Y}_{j,t^*+1:T}(\mathbf{w}_j) = \big(\mathrm{Y}_{j,t^*+1:T}^{(s)}(\mathbf{w}_j), \mathrm{Y}_{j,t^*+1:T}^{(c)}(\mathbf{w}_j)\big).$$

[1]Following Bojinov and Shephard (2019), we say the potential outcomes are nonanticipating if the outcomes at time t are not impacted by future treatment assignments. That is, the potential outcomes only depend on past or current treatment assignments. In our empirical setting, for $t < t^*$, this assumption would be violated if the knowledge of the upcoming price reduction changed present sales. For instance, consumers could have postponed their purchases, leading to a decrease in sales before the intervention. We can, however, safely exclude this, as the supermarket chain did not advertise the upcoming permanent discount in advance.

[2]Within our supermarket study, every store brand has its specific direct competitor; there are no cookies that belong to multiple pairs.

Note that, even though we dropped the t script from the assignment, our setup implicitly assumes that the outcomes at time $t > t^*$ are a function of the assignment path. This ensures that the potential outcomes at two different points in time correspond to two different treatment paths and are not directly comparable.

To connect the potential outcomes to the observed outcome, we assume that there is full compliance; that is, every pair receives the assigned treatment. In a causal inference setting for panel data, for each unit there is only one observed potential outcome time series, whereas the others are all unobserved. Generally, we will denote the observed treatment as $\mathbf{w}_j^{\text{obs}}$ which then leads to the observed outcome $\mathbf{Y}_{j,t^*+1:T} = \mathbf{Y}_{j,t^*+1:T}(\mathbf{w}_j^{\text{obs}})$. In our application, only the store brand receives the permanent price reduction making the observed outcome $\mathbf{Y}_{j,t^*+1:T} = \mathbf{Y}_{j,t^*+1:T}(1,0)$.

2.1.2. *Covariates*. Including good predictors of the preintervention outcome as covariates can greatly improve the prediction accuracy of the counterfactual series in the absence of intervention. However, suitable covariates should not be impacted by the treatment; otherwise, we would consider them as secondary outcomes. For each pair and time point, we, therefore, observe a vector of covariates $\mathbf{X}_{j,t} \in \mathcal{X}$ obeying to the following assumption.

ASSUMPTION 3 (Covariates-treatment independence). Let $\mathbf{X}_{j,t}$ be a vector of covariates; for all $t > t^*$ and for all assignments $\mathbf{w}_j, \mathbf{w}'_j \in \mathcal{W}^2$, we assume that

$$\mathbf{X}_{j,t}(\mathbf{w}_j) = \mathbf{X}_{j,t}(\mathbf{w}'_j) \quad \forall j \in \{1, \ldots, J\}.$$

Under Assumption 3 we can use the known covariates values posttreatment to improve the prediction of the outcome in the absence of intervention. For example, in the supermarket study we use the following covariates: weekend and holiday dummies, the prices of both goods before the intervention and the daily sales of products in categories unaffected by the price reduction. The latter are incorporated in the analysis as control cases because their postintervention pattern might provide useful information for predicting the counterfactual outcome in the absence of intervention. As fully detailed in Section 5.1, wines represent the only category that did not receive the permanent discount. The prior price is included as a covariate because it is a good predictor of sales had there not been an intervention (recall, the inclusion of the actual daily price after the reduction would have violated Assumption 3). For all of these covariates, Assumption 3 is likely to be satisfied but to check if the control series $\mathbf{X}_{j,1:T}$ are genuinely unaffected by the intervention, one could also test if the time series exhibit a change at the intervention time.

2.1.3. *Assignment mechanism*. We now define the class of assignment mechanism (i.e., conditional distributions of the assignment, given the set of potential outcomes, covariates and past assignments) that will allow us to estimate the causal effects, defined in the subsequent section. Our assumption has two parts. The first requires the assignment is individualistic; that is, the treatment of one pair has no bearing on another. The second requires the assignment is nonanticipating; that is, the assignment in a given period does not depend on future outcomes or covariates.

ASSUMPTION 4 (Nonanticipating individualistic treatment). The assignment mechanism at time $t^* + 1$ is independent across pairs and for the jth pair depends solely on its past outcomes and past covariates,

$$\begin{aligned}&\Pr(\mathbf{W}_{1:J,t^*+1} = \mathbf{w}_{1:J,t^*+1} | \mathbf{W}_{1:J,1:t^*}, \mathbf{W}_{1:J,t^*+2:T}, \mathbf{Y}_{1:J,1:T}(\mathbf{w}_{1:J,1:T}), \mathbf{X}_{1:J,1:T})\\ &\quad = \prod_{j=1}^{J} \Pr(\mathbf{W}_{j,t^*+1} = \mathbf{w}_{j,t^*+1} | \mathbf{Y}_{j,1:t^*}(\mathbf{w}_{j,1:t^*}), \mathbf{X}_{j,1:t^*}).\end{aligned}$$

The nonanticipating treatment assumption is the extension of the unconfounded assignment mechanism in a cross-sectional setting (Imbens and Rubin (2015), Bojinov, Rambachan and Shephard (2020)). Assumption 4 is essential in ensuring that, conditional on past outcomes and covariates, any differences in the outcomes are attributable to the intervention.

2.1.4. *Multivariate case*. Our framework easily generalizes to groups of size $d_j > 2$. For $j \in \{1, \ldots, J\}$, let $\mathrm{W}_j^i \in \mathcal{W}$ be the treatment status of the ith unit inside the jth group, and let $\mathbf{W}_j = (\mathrm{W}_j^{(1)}, \ldots, \mathrm{W}_j^{(d_j)}) \in \mathcal{W}^{d_j}$ be the treatment status of the jth group. Again, Assumption 1 allowed us to drop the subscript for time. We then define the outcome to be a d_j-variate vector, $\mathbf{Y}_t = (\mathrm{Y}_t^{(1)}, \ldots, \mathrm{Y}_t^{(d_j)})$, for $t \leq t^*$. Assuming that there is only partial interference, Assumption 2, the potential outcomes for $t > t^*$ for any $\mathbf{w}_j \in \{0, 1\}^{d_j}$ are

$$\mathbf{Y}_{j,t}(\mathbf{w}_j) = \big(\mathrm{Y}_t^{(1)}(\mathbf{w}_j), \ldots, \mathrm{Y}_t^{(d_j)}(\mathbf{w}_j)\big).$$

Again, we can use the more compact notation to denote the potential outcome time series, $\mathbf{Y}_{j,t^*+1:T}(\mathbf{w}_j)$. All other assumptions and definitions easily extend to the multivariate case.

2.2. *Causal estimands*. We now develop a new class of causal estimands for which we define a contemporaneous effect (i.e., an instantaneous effect at each time point after the intervention), a cumulative effect (i.e., a partial sum of the contemporaneous effect) and an average temporal effect (i.e., a normalization of the cumulative effect). To simplify our notation, we will drop the subscript j that tracks the group and focus on analyzing each multivariate time series separately; d will then indicate the group size. Even though our goal is to estimate the heterogeneous effect on each pair, the definitions below are given for a general multivariate case where units define groups of size $d > 2$. For simplicity, for the remainder of the paper we focus on the binary treatment case where $\mathcal{W} = \{0, 1\}$. Generalizing to multiple treatments is straightforward but makes the notation more cumbersome.

Since we are following the potential outcome approach to causal inference, we restrict $t > t^*$ so that the causal effects are defined as comparisons between two potential outcomes.

DEFINITION 1. For $\mathbf{w}, \tilde{\mathbf{w}} \in \mathcal{W}^d$, the *general causal effect* of an assignment $\mathbf{w}$, compared to an alternative assignment $\tilde{\mathbf{w}}$, is

$$\begin{aligned}\boldsymbol{\tau}_t(\mathbf{w}, \tilde{\mathbf{w}}) &= \big(\tau_t^{(1)}(\mathbf{w}, \tilde{\mathbf{w}}), \ldots, \tau_t^{(d)}(\mathbf{w}, \tilde{\mathbf{w}})\big) \\ &= \big(\mathbf{Y}_t^{(1)}(\mathbf{w}) - \mathbf{Y}_t^{(1)}(\tilde{\mathbf{w}}), \ldots, \mathbf{Y}_t^{(d)}(\mathbf{w}) - \mathbf{Y}_t^{(d)}(\tilde{\mathbf{w}})\big) \\ &= \big(\mathbf{Y}_t(\mathbf{w}) - \mathbf{Y}_t(\tilde{\mathbf{w}})\big).\end{aligned} \tag{1}$$

The cumulative general causal effect at time point $t' > t^*$ is

$$\Delta_{t'}(\mathbf{w}, \tilde{\mathbf{w}}) = \sum_{t=t^*}^{t'} \boldsymbol{\tau}_t(\mathbf{w}, \tilde{\mathbf{w}}). \tag{2}$$

The temporal average general causal effect at time point t' is

$$\bar{\boldsymbol{\tau}}_{t'}(\mathbf{w}, \tilde{\mathbf{w}}) = \frac{1}{t' - t^*} \sum_{t=t^*+1}^{t'} \boldsymbol{\tau}_t(\mathbf{w}, \tilde{\mathbf{w}}) = \frac{1}{t' - t^*} \Delta_{t'}(\mathbf{w}, \tilde{\mathbf{w}}). \tag{3}$$

In a general d-variate case, the total number of general causal effects that we can estimate is $C_{2^d,2}$.

