



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DOTTORATO DI RICERCA IN  
Matematica Informatica e Statistica  
Curriculum STATISTICA**

CICLO XXXIII


COORDINATORE Prof. Paolo Salani

Causal Inference in Time Series Settings under the Rubin Causal Model

Settore Scientifico Disciplinare SECS-S/03

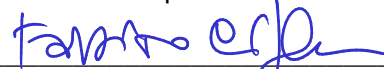
**Dottorando**

Dott. Menchetti Fiammetta

  
(firma)


**Tutore**

Prof. Cipollini Fabrizio

  
(firma)

**Co-tutore**

Prof. Mealli Fabrizia

  
(firma)

**Coordinatore**

Prof. Salani Paolo

  
(firma)

Anni 2017/2020

---

# Causal Inference in Time Series settings under the Rubin Causal Model

---

Author : Fiammetta Menchetti  
Supervisors: Fabrizio Cipollini, Fabrizia Mealli

March 15, 2021

## Abstract

The aim of this dissertation is to present novel approaches to estimate causal effects of interventions in time series settings under the Rubin Causal Model (RCM), which is a framework to define causal estimands, discuss assumptions and develop methods for the computation of causal effects. The dissertation is structured in three main sections describing different time series settings: (i) a panel situation where multiple time series are subject to a simultaneous intervention; (ii) a panel study where the time series interact with one another; (iii) a multi-intervention setting with a single time series subject to several interventions. Each section is connected to a research paper: (i) and (iii) are joint works with Fabrizio Cipollini (University of Florence) and Fabrizia Mealli (University of Florence); (ii) is a joint work with Iavor Bojinov (Harvard Business School) and it was developed during my visiting period at Harvard University.

## Acknowledgments

In the last three years I have gone really far from where I started. This dissertation and the works that will follow are my greatest achievements so far and they all were possible because many people contributed to them in the most unexpected and fantastic way.

First of all, I would like to thank my supervisors Fabrizia Mealli and Fabrizio Cipollini for their valuable advice, support and guidance during my PhD study. A special thank to Professor Cipollini for the weekly calls he's patiently engaged with me for the last 6 months: this thesis would certainly be sitting on a dusty shelf in my room without his help at a time when it was most needed.

I also thank Iavor Bojinov and Tommaso Proietti for their helpful comments to the previous version of this dissertation. Besides, I would like to express my deep and sincere gratitude to Iavor for giving me the opportunity to visit him in one of the greatest schools in the world: the time I spent at HBS was the most fruitful for my research and I feel that I have learned a lot from him and from the work we've done together.

A few months after the initial self-esteem boost provided by having passed the admission test, I would have bet against myself that I wasn't going to make it until the end. Luckily, I found the most wonderful colleagues and friends, Giulia and Cecilia, with whom I shared this experience and that never left me behind. Thanks for all you taught me and for the great time we spent together. I'm looking forward to our next wine tour to Greve in Chianti!

I would also like to thank a very special friend of mine. I learned so much from you, Michele: I am a true running enthusiast now! Thanks for having shared your passion with me, you cannot imagine how important this was since the spread of the global pandemic. Besides, you inspired me and gave me strength to reach my goals. I am honored to be your friend.

I cannot help to mention at this point that there were also very difficult moments when I thought I should give up. If I didn't, it was certainly because of my obstinacy and my love for lost causes but, mostly, for I have been gifted with the most special and warm family that one can desire. Thanks for always listening to my grievances and supporting me in any way you can. I know you got my back.

Last but not least, to my life companion Giulio, who's been bearing my bad temper for 15 years (poor soul): you supported me on far too many occasions that I can count and if it hadn't been for you, this thesis would have never seen the light. I know this PhD hasn't made things easier for you, but there you are.

# Contents

<b>1</b>	<b>Introduction &amp; relevant literature</b>	<b>5</b>
<b>2</b>	<b>Common theoretical background</b>	<b>12</b>
2.1	Assumptions . . . . .	12
2.1.1	Potential outcomes . . . . .	14
2.1.2	Covariates . . . . .	16
2.1.3	Assignment mechanism . . . . .	16
2.2	Potential outcome time series . . . . .	17
<b>3</b>	<b>Causal effect of an intervention on multiple non-interfering time series</b>	<b>19</b>
3.1	Background . . . . .	19
3.2	Causal estimands . . . . .	21
3.3	C-ARIMA . . . . .	22
3.3.1	Simplified framework . . . . .	22
3.3.2	General framework . . . . .	23
3.3.3	Comparison with REG-ARIMA . . . . .	24
3.3.4	Causal effect estimation . . . . .	25
3.4	Simulation study . . . . .	27
3.4.1	Design . . . . .	27
3.4.2	Results . . . . .	29
3.5	Empirical analysis . . . . .	32
3.5.1	Data & methodology . . . . .	32
3.5.2	Results . . . . .	33
3.6	Discussion . . . . .	37
<b>4</b>	<b>Causal effect of an intervention on multiple interfering time series</b>	<b>38</b>
4.1	Background . . . . .	38
4.2	Causal estimands . . . . .	39
4.3	Multivariate Bayesian Structural Time Series . . . . .	42
4.3.1	The model . . . . .	43
4.3.2	Prior elicitation . . . . .	45
4.3.3	Posterior Inference . . . . .	46
4.3.4	Prediction and estimation of causal effects . . . . .	46
4.3.5	Causal effect estimation . . . . .	47
4.3.6	Combining results . . . . .	48
4.4	Simulation study . . . . .	49
4.4.1	Design . . . . .	49

4.4.2	Results . . . . .	50
4.5	Empirical analysis . . . . .	54
4.5.1	Data & methodology . . . . .	54
4.5.2	Results . . . . .	56
4.6	Discussion . . . . .	59
<b>5</b>	<b>Causal effect of multiple interventions</b>	<b>60</b>
5.1	Background . . . . .	60
5.2	Causal estimands . . . . .	63
5.3	Multi-intervention C-ARIMA . . . . .	66
5.3.1	The model . . . . .	67
5.3.2	Causal effect estimation . . . . .	68
5.4	Empirical analysis . . . . .	70
5.4.1	Data & Methodology . . . . .	70
5.4.2	Results . . . . .	74
5.5	Discussion . . . . .	78
<b>6</b>	<b>Conclusions</b>	<b>79</b>
	<b>Appendices</b>	<b>90</b>
<b>A</b>		<b>90</b>
A.1	Additional tables . . . . .	90
A.2	Additional plots . . . . .	97
<b>B</b>		<b>125</b>
B.1	Proof of relations (10), (11) and (12) . . . . .	125
B.2	Causal effect estimation on the untransformed variable . . . . .	126
B.3	Proof of relations (23), (24) and (25) . . . . .	127
B.4	Unbiased causal effects . . . . .	130
B.5	Posterior predictive checks . . . . .	134
B.6	Sensitivity analysis . . . . .	135
B.7	Convergence diagnostics . . . . .	138
B.8	Proof of relations (53), (54) and (55) . . . . .	140

# 1 Introduction & relevant literature

One of the first lessons learned by students in any undergraduate course in Statistics is that “association is not causation” or, as a teacher with a good sense of humor might say “ice-cream sales do not cause summer droughts and the weather forecasters are definitely not responsible for the approaching hurricane”. Clearly, some relations are easier to understand than others; for example, a good statistician would not include ice-cream sales in a predictive model for droughts. However, what if we are asked to evaluate whether a promotion was successful in increasing daily sales? How can we detect and measure the effect of market news on the next two-week returns? These and other similar causal questions involving a temporal component are typically more challenging, due to added complications arising from serial dependence and seasonality.

The aim of this research is to provide an understanding of the methodologies that can be adopted to estimate causal effects of interventions (e.g., promotions or market news) in some common time series settings: i) a panel situation where multiple time series are subject to a simultaneous intervention; ii) a slightly more complex case where the time series show interactions with one another; iii) a multi-intervention setting with a single time series subject to several interventions. This research aims at providing both methodological and empirical contributions. In particular, we develop two novel approaches, C-ARIMA and CausalMBSTS, for the estimation of causal effects in the first two settings; we then extend C-ARIMA to uncover causal effects when multiple interventions take place. Furthermore, we illustrate the benefits of the proposed approaches by performing empirical analyses involving real sales data from a supermarket chain in Italy and, in our last application, Bitcoin volatility. In each of the three settings, we introduce specific causal estimands, framing them in a common theoretical background based on potential outcomes.

The potential outcomes approach to causal inference is a framework that allows to define the causal effect of a treatment (or “intervention”) as a contrast of potential outcomes, to discuss assumptions enabling to identify such causal effects from available data, as well as to develop methods for estimating causal effects under these assumptions ([Rubin, 1974, 1975, 1978](#); [Imbens and Rubin, 2015](#)). Following [Holland \(1986\)](#) we refer to this framework as the Rubin Causal Model (RCM). Under the RCM, the causal effect of a treatment is defined as a contrast of potential outcomes, only one of which will be observed while the others will be missing and become counterfactuals once the treatment is assigned.

Having its roots in the context of randomized experiments, several methods have been developed to define and estimate causal effects under the RCM for network data structures ([VanderWeele, 2010](#); [Forastiere et al., 2020](#); [Noirjean et al., 2020](#)), time series ([Robins, 1986](#); [Robins et al., 1999](#); [Bojinov and Shephard, 2019](#)) and panel data ([Rambachan and Shephard, 2019](#); [Bojinov et al., 2020](#)). Unlike randomized experiments, however, in an observational study the researcher has no

knowledge of, or no control on, the assignment mechanism, i.e., the process that determines the units receiving treatment and those under control. Thus, observational studies pose additional challenges to the definition and the estimation of causal effects, especially in a time series setting.

A different approach extensively used in the econometric literature to assess the impact of shocks occurring on a time series is *intervention analysis* (Box and Tiao, 1975, 1976). The effect is generally estimated by comparing observed data post-intervention with forecasts based on pre-intervention data; if the forecast deviates from the observed time series, an intervention component is included in the pre-intervention model, which is then re-estimated on the full series. In this way, however, rather than a properly defined “causal” effect, the estimated coefficient of the intervention variable describes an association between the outcome and the intervention component. Indeed, the definition of the effect and its estimation typically overlap, making this method prone to errors. Despite this drawback, intervention analysis is still among the most used methodologies to assess the effect of an intervention on a time series.

Closing the gap between causal inference under the RCM and intervention analysis, in Section 3 we propose the “Causal-ARIMA” (C-ARIMA) approach to estimate the effect of interventions in observational time series settings under the RCM. Laying its foundation in the potential outcomes framework, the proposed approach can successfully be used to estimate properly defined causal effects, whilst making use of ARIMA-type models that are so familiar to the intervention analysis literature. Our work is motivated by an analysis of the effectiveness of a new price policy introduced by an Italian supermarket chain. In particular, the main goal is to assess the effect of a permanent price reduction on a selected subset of store brands. Furthermore, since the supermarket chain sells competitor brands with the same characteristics as their store brand equivalent, to assess the overall effectiveness of the new policy we also investigate its impact on those products.

We need however point out that the C-ARIMA approach is suitable to assess the causal effect of an intervention only on a single time series at a time; hence, it may suffer from some limitations when the time series interact with one another. For example, in our application store and competitor brands are interconnected; thus, an intervention applied to a store brand is likely to produce an effect on its direct competitor. In the causal inference literature this situation is known as “interference”.

Therefore, in Section 4 we present “CausalMBSTS”, a novel approach to estimate the effects of interventions in panel settings with interference. Extending the univariate version of the popular method introduced by Brodersen et al. (2015), the proposed approach employ Multivariate Bayesian Structural Time Series (MBSTS) models to build a counterfactual from pre-intervention data. In the presence of cross-unit interactions, the proposed approach allows us to model the interference between the units by explicitly modeling their dependence struc-

ture. Furthermore, MBSTS models are flexible and can successfully estimate the effect that an intervention produces on each unit (the so-called “heterogeneous effect”).

Finally, in the last part of the dissertation we consider the situation where multiple interventions take place on the same time series. To estimate the heterogeneous effect of each intervention in such settings, in Section 5 we present an extension of the C-ARIMA approach. The motivating example is the estimation of the impact produced by the introduction of the first two regulated Bitcoin futures on Bitcoin volatility.

Our work bridges together several fields of research including Causal Inference, Econometrics and Finance. For each of them we now review the contributions that are closely related to our own.

The RCM was introduced by Rubin in a series of papers (Rubin, 1974, 1975, 1978, 2005; Imbens and Rubin, 2015). This approach defines the effect of an intervention by comparing the potential outcomes arising under different treatment allocations. Among these outcomes, only one is observed whereas all the others are “missing”, meaning that, once the treatment is assigned, they become counterfactual outcomes. For example, in a study investigating the effectiveness of a new statin, the cholesterol level of a patient assigned to treatment would be the observed outcome, whereas the level that the same patient would have had under a placebo drug is the counterfactual outcome. In particular, for a binary treatment (as in the case of the new statin and the placebo drug) there is only one counterfactual outcome.

This approach has recently been extended to a time series setting where an intervention is randomly allocated at any point in time (Bojinov and Shephard, 2019; Bojinov et al., 2020; Rambachan and Shephard, 2019). However, unlike randomized experiments where the randomization contributes to eliminate the differences between treated and control units, in an observational study the researcher has no knowledge of, or no control on, the assignment mechanism, i.e., the process that determines the units receiving treatment and those under control. Thus, such designs pose additional challenges to the definition and the estimation of causal effects. Furthermore, in a panel setting (multiple units observed over time) the temporal component brings added complications due to serial dependence and seasonality.

A method that has been extensively used to evaluate the effect of interventions in the absence of experimental data is the *Difference-in-Difference* (DiD) estimator (see e.g., Card and Krueger (1993); Meyer et al. (1995); Garvey and Hanka (1999); Angrist and Pischke (2008); Anger et al. (2011)). In its simplest formulation, this method requires to observe a treated and a control group at a single point in time before and after the intervention; the effect is then estimated by contrasting the change in the average outcome for the treated group with that of the control group. However, by focusing on a few time points, DiD is not able to also exploit information on the temporal dynamics. Furthermore, it relies on the often troubled assumption that, in the absence of treatment, the outcomes of the treated and control units would have followed



parallel paths (Abadie, 2005; Ryan et al., 2015; O’Neill et al., 2016).

Another popular method to infer the causal effect of an intervention from panel data under the RCM is constructing a *synthetic control* from a set of time series that are not directly impacted by the treatment and have pre-treatment variables matching those of the treated unit (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). For example, in a study investigating the impact of a new legislation to reduce pollution levels, a suitable set of control series could be the evolution of carbon emissions in neighboring states that did not activate the new regulation but with the same pre-intervention characteristics as the treated state (e.g., population density, number of industries). Unlike DiD, synthetic control methods measure the effect at each point in time after the intervention and rely on the less stringent assumption that the expected outcomes of the treated and control groups are the same in the absence of treatment, conditionally on past outcomes and covariates. Therefore, since their introduction, synthetic control methods have been successfully applied in a wide range of research areas, including healthcare (Kreif et al., 2016; Papadogeorgou et al., 2018; Viviano and Bradic, 2019), economics (Billmeier and Nannicini, 2013; Abadie et al., 2015; Dube and Zipperer, 2015; Gobillon and Magnac, 2016; Ben-Michael et al., 2018), marketing and online advertising (Brodersen et al., 2015; Li, 2019). Nevertheless, both the DiD estimator and the synthetic control method have one main drawback: in the impossibility to observe at least one suitable control unit none of them can be applied.

A recent approach overcoming this limitation is proposed by Brodersen et al. (2015). Their methodology share several features with DiD and synthetic control methods but, instead of using control units or external characteristics, it only requires to learn the dynamics of the treated unit prior to the intervention. In other words, it builds a synthetic control by forecasting the counterfactual series in the absence of intervention based on a model estimated on the pre-intervention data. In particular, the authors employ Bayesian Structural Time Series models (West and Harrison, 2006; Harvey, 1989) since they allow to add the components (e.g., trend, seasonality, cycle) that better describe the characteristics of the time series, whilst incorporating prior knowledge in the estimation process. Borrowing the name from the associated R package, from now on we refer to their method as “CausalImpact”.

Our work is closely related to synthetic control methods and to the methodology proposed by Brodersen et al. (2015). In the same vein as CausalImpact, we propose C-ARIMA as a novel approach to build a synthetic control for a time series subject to an intervention by learning its time dynamics in the pre-intervention period and then forecasting the series in the absence of intervention. Unlike CausalImpact, however, our methodology is based on ARIMA models and thus can be used as an alternative by researchers and practitioners in a wide range of fields that are not familiar to (or are not willing to adopt) Bayesian inference.