EXAMPLE 1. For the supermarket study, $d = 2$ and $\mathcal{W}^2 = \{(0,0), (0,1), (1,1), (1,0)\}$. The general causal effect, $\boldsymbol{\tau}_t((1,0),(0,0)) = \mathbf{Y}_t(1,0) - \mathbf{Y}_t(0,0)$, measures the units sold on day t when only the store brand receives a permanent discount, compared to the alternative scenario where neither receive a discount. The cumulative general causal effect $\Delta_{t'}((1,0),(0,0))$, obtained from summing the general causal effects, captures the total additional units sold, due to the price reduction up to time t'. Finally, the temporal average general effect $\bar{\tau}_t((1,0),(0,0))$ measures the average daily change in units sold, due to the new policy up to t'.

There are two natural extensions to the general causal effect: the marginal causal effect, which captures the impact of changing one unit's treatment averaged over other units' possible assignments, and the conditional causal effect which captures the effect of changing one unit's treatment fixing the other units' assignments. We provide the details in Appendix B.1, as they are not of primary interest in our supermarket study.

3. Multivariate Bayesian structural time series. We now outline our approach for estimation and inference of the causal effects defined in Section 2.2. We begin by deriving the multivariate Bayesian structural time series models (MBSTS) which are the multivariate extensions of the models used by Brodersen et al. (2015) and Papadogeorgou et al. (2018). Like their univariate versions, MBSTS models are flexible and allow for a transparent uncertainty incorporation. Flexibility comes from our ability to add subcomponents (e.g., trend, seasonality and cycle) that encapsulate the characteristics of the data. Uncertainty is quantified through the posterior distribution, which we derive and provide a sampling algorithm.

Estimation has two steps. First, we fit an MBSTS model for each pair in the period up to the intervention, $t \in \{1, \ldots, t^*\}$. Second, we estimate the target causal effects by forecasting the unobserved potential outcomes in the period following the intervention, $t \in \{t^* + 1, \ldots, T\}$. This section mirrors the two steps by first describing the model priors and posterior inference, followed by detailing the forecast and inference step. To improve the readability of the model equations, in Section 3.1 we drop the explicit dependence of the outcome on the treatment status (writing $\mathbf{Y}_t$ to indicate $\mathbf{Y}_t(\mathbf{w})$) because the model is fit using the data prior to the intervention when $\mathbf{W}_t = (0,0)$ for all store-competitor pairs and all $t \leq t^*$. We resume the usual notation in Section 3.2, where we derive the posterior distributions of the causal estimands defined in Section 2.2.

Throughout this section, we employ random matrices to simplify the notation and subsequent posterior inference by allowing us to avoid matrix vectorization. Recalling the notation introduced by Dawid (1981), let $\mathbf{Z}$ be an $(n \times d)$ matrix with standard normal entries, then $\mathbf{Z}$ follows a *standard matrix Normal distribution*, written $\mathbf{Z} \sim \mathcal{N}(I_n, I_d)$, where I_n and I_d are $(n \times n)$ and $(d \times d)$ identity matrices (the entries of $\mathbf{Z}$ are, therefore, independent). More generally, throughout the rest of paper, $\mathbf{Y} \sim \mathcal{N}(\mathbf{M}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})$ indicates that $\mathbf{Y}$ follows a matrix normal distribution with mean $\mathbf{M}$, row variance-covariance matrix $\boldsymbol{\Lambda}$ and column variance-covariance matrix $\boldsymbol{\Sigma}$. Finally, a d-dimensional vector ($n = 1$), following a multivariate standard Normal distribution, will be indicated as $\mathbf{Z} \sim N_d(\mathbf{0}, I_d)$, and $\mathcal{IW}(\nu, \boldsymbol{S})$ will denote an Inverse-Wishart distribution with ν degrees of freedom and scale matrix $\boldsymbol{S}$.

3.1. *The model.* Two equations define the MBSTS model. The first one is the "observation equation" that links the observed data $\mathbf{Y}_t$ to the state vector $\boldsymbol{\alpha}_t$ that models the different components in the data (such as trend, seasonality or cycle) and covariates which increase the counterfactual series' prediction accuracy. The second one is the "state equation" that

determines the state vector's evolution across time,

$$
\begin{aligned}
&\underbrace{\mathbf{Y}_t}_{1\times d} = \underbrace{\mathbf{Z}_t}_{1\times m}\underbrace{\boldsymbol{\alpha}_t}_{m\times d} + \underbrace{\mathbf{X}_t}_{1\times P}\underbrace{\boldsymbol{\beta}}_{P\times d} + \underbrace{\boldsymbol{\varepsilon}_t}_{1\times d}, \quad \boldsymbol{\varepsilon}_t \sim N_d(\mathbf{0}, H_t\boldsymbol{\Sigma}),\\
&\underbrace{\boldsymbol{\alpha}_{t+1}}_{m\times d} = \underbrace{\mathbf{T}_t}_{m\times m}\underbrace{\boldsymbol{\alpha}_t}_{m\times d} + \underbrace{\mathbf{R}_t}_{m\times r}\underbrace{\boldsymbol{\eta}_t}_{r\times d}, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_t, \boldsymbol{\Sigma}), \qquad \boldsymbol{\alpha}_1 \sim \mathcal{N}(\mathbf{a}_1, \boldsymbol{P}_1, \boldsymbol{\Sigma}).
\end{aligned} \tag{4}
$$

For all $t \le t^*$, $\boldsymbol{\alpha}_t$ is matrix of the m states of the d different time series, and $\boldsymbol{\alpha}_1$ is the starting value; $\mathbf{Z}_t$ is a vector selecting the states entering the observation equation; $\mathbf{X}_t$ is a vector of regressors;[3] $\boldsymbol{\beta}$ is matrix of regression coefficients, and $\boldsymbol{\varepsilon}_t$ is a vector of observation errors. For the state equation, $\boldsymbol{\eta}_t$ is a matrix of the r state errors (if all states have an error term, then $r = m$); $\mathbf{T}_t$ is a matrix defining the equation of the states components (e.g., in a simple local level model $\mathbf{T}_t = 1$), and $\mathbf{R}_t$ is a matrix selecting the rows of the state equation with nonzero error terms. Under our specification we assume that $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are mutually independent and independent of $\boldsymbol{\alpha}_1$. We denote the variance-covariance matrix of the dependencies between the time series by

$$
\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_d^2 \end{bmatrix}.
$$

Then, H_t is the variance of the observation error at time t; to simplify the notation, we can also define $\boldsymbol{\Sigma}_\varepsilon = H_t\boldsymbol{\Sigma}$. Finally, $\mathbf{C}_t$ is an $(r \times r)$ matrix of dependencies between the states' disturbances, and, since we are assuming that different states are independent, $\mathbf{C}_t$ is a diagonal matrix. Indeed, we can also write $\boldsymbol{\eta}_t \sim N_d(\mathbf{0}, \mathbf{Q}_t)$ where $\mathbf{Q}_t$ is the Kronecker product of $\mathbf{C}_t$ and $\boldsymbol{\Sigma}$, denoted by $\mathbf{Q}_t = \mathbf{C}_t \otimes \boldsymbol{\Sigma}$. Furthermore, different values in the diagonal elements of $\mathbf{C}_t$ allow each state disturbance to have its own $(d \times d)$ variance-covariance matrix $\boldsymbol{\Sigma}_r$.[4] In short,

$$
\mathbf{Q} = \mathbf{C}_t \otimes \boldsymbol{\Sigma}_\varepsilon = \begin{bmatrix} c_1\boldsymbol{\Sigma} & 0 & \cdots 0 & \\ 0 & c_2\boldsymbol{\Sigma} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_r\boldsymbol{\Sigma} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 & \cdots 0 & \\ 0 & \boldsymbol{\Sigma}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Sigma}_r \end{bmatrix}.
$$

To build intuition for the different components of the MBSTS model, we find it is useful to consider an example of a simple local level model.

EXAMPLE 2. The multivariate local level model is characterized by a trend component evolving according to a simple random walk without a seasonality component and Normally distributed disturbance terms,

$$
\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N_d(\mathbf{0}, H_t\boldsymbol{\Sigma}),\\
\boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \boldsymbol{\eta}_{t,\mu} \quad \boldsymbol{\eta}_{t,\mu} \sim N_d(\mathbf{0}, c_1\boldsymbol{\Sigma}).
\end{aligned} \tag{5}
$$

[3]Notice that this parametrization assumes the same set of regressors for each time series but still allows the coefficients to be different across the d time series.

[4]The notation $H_t\boldsymbol{\Sigma}$ and $c_r\boldsymbol{\Sigma}$ means that the dependence structure between the d series is the same for both $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$; furthermore, when H_t and $\mathbf{C}_t$ are known, the posterior distribution of $\boldsymbol{\alpha}_t$ is available in closed form (West and Harrison (2006)). Instead, we employ a simulation smoothing algorithm to sample from the posterior of the states, and in Section 3.1.2 we derive posterior distributions for $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_r$ in the general case of unknown H_t and $\mathbf{C}_t$.