Synthetic control methods often assume that the statistical units do not interact with one

another; the opposite situation, called “interference”, occurs when the treatment assigned to a unit affects the potential outcomes of other units. Even though the absence of interference assumption is a fundamental one under the RCM (Cox, 1958; Rubin, 1980), in many empirical applications it is violated (e.g., Hudgens and Halloran (2008), Tchetgen and VanderWeele (2012), Basse et al. (2019)). For example, in a study on the effectiveness of a vaccine, the treatment clearly affects the probability of the untreated population of getting the infection. However, if the patients receiving the vaccine and those taking the placebo drug live in two remote cities, the researcher can still compare the incidence of the disease between them. Indeed, whether the patients are treated or not affects the other citizens but has no impact across the two cities. This assumption, known as “partial interference” (Sobel, 2006), has been extensively studied in the cross-sectional literature (Rosenbaum, 2007; Hudgens and Halloran, 2008; Forastiere et al., 2020). Conversely, in a panel setting with multiple interfering units observed over time it has received relatively less attention (Cao and Dowd, 2019; Grossi et al., 2020).

We propose to address the interference issue by deriving the multivariate version of CausalImpact, which we denote as “CausalMBSTS”. Indeed, if we are able to group the units so that interference occurs within the group but not across them, we can rely on the partial interference assumption and estimate a multivariate model on each group. This allows to take into account the interference between units in the same group by explicitly modeling their dependence structure. Like CausalImpact and C-ARIMA, our proposed approach can be successfully employed to infer the heterogeneous causal effect on each unit in the group by constructing a synthetic control based on the dynamics learned in the pre-intervention period.

A different approach to uncover causal effects in time series settings is “Granger-Sims causality” (Granger, 1969; Sims, 1972). Formally, a variable  $X$  Granger-causes another variable  $Y$  if the prediction of  $Y$  based on its past values and past values of  $X$  is better than the prediction based on past values of  $Y$  alone. Similarly, Sims advocates that if causality runs from  $X$  to  $Y$  only, future values of  $X$  should have zero coefficients when  $Y$  is regressed on past and future values of  $X$ . Granger causality implies Sims causality while the other way around is generally not true (Sims, 1972; Chamberlain, 1982); however, starting from general definitions stated in terms of conditional distributions rather than linear predictors, Chamberlain (1982) shows that a stronger Sims causality conditioning also on past values of  $Y$  is equivalent to Granger causality. Thus, they are commonly denoted with the single name “Granger-Sims causality”. From an empirical perspective,  $X$  and  $Y$  are often continuous economic variables; hence, assessing the presence of Granger-Sims causality amounts to testing whether the coefficients of future values of  $X$  in the regression of  $Y$  (and those of future values of  $Y$  in the regression of  $X$ ) are statistically different from zero.

It should be clear by now that there is a fundamental difference between the RCM and Granger-

Sims approach: Granger-Sims view causality in terms of the predictive ability of a variable  $X$  toward  $Y$  and they also agree that a cause must precede any effect of it; conversely, under the RCM,  $X$  takes the form of a treatment (often a binary variable) and the effect measures what happens on  $Y$  when we switch the treatment from  $X$  to  $X'$ .

Nonetheless, some connections exist between the two approaches. Indeed, a fundamental assumption allowing the identification of causal effects in time series settings under the RCM is that the treatment is non-anticipating, i.e., the present assignment to treatment is not influenced by future outcomes, conditioning on past treatments. In Granger-Sims terminology, this amounts to saying that future outcomes do not Granger-cause the present treatment (Bojinov and Shephard, 2019; Rambachan and Shephard, 2019). Furthermore, Rambachan and Shephard (2019) establish a connection between their potential outcome time series framework and several common methods in econometrics, such as local projections and vector autoregressions. More precisely, they clarify the assumptions that are needed for those methods to have a causal content in the potential outcome framework.

Nonetheless, without the additional assumption that the treatment is non-anticipating, Granger-Sims causality does not imply causality under the RCM (Chamberlain, 1982). Moreover, in his overview of causation and causal inference, Holland (1986) points out that the result arising from an analysis based on Granger-Sims causality is temporary, since when new information is gathered and introduced in the predictive model, what was a causal effect might become a “spurious” association. Instead, the RCM is not subject to this drawback, since the identification and estimation steps are separated.

Another approach that has been extensively used in the econometric literature is *intervention analysis*, introduced by Box and Tiao (1975, 1976) to estimate the effect of an intervention on a time series. Since then, it has successfully been applied to uncover the effect of interventions in many fields, including economics (Larcker et al., 1980; Balke and Fomby, 1994), social science (Bhattacharyya and Layton, 1979; Murry et al., 1993) and terrorism (Cauley and Im, 1988; Enders and Sandler, 1993). Intuitively, the effect is estimated by comparing observed data post-intervention with forecasts based on pre-intervention data; if the forecast deviates from the observed time series, the pre-intervention model is modified to include an intervention component, whose structure is modeled based on the pattern followed by the resulting deviation (e.g., level shift, slope change and similar). Under this approach, the coefficient of the intervention component gives the size of the “effect”.

However, rather than a properly defined “causal” effect, such coefficient describes the association (sometimes spurious) between the response variable and the intervention component. For example, while investigating the impact of a new legislation to reduce carbon emissions, a positive association between the intervention component and the incidence of lung cancer does not clearly indicate that the new law caused lung cancer; conversely, this result suggests that

some confounders, such as record levels of air pollution, have not been included in the analysis. Hence, asserting that the uncovered effect is due to a specific intervention can only be done after an attentive causal analysis, which takes into account all possible confounders and lay out the assumptions underneath the identification and the attribution of causal effects.

A variant of this approach consists in testing the estimated deviation between the observed time series post-intervention with the forecasted series based on pre-intervention data (Box and Tiao, 1976). Even in this case, without a thorough discussion of the assumptions, the uncovered effect can not be attributed to the intervention.

Our work is closely related to intervention analysis, especially to the variant proposed in Box and Tiao (1976). Nonetheless, C-ARIMA overcomes the theoretical underpinnings of intervention analysis regarding the identification of causal effects, since it is based on a causal framework developed under the RCM. In other words, after an empirical analysis with C-ARIMA we are able to state that the uncovered effect is due to the intervention under the assumptions that we have explicitly stated.

In our last empirical analysis we extend C-ARIMA to a time series setting where multiple interventions occur. In particular, we want to determine whether the introduction of Bitcoin futures has affected the volatility of Bitcoin prices. The first two regulated Bitcoin futures were launched by the Chicago Board of Exchange (CBOE) and the Chicago Mercantile Exchange (CME) on, respectively, December 10 and December 18, 2017. Before that time, Bitcoin derivatives were only traded over-the-counter on unregulated exchanges; since their introduction in 2017, there has been a growing number of studies seeking to investigate the impact of regulated futures on Bitcoin volatility. Focusing on the first contract introduced by CBOE, Shi (2017) finds a significant decrease in the spot volatility, whereas later studies including CME future find that Bitcoin volatility increased right after the launch of the new contract (Kim et al., 2020) as well as around its announcement date (Corbet et al., 2018). Interestingly, the peak reached by Bitcoin price on December 16, 2017 matched the launch of the CME future and prices have experienced a sharp decline since then. This suggests that the newly introduced derivative instruments allowed the entrance of “pessimistic” traders willing to bet for a price drop but unable to do so until the creation of a derivative market (Hale et al., 2018); the downward pressure generated by the new traders might contribute to explain the short term increase in Bitcoin price volatility.

Albeit there are relatively few studies on the effect of Bitcoin futures, the impact of derivatives trading on the underlying spot volatility has been thoroughly investigated for instruments such as stocks and financial indexes. In particular, there are two conflicting theories about the effect of futures on the underlying spot markets and empirical evidence is mixed: some studies warn that speculative behaviors and information asymmetries brought by futures markets may increase volatility and jeopardize the value formation process in the spot market (Stein, 1987;

Figlewski, 1981; Antoniou and Holmes, 1995; Harris, 1989); other studies argue that futures trading enhances information flows and price discovery, thereby reducing volatility and stabilizing the underlying spot market prices (Danthine, 1978; Moriarty and Tosini, 1985; Edwards, 1988; Bessembinder and Seguin, 1992; Antoniou et al., 1998; Jochum and Kodres, 1998). Recent works on individual stock futures (McKenzie et al., 2001) and emerging economies (Baklaci and Tutek, 2006; Chen et al., 2013; Bohl et al., 2015) support the idea that futures trading acts as a stabilizing force. Among the aforementioned studies, only Kim et al. (2020) employ a causal approach — Difference-in-Difference (DiD)— in the attempt to attribute the uncovered effect to the introduction of futures. Nonetheless, as detailed before, DiD is impractical in all situations where, as in our case, there is no reliable control group available or the parallel trend assumption is troubled. We instead generalize the C-ARIMA approach to a multi-intervention setting: unlike intervention analysis, this method enables the computation of properly defined causal effects and, in contrast to DiD, it allows to estimate the effects when suitable controls are unavailable.

The remainder of the dissertation is organized as follows: Section 2 presents a common theoretical background; Section 3 illustrates the proposed C-ARIMA approach to estimate the causal effect of an intervention on multiple non-interfering time series; Section 4 describes the CausalMBSTS approach to infer the effect of an intervention on multiple interfering series; Section 5 extends C-ARIMA to a multi-intervention setting; Section 6 concludes.

## 2 Common theoretical background

### 2.1 Assumptions

For a generic statistical unit, let  $W_t \in \mathcal{W}$  denote the treatment assignment at time  $t \in \{1, \dots, T\}$ . Although multiple treatments are possible, in this research we focus on binary treatments, so that  $W_t \in \{0, 1\}$  where  $W_t = 1$  indicates that a treatment (or “intervention”) has taken place and  $W_t = 0$  denotes control, i.e., absence of treatment or an alternative form of treatment. Then,  $W_{1:T} = (W_1, \dots, W_T)$  is the sequence of treatments received by the statistical unit over time.

In a study where  $N$  statistical units are present, let  $W_{i,t} \in \{0, 1\}$  be the treatment status of unit  $i$  at time  $t$ , with  $i \in \{1, \dots, N\}$ . Then,  $W_{1:N,1:T} = (W_{1,1:T}, \dots, W_{N,1:T})$  is the treatment assigned to all units over time, sometimes denoted as “treatment panel” (Bojinov et al., 2020). A realization of  $W_{i,t}$  is denoted with the lower case letter  $w_{i,t}$ .

Now assume that the  $N$  statistical units can be divided in  $J$  equally sized groups with different characteristics and let  $d$  denote the group size.<sup>1</sup> Thus,  $W_{j,t}^{(d)} \in \{0, 1\}$  with  $j \in \{1, \dots, J\}$  is the

---

<sup>1</sup>By denoting with  $d_j$  the size of the  $j$ -th group, our causal framework can easily accommodate groups of

treatment assignment of the  $d$ -th unit inside group  $j$  at time  $t$ ;  $\mathbf{W}_{j,t} = (W_{j,t}^{(1)}, \dots, W_{j,t}^{(d)}) \in \{0, 1\}^d$  indicates the treatment allocation inside the group  $j$  at time  $t$  and, finally, the treatment panel indicating the assignments of all groups throughout the study is  $\mathbf{W}_{1:J,1:T} = (\mathbf{W}_{1,1:T}, \dots, \mathbf{W}_{J,1:T})$ . For example, in a simple study design where the units are grouped in pairs, the treatment allocation inside a generic group  $j$  at time  $t$  is  $\mathbf{W}_{j,t} = (W_{j,t}^{(1)}, W_{j,t}^{(2)})$ . Realization of random variables are indicated with the lower case letter, i.e.,  $\mathbf{w}_{j,t}$  is a realization of  $\mathbf{W}_{j,t}$ .

Notice that the above notation fully describes the three empirical applications included in the dissertation. In Section 3 we deal with multiple units (i.e., the products sold by the supermarket chain) therefore we use  $W_{i,t}$  to indicate their treatment assignment. In Section 4 we handle the interference issue by dividing the units in pairs, each assigned to one of four possible treatments  $\mathbf{W}_{j,t}$ : no permanent price reduction  $\mathbf{W}_{j,t} = (0; 0)$ , both receive a permanent price reduction  $\mathbf{W}_{j,t} = (1; 1)$ , store brand receive a permanent price reduction only  $\mathbf{W}_{j,t} = (1; 0)$ , or competitor brand receive a permanent reduction only  $\mathbf{W}_{j,t} = (0; 1)$ . Finally, in Section 5 we have a single statistical unit (Bitcoin cryptocurrency) and thus we can use the simplified notation  $W_t$  for the treatment assignment.<sup>2</sup>

In principle, the treatment can be administered at any point in time, as in the case of a randomized experiment where the execution of market orders is repeatedly assigned to one of two alternative methods (Bojinov and Shephard, 2019). However, it is not uncommon to observe a single persistent intervention, as in the case of a new law introduced by the government (Papadogeorgou et al., 2018) or an online advertising campaign run for several weeks in a row (Brodersen et al., 2015).

**Assumption 1 (Single persistent intervention)** *We say group  $j$  received a single intervention, if there exists a  $t_j^* \in \{1, \dots, T\}$  such that for all  $t \leq t_j^*$  we have  $\mathbf{W}_{j,t} = (0, \dots, 0)$  and for all  $t, t' > t_j^*$  we have  $\mathbf{W}_{j,t} = \mathbf{W}_{j,t'}$ . If all groups receive a single intervention, then we say the study is single intervention panel study. If the intervention happens simultaneously on all groups, that is,  $t_j^* = t_{j'}^* = t^*$  we say the study is a simultaneous intervention panel study.*

In general, the groups may receive the intervention at different times, that is,  $t_j^* \neq t_{j'}$  for all  $j, j' \in \{1, \dots, J\}$ , a situation commonly referred to as “staggered adoption” (Athey and Imbens, 2018; Ben-Michael et al., 2019). Instead, in our empirical applications the treatments are assigned simultaneously and thus Assumption 1 means that there is a single treatment administered at time  $t^*$  producing its effects on all groups starting from time  $t^* + 1$ .

Notice that Assumption 1 can be stated even in a setting where the  $N$  units form a single group; in this case the notation simplifies to  $W_{i,t} = W_{i,t'}$  for all  $i \in \{1, \dots, N\}$  and, in the special situation  $N = 1$  we write  $W_t = W_{t'}$ .

---

different sizes.

<sup>2</sup>Notice that the subscripts of the treatment indicator  $W_{i,t}$  denote the unit of analysis and the time of the treatment. Thus, when the units are grouped according to their characteristics, the notation  $\mathbf{W}_{j,t}$  makes clear that the unit of analysis is the entire group.



The same persistent intervention may occur several times during the analysis period. For example, supermarket managers often schedule temporary promotions on selected goods at regular intervals during the year; every time this happens, the price is lowered for some weeks and then it bounces back at the previous level. Furthermore, a persistent intervention may occur even while the previous one is still in place. In financial markets, for instance, future contracts have pre-specified expiration dates; thus, it may happen that a future is issued before a previous one expires on the same underlying asset.

**Assumption 2 ( $M$  persistent interventions)** *Indicating with  $\Lambda = \{t_1, \dots, t_M\}$  the subset of time points at which the interventions take place, we say group  $j$  received  $M$  persistent interventions, if for all  $t < t_1$  we have  $\mathbf{W}_{j,t} = (0, \dots, 0)$  and for all  $t, t' \in \{t_m, \dots, t_{m+1} - 1\}$  we have  $\mathbf{W}_{j,t} = \mathbf{W}_{j,t'}$ . If the set  $\Lambda$  is the same for all groups, we say this study is a simultaneous intervention panel study.*

Under Assumption 2, we allow the  $m$ -th intervention assigned at time  $t_m$  to produce a contemporaneous effect at the same time  $t_m$ . This is a convenient choice, since in the last empirical analysis we deal with interventions that are able to produce instantaneous effects.<sup>3</sup> Again, Assumption 2 can be stated in a setting where the  $N$  units form a single group, in which case the notation simplifies to  $W_{i,t} = W_{i,t'}$  for all  $i \in \{1, \dots, N\}$ , while in the limiting case  $N = 1$  we have  $W_t = W_{t'}$ .

Those described by Assumptions 1 and 2 are special situations where a single intervention or few persistent interventions occur. In the remainder of this section we present additional assumptions on the potential outcomes, covariates and the assignment mechanism that should hold irrespective of the number of treatments. Thus, they are given for a general panel setting where the intervention may occur at any point in time. If needed, the limiting case with a single unit can be recovered easily; then, in Section 2.2 we discuss the special cases where also Assumptions 1 and 2 hold.

### 2.1.1 Potential outcomes

Whether a unit is assigned to treatment or control may impact its outcome. For example, if the units are products subject to a promotion, a product will likely sell more under a 50% price discount compared to a situation where it is not discounted. Under the RCM the sales under these two alternative scenarios are known as “potential outcomes”.

---

<sup>3</sup>If needed, a notation in line with Assumption 1 where the intervention produces an effect starting from time  $t^* + 1$  can be easily recovered by writing  $\mathbf{W}_{j,t} = \mathbf{W}_{j,t'}$  for all  $t, t' \in \{t_m + 1, \dots, t_{m+1}\}$ . We advocate that the choice between the two notations should be based on the empirical application, namely, it depends on whether the researcher wants to allow or exclude a contemporaneous effect.

Typically, the potential outcomes of a unit  $i$  at time  $t$  depend the full treatment panel, i.e.  $Y_{i,t}(\mathbf{w}_{1:N,1:T})$  (Bojinov et al., 2020). However, in many applications we are able to restrict this dependence structure by focusing on non-anticipating potential outcomes.

**Assumption 3 (Non-anticipating potential outcomes)** For all  $t \in \{1, \dots, T\}$  and all  $i \in \{1, \dots, N\}$ , for any two alternative treatment paths  $\mathbf{w}_{1:N,t+1:T}, \mathbf{w}'_{1:N,t+1:T}$  the outcome of unit  $i$  at time  $t$  is independent of future treatments

$$Y_{i,t}(\mathbf{w}_{1:N,1:t}; \mathbf{w}_{1:N,t+1:T}) = Y_{i,t}(\mathbf{w}_{1:N,1:t}; \mathbf{w}'_{1:N,t+1:T}).$$

In words, present outcomes can be function of present and past treatments but they are not impacted by future treatment assignments.

Under Assumption 3 we can write  $Y_{i,t}(\mathbf{w}_{1:N,1:t})$  but we can further restrict the set of potential outcomes by ruling out any form of interference between the statistical units.

**Assumption 4 (Temporal no-interference)** For all  $t \in \{1, \dots, T\}$  and all  $i \in \{1, \dots, N\}$ , we assume that for any  $\mathbf{w}_{1:N,t}, \mathbf{w}'_{1:N,t}$  such that  $w_{i,t} = w'_{i,t}$ ,

$$Y_{i,t}(\mathbf{w}_{1:N,t}) = Y_{i,t}(\mathbf{w}'_{1:N,t}).$$

It means that the treatments assigned to the other units do not affect unit  $i$ 's potential outcomes at the same time. This assumption, which is also known as Temporal Stable Unit Treatment Value Assumption or TSUTVA (Bojinov and Shephard, 2019; Bojinov et al., 2020), is the time series equivalent of the cross-sectional SUTVA (Rubin, 1974) and contributes to reduce the number of potential outcomes. As a result, if both Assumption 3 and 4 hold we can write  $Y_{i,t}(w_{i,1:t})$ , indicating that the potential outcomes of the  $i$ -th unit at time  $t$  only depend on its treatment path up to time  $t$ .

Nevertheless, there are many applications in which the statistical units interfere with one another, meaning that the treatment has an impact also on the units assigned to the control group (Hudgens and Halloran, 2008; Tchetgen and VanderWeele, 2012; Basse et al., 2019). In a panel setting where multiple units are observed over time, this is the case, for example, of a new light rail line that impacts the sales of both the shops facing the construction sites and those located in the neighboring streets (Grossi et al., 2020); in the same vein, a promotion on selected goods is likely to influence the sales of their direct substitutes. In such situations, Assumption 4 is clearly violated, but we might be able to group the statistical units so that the groups do not interfere with one another.

**Assumption 5 (Partial temporal no-interference)** For all  $t \in \{1, \dots, T\}$  and all  $j \in \{1, \dots, J\}$ , we assume that for any  $\mathbf{w}_{1:J,t}, \mathbf{w}'_{1:J,t}$  such that  $\mathbf{w}_{j,t} = \mathbf{w}'_{j,t}$ ,

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



In words, we are ruling out any form of interference between the groups whilst allowing cross-unit interactions within them. In the above example, we may re-define our statistical unit to be the group formed by the discounted good and its direct substitute.

Partial interference (Sobel, 2006) has been extensively studied within the cross-sectional literature (e.g., Rosenbaum (2007), Hudgens and Halloran (2008), Forastiere et al. (2020)) and Assumption 5 constitutes its extension to the panel setting.

### 2.1.2 Covariates

In many applications, the potential outcomes of a unit are likely influenced by many variables. Including covariates in the estimation process can improve the accuracy of the estimated causal effect or, conversely, can produce biased estimates if the covariates are influenced by the treatment. Therefore, we should select a set of covariates for which the following assumption is plausible.

**Assumption 6 (Covariates-treatment independence)** *Let  $X_{i,t}$  be a vector of covariates that are predictive of the outcome of unit  $i$ ; for all  $t \in \{1, \dots, T\}$ , all  $i \in \{1, \dots, N\}$  and for any two alternative treatment paths  $w_{i,1:t}, w'_{i,1:t}$  such covariates are not affected by the interventions*

$$X_{i,t}(w_{i,1:t}) = X_{i,t}(w'_{i,1:t}).$$

As a result, we can use the known covariates values post-treatment to improve the prediction of the outcome in the absence of intervention.

### 2.1.3 Assignment mechanism

A final assumption in our causal framework regards the assignment mechanism. To understand the importance of such an assumption, consider as an example a research on a new drug where doctors are asked to select the participants among their patients. If doctors assign to treatment only those patients they believe have better chance to complete the treatment successfully and without side effects, contrasting treated and controls would provide biased evidence of the true effect of the treatment. Thus, we need to ensure that the treatment assigned to a unit only depends on its past outcomes and covariates; in addition, it should be independent of the treatment assigned to the other units.

**Assumption 7 (Non-anticipating treatment)** *The assignment mechanism at time  $t + 1$  for the  $i$ -th unit depends solely on past outcomes and past covariates.*

$$\Pr(W_{1:N,t+1} = w_{1:N,t+1} \mid W_{1:N,1:t}, Y_{1:N,1:T}(w_{1:N,1:T}), X_{1:N,1:T}) = \prod_{i=1}^N \Pr(W_{i,t+1} = w_{i,t+1} \mid Y_{i,1:t}(w_{i,1:t}), X_{i,1:t}).$$

A non-anticipating treatment in a time series setting is the analogous of the unconfounded assignment mechanism in the cross-sectional setting (Imbens and Rubin, 2015). Whilst a classical randomized experiment is unconfounded by design, we focus on observational studies where we have no control on the assignment mechanism. Thus, the non-anticipating treatment assumption is essential to ensure that any difference in the potential outcomes is due to the treatment.

## 2.2 Potential outcome time series

Above assumptions, although playing a partially different role, are crucial to estimate properly defined causal effects: Assumptions 4 and 5 restrict the set of potential outcomes; Assumption 6 allows to include covariates to improve the estimation accuracy; finally, Assumptions 3 and 7 ensure that the effect is identifiable, meaning that the uncovered effect can be attributed to the treatment. Even so, notice that the number of potential outcomes may still be very large and, among them, only one is actually observed whereas the others are commonly referred as “missing”.

Let  $w_{i,1:t}^{obs}$  and  $w_{i,1:t}^{mis}$  denote, respectively, the observed treatment path and the missing treatment path of unit  $i$  up to time  $t$ . Then, under the above assumptions, the observed and missing potential outcome time series can be denoted as  $Y_{i,1:t}(w_{i,1:t}^{obs})$  and  $Y_{i,1:t}(w_{i,1:t}^{mis})$ . Figure 1 provides an illustration of the observed and missing outcome time series in a simple multi-intervention setting with two treatments. Instead, in a special setting where also Assumption 1 holds, there are only two possible treatment paths for each unit  $i$ ,

$$w_{i,1:T} = \left( \underbrace{0, \dots, 0}_{t \in \{1, \dots, t^*\}}, \underbrace{1, \dots, 1}_{t \in \{t^*+1, \dots, T\}} \right); \quad w'_{i,1:T} = \left( \underbrace{0, \dots, 0}_{t \in \{1, \dots, t^*\}}, \underbrace{0, \dots, 0}_{t \in \{t^*+1, \dots, T\}} \right)$$

where  $w_{i,1:T}$  indicates that the unit receives the persistent treatment starting from  $t^* + 1$ , whereas under the alternative path  $w'_{i,1:T}$  the unit gets the persistent control. Thus, focusing on the time periods following the intervention, the observed potential outcome time series is  $Y_{i,t^*+1:T}(w_{i,t^*+1:T})$ , whereas  $Y_{i,t^*+1:T}(w'_{i,t^*+1:T})$  is denoted as the *missing* or *counterfactual* potential outcome time series.

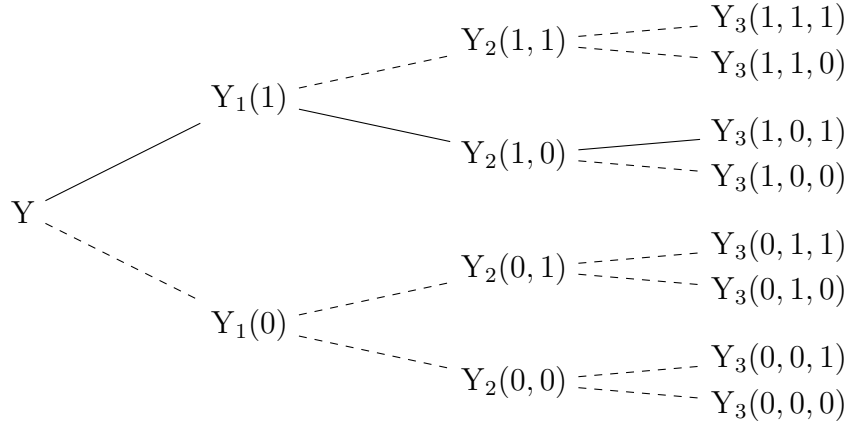
Under Assumption 2 the same applies in any time interval between two subsequent interventions. Indeed, if Assumption 7 holds, we can condition on past treatments, so that in the time interval between two subsequent interventions we only have two possible paths,

$$w_{i,1:t} = \left( \underbrace{0, \dots, 0}_{t \in \{1, \dots, t_m-1\}}, \underbrace{1, \dots, 1}_{t \in \{t_m, \dots, t_{m+1}-1\}} \right); \quad w'_{i,1:t} = \left( \underbrace{0, \dots, 0}_{t \in \{1, \dots, t_m-1\}}, \underbrace{0, \dots, 0}_{t \in \{t_m, \dots, t_{m+1}-1\}} \right)$$

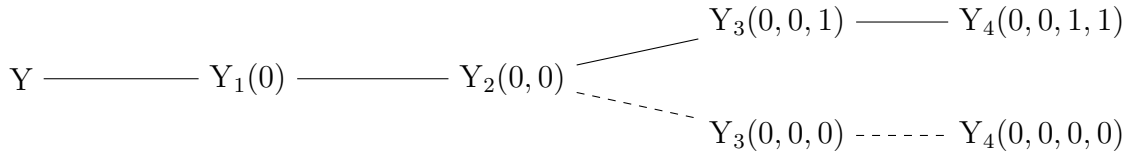
where  $w_{i,1:t}$  denotes that, holding fixed past treatment paths, before  $t_m$  the unit does not

experience the  $m$ -th persistent intervention, whereas  $w'_{i,1:t}$  indicates that the unit never receives the  $m$ -th intervention. Figure 2 illustrates the potential outcome series in a simple case with a single persistent intervention and Figure 3 depicts a situation with  $M$  persistent interventions.

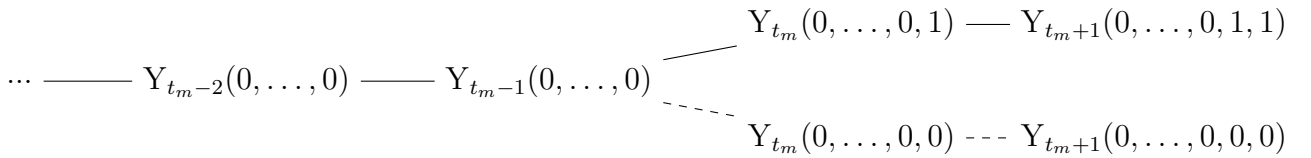
**Figure 1:** Potential outcome time series of a generic statistical unit when  $T = 3$ . The chart shows two treatments producing their effects at times  $t = 1$  and  $t = 3$ . The solid line represents the observed series while the dashed line depicts the missing potential outcome series.



**Figure 2:** Potential outcome time series of a generic statistical unit in case of a persistent intervention producing its effects starting from time  $t = 3$ . The solid line represents the observed series while the dashed line depicts the missing potential outcome series.



**Figure 3:** Potential outcome time series of a generic statistical unit for the  $m$ -th persistent interventions, conditioning on previous treatment paths and potential outcomes,  $Y_{1:t_m-1}(w_{1:t_m-1}^{obs})$ .



### 3 Causal effect of an intervention on multiple non-interfering time series

In this section we describe a novel approach, C-ARIMA, to estimate the causal effect of an intervention in a panel setting under the assumption of temporal no-interference. This approach is motivated by the analysis of a new price policy introduced by an Italian supermarket chain on a selected subset of store brands. Without loss of generality, this method can also be used in simple settings where only one time series is present, thereby constituting a valid alternative to CausalImpact. Indeed, building over the same models used by standard intervention analysis, C-ARIMA has the advantage to be accessible to a wide range of practitioners and researchers in many fields, whilst allowing the estimation of causal effects under the RCM.

The remainder of this section is organized as follows: Section 3.1 introduces the background of the empirical analysis and discusses the assumptions needed to define and estimate causal effects; Sections 3.2 and 3.3 illustrate the causal estimands and the proposed C-ARIMA approach; Section 3.4 presents a simulation study on the ability of C-ARIMA to uncover causal effects as compared to standard intervention analysis; finally, Section 3.5 presents the results of the empirical analysis.

#### 3.1 Background

On October 4, 2018 the Florence branch of an Italian supermarket chain introduced a new price policy that permanently lowered the price of 707 store brands. In particular, we focus on the goods belonging to the “cookies” category and the main goal is to estimate the causal effect of the price reduction on the sales of store brands cookies. Furthermore, the supermarket chain also sells competitor brand cookies with the same characteristics (e.g., ingredients, flavor, shape) as their store brand equivalent; these products might be influenced by the price policy as well, since consumers may perceive non-discounted goods as more expensive and modify their consumption habits accordingly. Therefore, to assess the overall impact of the price policy, we perform separate analyses on two subgroups of products: the store and the competitor brand cookies.

Among the Assumptions presented in Section 2.1 we now recall those that are needed to estimate causal effects and we discuss whether they are reliable in the context of our empirical application.

**Assumption 1** The single persistent intervention is the permanent price reduction introduced on October 4, 2018 and the statistical units included in the analysis are the store brand cookies and their direct substitutes, i.e., competitor brand cookies with the same characteristics (e.g., ingredients, flavor, shape) as their store brand equivalent.

**Assumption 3** Since the permanent price discount is a persistent intervention, the future treatment path is known and has no bearing on present outcomes (this intervention was advertised as a permanent price reduction, so we can exclude the possibility that the customers perceived it as a transitory change). Conversely, before the new policy became effective, 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, but we can safely exclude this, since in our empirical setting the supermarket chain did not advertise the price reduction in advance.

**Assumption 4** In our empirical setting, the store-brand cookies selected for the permanent price reduction differ on many characteristics, thus we can safely assume that they appeal to different customers. Therefore, the temporal no-interference assumption is plausible within the subgroup of store brands and the same applies within the subgroup of competitor brands. Then, to handle the between-group interference we perform two separate analyses under two different definitions of treatment: the permanent price reduction for the store brands; the relative price increase for competitor brands resulting inevitably from having discounted the corresponding store brand. In this way, we account for any possible interference arising from store to competitor brands when the former get a price reduction, since we consider competitor cookies to be under active treatment as well.