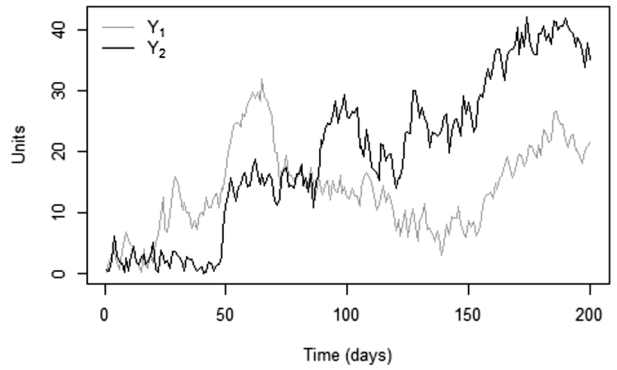

FIG. 1. *The figure shows* 200 *observations sampled from a multivariate local level model with* $d = 2$. *In our empirical application,* Y_1 *and* Y_2 *would denote the number of units sold of the store and competitor brands.*

We can recover the general formulation, outlined in (4), by setting $\boldsymbol{\alpha}_t = \boldsymbol{\mu}_t$ and $\mathbf{Z}_t = \mathbf{T}_t = \mathbf{R}_t = 1$. Figure 1 provides a graphical representation of what a sample from this model would look like when $d = 2$.

Unlike the previous example, the data in our supermarket study exhibit a weekly pattern (see Section 5). The following MBSTS models is an extension of Example 2 that includes a seasonal component:

$$
\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\gamma}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N_d(\mathbf{0}, H_t\boldsymbol{\Sigma}),\\
\boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \boldsymbol{\eta}_{t,\mu} \quad \boldsymbol{\eta}_{t,\mu} \sim N_d(\mathbf{0}, c_1\boldsymbol{\Sigma}),\\
\boldsymbol{\gamma}_{t+1} &= -\sum_{s=0}^{S-2} \boldsymbol{\gamma}_{t-s} + \boldsymbol{\eta}_{t,\gamma}\boldsymbol{\eta}_{t,\gamma} \sim N_d(\mathbf{0}, c_2\boldsymbol{\Sigma}),
\end{aligned}
\tag{6}
$$

where $\mathbf{Y}_t = (\mathrm{Y}_t^{(s)}, \mathrm{Y}_t^{(c)})$ is a bivariate vector of the units sold by the store brand, $\mathrm{Y}_t^{(s)}$, and the units sold by the corresponding competitor brand, $\mathrm{Y}_t^{(c)}$; $\boldsymbol{\mu}_t$ and $\boldsymbol{\gamma}_t$ denote, respectively, the trend and seasonal components, and $\mathbf{X}_t$ is the vector of covariates satisfying Assumption 3. Finally, $\boldsymbol{\eta}_{t,\mu}$, $\boldsymbol{\eta}_{t,\gamma}$ are the state errors having variance-covariance matrices $\boldsymbol{\Sigma}_1 = c_1\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_2 = c_2\boldsymbol{\Sigma}$, and $S = 7$ is the weekly seasonal period. We selected the trend plus seasonal model based on the results of our posterior predictive checks; as detailed in Appendix B.5, posterior predictive checks are a viable tool to assess model performance as well as to check if the covariates used in the MBSTS model are suitable controls in the absence of treatment.

3.1.1. *Prior elicitation.* The unknown parameters of Model (4) are the variance-covariance matrices of the error terms and the matrix of regression coefficients $\boldsymbol{\beta}$. Since both the observation and state errors are normally distributed, we use a conjugate Inverse-Wishart prior for their variance-covariance matrices. Generally, the MBSTS model can handle dynamic covariate coefficients. However, in our supermarket study the relationship between covariates and the outcome is likely stable over time, and so we use a matrix normal prior, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}_0, \mathbf{H}, \boldsymbol{\Sigma}_\varepsilon)$.

In our application we have a large pool of possible controls but believe that only a small subset is useful. We can incorporate such a sparsity assumption by setting $\mathbf{b}_0 = 0$ and introducing a selection vector $\boldsymbol{\varrho} = (\varrho_1, \ldots, \varrho_P)'$, with $\varrho_p \in \{0, 1\}$, $p \in [1, \ldots, P]$. Then, $\boldsymbol{\beta}_p = 0$ when $\varrho_p = 0$, meaning that the corresponding row of $\boldsymbol{\beta}$ is set to zero and the regressor X_p is excluded from our model; when $\varrho_p = 1$ then we include X_p in our model. This is known as spike and slab prior, and it can be written as

$$
\Pr(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}) = \Pr(\boldsymbol{\beta}_\varrho | \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}) \Pr(\boldsymbol{\Sigma}_\varepsilon | \boldsymbol{\varrho}) \Pr(\boldsymbol{\varrho}).
$$

We model each element of $\boldsymbol{\varrho}$ as an independent Bernoulli random variable with parameter π.

Let $\boldsymbol{\theta} = (\nu_\varepsilon, \nu_r, \mathbf{S}_\varepsilon, \mathbf{S}_r, \mathbf{X}_{1:t^*})$ be the vector of known parameters and matrices, and $\mathbf{X}_\varrho$ and $\mathbf{H}_\varrho$ be the selected regressors and the variance-covariance matrix of the corresponding rows of $\boldsymbol{\beta}$; the full set of prior distributions at time $t \leq t^*$ is

$$\begin{aligned}
\boldsymbol{\varrho}|\boldsymbol{\theta} &\sim \prod_{p=1}^{P} \pi^{\varrho_p}(1-\pi)^{1-\varrho_p},\\
\boldsymbol{\Sigma}_\varepsilon|\boldsymbol{\varrho}, \boldsymbol{\theta} &\sim \mathcal{IW}(\nu_\varepsilon, \mathbf{S}_\varepsilon),\\
\boldsymbol{\beta}_\varrho|\boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0}, \mathbf{H}_\varrho, \boldsymbol{\Sigma}_\varepsilon),\\
\boldsymbol{\alpha}_t|\mathbf{Y}_{1:t-1}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\Sigma}_r, \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{a}_t, \boldsymbol{P}_t, \boldsymbol{\Sigma}),\\
\boldsymbol{\Sigma}_r|\boldsymbol{\theta} &\sim \mathcal{IW}(\nu_r, \mathbf{S}_r).
\end{aligned}$$

For setting the prior hyperparameters, Brown, Vannucci and Fearn (1998) suggest using $\nu_\varepsilon = d + 2$, the smallest integer value such that the expectation of $\boldsymbol{\Sigma}_\varepsilon$ exists. We use a similar strategy for ν_r. As for the scale matrices of the Inverse-Wishart distributions, in our empirical analysis we set

$$\mathbf{S}_\varepsilon = \mathbf{S}_r = \begin{bmatrix} s_1^2 & s_1 s_2 \rho \\ s_1 s_2 \rho & s_2^2 \end{bmatrix},$$

where, s_1^2, s_2^2 are the sample variances of the store and the competitor brand, respectively, and ρ is a correlation coefficient that can be elicited by incorporating our prior belief on the dependence structure of the two series. Finally, we set $\mathbf{H}_\varrho = (\mathbf{X}_\varrho' \mathbf{X}_\varrho)$ which is the Zellner's g-prior (Zellner and Siow (1980)).

3.1.2. *Posterior inference.* Let $\tilde{\mathbf{Y}}_{1:t^*} = \mathbf{Y}_{1:t^*} - \mathbf{Z}_{1:t^*}\boldsymbol{\alpha}_{1:t^*}$ indicate the observations up to time t^* with the time series component subtracted out. The full conditional distributions are given by

$$\boldsymbol{\beta}_\varrho|\tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{M}, \mathbf{W}, \boldsymbol{\Sigma}_\varepsilon), \tag{7}$$

$$\boldsymbol{\Sigma}_\varepsilon|\tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\varrho}, \boldsymbol{\theta} \sim \mathcal{IW}\big(\nu_\varepsilon + t^*, \mathbf{SS}_\varepsilon\big), \tag{8}$$

$$\boldsymbol{\Sigma}_r|\boldsymbol{\eta}_{1:t^*}^{(r)}, \boldsymbol{\theta} \sim \mathcal{IW}\big(\nu_r + t^*, \mathbf{SS}_r\big), \tag{9}$$

where $\mathbf{M} = (\mathbf{X}_\varrho'\mathbf{X}_\varrho + \mathbf{H}_\varrho^{-1})^{-1}\mathbf{X}_\varrho'\tilde{\mathbf{Y}}_{1:t^*}$, $\mathbf{W} = (\mathbf{X}_\varrho'\mathbf{X}_\varrho + \mathbf{H}_\varrho^{-1})^{-1}$, $\mathbf{SS}_\varepsilon = \mathbf{S}_\varepsilon + \tilde{\mathbf{Y}}_{1:t^*}'\tilde{\mathbf{Y}}_{1:t^*} - \mathbf{M}'\mathbf{W}^{-1}\mathbf{M}$, $\mathbf{SS}_r = \mathbf{S}_r + \boldsymbol{\eta}_{1:t^*}^{\prime(r)}\boldsymbol{\eta}_{1:t^*}^{(r)}$ and $\boldsymbol{\eta}_{1:t^*}^{(r)}$ indicates the disturbances up to time t^* of the rth state. Full proof of relations (7), (8) and (9) is given in Appendix B.3.

To sample from the joint posterior distribution of the states and model parameters, we employ a Gibbs sampler in which we alternate sampling from the distribution of the states, given the parameters and sampling from the distribution of the parameters, given the states (see Algorithm 1 in Appendix B.3).

3.1.3. *Prediction and estimation of causal effects.* Let $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}_{1:t^*}, \boldsymbol{\beta}_\varrho, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\Sigma}_r, \boldsymbol{\varrho})$ be the vector of states and model parameters. We can use the joint posterior distribution $\Pr(\boldsymbol{\vartheta}|\mathbf{Y}_{1:t^*})$ to make in-sample and out-of-sample forecasts by drawing from the posterior predictive distribution. This process is particularly straightforward for in-sample forecasts.