**Assumption 6** The sales of supermarket goods are likely influenced by many variables. More specifically, our set of covariates includes a holiday dummy, some day-of-the-week dummies and the price per unit. While it is quite obvious that all the dummies are unaffected by the intervention, things get trickier for price. In the analysis on competitor brands we use their absolute price, since it is not directly affected by the intervention that is defined on the relative price; conversely, in the analysis on store brands we consider a “modified” price that, starting from the intervention date, assumes a constant value equal to the price of the day before the intervention, which is the most likely price that the item would have had in the absence of treatment.<sup>4</sup>

**Assumption 7** We assume that the decision of lowering the price of a product is informed only by its past sales performances and past covariates or, at most, by general beliefs on the sales evolution under active treatment. Instead, the assumption would be violated, for example, if prices are lowered to discourage the opening of a competing supermarket chain store in the

---

<sup>4</sup>The supermarket chain sometimes run temporary promotions reducing the price of selected goods for a limited period of time. The time interval after the permanent price discount spans from October 4, 2018 to April 30, 2019 and in the corresponding period of the year before the intervention (October 4, 2017 - April 30, 2018) there were no temporary promotions on the store brands that are part of this analysis.

neighborhood: in this case we would be uncertain if a positive effect on sales was due to the price reduction or to the deferred market entrance of the competitor.

### 3.2 Causal estimands

We now introduce one class of causal estimands that we denote as “pointwise effects”. This class consists of three related estimands: the *point* effect (an instantaneous effect at each point in time after the intervention), the *cumulative* effect (a partial sum of the point effects) and the *temporal average* effect (the average of the point effects in a given time period).

Assumption 1 allows us to drop the  $t$  subscript from the treatment assignment and to write  $W_{i,t} = w_{i,t} = 0$  for  $t \leq t^*$  and  $W_{i,t} = W_i$ ,  $w_{i,t} = w_i$  for  $t > t^*$ . Furthermore, throughout this section we make the assumption that the units do not interfere with one another (Assumption 4); thus, we can also drop the  $i$  subscript from both the assignment indicator and the potential outcomes.

**Definition 1 (Pointwise effects)** *For the two treatment paths  $w$  and  $w'$ , the point effect at time  $t > t^*$  is,*

$$\tau_t(w; w') = Y_t(w) - Y_t(w') \quad (1)$$

*then, the cumulative effect at time  $t$  is the pointwise sum of the effects up to time  $t$ ,*

$$\Delta_t(w; w') = \sum_{s=t^*+1}^t \tau_s(w; w') \quad (2)$$

*and, finally, the temporal average effect at time  $t$  is,*

$$\bar{\tau}_t(w; w') = \frac{1}{t - t^*} \sum_{s=t^*+1}^t \tau_s = \frac{\Delta_t(w; w')}{t - t^*}. \quad (3)$$

**Example 1** *Assume that the single treatment occurring at time  $t^* = 2$  in Figure 2 is a persistent price reduction on a cookie brand. The two point effects are  $\tau_3(1; 0) = Y_3(0, 0, 1) - Y_3(0, 0, 0)$  and  $\tau_4((1, 1); (0, 0)) = Y_4(0, 0, 1, 1) - Y_4(0, 0, 0, 0)$ , indicating the additional sales due to the price reduction at time  $t = 3$  and  $t = 4$ , respectively. Then, the cumulative effect at time  $t = 4$  is given by  $\Delta_4((1, 1); (0, 0)) = \tau_3(1; 0) + \tau_4((1, 1); (0, 0))$  and it is the total number of cookies sold due to the intervention. Finally, the temporal average effect is  $\bar{\tau}_4((1, 1); (0, 0)) = \frac{1}{2}\Delta_4((1, 1); (0, 0))$ , denoting the number of cookies sold daily, on average, due to the permanent price reduction.*

In words, the point effect measures the causal effect at a specific point in time and can be defined at every  $t \in \{t^* + 1, \dots, T\}$ , thereby originating a vector of effects. The cumulative and temporal average effects are obtained by summing or averaging the point effects in a given time

period. Notice that the pointwise effects are analogous to the general causal effects defined in [Bojinov and Shephard \(2019\)](#), with the only difference that this class of effects is re-defined for a special setting where the units are subject to a single persistent treatment occurring at time  $t^*$ .

### 3.3 C-ARIMA

We propose a causal version of the widely used ARIMA model, which we indicate as C-ARIMA. We first introduce a simplified version of this model for stationary data generating processes; then, we relax the stationarity assumption and we extend the model to encompass seasonality and external regressors. Finally, we provide a theoretical comparison of the proposed approach with a standard ARIMA model with no causal connotation, from now on denoted as REG-ARIMA.

#### 3.3.1 Simplified framework

We start with a simplified model for stationary data generating processes: this allows us to illustrate the building blocks of our approach with a clear and easy-to-follow notation. So, let us assume  $\{Y_t\}$  evolving as

$$Y_t(\mathbf{w}) = c + \frac{\theta_q(L)}{\phi_p(L)}\varepsilon_t + \tau_t\mathbb{1}_{\{\mathbf{w}=1\}} \quad (4)$$

where  $\phi_p(L)$  and  $\theta_q(L)$  are lag polynomials with  $\phi_p(L)$  having all roots outside the unit circle;  $c$  is a constant term;  $\tau_t = 0 \forall t \leq t^*$  and  $\mathbb{1}_{\{\mathbf{w}=1\}}$  is an indicator function which is one if  $\mathbf{w} = 1$ . As a result,  $\tau_t$  can be interpreted as the point causal effect at time  $t > t^*$ , since it is defined as a contrast of potential outcomes,

$$\tau_t(\mathbf{w} = 1; \mathbf{w} = 0) = Y_t(\mathbf{w} = 1) - Y_t(\mathbf{w} = 0) = \tau_t.$$

Notice that under the Assumptions introduced in [Section 2](#),  $\tau_t$  is a properly defined causal effect in the RCM and, as such, it should not be confused with additive outliers or any other kind of intervention component typically used in the econometric literature (e.g., [Box and Tiao \(1975\)](#), [Chen and Liu \(1993\)](#)). Indeed, we can show that [Equation \(4\)](#) encompasses all types of interventions. For example, consider the following model specification (innovation-type effect),

$$Y_t(\mathbf{w}) = c + \frac{\theta_q(L)}{\phi_p(L)}(\varepsilon_t + \tau_t\mathbb{1}_{\{\mathbf{w}=1\}})$$

and define  $\tau_t^\dagger = \frac{\theta_q(L)}{\phi_p(L)}\tau_t$ . Then, we have

$$Y_t(\mathbf{w}) = c + \frac{\theta_q(L)}{\phi_p(L)}\varepsilon_t + \tau_t^\dagger \mathbb{1}_{\{\mathbf{w}=1\}}$$

where  $\tau_t^\dagger(1;0) = Y_t(1) - Y_t(0) = \tau_t^\dagger$  is the point causal effect at time  $t$ . As it will be clear in Section 3.3.4, our model is estimated on the pre-intervention data, thus in the C-ARIMA approach we do not need to find the structure that better represents the effect of the intervention (e.g., additive outlier, transient change, innovation outlier); conversely, such effect emerges as a contrast of potential outcomes in the post-intervention period. In other words, the proposed approach allows us to estimate  $\tau_t^\dagger$  (whatever structure it has).

To improve readability of the model equations, from now on we use  $Y_t$  to indicate  $Y_t(\mathbf{w})$ ; the usual notation is resumed in Section 3.3.4. Thus, Equation (4) can be written as,

$$Y_t = c + \frac{\theta_q(L)}{\phi_p(L)}\varepsilon_t + \tau_t. \quad (5)$$

Setting

$$z_t = \frac{\theta_q(L)}{\phi_p(L)}\varepsilon_t, \quad (6)$$

Equation (5) becomes,

$$Y_t = c + z_t + \tau_t.$$

Assuming perfect knowledge of the parameters ruling  $\{z_t\}$ , indicating with  $\mathcal{I}_{t^*}$  the information up to time  $t^*$  and denoting with  $H_0$  the situation where the intervention has no effect (namely,  $\tau_t = 0$  for all  $t > t^*$ ) we have that for a positive integer  $k$ , the  $k$ -step ahead forecast of  $Y_t$  under  $H_0$  conditionally on  $\mathcal{I}_{t^*}$  is

$$\hat{Y}_{t^*+k} = E[Y_{t^*+k} | \mathcal{I}_{t^*}, H_0] = c + \hat{z}_{t^*+k|t^*} \quad (7)$$

where  $\hat{z}_{t^*+k|t^*} = E[z_{t^*+k} | \mathcal{I}_{t^*}, H_0]$ . Thus,  $\hat{z}_{t^*+k|t^*}$  represents our estimate of the missing potential outcomes in the absence of intervention, i.e.,  $\hat{Y}_{t^*+k}(0) = \hat{z}_{t^*+k|t^*}$ .

### 3.3.2 General framework

We now generalize the above framework to a setting where  $\{Y_t\}$  is non-stationary and possibly includes seasonality as well as external regressors.

Let  $\{Y_t\}$  follow a regression model with ARIMA errors and the addition of the point effect  $\tau_t$ ,

$$(1 - L^s)^D(1 - L)^d Y_t = \frac{\Theta_Q(L^s)\theta_q(L)}{\Phi_P(L^s)\phi_p(L)}\varepsilon_t + (1 - L^s)^D(1 - L)^d X_t' \beta + \tau_t \quad (8)$$



where  $\Theta_Q(L^s)$ ,  $\Phi_P(L^s)$  are the lag polynomials of the seasonal part of the model with  $\Phi_P(L^s)$  having roots all outside the unit circle;  $X_t$  is a set of external regressors;  $(1 - L^s)^D$  and  $(1 - L)^d$  are contributions of the differencing operators to ensure stationarity, and  $s$  is the seasonal period. Notice that the intercept defined in model (5) is now included in the set of regressors. To ease notation, defining

$$z_t = \frac{\Theta_Q(L^s)\theta_q(L)}{\Phi_P(L^s)\phi_p(L)}\varepsilon_t$$

and indicating with  $T(\cdot)$  the transformation of  $Y_t$  needed to achieve stationarity, i.e.  $T(Y_t) = (1 - L^s)^D(1 - L)^d Y_t$ , model (8) becomes

$$S_t = T(Y_t) - T(X_t)'\beta = z_t + \tau_t,$$

where  $T(X_t)' = (1 - L^s)^D(1 - L)^d X_t'$  indicates that the same transformation is applied to the vector of regressors. Thus, the  $k$ -step ahead forecast of  $S_t$  under  $H_0$ , given the information up to time  $t^*$  is

$$\hat{S}_{t^*+k} = E[S_{t^*+k}|\mathcal{I}_{t^*}, H_0] = E[T(Y_{t^*+k}) - T(X_{t^*+k})'\beta|\mathcal{I}_{t^*}, H_0] = \hat{z}_{t^*+k|t^*}.$$

### 3.3.3 Comparison with REG-ARIMA

To measure the effect of an intervention on an outcome repeatedly observed over time, a widely used approach is fitting a linear regression with ARIMA errors (REG-ARIMA). In particular, this method uses the entire time series and a dummy variable activating after the intervention; then, SARIMA-type errors are added to the model to account for autocorrelation and possible seasonality. In its simplest formulation, such a model can be written as,

$$\begin{aligned} Y_t &= c + D_t\beta_0 + z_t \\ z_t &= \frac{\theta_q(L)}{\phi_p(L)}\varepsilon_t \end{aligned}$$

where  $z_t$  is a stationary ARMA( $p, q$ );  $D_t$  is a dummy variable taking value 1 after the intervention and 0 otherwise and  $\beta_0$  is the regression coefficient. Generalizing to a non-stationary ARIMA( $p, d, q$ ), above model can be re-written as,

$$\begin{aligned} Y_t &= X_t'\beta + z_t \\ z_t &= \frac{\theta_q(L)\Theta_Q(L^s)}{(1 - L)^d(1 - L^s)^D\phi_p(L)\Phi_P(L^s)}\varepsilon_t \end{aligned} \tag{9}$$

where  $X_t$  is a set of regressors, including the intercept and the dummy variable and  $\beta$  is a vector of regression coefficients.

Essentially, REG-ARIMA is a standard intervention analysis approach that is used when the intervention is supposed to have produced a level shift on the outcome, i.e. a fixed change in the level of the outcome during the post-intervention period. Thus, there are two main differences between C-ARIMA and REG-ARIMA. First, without a critical discussion of the fundamental assumptions, the effect grasped by  $\beta_0$  can not be attributed with certainty to the intervention. For example, it might be driven by an undetected confounder, biased by the inclusion of a regressor linked to the treatment, or even be the anticipated result of a future intervention. Second, the size of the effect is given by the estimated coefficient of a dummy variable activating after the intervention, so that REG-ARIMA can only capture effects in the form of level shifts. Conversely, C-ARIMA assumes no structure on  $\tau_t$  and, as such, it can capture any form of effects (level shift, slope change and even irregular time-varying effects). Furthermore, the estimation of the effect is done in a very natural way under C-ARIMA. Indeed, intervention analysis requires the estimation of two models: the first learns the structure of the effect and the second measures its size; by letting the intervention component free to vary, C-ARIMA can instead estimate any form of effect in only one step.

In Section 3.4 we report a simulation study where we compare the empirical performance of both approaches (C-ARIMA and REG-ARIMA) in inferring causal effects.

### 3.3.4 Causal effect estimation

We now derive estimators for the causal effects defined in Section 3.2 based on the C-ARIMA model and we discuss their properties.

**Definition 2 (Pointwise effects estimators)** *For any integer  $k$ , let  $S_{t^*+k}(w)$  be the observed potential outcome time series and let  $\hat{S}_{t^*+k}(w') = \hat{z}_{t^*+k|t^*}(w')$  be the corresponding estimate of the missing potential outcomes under model (8). Then, estimators of the point, cumulative and temporal average effects are, respectively,*

$$\hat{\tau}_{t^*+k}(w; w') = S_{t^*+k}(w) - \hat{S}_{t^*+k}(w') = S_{t^*+k}(w) - \hat{z}_{t^*+k|t^*}(w')$$

$$\hat{\Delta}_{t^*+k}(w; w') = \sum_{h=1}^k \hat{\tau}_{t^*+h}(w; w')$$

$$\hat{\bar{\tau}}_{t^*+k}(w; w') = \frac{1}{k} \sum_{h=1}^k \hat{\tau}_{t^*+h}(w; w') = \frac{\hat{\Delta}_{t^*+k}(w; w')}{k}.$$

Considering this setup we can show that

$$\hat{\tau}_{t^*+k}(w; w') \sim \left[ \tau_{t^*+k}(w; w'), \sigma_\varepsilon^2 \sum_{i=0}^{k-1} \psi_i^2 \right] \quad (10)$$

$$\hat{\Delta}_{t^*+k}(w; w') \sim \left[ \Delta_{t^*+k}(w; w'), \sigma_\varepsilon^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_i \right)^2 \right] \quad (11)$$

$$\hat{\bar{\tau}}_{t^*+k}(w; w') \sim \left[ \bar{\tau}_{t^*+k}(w; w'), \frac{1}{k^2} \sigma_\varepsilon^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_i \right)^2 \right] \quad (12)$$

where, the  $\psi_i$ 's are the coefficients of a moving average of order  $k - 1$  whose values are functions of the ARMA parameters ruling  $z_t$  (as defined in Equation (6)). Indeed, starting from the MA( $\infty$ ) representation of the stationary component  $z_{t^*+k}$ , we can show that  $\hat{\tau}_{t^*+k}(w; w')$  is MA( $k - 1$ ).<sup>5</sup> Equations (10), (11), (12) can be used to derive confidence intervals for the corresponding causal estimands.

Summarizing, in order to estimate the causal effects (1), (2) and (3) we need to follow a four-step process: i) split the analysis period in two time intervals: the pre-intervention and post-intervention periods, respectively  $\{1, \dots, t^*\}$  and  $\{t^*+1, \dots, T\}$ ; ii) estimate the ARIMA model only in the pre-intervention period, so as to learn the dynamics of the dependent variable and the links with the covariates without being influenced by the treatment; iii) based on the process learned in the pre-intervention period, perform a prediction step and obtain the counterfactual outcome during the post-intervention period in the absence of intervention; iv) by comparing the observations with the corresponding forecasts at any time point after the intervention, evaluate the resulting differences, which represent the estimated point causal effects.

Conversely, REG-ARIMA model is fitted to the full time series (pre- and post-intervention) and the estimated coefficient for the dummy variable  $D_t$  gives a measure of the association between the intervention (in the form of a level shift) and the outcome.

The C-ARIMA shares many features with the approach described in [Box and Tiao \(1976\)](#), where the authors suggest to compare the observed data after an intervention with the forecasts from a model fitted to the pre-intervention period. However, the interpretation of the resulting difference as a causal effect must follow from a thorough discussion of the assumptions underneath such attribution: this is the main building block of C-ARIMA, making it different from other popular approaches based on ARIMA models, like REG-ARIMA and intervention analysis.

---

<sup>5</sup>Proof and additional details on how to recover the effect on the untransformed variable  $Y_t$  are given in Appendix [B.1](#) and [B.2](#).

### 3.4 Simulation study

We perform a simulation study to check the ability of the C-ARIMA approach to uncover causal effects. Furthermore, in order to show its merits over a more standard approach, we also assess the performance of REG-ARIMA. We remark, however, that the comparison is purely methodological, since the theoretical limitations of REG-ARIMA do not allow the attribution of such effects to the intervention. Sections 3.4.1 and 3.4.2 illustrate, respectively, the simulations design and the results.

#### 3.4.1 Design

We generate 1000 replications from the following ARIMA(1, 0, 1)(1, 0, 1)<sub>7</sub> model,

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + z_t$$

$$z_t = \frac{\theta_q(L)\Theta_Q(L^s)}{\phi_p(L)\Phi_P(L^s)}\varepsilon_t.$$

The two covariates of the regression equation are generated as  $X_{1,t} = \alpha_1 t + u_{1,t}$  and  $X_{2,t} = \sin(\alpha_2 t) + u_{2,t}$ , with  $\alpha_1 = \alpha_2 = 0.01$ ,  $u_{1,t} \sim N(0, 0.02)$ ,  $u_{2,t} \sim N(0, 0.5)$  and coefficients  $\beta_1 = 0.7$  and  $\beta_2 = 2$ , respectively; regarding the ARIMA parameters, they are set to  $\phi_1 = 0.7$ ,  $\Phi_1 = 0.6$ ,  $\theta_1 = 0.6$  and  $\Theta_1 = 0.5$ . Finally,  $\varepsilon_t \sim N(0, \sigma)$  with  $\sigma = 5$ . Figure 4 shows the evolution of the generated covariates and their linear combination according to the above model.

We assume that each generated time series starts on January 1, 2017 and ends on December 31, 2019 and that a fictional intervention takes place on June 30, 2019. In particular, we tested two types of intervention: i) a level shift with 5 different magnitudes, i.e., +1%, +10%, +25%, +50%, +100% ; ii) an intervention producing an immediate shock of +10% followed by a steady increase up to +40%, a regular decline afterwards and a second increase near the end of the analysis period. As an example, Figure 5 provides a graphical representation of the two interventions for one of the simulated time series.

The estimation of the causal effect is performed under two different models: the proposed C-ARIMA approach and REG-ARIMA, i.e, a linear regression with ARIMA errors and the addition of a dummy variable, as in Equation (9). Recall from previous Section 3.3.4 that the C-ARIMA approach requires that the model is estimated on the pre-intervention data and the effect is given by direct comparison of the observed series and the corresponding forecasts post-intervention. Conversely, REG-ARIMA is fitted on the full time series and the estimated coefficient of the dummy variable provides a measure of the impact of the intervention. In addition, we estimate two versions of each model: a correctly specified model, denoted respectively with C-ARIMA<sup>TRUE</sup> and REG-ARIMA<sup>TRUE</sup>, and the best fitting model selected

by BIC minimization, denoted with C-ARIMA<sup>BIC</sup> and REG-ARIMA<sup>BIC</sup>. Finally, in order to evaluate the performance of both approaches in uncovering causal effects at longer time horizons, we perform predictions at 1 month, 3 months and 6 months from the intervention. As a result, the total number of estimated models in the pre-intervention period is 4000 and the total number of estimated causal effects is 72,000 (one for each time series, model, tested intervention and time horizon).

We measure the performance of the four models in terms of three indicators:

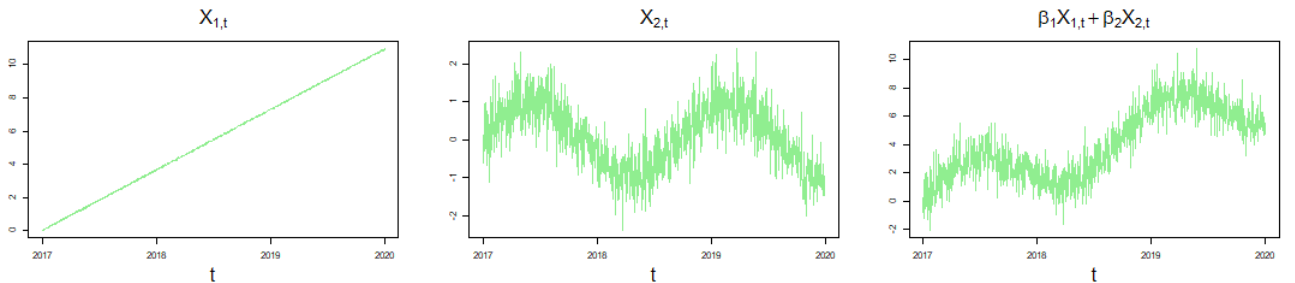
1. the length of the confidence interval around the true temporal average effect  $\bar{\tau}_t$  for C-ARIMA<sup>TRUE</sup> and C-ARIMA<sup>BIC</sup> and around  $\beta_0$  for REG-ARIMA<sup>TRUE</sup> and REG-ARIMA<sup>BIC</sup>;
2. the absolute percentage error, defined for each model and intervention type as,

$$a_{i,th} = \frac{|\hat{\tau}_{i,h} - \bar{\tau}_{i,h}|}{\bar{\tau}_{i,h}} ; a'_{i,h} = \frac{|\hat{\beta}_{0,i,h} - \bar{\tau}_{i,h}|}{\bar{\tau}_{i,h}} \quad i = 1, \dots, 1000;$$

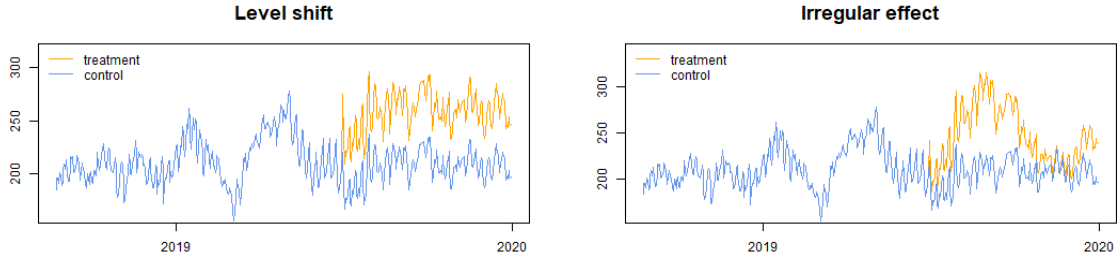
where,  $\hat{\tau}_{i,h}$  and  $\hat{\beta}_{0,i,h}$  denote, respectively, the estimated causal effect and the estimated coefficient of the intervention dummy for the  $i$ -th simulated time series at time horizon  $h$ , where  $h = \{1, 2, 3\}$  indicate the months after the intervention; instead,  $\bar{\tau}_{i,t}$  denotes the true temporal average effect (always positive);

3. the interval coverage, computed for the REG-ARIMA as the proportion of true effects in the estimated 95% confidence intervals over the 1000 simulated series, whereas for the C-ARIMA it is obtained as the proportion of the true point effects within the estimated confidence intervals, which is then averaged over the simulated series.

**Figure 4:** Evolution of the generated covariates,  $X_{1,t}$  and  $X_{2,t}$  and their combination according to the simulated model,  $\beta_1 X_{1,t} + \beta_2 X_{2,t}$ .



**Figure 5:** For the same simulated time series (denoted as “control”) the plots display two different types of effect: on the left, a level shift of 25%; on the right, the time series under treatment follows an irregular pattern.



### 3.4.2 Results

Table 1 shows the simulation results in terms of the length of the 95% confidence intervals around  $\tau_t(1; 0)$  and  $\beta_0$ , respectively. As expected, for the C-ARIMA models the interval length is independent of the impact; we can also notice that it reduces as the time horizon increases, whereas the interval length estimated under REG-ARIMA is stable over time. Finally, we can observe that REG-ARIMA yields shorter confidence intervals than C-ARIMA.

Table 2 reports the absolute percentage errors resulting from the simulations. When the intervention takes the form of a level shift, the error decreases with the size of the effect and, unsurprisingly, REG-ARIMA yields slightly better results than C-ARIMA. Indeed, the former model is especially suited for interventions in the form of level shifts. However, when we consider a level shift  $> +50\%$  or an irregular intervention, the estimation errors of REG-ARIMA are 2 to 4 times higher than those coming from C-ARIMA.

The interval coverage is reported in Table 3. Again, the coverage of the C-ARIMA approach does not vary with the impact size and it is very close to the nominal 95% level. Instead, the coverage of REG-ARIMA decreases with the impact size and, with the only exception of the first two impacts, the results are quite far from the nominal 95% level. This can be explained by the short confidence intervals achieved by REG-ARIMA, suggesting that even though the estimation error is small, the confidence intervals are not wide enough to contain the true effect. More importantly, when the effect is irregular, the estimated confidence intervals never contain the true effect.

Concluding, REG-ARIMA approach fails to detect irregular interventions and most of the times it does not achieve the desired interval coverage. As expected, REG-ARIMA model is suited only when there is reason to believe that the intervention produced a fixed change in the outcome level. Otherwise, should the researcher fail to identify the structure of the effect, using REG-ARIMA on irregular patterns produces biased estimates. Conversely, the C-ARIMA approach does a reasonably good job in detecting both type of interventions. Moreover, C-ARIMA does

not require an investigation of the effect type prior to the estimation step; in addition, when the intervention is in the form of a level shift, the reliability of the estimates increases with the impact size. Finally, we can observe that the results of the models selected through BIC minimization are very similar to those of the correct model specifications (indeed, the BIC criterion correctly identifies 74% of the models).

**Table 1:** Length of the 95% confidence intervals around the estimated intervention effect  $\tau_t(1;0)$  (for C-ARIMA) and  $\beta_0$  (for REG-ARIMA). The different impact sizes ranging from +1% to +100% in the rows denote estimated effects in the form of level shifts, whereas NS stands for “no structure”, thereby indicating the irregular effect. For each generated time series, impact size and time horizon (1, 3 and 6 months), the estimates are performed under two model specifications: the true model and the best fitting model based on BIC, denoted, respectively, with the superscripts *TRUE* and *BIC*.

$\tau_t(1;0)$	C-ARIMA <sup>BIC</sup>			REG-ARIMA <sup>BIC</sup>		
	1 month	3 months	6 months	1 month	3 months	6 months
+1%	42.029	34.479	26.330	10.620	10.458	10.340
+10%	42.029	34.479	26.330	10.656	10.555	10.511
+25%	42.029	34.479	26.330	10.734	10.751	10.849
+50%	42.029	34.479	26.330	10.916	11.164	11.528
+100%	42.029	34.479	26.330	11.454	12.259	13.233
NS	42.027	34.474	26.325	10.709	10.931	10.891
$\tau_t(1;0)$	C-ARIMA <sup>TRUE</sup>			REG-ARIMA <sup>TRUE</sup>		
	1 month	3 months	6 months	1 month	3 months	6 months
+1%	42.055	34.536	26.381	10.604	10.451	10.334
+10%	42.055	34.536	26.381	10.639	10.548	10.506
+25%	42.055	34.536	26.381	10.716	10.743	10.842
+50%	42.055	34.536	26.381	10.895	11.152	11.520
+100%	42.055	34.536	26.381	11.424	12.240	13.225
NS	42.058	34.533	26.377	10.691	10.913	10.876

**Table 2:** Absolute percentage error for the estimated intervention effect  $\tau_t(1;0)$  (for C-ARIMA) and  $\beta_0$  (for REG-ARIMA). The different impact sizes ranging from +1% to +100% in the rows denote estimated effects in the form of level shifts, whereas NS stands for “no structure”, thereby indicating the irregular effect. For each generated time series, impact size and time horizon (1, 3 and 6 months), the estimates are performed under two model specifications: the true model and the best fitting model based on BIC, denoted, respectively, with the superscripts *TRUE* and *BIC*.

$\tau_t(1;0)$	C-ARIMA <sup>BIC</sup>			REG-ARIMA <sup>BIC</sup>		
	1 month	3 months	6 months	1 month	3 months	6 months
+1%	4.185	3.783	3.405	0.970	0.955	0.950
+10%	0.418	0.378	0.340	0.116	0.118	0.117
+25%	0.167	0.151	0.136	0.074	0.080	0.078
+50%	0.084	0.076	0.068	0.065	0.072	0.071
+100%	0.042	0.038	0.034	0.062	0.070	0.069
NS	0.237	0.138	0.173	0.423	0.610	0.463

$\tau_t(1;0)$	C-ARIMA <sup>TRUE</sup>			REG-ARIMA <sup>TRUE</sup>		
	1 month	3 months	6 months	1 month	3 months	6 months
+1%	4.182	3.775	3.398	0.963	0.946	0.943
+10%	0.418	0.378	0.340	0.115	0.118	0.116
+25%	0.167	0.151	0.136	0.074	0.079	0.078
+50%	0.084	0.076	0.068	0.065	0.072	0.071
+100%	0.042	0.038	0.034	0.062	0.070	0.068
NS	0.237	0.137	0.172	0.424	0.610	0.463

**Table 3:** Interval coverage in percentage of the true effects within the estimated intervals around  $\tau_t(1;0)$  (for C-ARIMA) and  $\beta_0$  (for REG-ARIMA). The different impact sizes ranging from +1% to +100% in the rows denote estimated effects in the form of level shifts, whereas NS stands for “no structure”, thereby indicating the irregular effect. For each generated time series, impact size and time horizon (1, 3 and 6 months), the estimates are performed under two model specifications: the true model and the best fitting model based on BIC, denoted, respectively, with the superscripts *TRUE* and *BIC*.

$\tau_t(1;0)$	C-ARIMA <sup>BIC</sup>			REG-ARIMA <sup>BIC</sup>		
	1 month	3 months	6 months	1 month	3 months	6 months
+1%	94.25	93.68	93.15	95.20	94.78	95.62
+10%	94.25	93.68	93.15	90.74	90.40	91.25
+25%	94.25	93.68	93.15	71.89	68.01	69.44
+50%	94.25	93.68	93.15	45.62	40.66	42.59
+100%	94.25	93.68	93.15	24.83	23.91	27.02
NS	94.21	93.66	93.16	0.17	0.00	0.00

$\tau_t(1;0)$	C-ARIMA <sup>TRUE</sup>			REG-ARIMA <sup>TRUE</sup>		
	1 month	3 months	6 months	1 month	3 months	6 months
+1%	94.27	93.70	93.19	95.12	94.95	95.71
+10%	94.27	93.70	93.19	90.99	90.40	91.67
+25%	94.27	93.70	93.19	71.89	67.93	69.70
+50%	94.27	93.70	93.19	45.62	40.74	43.10
+100%	94.27	93.70	93.19	24.92	23.99	27.02
NS	94.23	93.69	93.20	0.17	0.00	0.00



### 3.5 Empirical analysis

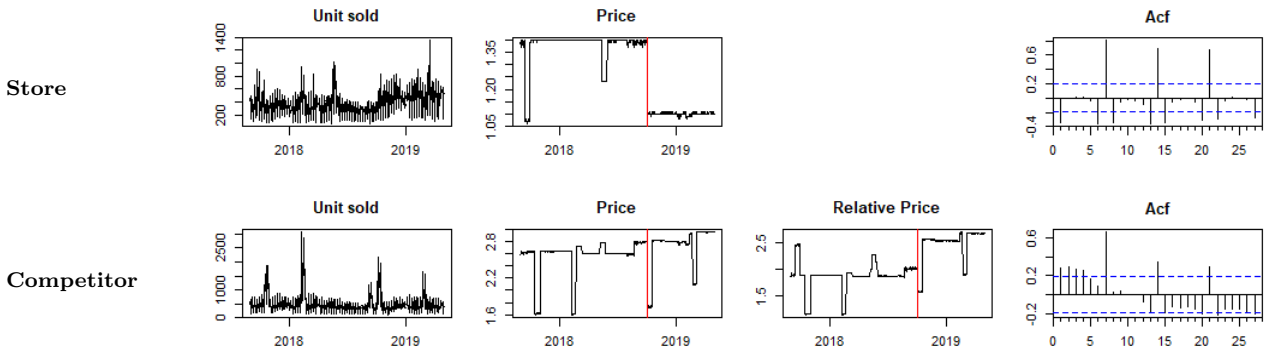
In this section we describe the results of our empirical application; the goal is estimating the impact of the permanent price reduction performed by an Italian supermarket chain.