To sample a new vector of observations $\mathbf{Y}_{1:t^*}^{\text{new}}$, given the observed preintervention data $\mathbf{Y}_{1:t^*}$, we note that

$$\Pr\big(\mathbf{Y}_{1:t^*}^{\text{new}}|\mathbf{Y}_{1:t^*}\big) = \int \Pr\big(\mathbf{Y}_{1:t^*}^{\text{new}}, \boldsymbol{\vartheta}|\mathbf{Y}_{1:t^*}\big)\, d\vartheta$$

$$
\begin{aligned}
&= \int \Pr(\mathbf{Y}^{\text{new}}_{1:t^*} | \mathbf{Y}_{1:t^*}, \boldsymbol{\vartheta}) \Pr(\theta | \mathbf{Y}_{1:t^*}) \, d\boldsymbol{\vartheta} \\
&= \int \Pr(\mathbf{Y}^{\text{new}}_{1:t^*} | \boldsymbol{\vartheta}) \Pr(\boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*}) \, d\boldsymbol{\vartheta},
\end{aligned} \tag{10}
$$

where the last equality follows because $\mathbf{Y}^{\text{new}}_{1:t^*}$ is independent of $\mathbf{Y}_{1:t^*}$, conditional on $\boldsymbol{\vartheta}$. We then obtain in-sample forecasts from the posterior predictive distribution by substituting the Gibbs draws from $\Pr(\boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*})$ into the model equations (4). We typically use in-sample forecasting for model checking.

To predict the counterfactual time series in the absence of an intervention, we need out-of-sample forecasts. Forecasting future observations, given the model estimated on the preintervention data, is still relatively straightforward; except, the new values are no longer independent of $\mathbf{Y}_{1:t^*}$, given $\boldsymbol{\vartheta}$. To see this, consider the vector $\boldsymbol{\vartheta}' = (\boldsymbol{\alpha}_{t^*+k}, \ldots, \boldsymbol{\alpha}_{t^*+1}, \boldsymbol{\vartheta})$, and let $\mathbf{Y}^{\text{new}}_{t^*+k}$ denote the k-step ahead forecast after the intervention. Then,

$$
\begin{aligned}
&\Pr(\mathbf{Y}^{\text{new}}_{t^*+k} | \mathbf{Y}_{1:t^*}) \\
&\quad = \int \Pr(\mathbf{Y}^{\text{new}}_{t^*+k}, \boldsymbol{\vartheta}' | \mathbf{Y}_{1:t^*}) \, d\boldsymbol{\vartheta}' \\
&\quad = \int \Pr(\mathbf{Y}^{\text{new}}_{t^*+k}, \boldsymbol{\alpha}_{t^*+k}, \ldots, \boldsymbol{\alpha}_{t^*+1}, \boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*}) \, d\boldsymbol{\vartheta}' \\
&\quad = \int \Pr(\mathbf{Y}^{\text{new}}_{t^*+k} | \boldsymbol{\alpha}_{t^*+k}, \ldots, \boldsymbol{\alpha}_{t^*+1}, \boldsymbol{\vartheta}, \mathbf{Y}_{1:t^*}) \Pr(\boldsymbol{\alpha}_{t^*+k} | \boldsymbol{\alpha}_{t^*+k-1}, \ldots, \boldsymbol{\alpha}_{t^*+1}, \boldsymbol{\vartheta}, \mathbf{Y}_{1:t^*}) \\
&\qquad \cdots \Pr(\boldsymbol{\alpha}_{t^*+1} | \mathbf{Y}_{1:t^*}, \boldsymbol{\vartheta}) \Pr(\boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*}) \, d\boldsymbol{\vartheta}'.
\end{aligned}
$$

To make out-of-samples forecasts, respecting the dependence structure highlighted above, we substitute the existing draws from $\Pr(\boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*})$, obtained by the Gibbs sampler, into the model equations (4), thereby updating the states and sampling the new sequence $\mathbf{Y}^{\text{new}}_{t^*+1}, \ldots, \mathbf{Y}^{\text{new}}_{t^*+k}$.

3.2. *Causal effect estimation.* We can now estimate the causal effects, defined in Section 2.2, by using the MBSTS model to predict the counterfactual outcomes. Below, we focus on the general causal effect given in equation (1); the details for the marginal and conditional effects are in Appendix B.1.

Recall that $\mathbf{Y}_{1:t^*}(0,0)$ is the observed preintervention data. For two treatments $\mathbf{w}, \tilde{\mathbf{w}} \in \{0,1\}^d$, let $\Pr(\mathbf{Y}_{t^*+k}(\mathbf{w}) | \mathbf{Y}_{1:t^*}(0,0))$ and $\Pr(\mathbf{Y}_{t^*+k}(\tilde{\mathbf{w}}) | \mathbf{Y}_{1:t^*}(0,0))$ be the posterior predictive distributions of the outcome at time $t^* + k$ under the two treatment assignments.

Then, for each draw from the posterior predictive distributions we set

$$
\boldsymbol{\tau}^{\text{new}}_{t^*+k}(\mathbf{w}, \tilde{\mathbf{w}}) = \mathbf{Y}^{\text{new}}_{t^*+k}(\mathbf{w}) - \mathbf{Y}^{\text{new}}_{t^*+k}(\tilde{\mathbf{w}}), \tag{11}
$$

yielding samples from the posterior distribution of the general causal effect. Samples from the posterior distributions of the cumulative general effect and the temporal average general effect at $t' > t^*$ can be derived from (11) as follows:

$$
\Delta^{\text{new}}_{t'}(\mathbf{w}, \tilde{\mathbf{w}}) = \sum_{t=t^*+1}^{t'} \boldsymbol{\tau}^{\text{new}}_{t}(\mathbf{w}, \tilde{\mathbf{w}}), \tag{12}
$$

$$
\bar{\boldsymbol{\tau}}^{\text{new}}_{t}(\mathbf{w}, \tilde{\mathbf{w}}) = \frac{1}{t' - t^*} \Delta^{\text{new}}_{t'}(\mathbf{w}, \tilde{\mathbf{w}}). \tag{13}
$$

Having samples from posterior distributions of the causal effects, we can easily compute posterior means and 95% credible intervals.

EXAMPLE 3. In our supermarket study we are interested in estimating the general causal effect of the permanent price reduction on the store-competitor pair, $\boldsymbol{\tau}_t((1,0),(0,0)) = \mathbf{Y}_t(1,0) - \mathbf{Y}_t(0,0)$, with $t > t^*$. For a positive integer k, $\mathbf{Y}_{t^*+k}(1,0)$ is the observed outcome postintervention, and the predictive posterior distribution of the counterfactual outcome in the absence of intervention is $\Pr(\mathbf{Y}_{t^*+k}(0,0)|\mathbf{Y}_{1:t^*}(0,0))$. To get samples from the posterior distribution of the general causal effect at time t^*+k, we draw multiple times from $\Pr(\mathbf{Y}_{t^*+k}(0,0)|\mathbf{Y}_{1:t^*}(0,0))$, that is, $\boldsymbol{\tau}^{\text{new}}_{t^*+k}((1,0),(0,0)) = \mathbf{Y}_{t^*+k}(1,0) - \mathbf{Y}^{\text{new}}_{t^*+k}(0,0)$.

Notice that (11), (12) and (13) do not require $\mathbf{Y}_t(\mathbf{w})$ or $\mathbf{Y}_t(\tilde{\mathbf{w}})$ to be observed in the postintervention period. However, estimation of unobserved potential outcomes, other than $\mathbf{Y}_t(0,0)$, requires a stronger set of modelling assumptions, making the inference less reliable.

In practice, to obtain reliable estimates of the causal effects, the assumed model has to describe the data adequately. Therefore, we recommend checking model adequacy through the use of posterior predictive checks (Rubin (1981), Rubin (1984), Gelman et al. (2013)). Under our setup we can also show that the above procedure yields unbiased estimates of the general causal effect and, in turn, the marginal and conditional effects. A detailed description of posterior predictive checks and the discussion of our estimators' frequentist properties are given, respectively, in Appendix B.5 and B.4.

3.2.1. *Combining results*. To estimate an average across the various store-competitor pairs, we can combine the separate estimates through a meta-analysis.[5] For example, denote the temporal average causal effect of the permanent price reduction on the jth cookie pair is $\bar{\boldsymbol{\tau}}_{j,t}((1,0),(0,0))$ with posterior distribution $\Pr(\bar{\boldsymbol{\tau}}_{j,t}((1,0),(0,0))|\mathbf{Y}_{1:t^*}(0,0))$ given in (13). We can define the summary temporal average effect across all j pairs,

$$\bar{\bar{\boldsymbol{\tau}}}_t\big((1,0),(0,0)\big) = \frac{1}{J}\sum_{j=1}^{J} \bar{\boldsymbol{\tau}}_{j,t}\big((1,0),(0,0)\big). \tag{14}$$

To obtain samples from the posterior distribution of $\bar{\bar{\boldsymbol{\tau}}}_t((1,0),(0,0))$, we aggregate the posterior samples from each of the j temporal average causal effect.

4. Simulation study. We now describe a simulation study exploring the frequentist properties of our proposed approach for correctly specified models and a misspecified model. The results suggest that the misspecification only leads to a minor drop in performance and that posterior predictive checks are viable approaches to assess model adequacy.