#### 3.5.1 Data & methodology

Data consists of daily sales counts of 11 store brands and their corresponding competitor brand cookies in the period September 1, 2017, April 30, 2019.<sup>6</sup> The permanent price reduction on the store brand cookies was introduced by the supermarket chain on October 4, 2018.

As an example, Figure 6 shows the time series of units sold, the evolution of price per unit and the autocorrelation function of one store brand and its direct competitor. The plots for the remaining store-brand and competitor-brand cookies are provided in Appendix A.2. The occasional price drops before the intervention date indicate temporary promotions run regularly by the supermarket chain. The products exhibit a clear weekly seasonal pattern, illustrated by the spikes in the autocorrelation functions. In the panel referred to the direct competitor brand, we can also observe the evolution of the relative price per unit (the ratio between the prices of the competitor brand and the corresponding store brand). Unsurprisingly, despite the occasional drops due to the usual promotions, the price of the competitor brand relative to the corresponding store brand has increased after the intervention.

**Figure 6:** Time series of unit sold daily, price per unit and autocorrelation function for two selected items (i.e., store brand 6 and the corresponding competitor brand). For the competitor, the relative price plot shows the ratio between its unit price and the price of the store brand.



To determine the causal effect of the permanent price discount on the sales of store-brands cookies we follow the approach outlined in Section 3.3. In particular, under Assumption 4, we analyze each cookie separately, thereby fitting 11 independent models. In order to improve model diagnostics, the dependent variable is the natural log of the daily sales count. This also means that we are postulating the existence of a multiplicative effect of the new price policy on

<sup>6</sup>We excluded the last competitor brand because 62% of observations were missing. Thus, we analyzed 11 store and 10 competitor brands.

the sales of cookies. Since in terms of the original variable the cumulative sum of daily effects is equivalent to their product, we focused our attention on estimating the temporal average causal effect, which can still be interpreted as an average multiplicative effect. Furthermore, we included covariates to improve prediction of the missing potential outcomes in the absence of intervention. In particular, to take care of the seasonality we included six dummy variables corresponding to the day of the week and one dummy denoting December Sundays.<sup>7</sup> Indeed, the policy of the supermarket chain implies that all shops are closed on Sunday afternoon except during Christmas holidays. Thus, we may have two opposite “Sunday effects”: a positive effect in December, when the shops are open all the day; a negative effect during the rest of the year, since all shops are closed in the afternoon. We also included a holiday dummy taking value 1 before and after a national holiday and 0 otherwise. This is to account for consumers’ tendency to increase purchases before and after a closure day.<sup>8</sup> Finally, we included a modified version of the unit price, that after the intervention day and during all the post-period is taken equal to the last price before the permanent discount. As explained in discussing Assumption 6, this is the most likely price that the unit would have had in the absence of intervention. In addition, to estimate the average causal effect of the intervention on store brands, we are also interested in evaluating how this effect evolves with time. Thus, we repeated the analysis by making predictions at three different time horizons: 1 month, 3 months and 6 months after the intervention.

The same methodology is applied to the competitor brands, with a slight modification on the set of covariates. Indeed, this time the unit price is not directly influenced by the intervention; so, to forecast competitor sales in the absence of intervention we directly used the actual price. Again, to illustrate the merits of our causal approach, the results obtained from C-ARIMA are then compared to those of REG-ARIMA, as described by Equation (9). More specifically, we fitted independent linear regressions with ARIMA errors for each of the 11 store brands and their competitors.

### 3.5.2 Results

Table 4 shows the results of the C-ARIMA and the REG-ARIMA approaches applied to the store brands. Figure 7 illustrates the causal effect, the observed time series and the forecasted series in the absence of intervention for one selected item.<sup>9</sup> At the 1-month time horizon, the causal effect is significantly positive for 8 out of 11 items; three months after the intervention, the causal effect is significantly positive for 10 items; after six months, the effect is significant and

---

<sup>7</sup>In principle, we may also have a monthly seasonal pattern on top of the weekly cycle but the reduced length of the pre-intervention time series (398 observations) does not allow us to assess whether a double seasonality is present.

<sup>8</sup>To be precise, on the day of a national holiday we have a missing value (so there is no holiday effect), whereas the dummy variable should capture the effect of additional purchases before and after the closure day(s).

<sup>9</sup>The same plots for the remaining store-brand and competitor-brand cookies are provided in Appendix A.2.

positive for all items. Conversely, REG-ARIMA fails to detect some of the effects: compared to the C-ARIMA results, the effect on items 4 and 11 at the first time horizon, on item 11 at the second horizon and on items 5 and 11 at the third horizon are not significant.

Table 5 reports the results for the competitor brands and Figure 8 plots the causal effect, the observed series and the forecasted series for one selected item. Again, the causal effect seems to strengthen as we proceed far away from the intervention. At 1-month horizon no significant effect is observed; three months after the intervention, we find a significant and negative effect on item 10; at 6-month horizon we find significant negative effects on items 8 and 10 and a significant positive effect on item 5. A negative effect suggests that following the permanent price discount, consumers have changed their behavior by privileging the cheaper store brand. Instead, a positive effect might indicate that the price policy has determined an increase in the customer base, i.e. new clients have entered the shop and eventually bought the items at full price. Again, REG-ARIMA model leads to partially different results: at 6-month horizon, a positive effect is found on item 6 and no effect is detected on item 8.

Summarizing, the intervention seems to have produced a significant and positive effect on the sales of store brand cookies. Conversely, we do not find considerable evidence of a detrimental effect on competitor cookies (the only exceptions being items 8 and 10). This indicates that, even though each store-competitor pair is formed by perfect substitutes, price might not be the only factor driving sales. For example, unobserved factors such as individual preferences or brand faithfulness may have a role as well.

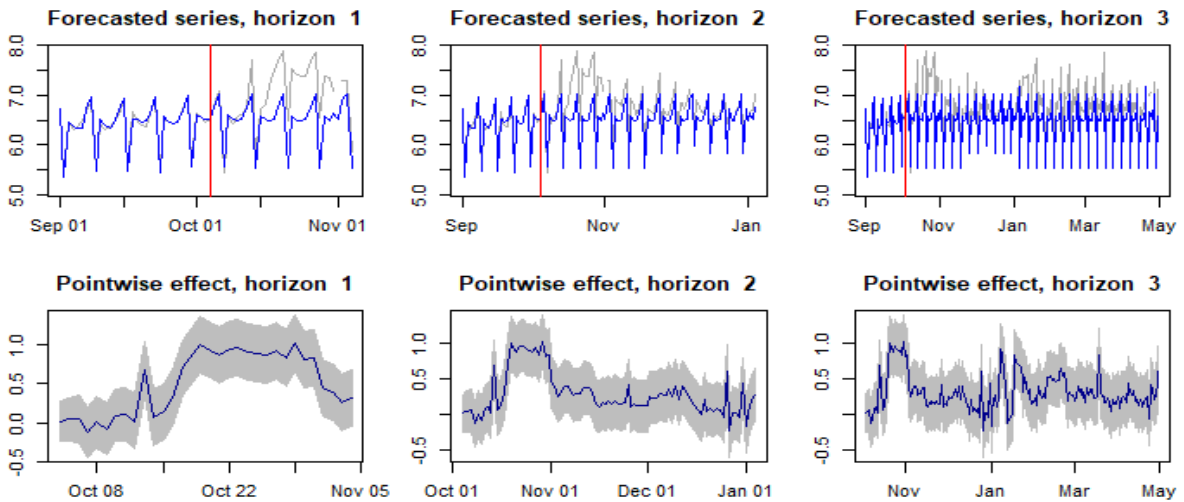
**Table 4:** Causal effect estimates of the permanent price rebate on sales of store-brand cookies after one month, three months and six months from the intervention. In this table,  $\hat{\tau}_t$  is the estimated temporal average effect ( $\hat{\tau}_t = 0$  implies no effect), while  $\hat{\beta}_0$  is the coefficient estimate of the intervention dummy according to REG-ARIMA ( $\hat{\beta}_0 = 0$  implies absence of association).

Item	Time horizon:					
	1 month		3 months		6 months	
	$\hat{\tau}_t$	$\hat{\beta}_0$	$\hat{\tau}_t$	$\hat{\beta}_0$	$\hat{\tau}_t$	$\hat{\beta}_0$
1	0.14 (0.12)	0.14 (0.09)	0.15 <sup>*</sup> (0.08)	0.12 <sup>*</sup> (0.07)	0.18 <sup>***</sup> (0.06)	0.16 <sup>**</sup> (0.06)
2	0.14 (0.12)	0.10 (0.14)	0.13 <sup>*</sup> (0.08)	0.12 <sup>*</sup> (0.07)	0.14 <sup>**</sup> (0.05)	0.13 <sup>**</sup> (0.05)
3	0.19 <sup>*</sup> (0.11)	0.15 <sup>*</sup> (0.08)	0.21 <sup>**</sup> (0.07)	0.15 <sup>*</sup> (0.07)	0.25 <sup>***</sup> (0.05)	0.24 <sup>***</sup> (0.04)
4	0.49 <sup>***</sup> (0.09)	0.00 (0.13)	0.30 <sup>***</sup> (0.06)	0.19 <sup>*</sup> (0.08)	0.32 <sup>***</sup> (0.04)	0.28 <sup>***</sup> (0.05)
5	-0.02 (0.12)	-0.06 (0.12)	0.07 (0.08)	-0.07 (0.11)	0.11 <sup>*</sup> (0.06)	-0.06 (0.10)
6	0.24 <sup>*</sup> (0.12)	0.26 <sup>*</sup> (0.14)	0.34 <sup>***</sup> (0.08)	0.24 <sup>*</sup> (0.12)	0.37 <sup>***</sup> (0.06)	0.23 <sup>*</sup> (0.11)
7	0.55 <sup>***</sup> (0.10)	0.75 <sup>***</sup> (0.11)	0.34 <sup>***</sup> (0.07)	0.70 <sup>***</sup> (0.10)	0.30 <sup>***</sup> (0.05)	0.77 <sup>***</sup> (0.11)
8	0.26 <sup>***</sup> (0.08)	0.29 <sup>**</sup> (0.09)	0.25 <sup>***</sup> (0.07)	0.29 <sup>**</sup> (0.10)	0.14 <sup>**</sup> (0.05)	0.28 <sup>**</sup> (0.09)
9	0.47 <sup>***</sup> (0.06)	0.70 <sup>***</sup> (0.10)	0.20 <sup>***</sup> (0.04)	0.29 <sup>***</sup> (0.09)	0.21 <sup>***</sup> (0.03)	0.26 <sup>***</sup> (0.06)
10	0.66 <sup>***</sup> (0.11)	0.85 <sup>***</sup> (0.14)	0.57 <sup>***</sup> (0.08)	0.82 <sup>***</sup> (0.15)	0.33 <sup>***</sup> (0.06)	0.85 <sup>***</sup> (0.15)
11	0.12 <sup>*</sup> (0.06)	0.02 (0.11)	0.16 <sup>**</sup> (0.05)	0.04 (0.11)	0.14 <sup>***</sup> (0.04)	0.08 (0.13)

Note:

<sup>\*</sup>p<0.1; <sup>\*</sup>p<0.05; <sup>\*\*</sup>p<0.01; <sup>\*\*\*</sup>p<0.001

**Figure 7:** First row: observed sales (gray) and forecasted sales (blue) of store brand 4 at 1 month (horizon 1), 3 months (horizon 2) and 6 months (horizon 3) from the intervention; the vertical bar indicates the intervention date. Second row: pointwise causal effect, computed as the difference between observed and forecasted sales, with its 95% confidence interval.



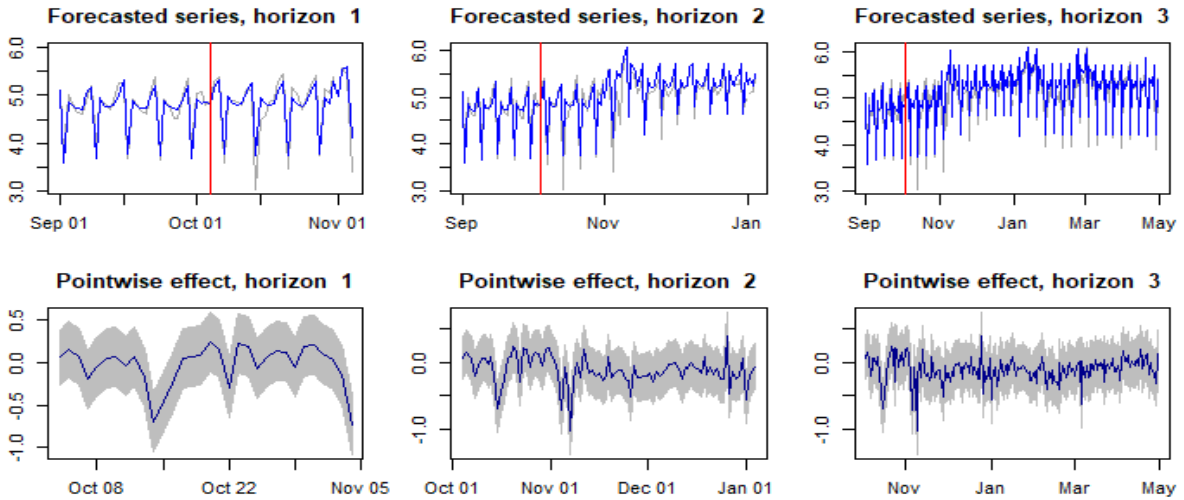
**Table 5:** Causal effect estimates of the permanent price rebate on sales of competitor-brand cookies after one month, three months and six months from the intervention. In this table,  $\hat{\tau}_t$  is the estimated temporal average effect ( $\hat{\tau}_t = 0$  implies no effect), while  $\hat{\beta}_0$  is the coefficient estimate of the intervention dummy according to REG-ARIMA ( $\hat{\beta}_0 = 0$  implies absence of association).

Item	Time horizon:					
	1 month		3 months		6 months	
	$\hat{\tau}_t$	$\hat{\beta}_0$	$\hat{\tau}_t$	$\hat{\beta}_0$	$\hat{\tau}_t$	$\hat{\beta}_0$
1	-0.03 (0.55)	0.02 (0.19)	0.02 (0.46)	-0.16 (0.25)	0.04 (0.34)	-0.12 (0.22)
2	-0.13 (0.50)	-0.18 (0.22)	-0.07 (0.47)	-0.13 (0.20)	-0.15 (0.36)	-0.13 (0.19)
3	0.04 (0.38)	-0.06 (0.22)	0.09 (0.23)	-0.03 (0.20)	0.17 (0.11)	0.03 (0.17)
4	0.00 (0.29)	0.08 (0.21)	-0.13 (0.21)	0.02 (0.22)	-0.04 (0.14)	0.01 (0.13)
5	-0.03 (0.10)	-0.01 (0.10)	0.05 (0.06)	0.06 (0.06)	0.12** (0.04)	0.12* (0.05)
6	-0.05 (0.12)	-0.01 (0.10)	0.03 (0.09)	0.06 (0.06)	0.09 (0.07)	0.10* (0.05)
7	0.04 (0.54)	-0.11 (0.29)	0.11 (0.33)	-0.05 (0.26)	0.40* (0.23)	0.02 (0.23)
8	-0.09 (0.07)	-0.02 (0.07)	-0.06 (0.05)	-0.10 (0.10)	-0.08* (0.04)	-0.12 (0.10)
9	-0.09 (0.13)	-0.08 (0.13)	-0.11 (0.09)	-0.11 (0.08)	-0.10 (0.06)	-0.09 (0.06)
10	-0.03 (0.06)	-0.02 (0.05)	-0.12** (0.04)	-0.09* (0.04)	-0.11*** (0.03)	-0.08* (0.04)

Note:

$\cdot$ p<0.1; \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

**Figure 8:** First row: observed sales (grey) and forecasted sales (blue) of competitor brand 10 at 1 month (horizon 1), 3 months (horizon 2) and 6 months (horizon 3) from the intervention; the vertical bar indicates the intervention date. Second row: pointwise causal effect, computed as the difference between observed and forecasted sales, with its 95% confidence interval.



### 3.6 Discussion

In this section, we presented C-ARIMA, a novel approach to estimate the effect of interventions in a time series setting under the RCM. We reported a detailed discussion of the assumptions underneath the causal framework, we defined three related estimands and we introduced a methodology to perform inference. To measure the performance of C-ARIMA in uncovering causal effects, we detailed a simulation study showing that this approach performs well in comparison with a standard intervention analysis approach when the true effect is in the form of a level shift; it also outperforms the latter in case of irregular, time-varying effects.

We believe that C-ARIMA can successfully be used as an alternative to CausalImpact to estimate the effect of an intervention on a single time series as well as in the context of multiple non-interfering series. Indeed, our methodology is accessible to all those researchers and practitioners that are not accustomed to (or are not willing to adopt) the Bayesian framework, at the same time providing several improvements over the standard intervention analysis approach.

In our empirical application, we estimated the causal effect of a new price policy introduced by an Italian supermarket chain, which addressed a selected subset of store brands by permanently lowering their price. Furthermore, we also assessed the indirect effect on competitor-brand products only differing in the brand name. To do that, we handled between-group interference by considering the competitors to be treated as well and defining the treatment in terms of the relative price increase. In this way we can rule out any interference originating from the permanent price reduction and spilling over from store to competitor brands.

Nevertheless, when used in our empirical context, this approach suffers from some limitations. Indeed, we are not able to control for possible interactions beyond those stemming from price or going the other way round. For example, if consumers can collect coupons to buy a competitor brand cookie, we reasonably expect a negative impact on the sales of the corresponding store brand. However, aside from their price, we do not have information on individual cookies and we are not able to include in the analysis our general belief that the two goods are perfect substitutes.

In the next section we overcome these limitations by developing a novel approach to deal with multiple interfering time series and we use it to re-analyze our data.

## 4 Causal effect of an intervention on multiple interfering time series

This section presents a novel approach to estimate the causal effect of an intervention in panel settings where the statistical units interfere with one another. Motivated by the analysis of the price policy change introduced by the Italian supermarket chain, we now treat jointly the pair formed by the store brand and the corresponding competitor brand, allowing interactions within the pairs but not between them and thereby relaxing, at least partially, the temporal no-interference assumption.

After presenting three new classes of causal estimands in the potential outcomes causal framework, we derive the multivariate extension of the popular Bayesian structural time series model for causal inference introduced by Brodersen et al. (2015). Like its univariate counterpart, MBSTS model is flexible due to its ability to incorporate trends and seasonality effects and the underlying distributional assumptions can be tested in very natural way by posterior predictive checks. In addition, our methodology allows to model the interference between units in the same group by explicitly modeling their dependence structure.

The proposed approach is implemented in the CausalMBSTS R package.

The remainder of this section is organized as follows: we begin by discussing the assumptions that are needed to estimate causal effects in this setting; Sections 4.2 and 4.3 illustrate the causal estimands and the proposed approach; in Section 4.4 we detail a simulation study to investigate the performance of our method; finally, Section 4.5 presents the results of the empirical analysis.

### 4.1 Background

Recalling our empirical application from previous Section 3, among the 284 items in the “cookies” category, there are 28 store brands, of which 11 were selected for a permanent price reduction ranging from  $-3.5\%$  to  $-23.2\%$ . For each store brand, the supermarket chain identified a direct competitor brand, thereby defining 11 pairs of cookies. Those in the same pair are almost identical except for their brand name.

We now review the set of assumptions for a causal framework where multiple units are grouped based on shared characteristics and we discuss them in the context of our application.

**Assumption 1** The single persistent intervention is the permanent price reduction introduced by the supermarket chain and the groups are the store-competitor pairs formed by one store brand and its direct competitor.

**Assumption 3** Since in our empirical setting the supermarket chain did not advertise the price reduction in advance and the treatment is a persistent intervention, this is a solid as-

sumption.

**Assumption 5** 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 only directly impact its sales and its direct competitor’s sales. Essentially, in our empirical context this is an assumption about consumer behavior; since each pair represents a different type of cookie (e.g., chocolate, whole grain, cream cookie), we are assuming that consumers’ choice of one pair or the other is not driven by price, but rather by individual preferences.<sup>10</sup>

**Assumption 6** Our set of covariates for each pair includes: i) weekend and holiday dummies; ii) daily sales of products that are in categories that did not receive the price reduction; iii) the prices of both goods before the intervention, as they are good predictors of sales. For all of these covariates, the assumption is likely to be satisfied. Indeed, we modify the price of store brands to remain constant after the intervention date. Note that the inclusion of the actual daily price after the reduction would have violated this assumption.

**Assumption 7** We assume that the assignment is individualistic — the treatment allocation of each pair has no bearing on others — and informed only by past sales performances and past covariates or, at most, by general beliefs on the sales evolution under active treatment.

## 4.2 Causal estimands

In a panel setting, the number of causal estimands increases substantially, as any contrast of potential outcomes has a causal interpretation. In this section, we develop three classes of causal effects; for each, we can define a *point* 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 a *temporal average* effect (i.e., a normalization of the cumulative effect). Our primary objective is to obtain an estimate for each group. Under Assumption 5 we can drop the subscript  $j$  that identifies the group and focus on analyzing each multivariate time series separately. Even though the goal of the empirical analysis is to estimate the heterogeneous effect on each pair of products, the definitions below are given for a general multivariate case where units define groups of size  $d > 2$ . Furthermore, under Assumption 1 we can restrict to  $t > t^*$  and drop the subscript  $t$  from the treatment assignment.

---

<sup>10</sup>Also, every store brand has its own specific direct competitor, thus the possibility that the same good belongs to more than one pair is ruled out.



**Definition 3 (General effects)** For  $\mathbf{w}, \mathbf{w}' \in \{0, 1\}^d$ , the general causal effect of an assignment  $\mathbf{w}$  compared to an alternative assignment  $\mathbf{w}'$  at a given time point is

$$\begin{aligned}\boldsymbol{\tau}_t(\mathbf{w}; \mathbf{w}') &= (\tau_t^{(1)}(\mathbf{w}; \mathbf{w}'), \dots, \tau_t^{(d)}(\mathbf{w}; \mathbf{w}')) \\ &= (\mathbf{Y}_t^{(1)}(\mathbf{w}) - \mathbf{Y}_t^{(1)}(\mathbf{w}'), \dots, \mathbf{Y}_t^{(d)}(\mathbf{w}) - \mathbf{Y}_t^{(d)}(\mathbf{w}')) = (\mathbf{Y}_t(\mathbf{w}) - \mathbf{Y}_t(\mathbf{w}')).\end{aligned}\quad (13)$$

The cumulative general causal effect at time point  $t > t^*$  is

$$\Delta_t(\mathbf{w}; \mathbf{w}') = \sum_{s=t^*}^t \boldsymbol{\tau}_s(\mathbf{w}; \mathbf{w}'). \quad (14)$$

The temporal average general causal effect at time point  $t$  is

$$\bar{\boldsymbol{\tau}}_t(\mathbf{w}; \mathbf{w}') = \frac{1}{t - t^*} \sum_{s=t^*+1}^t \boldsymbol{\tau}_s(\mathbf{w}; \mathbf{w}') = \frac{\Delta_t(\mathbf{w}; \mathbf{w}')}{t - t^*}. \quad (15)$$

**Example 2** In our empirical application we have a bivariate outcome, with  $d = 2$ , and  $\{0, 1\}^2 = \{(0, 0), (0, 1), (1, 1), (1, 0)\}$ . Then,  $\boldsymbol{\tau}_t((1, 0); (0, 0)) = \mathbf{Y}_t(1, 0) - \mathbf{Y}_t(0, 0)$  is the change in units sold when only the store brand gets a discount compared to the alternative scenario where none of them receive a discount.

The general effects can be viewed as the multivariate versions of the pointwise effects defined in Section 3.2. We can combine the general causal effects to define the marginal causal effect that captures the impact of changing a single unit within a group across all possible treatment combinations the group could have received.

**Definition 4 (Marginal effects)** Let  $\mathcal{A}_i \subset \{0, 1\}^d$  be the subset of all treatment paths  $\mathbf{w}$  such that  $w^{(i)} = 1$  and  $\mathcal{B}_i \subset \{0, 1\}^d$  be the subset of all treatment paths  $\mathbf{w}'$  such that  $w^{(i)} = 0$ . The marginal causal effect on the  $i^{\text{th}}$  series at a given time point is the sum of the  $i^{\text{th}}$  elements of  $\boldsymbol{\tau}_t(\mathbf{w}; \mathbf{w}')$  computed across all the possible realizations in  $\mathcal{A}_i \times \mathcal{B}_i$ ,

$$\tau_t(i) = \sum_{(\mathbf{w}, \mathbf{w}') \in \mathcal{A}_i \times \mathcal{B}_i} \tau_t^{(i)}(\mathbf{w}; \mathbf{w}'). \quad (16)$$

The cumulative marginal causal effect at time point  $t > t^*$  is

$$\Delta_t(i) = \sum_{s=t^*+1}^t \tau_s(i). \quad (17)$$

The temporal average marginal causal effect at time point  $t$  is

$$\bar{\tau}_t(i) = \frac{1}{t - t^*} \sum_{s=t^*+1}^t \tau_s(i) = \frac{1}{t - t^*} \Delta_t(i). \quad (18)$$

Now, let  $N_{\mathcal{A}_i \times \mathcal{B}_i}$  denote the total number of possible assignments in  $\mathcal{A}_i \times \mathcal{B}_i$ ; the **mean marginal causal effect** can be defined as,

$$\tau_t(i, N_{\mathcal{A}_i \times \mathcal{B}_i}) = \frac{1}{N_{\mathcal{A}_i \times \mathcal{B}_i}} \sum_{(\mathbf{w}, \mathbf{w}') \in \mathcal{A}_i \times \mathcal{B}_i} \tau_t^{(i)}(\mathbf{w}; \mathbf{w}'). \quad (19)$$

The cumulative and temporal average mean marginal effects can be then derived as in equations (17) and (18).

The marginal causal effect captures the impact of assigning the  $i^{\text{th}}$  unit to treatment, averaged over all possible interventions that could have been applied to the other units. Thus, the marginal effect can be considered an extension to the time series setting of the average distributional shift effect in Sävje et al. (2020), with one main difference: the average distributional shift effect is averaged across units whereas the marginal effect is individual-specific and, in its temporal average version, it is averaged across times. We could make this effect slightly more general by introducing non-stochastic weights in the summation to up-weight or down-weight particular treatment combinations. However, this makes the notation somewhat more cumbersome without adding new insights.

**Example 3** Suppose that we are interested in estimating the marginal effect of the active treatment on the store brand, then  $\mathcal{A} = \{(1, 0), (1, 1)\}$ ,  $\mathcal{B} = \{(0, 0), (0, 1)\}$ , and  $\mathcal{A} \times \mathcal{B} = \{(1, 0)(0, 0); (1, 0)(0, 1); (1, 1)(0, 0); (1, 1)(0, 1)\}$ . Furthermore, denoting the store and competitor brand with  $i = s$  and  $i = c$ ,  $\tau_t(\mathbf{w}; \mathbf{w}') = (\tau_t^{(s)}(\mathbf{w}; \mathbf{w}'), \tau_t^{(c)}(\mathbf{w}; \mathbf{w}'))$  and hence,  $\tau_t(s) = \tau_t^{(s)}((1, 0); (0, 0)) + \tau_t^{(s)}((1, 0); (0, 1)) + \tau_t^{(s)}((1, 1); (0, 0)) + \tau_t^{(s)}((1, 1); (0, 1))$ . Finally, the mean marginal effect of the active treatment on the store brand is  $\tau_t(s, 4) = 1/4 \cdot \tau_t(s)$ .

A special case of the general causal effect is the conditional causal effect that fixes the treatments for all units within the group except for the  $i^{\text{th}}$  unit.

**Definition 5 (Conditional effects)** For  $\mathbf{w} \in \{0, 1\}^{d-1}$ , the conditional causal effect at a given time point is the effect of assigning the  $i^{\text{th}}$  series to treatment as opposed to control, fixing the treatments of the other series to equal  $\mathbf{w}$

$$\tau_t^\dagger(i, \mathbf{w}) = \mathbf{Y}_t((w_1, \dots, w_{i-1}, 1, w_i, \dots, w_{d-1})) - \mathbf{Y}_t((w_1, \dots, w_{i-1}, 0, w_i, \dots, w_{d-1})). \quad (20)$$

Similar to the marginal and mean marginal causal effects, we can define the cumulative and temporal average conditional causal effect at time point  $t > t^*$ .

The conditional effect can also be seen as the generalization to the time-series setting of the assignment-conditional unit-level treatment effect in [Sävje et al. \(2020\)](#).

**Example 4** *The general effect defined in Example 2 is already a conditional effect, since it measures the impact of the permanent reduction on the store brand given that the competitor is always assigned to control. However, we may also be interested in the conditional effect of the permanent price reduction on the store brand when the competitor brand is discounted as well, that is,  $\mathbf{w} = (1, 1)$ ,  $\mathbf{w}' = (0, 1)$  and  $\tau_t^\dagger(s, (1, 1)) = \mathbf{Y}_t(1, 1) - \mathbf{Y}_t(0, 1)$ .*

### 4.3 Multivariate Bayesian Structural Time Series

We now outline our approach for estimation and inference of the causal effects defined in Section 4.2. 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 way to deal with uncertainty. Flexibility comes from our ability to add sub-components (e.g., trend, seasonality, and cycle) that encapsulate the characteristics of a data set. Uncertainty is quantified through the posterior distribution, which we derive and provide a sampling algorithm.

Estimation in this approach has two steps: first, we estimate an MBSTS model for each pair in the period up to the intervention,  $t \in \{1, \dots, t^*\}$ ; then, we estimate the target causal effects by forecasting the unobserved potential outcomes in the period following the intervention,  $t \in \{t^* + 1, \dots, T\}$ . This section mirrors the two steps by first describing the model priors and posterior inference followed by the forecast and inference step.

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). Throughout this section,  $\mathbf{Y} \sim \mathcal{N}(\mathbf{M}, \mathbf{\Lambda}, \mathbf{\Sigma})$  indicates that  $\mathbf{Y}$  follows a matrix normal distribution with mean  $\mathbf{M}$ , row variance-covariance matrix  $\mathbf{\Lambda}$  and column variance-covariance matrix  $\mathbf{\Sigma}$ . 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  $\mathcal{IW}(\nu, \mathbf{S})$  will denote an Inverse-Wishart distribution with  $\nu$  degrees of freedom and scale matrix  $\mathbf{S}$ .

To improve readability of the model equations, we use  $\mathbf{Y}_t$  to indicate  $\mathbf{Y}_t(\mathbf{w})$ . We resume the usual notation in Section 4.3.5.

### 4.3.1 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, seasonal, or cycle). We also allow for covariates’ presence to increase the counterfactual series’ prediction accuracy in the absence of intervention. The second one is the “state equation” that determines the state vector’s evolution across time.

$$\underbrace{\mathbf{Y}_t}_{1 \times d} = \underbrace{\mathbf{Z}_t}_{1 \times m} \underbrace{\boldsymbol{\alpha}_t}_{m \times d} + \underbrace{\mathbf{X}_t}_{1 \times P} \underbrace{\boldsymbol{\beta}}_{P \times d} + \underbrace{\boldsymbol{\varepsilon}_t}_{1 \times d}, \quad \boldsymbol{\varepsilon}_t \sim N_d(\mathbf{0}, H_t \boldsymbol{\Sigma})$$

$$\underbrace{\boldsymbol{\alpha}_{t+1}}_{m \times d} = \underbrace{\mathbf{T}_t}_{m \times m} \underbrace{\boldsymbol{\alpha}_t}_{m \times d} + \underbrace{\mathbf{R}_t}_{m \times r} \underbrace{\boldsymbol{\eta}_t}_{r \times d}, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_t, \boldsymbol{\Sigma}), \quad \boldsymbol{\alpha}_1 \sim \mathcal{N}(\mathbf{a}_1, \mathbf{P}_1, \boldsymbol{\Sigma}) \quad (21)$$

Where, for all  $t \leq t^*$ ,  $\boldsymbol{\alpha}_t$  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;<sup>11</sup>  $\boldsymbol{\beta}$  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  $\mathbf{T}_t = \mathbf{1}$ ); and  $\mathbf{R}_t$  is a matrix selecting the rows of the state equation with non-zero 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 variance-covariance matrix of the dependencies between the time series by

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

$H_t$  is the variance of the observation error at time  $t$ ; to simplify 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  $\mathbf{C}_t$  allows each state disturbance to have its own  $(d \times d)$  variance-covariance matrix  $\boldsymbol{\Sigma}_r$ .<sup>12</sup> In short,

<sup>11</sup>Notice that this parametrization assumes the same set of regressors for each time series but still ensures that the coefficients are different across the  $d$  time series.