4.1. *Design*. The simulation study is specifically designed to resemble our supermarket example. As described in Section 3.1, we use an MBSTS model with both a trend and a seasonal component. The simulated data is then generated according to model (6),

$$\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\gamma}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N_d(\mathbf{0}, H_t\boldsymbol{\Sigma}),\\
\boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \boldsymbol{\eta}_{t,\mu} \quad \boldsymbol{\eta}_{t,\mu} \sim N_d(\mathbf{0}, c_1\boldsymbol{\Sigma}),\\
\boldsymbol{\gamma}_{t+1} &= -\sum_{s=0}^{S-2} \boldsymbol{\gamma}_{t-s} + \boldsymbol{\eta}_{t,\gamma} \quad \boldsymbol{\eta}_{t,\gamma} \sim N_d(\mathbf{0}, c_2\boldsymbol{\Sigma}),
\end{aligned}$$

[5]We avoid using a full joint model across different pairs for computational feasibility. Moreover, as there is no interference across the store-competitor pairs, we can factor the joint distribution into a product of marginals that can be analyzed separately.

where $\mathbf{Y}_t = (\mathrm{Y}_1, \mathrm{Y}_2)$ is a bivariate time series, $\boldsymbol{\mu}_t$ is a trend component evolving according a random walk and $\boldsymbol{\gamma}_t$ is a seasonal component with period $S = 7$. We further set $H_t = 1$, $c_1 = 3$, $c_2 = 2$ and $\boldsymbol{\Sigma} = \left[\begin{smallmatrix} 1 & -0.3 \\ -0.3 & 1 \end{smallmatrix}\right]$. We then assume a regression component formed by two covariates, $X_1 \sim f(x)$, with $f(x) = 1 - x + N(0, 0.5)$, and $X_2 \sim N(2, 0.3)$, with coefficient $\boldsymbol{\beta}$ sampled from a matrix-normal distribution with mean $\mathbf{b}_0 = \mathbf{0}$ and $\mathbf{H} = I_P$.

To estimate the causal effect, we use two different models for inference: a correctly specified model with both trend and seasonal components (M1) and a misspecified model with only the seasonal part (M2). For both models we choose the following set of hyperparameters: $\nu_\varepsilon = \nu_r = 4$; $\mathbf{S}_\varepsilon = \mathbf{S}_r = 0.2\left[\begin{smallmatrix} s_1^2 & s_1 s_2 \rho \\ s_1 s_2 \rho & s_2^2 \end{smallmatrix}\right]$, where s_1^2 and s_2^2 are the sample variances of Y_1 and Y_2, respectively, $\rho = -0.8$ is a correlation coefficient reflecting our prior belief of their dependence structure and Zellner's g-prior for the variance-covariance matrix of $\boldsymbol{\beta}$.

To make our simulation close to our empirical application, we generated 1000 data sets in a fictional time period, starting January 1, 2018, and ending June 30, 2019. We model the intervention as taking place on January 2, 2019, and assume a fixed persistent contemporaneous effect; for example, the series goes up by $+10\%$ and stays at this level throughout. To study the empirical power and coverage, we tried *five* different impact sizes, ranging from $+1\%$ to $+100\%$ on Y_1 and from -1% to -90% on Y_2. After generating the data, we estimated the effects, using both M1 and M2, for a total of $10,000$ estimated models (one for each data set, impact size and model type), each having 1000 draws from the resulting posterior distribution. Finally, we predicted the counterfactual series in the absence of intervention for three-time horizons, namely, after *one* month, *three* months and *six* months from the intervention.

We evaluate the performance of the models in terms of:

1. length of the credible intervals around the temporal average general effect $\bar{\tau}_t((1,0),(0,0))$;
2. absolute percentage estimation error, computed as $\frac{|\hat{\bar{\tau}}_t((1,0),(0,0)) - \bar{\tau}_t((1,0),(0,0))|}{\bar{\tau}_t((1,0),(0,0))}$;
3. interval coverage, namely, the proportion of the true pointwise effects covered by the estimated 95% credible intervals.

We focus on the percentage estimation error because different effect sizes are not immediately comparable without normalizing. For example, a small bias when estimating a large effect is better than the same bias when estimating a much smaller effect.

4.2. *Results*. Table 1 reports the average interval length for M1 and M2 across the different effect sizes and time horizons. As expected, the length of credible intervals, estimated under M1, increases with the time horizon. In contrast, for M2 the interval length is stable across time, as the model lacks a trend component. Figure 2 shows the absolute percentage error decreases as the effect size increases because small effects are more difficult to detect. To confirm this claim, in Figure 3 we report the percentage of times we detect a causal effect over the 1000 simulated data sets. Under M1 for the two smallest effect sizes, which exhibit the highest estimation errors, we rarely correctly conclude that a causal effect is present. However, when the effect size increases, we can detect the presence of a causal effect at a much higher rate. The results under M2 are somewhat counterintuitive as, even though the model is misspecified, smaller effects are more readily detected. This phenomenon occurs because of the smaller credible intervals; that is, for small effect sizes our results are biased with low variance which means we often conclude there is an effect.

Table 2 reports the average interval coverage under M1 and M2. The coverage under M2 ranges from 82.0% to 88.6% which is lower than the desired 95%. In contrast, the frequentists coverage under M1 is at the nominal 95% for both Y_1 and Y_2.

TABLE 1

Length of credible intervals around the temporal average general effect, $\bar{\tau}_t((1,0),(0,0))$, estimated under M1 and M2, for each effect size and time horizon

		1 month		3 months		6 months	
	$\bar{\tau}_t((1,0),(0,0))$	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
M1	(1.01, 0.99)	20.93	21.10	27.62	27.80	46.58	46.28
	(1.10, 0.90)	21.34	21.37	28.09	28.15	46.98	46.89
	(1.25, 0.75)	21.33	21.30	28.18	28.09	47.11	46.97
	(1.50, 0.50)	21.30	21.31	28.11	28.11	47.02	46.91
	(2.00, 0.10)	21.38	21.25	28.24	28.06	47.12	46.90
M2	(1.01, 0.99)	30.39	30.39	30.40	30.41	30.48	30.47
	(1.10, 0.90)	30.48	30.48	30.50	30.50	30.57	30.58
	(1.25, 0.75)	30.48	30.46	30.51	30.49	30.60	30.58
	(1.50, 0.50)	30.45	30.43	30.47	30.46	30.55	30.54
	(2.00, 0.10)	30.49	30.49	30.52	30.51	30.60	30.57

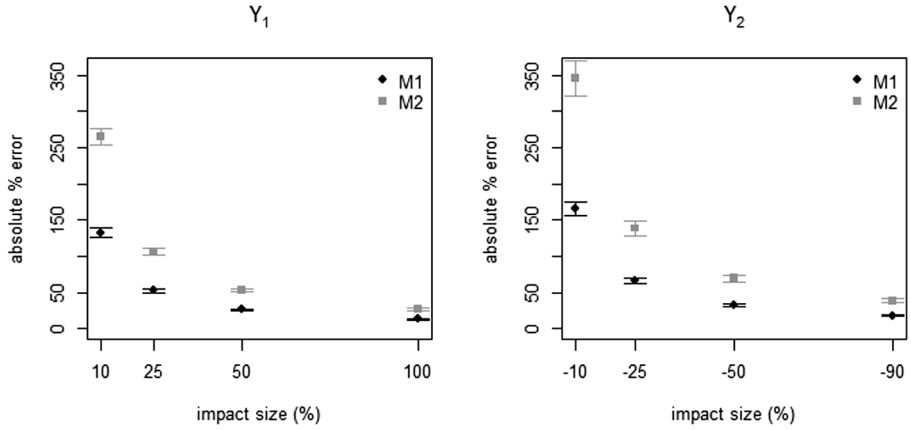

FIG. 2. *Average absolute percentage error* ($\pm$ 2 *s.e.m*) *at the first time horizon under M1 and M2 for the impact sizes* $\geq 10\%$ *(*Y_1*) and* $\leq -10\%$ *(*Y_2*).*

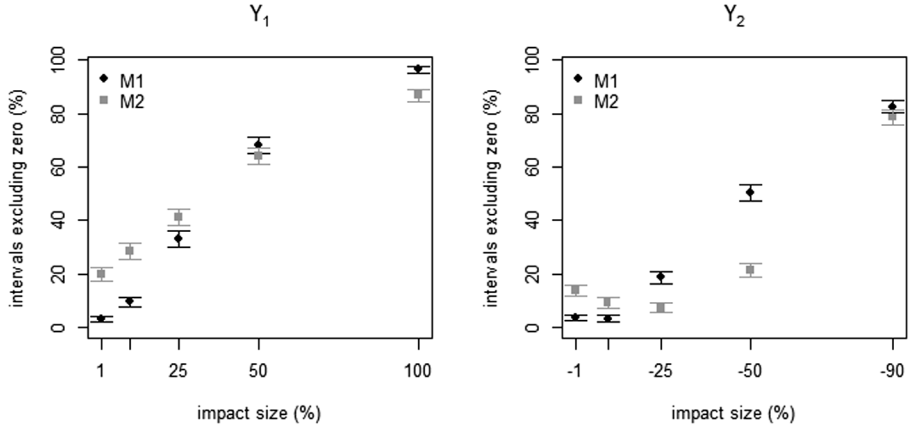

FIG. 3. *Average proportion of credible intervals excluding zero* ($\pm$ 2 *s.e.m*) *at the first time horizon under M1 and M2 for all impact sizes.*

TABLE 2
Interval coverage under M1 and M2 for each effect size and time horizon

		1 month		3 months		6 months	
	$\bar{\tau}_t((1,0),(0,0))$	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
M1	(1.01, 0.99)	96.0	95.0	96.1	95.3	96.0	96.3
	(1.10, 0.90)	95.9	94.9	96.0	95.2	95.9	96.3
	(1.25, 0.75)	96.0	95.0	96.0	95.3	96.0	96.2
	(1.50, 0.50)	96.1	94.9	96.1	95.2	96.1	96.2
	(2.00, 0.10)	95.9	95.0	96.1	95.3	96.0	96.3
M2	(1.01, 0.99)	86.8	88.4	85.5	87.2	82.0	84.6
	(1.10, 0.90)	87.0	88.5	85.7	87.3	82.1	84.7
	(1.25, 0.75)	87.0	88.6	85.7	87.3	82.1	84.7
	(1.50, 0.50)	86.9	88.6	85.6	87.3	82.0	84.7
	(2.00, 0.10)	86.9	88.6	85.7	87.3	82.1	84.6

Overall, the simulation results suggest that, when the model is correctly specified, the proposed approach performs well in estimating the causal effect of an intervention. Conversely, when the model is misspecified, the estimation error increases and the credible intervals do not achieve the required coverage; however, the results are likely to provide practitioners with useful insights.

In practice, we recommend assessing our model's adequacy before performing substantive analysis by using posterior predictive checks. In Appendix A.1 we provide examples results obtained under M1 (Figures 3 and 4) and posterior predictive checks under both M1 and M2. From the observation of Figures 5 and 6, we can immediately see that M1 yields a better approximation of the empirical density of the simulated data and lower residual autocorrelation than M2.

5. Empirical analysis. We now describe the results of our empirical application, where we analyze the efficacy of a strategic shift by an Italian supermarket chain to permanently reduce the price of a selected subset of store brands in its Florence stores. The firm's primary objective was to increase the customer base and sales. The policy change affected 707 products in several categories; below, we provide the details for the "cookies" category.

5.1. *Data & methodology.* Among the 284 items in the "cookies" category, there are 28 store brands, of which 10 were selected for a permanent price reduction, ranging from -3.5% to -23.2% (the median was -11.8%). For each store brand the supermarket chain identified a direct competitor brand, thereby defining 10 pairs of cookies.

Those in the same pair are almost identical, except for their brand name. In contrast, cookies belonging to different pairs differ on many characteristics (e.g., ingredients, target market and weight). As discussed earlier, in this setup the permanent discount on a store brand is likely to impact its direct competitor but is unlikely to affect the sales of the cookies in different pairs, allowing us to justify the partial temporal no-interference assumption.

Our data consists of daily sales for all cookies from September 1, 2017, until April 30, 2019. Our outcome variable is the average units sold per hour, computed as the number of units sold daily divided by the number of hours that the stores stay open. We focus on hourly average sales because Italian regulations dictate that the supermarket chain only operates for a limited number of hours on Sundays; this discrepancy leads to a considerable difference in daily sales. As an example, Figure 4 shows the time series of daily units sold by two store

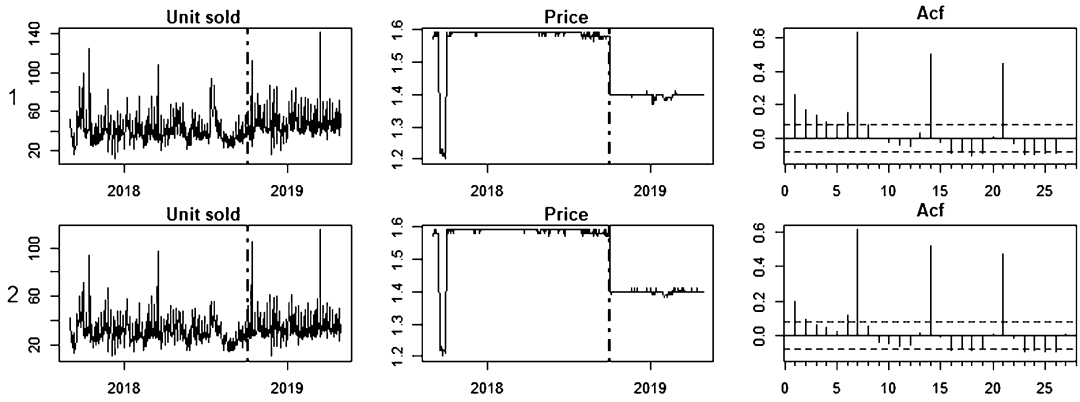

FIG. 4. Store brands. *Starting from the left*: *Time series of the average unit sold per hour*; *evolution of price per unit*; *autocorrelation function*. *The price plot shows the permanent price reduction after the intervention date* (*indicated by the vertical dashed line*).

brands, their price and the autocorrelation function. The plots show a strong weekly seasonal pattern. Figure 5 exhibits the same plots for two competitor brands.[6] The occasional drops in the price series are from temporary promotions run regularly by the supermarket chain. In our data the competitor brands are subject to several promotions during the analysis period. However, those differ from the permanent price reduction on their temporary nature and the regular frequency. As our goal is to evaluate the effectiveness of the store's policy change—a permanent price reduction; we will not consider temporary promotions as interventions. There is also considerable visual evidence that the store brands' intervention influenced the competitor cookies' pricing strategy. Indeed, all competitor brands (except for brand 10) received a temporary promotion matching the time of the intervention, suggesting that competitors may have reacted to the new policy.[7]

Under partial temporal no-interference, we fit an MBSTS model for each pair; we also use covariates to improve the prediction of the counterfactual series. In particular, the set of re-

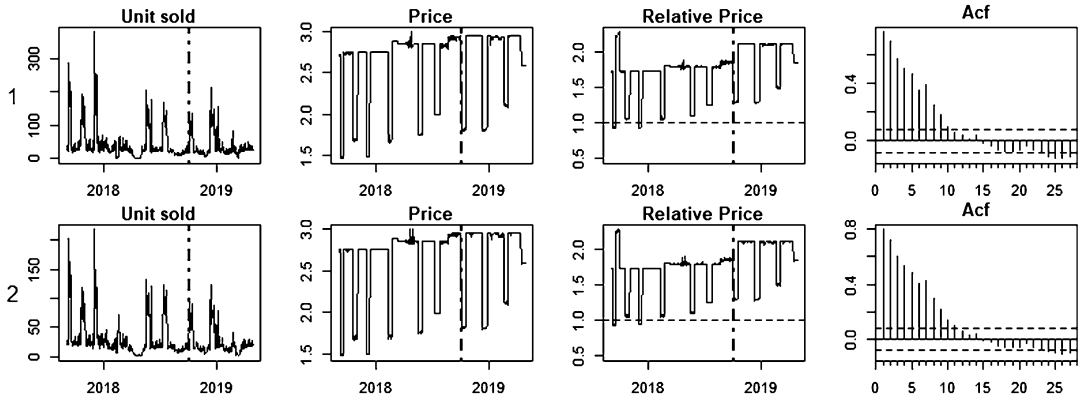

FIG. 5. Competitor brands. *Starting from the left*: *Time series of the average unit sold per hour*; *evolution of price per unit*; *evolution of price relative to the store brand* (*the dashed horizontal line indicates a relative price equal to* 1); *autocorrelation function*. *The price plot shows the temporary promotions these brands are subject to*, *that is*, *both before and after the intervention date* (*indicated by the vertical dashed line*) *the price of competitor brands is reduced for a while and then bounces back to the original level*.

[6]The equivalent plots for all the remaining store and competitor brands are provided in Appendix A.1.

[7]See Figure 2 in Appendix A.1.

gressors include: (1) two dummies taking value 1 on Saturday and Sunday, the former being the most profitable day of the week, whereas, on the latter, stores operate reduced hours; a holiday dummy taking value 1 on the day before and after a national holiday, accounting for consumers' tendency to shop more before and after a closure day, and (2) a set of synthetic controls selected among one category (e.g., wine sales) that did not receive active treatment. Including covariates should increase prediction accuracy in the absence of intervention, but suitable covariates must respect two conditions: they should be good predictors of the outcome before the intervention, and they must satisfy Assumption 3. As a result, the unit prices can not be part of our models; nevertheless, they are important drivers of sales, especially during promotions (Blattberg, Briesch and Fox (1995), Neslin, Henderson and Quelch (1985), Pauwels, Hanssens and Siddarth (2002)). We solved this issue by using the "prior price," which is equal to the actual price up to the intervention, and then it is set equal to the last price before intervention (which is the most reliable estimate of the price without an intervention).

Finally, to speed up computations, the set of synthetic controls is selected in two steps: first, we select the best 10 matches among the 260 possible control series in the "wines" category by dynamic time warping;[8] then, we group them with the other predictors and perform multivariate Bayesian variable selection.

Each model is estimated in the period before the intervention; then, as described in Section 3.1.3, we predict the counterfactual series in the absence of intervention by performing out-of-sample forecasts. Next, we estimate the intervention's causal effect at three different time horizons—one month, three months and six months from the treatment day. This allows us to determine whether the effect persists over time or quickly disappears.

5.2. *Results*. We now present the results for the best MBSTS model with both a trend and seasonality component. The model was selected amongst an array of possible alternatives, using posterior predictive checks; see Appendix A.1 for the details and Appendix A.2 for a description of the other models tried. Convergence diagnostics are provided in Appendix B.7.

The estimates of the temporal average general effect, reported in Table 3, reveal the presence of three significant causal effects—where the 95% credible intervals do not include 0—on the store brands belonging to pairs 4, 7 and 10 at the first time horizon. Interestingly, we do not find a significant effect on the competitor brands in the same pairs, most likely because, during the intervention period, competitor brands were subject to multiple temporary promotions that might have reduced the negative impact of the permanent discount on store brands. Furthermore, Italian supermarket chains have introduced store brands products only in recent years; so, despite the price reduction on store brand cookies, some consumers may still prefer the competitor cookie because of subjective factors, such as brand loyalty. Another important result is that after the initial surge in sales, we cannot detect a significant effect for longer time horizons. Figure 6 plots the pointwise general effect $\hat{\boldsymbol{\tau}}_t((1,0),(0,0))$ for the fourth pair at each time horizon, that is, the difference between the observed series and the predicted counterfactual computed at every time point; see Appendix A.1 for additional plots.

Overall, these results suggest that the firm's strategic change had a minor impact on the store brands' sales. Furthermore, since we do not detect an effect after the first month, it

[8]Dynamic time warping (DTW) is a technique for finding the optimal alignment between two time series. Instead of minimizing the Euclidean distance between the two sequences, it finds the minimum-distance warping path, i.e., given a matrix of distances between each point of the first series with each point of the second series, the path through the matrix minimizing the total cumulative distance between the two sequences. For further details, see Keogh and Ratanamahatana (2005), Salvador and Chan (2007). Implementation of DTW has been done with the R package `MarketMatching` (Larsen (2019)).

TABLE 3

Temporal average general causal effects of the new price policy on the 10 *store* (*s*) *- competitor* (*c*) *pairs computed at three time horizons. In this table,* $\hat{\tilde{\boldsymbol{\tau}}}_t$ *stands for the general effect* $\hat{\tilde{\boldsymbol{\tau}}}_t((1,0),(0,0))$

		1 month			3 months			6 months		
		$\hat{\tilde{\boldsymbol{\tau}}}_t$	2.5%	97.5%	$\hat{\tilde{\boldsymbol{\tau}}}_t$	2.5%	97.5%	$\hat{\tilde{\boldsymbol{\tau}}}_t$	2.5%	97.5%
(1)	s	6.97	−24.25	38.47	4.68	−44.00	53.61	6.99	−65.91	79.55
	c	24.89	−101.30	153.64	17.49	−193.06	219.08	5.09	−307.48	309.00
(2)	s	7.02	−14.79	28.90	4.92	−30.20	38.56	6.56	−44.17	58.01
	c	14.71	−62.26	99.44	8.92	−119.33	144.72	0.92	−205.51	201.82
(3)	s	7.94	−14.08	32.26	5.30	−31.95	41.38	7.82	−48.46	62.50
	c	15.42	−62.17	90.81	11.06	−113.64	132.60	4.84	−189.44	197.55
(4)	s	**47.84**	**4.71**	**96.82**	22.65	−52.13	96.38	23.73	−88.10	131.67
	c	28.86	−77.93	135.93	20.91	−151.05	190.01	11.20	−256.88	279.74
(5)	s	4.11	−46.65	54.64	7.57	−76.37	91.02	11.75	−111.67	136.65
	c	45.47	−63.13	154.24	16.68	−156.03	188.67	9.42	−263.47	280.16
(6)	s	9.53	−14.45	33.68	11.76	−28.33	51.70	13.58	−45.97	74.20
	c	25.64	−37.88	93.36	6.71	−104.80	113.12	4.13	−163.82	164.96
(7)	s	**78.19**	**0.15**	**154.08**	34.45	−82.11	151.65	29.48	−149.12	206.10
	c	182.70	−221.16	600.08	102.61	−581.90	769.52	80.62	−951.26	1069.94
(8)	s	25.23	−28.60	78.16	23.34	−67.87	109.37	17.07	−115.20	145.12
	c	15.91	−15.15	47.53	6.03	−44.60	60.30	3.82	−73.60	82.80
(9)	s	40.29	−9.84	90.38	15.37	−64.38	97.76	12.07	−108.11	136.44
	c	17.17	−30.76	68.56	1.20	−79.88	84.48	2.81	−118.55	127.05
(10)	s	**12.43**	**1.35**	**23.64**	9.64	−8.07	27.98	5.30	−22.02	32.67
	c	0.04	−9.36	9.79	1.92	−13.22	17.72	4.00	−18.33	27.03

seems that this intervention failed to significantly and permanently impact sales. Of course, as we showed in the simulation study, there could have been a small effect that our model was unable to detect. However, since the firm needed a significant boost in sales to make up for the loss in profits due to the price reduction, we can conclude that this policy was ineffective. This result is robust to different prior assumptions (see Appendix B.6 for detailed sensitivity analysis) and to modifications in the set of covariates. In particular, we obtain similar results

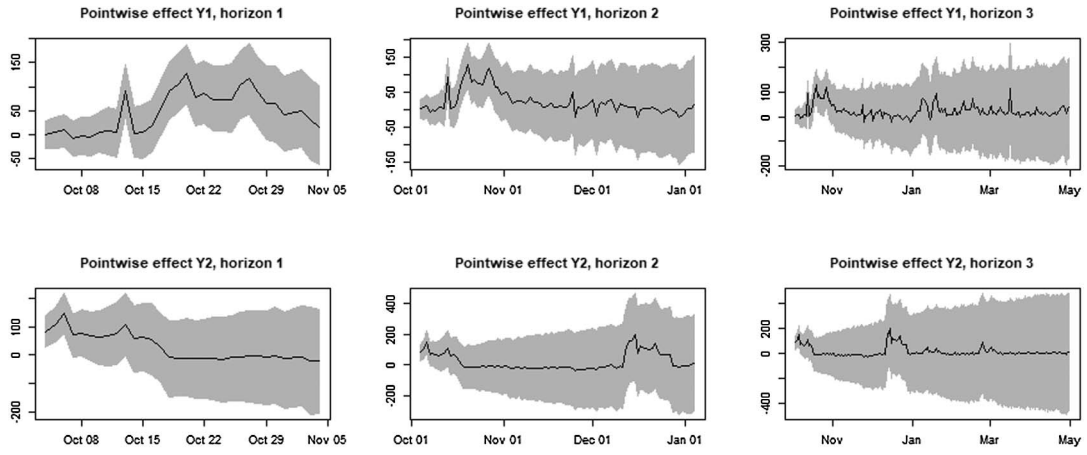

FIG. 6. *Pointwise causal effect of the permanent price reduction on the fourth store-competitor pair at one month, three months and six months after the intervention.*

when, instead of using the individual prices, we include among the predictors the difference in price or the price ratio between the store and competitor brand (see Tables 3 and 4 in Appendix B.2).

An alternative analysis strategy is to aggregate the sales of store and competitor brands and treating each aggregate as a univariate time series. However, this procedure leads to a loss of information, providing misleading results that could drive the analyst to make the wrong decision. To show that, we estimated the causal effect using the univariate BSTS models on a range of different aggregated sales. We report the results for three: the average sales of the brands in the same pair, the average sales of all store brands and the average sales of all store and competitor brands. The average is computed as the total number of units sold daily by all products in the aggregate divided by the opening hours. Notice that we did not consider the aggregate of the competitor brands alone. This is because it would have required the prediction of the counterfactual series under treatment.

Like the multivariate analysis, for each aggregate we used a model that contained a trend and seasonality component as well as a set of covariates. The covariates included the three dummies (described earlier), aggregate sales of all wines and the prior price computed by averaging the prior prices of all cookies in each aggregate. Table 4 shows the results of the univariate analysis. We find evidence of a positive effect on the tenth pair at the first and second-time horizons and a positive effect on the eighth pair at the first horizon. In addition, the estimated effects on the store brands aggregate and the store-competitor aggregate are both positive and significant for the first time horizon. To provide a comparison with these last two aggregates, Table 4 reports the summary temporal average effect on all cookie pairs obtained by combining the individual estimates with a meta-analysis, as described in Section 3.2.1. The summary effect on the store brands is positive and significant at the first time horizon, and, interestingly, it is in line with the estimated effect on the store brands aggregate from the univariate analysis. However, with the univariate analysis we cannot isolate the effect on the competitor brands, and we would have erroneously concluded that the new policy had a positive impact on the store-competitor aggregate. In contrast, the meta-analysis shows

TABLE 4

Univariate temporal average causal effect ($\hat{\bar{\tau}}_t$) *at three time horizons of the new price policy on*: (*i*) *aggregated sales* (*pairs* 1–10); (*ii*) *the store brands aggregate* (*SA*); (*iii*) *the store - competitor aggregate* (*SCA*). *The last two lines show, separately for the store brands* (*META-S*) *and the competitor brands* (*META-C*), *the summary temporal average effect combined with a meta-analysis*

	1 month			3 months			6 months		
	$\hat{\bar{\tau}}_t$	2.5%	97.5%	$\hat{\bar{\tau}}_t$	2.5%	97.5%	$\hat{\bar{\tau}}_t$	2.5%	97.5%
Pair 1	16.65	−36.89	64.97	12.46	−73.66	93.47	6.97	−115.80	130.39
Pair 2	9.85	−25.50	42.76	4.56	−54.77	62.29	−0.24	−85.55	85.37
Pair 3	11.20	−29.89	48.21	8.66	−58.13	73.73	6.25	−90.95	107.34
Pair 4	36.86	−4.18	75.70	22.78	−46.31	87.32	18.50	−76.66	119.12
Pair 5	29.05	−40.13	88.51	11.51	−102.42	121.54	10.70	−158.37	186.19
Pair 6	16.86	−14.59	44.80	4.09	−50.47	57.12	5.40	−74.01	88.53
Pair 7	120.86	−129.59	352.65	75.54	−272.11	393.52	57.87	−568.82	687.77
Pair 8	**20.06**	**4.95**	**34.39**	12.59	−11.39	36.03	8.91	−25.75	42.42
Pair 9	28.58	−0.03	55.95	8.51	−38.36	54.54	9.53	−56.66	78.61
Pair 10	**7.29**	**4.19**	**10.00**	**6.63**	**1.64**	**10.94**	5.75	−1.49	12.17
SA	**25.01**	**10.08**	**39.04**	15.04	−8.80	37.56	15.52	−19.30	49.19
SCA	**34.56**	**8.55**	**58.78**	19.98	−20.53	58.62	16.16	−44.40	78.19
META-S	**23.95**	**3.62**	**45.32**	13.97	−18.39	47.89	13.43	−34.05	67.37
META-C	37.08	−34.98	106.39	19.35	−100.10	133.78	12.68	−163.61	184.61

that the effect on competitor brands is not significant. Overall, despite a similar result for the tenth pair, we would have reached the wrong conclusions for pairs 4, 7 and 8 and would have reported the misleading finding of an overall positive impact on the sales.

To further illustrate the range of possible causal estimands in a multivariate setting, we also estimated the marginal and the conditional effects. The results, given in Table 1 and Table 2 in Appendix B.1, show three significant marginal effects on the sales of store brands and little evidence of a conditional effect.

6. Conclusion. This paper presents a causal analysis of the effectiveness of a new pricing strategy implemented by an Italian supermarket chain. The results suggest that the policy change had a minor impact on the store brands' sales and little evidence of a detrimental effect on competitor brands. Our findings relied on a new methodology for analyzing the effectiveness of a single persistent intervention in the presence of partial interference. Interestingly, we showed that methods that fail to account for the interference lead to incorrect results that overestimate the price reduction's effectiveness.

We believe that our approach brings several contributions to the nascent stream of literature on synthetic control methods in panel settings with interference. First, we derived a wide class of new causal estimands. Second, MBSTS allows us to model the interference between units in the same group by explicitly modeling their dependence structure and, simultaneously, ensuring a transparent way to deal with the surrounding uncertainty. Finally, the approach is flexible, and the underlying distributional assumptions can be tested in a very natural way by posterior inference.

SUPPLEMENTARY MATERIAL

Appendices.pdf (DOI: 10.1214/21-AOAS1498SUPPA; .pdf). A pdf file that contains Appendix A and Appendix B.

Rcodes_and_data.zip (DOI: 10.1214/21-AOAS1498SUPPB; .zip). A zip file that contains the code and the data set to replicate the analysis shown in the paper.

REFERENCES

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* **105** 493–505. MR2759929 https://doi.org/10.1198/jasa.2009.ap08746

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2015). Comparative politics and the synthetic control method. *Amer. J. Polit. Sci.* **59** 495–510.

ABADIE, A. and GARDEAZABAL, J. (2003). The economic costs of conflict: A case study of the Basque country. *Am. Econ. Rev.* **93** 113–132.

BASSE, G. W., FELLER, A. and TOULIS, P. (2019). Randomization tests of causal effects under interference. *Biometrika* **106** 487–494. MR3949317 https://doi.org/10.1093/biomet/asy072

BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2018). The augmented synthetic control method. Preprint. Available at arXiv:1811.04170.

BILLMEIER, A. and NANNICINI, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Rev. Econ. Stat.* **95** 983–1001.

BLATTBERG, R. C., BRIESCH, R. and FOX, E. J. (1995). How promotions work. *Mark. Sci.* **14** G122–G132.

BOJINOV, I., CHEN, A. and LIU, M. (2020). The importance of being causal. *Harvard Data Science Review*.

BOJINOV, I. and MENCHETTI, F. (2020). CausalMBSTS: MBSTS Models for Causal Inference and Forecasting. R package version 0.1.0.

BOJINOV, I., RAMBACHAN, A. and SHEPHARD, N. (2020). Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective. Preprint. Available at arXiv:2003.09915.

BOJINOV, I. and SHEPHARD, N. (2019). Time series experiments and causal estimands: Exact randomization tests and trading. *J. Amer. Statist. Assoc.* **114** 1665–1682. MR4047291 https://doi.org/10.1080/01621459.2018.1527225

BRODERSEN, K. H., GALLUSSER, F., KOEHLER, J., REMY, N. and SCOTT, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Ann. Appl. Stat.* **9** 247–274. MR3341115 https://doi.org/10.1214/14-AOAS788

BROWN, P. J., VANNUCCI, M. and FEARN, T. (1998). Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 627–641. MR1626005 https://doi.org/10.1111/1467-9868.00144

CAO, J. and DOWD, C. (2019). Estimation and inference for synthetic control methods with spillover effects. Preprint. Available at arXiv:1902.07343.

COX, D. R. (1958). *Planning of Experiments. A Wiley Publication in Applied Statistics*. Wiley, New York. MR0095561

DAWID, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68** 265–274. MR0614963 https://doi.org/10.1093/biomet/68.1.265

DUBE, A. and ZIPPERER, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies. IZA Discussion Paper 8944.

FORASTIERE, L., AIROLDI, E. M. and MEALLI, F. (2021). Identification and Estimation of Treatment and Interference Effects in Observational Studies on Networks. *J. Amer. Statist. Assoc.* **116** 901–918. MR4270033 https://doi.org/10.1080/01621459.2020.1768100

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*. CRC press, Boca Raton.

GOBILLON, L. and MAGNAC, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Rev. Econ. Stat.* **98** 535–551.

GROSSI, G., LATTARULO, P., MARIANI, M., MATTEI, A. and ÖNER, Ö. (2020). Synthetic Control Group Methods in the Presence of Interference: The Direct and Spillover Effects of Light Rail on Neighborhood Retail Activity. Preprint. Available at arXiv:2004.05027.

HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 https://doi.org/10.1198/016214508000000292

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge Univ. Press, Cambridge.

KEOGH, E. and RATANAMAHATANA, C. A. (2005). Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **7** 358–386.

KREIF, N., GRIEVE, R., HANGARTNER, D., TURNER, A. J., NIKOLOVA, S. and SUTTON, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Econ.* **25** 1514–1528. https://doi.org/10.1002/hec.3258

LARSEN, K. (2019). MarketMatching Package Vignette. R package version 1.1.2.

LI, K. T. (2020). Statistical inference for average treatment effects estimated by synthetic control methods. *J. Amer. Statist. Assoc.* **115** 2068–2083. MR4189777 https://doi.org/10.1080/01621459.2019.1686986

MENCHETTI, F. and BOJINOV, I. (2022). Supplement to "Estimating the effectiveness of permanent price reductions for competing products using multivariate Bayesian structural time series models." https://doi.org/10.1214/21-AOAS1498SUPPA, https://doi.org/10.1214/21-AOAS1498SUPPB

NESLIN, S. A., HENDERSON, C. and QUELCH, J. (1985). Consumer promotions and the acceleration of product purchases. *Mark. Sci.* **4** 147–165.

NICHOLSON, W. and SNYDER, C. M. (2012). *Microeconomic Theory: Basic Principles and Extensions*. Nelson Education.

O'NEILL, S., KREIF, N., GRIEVE, R., SUTTON, M. and SEKHON, J. S. (2016). Estimating causal effects: Considering three alternatives to difference-in-differences estimation. *Health Serv. Outcomes Res. Methodol.* **16** 1–21. https://doi.org/10.1007/s10742-016-0146-8

PAPADOGEORGOU, G., MEALLI, F., ZIGLER, C. M., DOMINICI, F., WASFY, J. H. and CHOIRAT, C. (2018). Causal Impact of the Hospital Readmissions Reduction Program on Hospital Readmissions and Mortality. Preprint. Available at arXiv:1809.09590.

PAUWELS, K., HANSSENS, D. M. and SIDDARTH, S. (2002). The long-term effects of price promotions on category incidence, brand choice, and purchase quantity. *J. Mark. Res.* **39** 421–439.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512. MR0877758 https://doi.org/10.1016/0270-0255(86)90088-6

ROBINS, J. M., GREENLAND, S. and HU, F.-C. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *J. Amer. Statist. Assoc.* **94** 687–712. MR1723276 https://doi.org/10.2307/2669978

ROSENBAUM, P. R. (2007). Interference between units in randomized experiments. *J. Amer. Statist. Assoc.* **102** 191–200. MR2345537 https://doi.org/10.1198/016214506000001112

RUBIN, D. B. (1981). Estimation in parallel randomized experiments. *J. Educ. Stat.* **6** 377–401.

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172. MR0760681 https://doi.org/10.1214/aos/1176346785

SALVADOR, S. and CHAN, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **11** 561–580.

SOBEL, M. E. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *J. Amer. Statist. Assoc.* **101** 1398–1407. MR2307573 https://doi.org/10.1198/016214506000000636

TCHETGEN TCHETGEN, E. J. and VANDERWEELE, T. J. (2012). On causal inference in the presence of interference. *Stat. Methods Med. Res.* **21** 55–75. MR2867538 https://doi.org/10.1177/0962280210386779

VANDERWEELE, T. J. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociol. Methods Res.* **38** 515–544. MR2758165 https://doi.org/10.1177/0049124110366236

VIVIANO, D. and BRADIC, J. (2019). Synthetic learner: Model-free inference on treatments over time. Preprint. Available at arXiv:1904.01490.

WEST, M. and HARRISON, J. (2006). *Bayesian Forecasting and Dynamic Models*. Springer, Berlin.

ZELLNER, A. and SIOW, A. (1980). Posterior odds ratios for selected regression hypotheses. *Trab. Estad. Investig. Oper.* **31** 585–603.