<sup>12</sup>The notation  $H_t \boldsymbol{\Sigma}$  and  $c_r \boldsymbol{\Sigma}$  allows to understand 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 4.3.3 we derive posterior distributions for  $\boldsymbol{\Sigma}_\varepsilon$  and  $\boldsymbol{\Sigma}_r$  in the general

$$\mathbf{Q} = \mathbf{C}_t \otimes \Sigma_\varepsilon = \begin{bmatrix} c_1 \Sigma & 0 & \cdots & 0 \\ 0 & c_2 \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_r \Sigma \end{bmatrix} = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_r \end{bmatrix}.$$

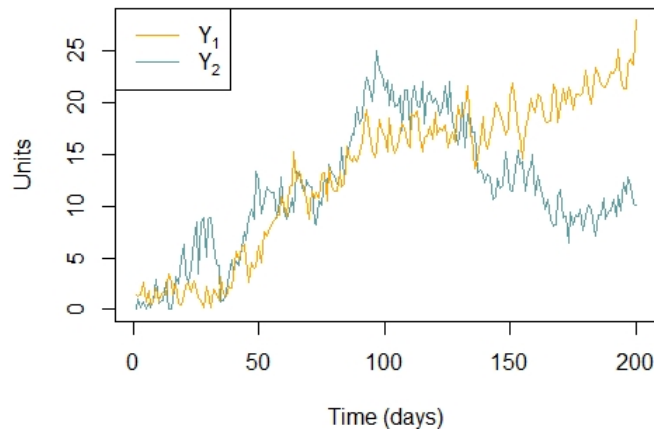
To build intuition for the different components of the MBSTS model, we find it is useful to consider an example of a simple local level model.

**Example 5** *The multivariate local level model is characterized by a trend component evolving according to a simple random walk, no seasonal is present, and both the disturbance terms are assumed to be Normally distributed.*

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t & \boldsymbol{\varepsilon}_t &\sim N_d(\mathbf{0}, H_t \Sigma) \\ \boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \boldsymbol{\eta}_{t,\mu} & \boldsymbol{\eta}_{t,\mu} &\sim N_d(\mathbf{0}, c_1 \Sigma) \end{aligned} \quad (22)$$

We can recover the general formulation outlined in (21) by setting  $\boldsymbol{\alpha}_t = \boldsymbol{\mu}_t$  and  $\mathbf{Z}_t = \mathbf{T}_t = \mathbf{R}_t = \mathbf{1}$ . Figure 9, provides a graphical representation of what a sample from this model would look like when  $d = 2$ .

**Figure 9:** The plot shows 200 observations (e.g., number of units sold of a specific item) sampled from a multivariate local level model with  $d = 2$ .




---

case of unknown  $H_t$  and  $\mathbf{C}_t$ .

### 4.3.2 Prior elicitation

The unknown parameters of Model (21) are the variance-covariance matrices of the error terms and the matrix of regression coefficients  $\boldsymbol{\beta}$ . Since we assume that both the observation and state errors are normally distributed, for their variance-covariance matrices, we choose the conjugate Inverse-Wishart distributions.

Generally, the MBSTS model can handle dynamic covariate coefficients. However, in our empirical application we believe that the relationship between covariate and the outcome is 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 many applications, 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  $\mathbf{b}_0 = \mathbf{0}$  and introducing a selection vector  $\boldsymbol{\varrho} = (\varrho_1, \dots, \varrho_P)'$  such that  $\varrho_p \in \{0, 1\}$ ,  $p \in \{1, \dots, P\}$ . Then,  $\boldsymbol{\beta}_p = 0$  when  $\varrho_p = 0$ , meaning that the corresponding row of  $\boldsymbol{\beta}$  is set to zero and that we are eliminating regressor  $X_p$  from our model. When  $\varrho_p = 1$  then  $\boldsymbol{\beta}_p \neq 0$ , meaning that we are including regressor  $X_p$  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 assume each element in  $\boldsymbol{\varrho}$  to be an independent Bernoulli distributed random variable with parameter  $\pi$ .

Indicating with  $\boldsymbol{\theta} = (\nu_\varepsilon, \nu_r, \mathbf{S}_\varepsilon, \mathbf{S}_r, \mathbf{X}_{1:t^*})$  the vector of known parameters and matrices and denoting with  $\mathbf{X}_\varrho$  and  $\mathbf{H}_\varrho$  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,

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

For the elicitation of prior hyperparameters, [Brown et al. \(1998\)](#) suggest setting  $\nu_\varepsilon = d + 2$ , which is the smallest integer value such that the expectation of  $\boldsymbol{\Sigma}_\varepsilon$  exists. We use a similar strategy for  $\nu_r$ . 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  $\rho$  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).

### 4.3.3 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. We can derive the following full conditional distributions as,

$$\boldsymbol{\beta}_\varrho | \tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{M}, \mathbf{W}, \boldsymbol{\Sigma}_\varepsilon), \quad (23)$$

$$\boldsymbol{\Sigma}_\varepsilon | \tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\varrho}, \boldsymbol{\theta} \sim \mathcal{IW}(\nu_\varepsilon + t^*, \mathbf{SS}_\varepsilon), \quad (24)$$

$$\boldsymbol{\Sigma}_r | \boldsymbol{\eta}_{1:t^*}^{(r)}, \boldsymbol{\theta} \sim \mathcal{IW}(\nu_r + t^*, \mathbf{SS}_r), \quad (25)$$

where  $\mathbf{M} = (\mathbf{X}'_\varrho \mathbf{X}_\varrho + \mathbf{H}_\varrho^{-1})^{-1} \mathbf{X}'_\varrho \tilde{\mathbf{Y}}_{1:t^*}$ ,  $\mathbf{W} = (\mathbf{X}'_\varrho \mathbf{X}_\varrho + \mathbf{H}_\varrho^{-1})^{-1}$ ,  $\mathbf{SS}_\varepsilon = \mathbf{S}_\varepsilon + \tilde{\mathbf{Y}}'_{1:t^*} \tilde{\mathbf{Y}}_{1:t^*} - \mathbf{M}' \mathbf{W}^{-1} \mathbf{M}$ ,  $\mathbf{SS}_r = \mathbf{S}_r + \boldsymbol{\eta}'_{1:t^*} \boldsymbol{\eta}_{1:t^*}^{(r)}$  and  $\boldsymbol{\eta}_{1:t^*}^{(r)}$  indicates the disturbances up to time  $t^*$  of the  $r$ -th state. Full proof of relations (23), (24) and (25) 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).

### 4.3.4 Prediction and estimation of causal effects

Given the draws from the joint posterior distribution of states and model parameters, we can use them 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.

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

$$\Pr(\mathbf{Y}_{1:t^*}^{new} | \mathbf{Y}_{1:t^*}) = \int \Pr(\mathbf{Y}_{1:t^*}^{new}, \boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*}) d\boldsymbol{\vartheta} = \int \Pr(\mathbf{Y}_{1:t^*}^{new} | \mathbf{Y}_{1:t^*}, \boldsymbol{\vartheta}) \Pr(\boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*}) d\boldsymbol{\vartheta} \quad (26)$$

$$= \int \Pr(\mathbf{Y}_{1:t^*}^{new} | \boldsymbol{\vartheta}) \Pr(\boldsymbol{\vartheta} | \mathbf{Y}_{1:t^*}) d\boldsymbol{\vartheta} \quad (27)$$

where the last equality follows because  $\mathbf{Y}_{1:t^*}^{new}$  is independent of  $\mathbf{Y}_{1:t^*}$  conditional on  $\boldsymbol{\vartheta}$ . We can, therefore, obtain in-sample forecasts from the posterior predictive distribution by using the draws from  $\Pr(\boldsymbol{\vartheta}|\mathbf{Y}_{1:t^*})$  that were obtained through the Gibbs sampler and substitute them in the model equations (21). We typically use in-sample forecasting for performing model checking.

To predict the counterfactual time series in the absence of an intervention, we need out-of-sample forecasts. Drawing from the predictive posterior distribution is still relative straightforward, except the new samples are no longer independent of  $\mathbf{Y}_{1:t^*}$  given  $\boldsymbol{\vartheta}$ . To see this, consider the vector  $\boldsymbol{\vartheta}' = (\boldsymbol{\alpha}_{t^*+k}, \dots, \boldsymbol{\alpha}_{t^*+1}, \boldsymbol{\vartheta})$ . Then,

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

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 (21), thereby updating the states and sampling the new sequence  $\mathbf{Y}_{t^*+1}, \dots, \mathbf{Y}_{t^*+k}$ .

#### 4.3.5 Causal effect estimation

We can now estimate the causal effects defined in Section 4.2 by using the MBSTS models to predict the missing potential outcomes. In particular, we derive the posterior distribution of the general causal effect given in equation (13); the other two effects are simply functions or special cases of the general causal effect.

Let  $\Pr(\mathbf{Y}_t(\mathbf{w})|\mathbf{Y}_{t^*}(\mathbf{w}))$  and  $\Pr(\mathbf{Y}_t(\mathbf{w}')|\mathbf{Y}_{t^*}(\mathbf{w}'))$  with  $t > t^*$  be the out-of-samples draws from the posterior predictive distribution of the outcome under the treatment assignments  $\mathbf{w}, \mathbf{w}' \in \{0, 1\}^d$ . Then,

$$\Pr(\boldsymbol{\tau}_t(\mathbf{w}; \mathbf{w}')|\mathbf{Y}_{1:t^*}(\mathbf{w}), \mathbf{Y}_{1:t^*}(\mathbf{w}')) = \Pr(\mathbf{Y}_t(\mathbf{w})|\mathbf{Y}_{1:t^*}(\mathbf{w})) - \Pr(\mathbf{Y}_t(\mathbf{w}')|\mathbf{Y}_{1:t^*}(\mathbf{w}')) \quad (28)$$

is the posterior distribution of the general causal effect  $\boldsymbol{\tau}_t(\mathbf{w}; \mathbf{w}')$  and it is the difference between the posterior predictive distributions of the outcome under the two alternative treatment paths. Then, the posterior distributions of the cumulative general effect and the temporal average general effect at  $t > t^*$  can be derived from (28) as follows:

$$\Pr(\Delta_t(\mathbf{w}; \mathbf{w}')|\mathbf{Y}_{1:t^*}(\mathbf{w}), \mathbf{Y}_{1:t^*}(\mathbf{w}')) = \sum_{s=t^*+1}^t \Pr(\boldsymbol{\tau}_s(\mathbf{w}; \mathbf{w}')|\mathbf{Y}_{1:t^*}(\mathbf{w}), \mathbf{Y}_{1:t^*}(\mathbf{w}')), \quad (29)$$



$$\Pr(\bar{\tau}_t(\mathbf{w}; \mathbf{w}') | \mathbf{Y}_{t^*}(\mathbf{w}), \mathbf{Y}_{t^*}(\mathbf{w}')) = \frac{1}{t - t^*} \Pr(\Delta_t(\mathbf{w}; \mathbf{w}') | \mathbf{Y}_{t^*}(\mathbf{w}), \mathbf{Y}_{t^*}(\mathbf{w}')). \quad (30)$$

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

Notice that 28 - 30 do not require  $\mathbf{Y}_t(\mathbf{w})$  or  $\mathbf{Y}_t(\mathbf{w}')$  to be observed. However, estimation of unobserved potential outcomes other than  $\mathbf{Y}_t(0, \dots, 0)$  requires a strong set of model assumptions, and as such is often less reliable. In our application, we are mostly interested in estimating the general effect  $\hat{\tau}_t((1, 0); (0, 0)) = \mathbf{Y}_t(1, 0) - \mathbf{Y}_t(0, 0)$ , where  $\mathbf{Y}_t(1, 0)$  is the observed outcome. The marginal and the conditional effects are of secondary importance and are included in the latter part of the analysis.

In practice, to obtain reliable estimates of the causal effects, the assumed model has to adequately describe the data. We therefore recommend to check model adequacy through the use of posterior predictive checks (Rubin, 1981, 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, of the marginal and conditional effects. A detailed description of posterior predictive checks and the discussion of the frequentist properties of our estimators are given, respectively, in Appendix B.5 and B.4.

### 4.3.6 Combining results

Even though the main goal of the empirical analysis is estimating the heterogeneous effect on each cookie pair, it is possible to combine the results of all pairs and estimate an average effect. One way to accomplish this goal is through the use of meta-analysis. Indeed, as the number of time series increases, the estimation of a multivariate Bayesian model becomes computationally inefficient.

Meta-analysis is the statistical synthesis of the results obtained from multiple scientific studies and is often applied in the setting of a single study with multiple independent subgroups (Borenstein et al., 2011). For example, in a study investigating the effect of a drug, the researcher may divide the participants in different groups according to the stage of the disease; in our application, the subgroups are the cookie pairs. We then treat each pair as an independent study and follow the standard steps in a meta-analysis, described below.

One basic meta-analysis approach is computing the summary effect as a weighted average of point estimates (i.e., the results of the individual studies) with weights based on the estimated standard errors. The main caveat of this approach is the inherent dependence on the sample size: a small number of studies would result in a loss of precision of the estimated between-study variance. In such case, we can resort to a fully Bayesian meta-analysis (Smith et al., 1995; Sutton and Abrams, 2001; Sutton and Higgins, 2008). This approach is based on hierarchical Bayesian models that assume a distribution on the true effect and place suitable priors on its

hyperparameters.

The above described methodologies can be used to combine point estimates of multiple independent studies. However, by following the estimation process described in this section, we obtain a posterior distribution of the general causal effect for each analyzed cookie pair. As a result, combining the estimates of the individual pairs is a lot more intuitive.

For example, let  $\bar{\tau}_{j,t}(\mathbf{w}; \mathbf{w}')$  be the temporal average causal effect on the  $j$ -th cookie pair and assume we estimated a posterior distribution for each  $j$  as in (30). Then, we can define the summary temporal average effect across all  $j$  pairs and its posterior distribution as,

$$\bar{\tau}_t(\mathbf{w}; \mathbf{w}') = \frac{1}{J} \sum_{j=1}^J \bar{\tau}_{j,t}(\mathbf{w}; \mathbf{w}'), \quad (31)$$

$$\Pr(\bar{\tau}_t(\mathbf{w}; \mathbf{w}') | \mathbf{Y}_{1:t^*}(\mathbf{w}), \mathbf{Y}_{1:t^*}(\mathbf{w}')) = \frac{1}{J} \sum_{j=1}^J \Pr(\bar{\tau}_{j,t}(\mathbf{w}; \mathbf{w}') | \mathbf{Y}_{1:t^*}(\mathbf{w}), \mathbf{Y}_{1:t^*}(\mathbf{w}')). \quad (32)$$

In words, to combine the estimated temporal average effect of the individual cookie pairs we can directly average across their posterior distributions.

## 4.4 Simulation study

We now describe a simulation study exploring the frequentist properties of our proposed approach for correctly specified models and a misspecified model.

### 4.4.1 Design

We generate simulated data according to the following MBSTS model:

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\gamma}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t & \boldsymbol{\varepsilon}_t &\sim N_d(\mathbf{0}, H_t \boldsymbol{\Sigma}) \\ \boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \boldsymbol{\eta}_{t,\mu} & \boldsymbol{\eta}_{t,\mu} &\sim N_d(\mathbf{0}, c_1 \boldsymbol{\Sigma}) \\ \boldsymbol{\gamma}_{t+1} &= - \sum_{s=0}^{S-2} \boldsymbol{\gamma}_{t-s} + \boldsymbol{\eta}_{t,\gamma} & \boldsymbol{\eta}_{t,\gamma} &\sim N_d(\mathbf{0}, c_2 \boldsymbol{\Sigma}) \end{aligned} \quad (33)$$

Where  $\mathbf{Y}_t = (Y_1, Y_2)$  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  $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.3 \\ -0.3 & 1 \end{bmatrix}$ . We then assume a regression component formed by two covariates,  $X_{1,t} = 1 - \gamma_1 t + u_{1,t}$ , where  $\gamma_1 = 0.01$ ,  $u_{1,t} \sim N(0, 0.5)$ , and  $X_{2,t} = u_{2,t}$ , where  $u_{2,t} \sim N(2, 0.3)$  while  $\boldsymbol{\beta}$  is 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:  $\nu_\varepsilon = \nu_r = 4$ ;  $\mathbf{S}_\varepsilon = \mathbf{S}_r = 0.2 \begin{bmatrix} s_1^2 & s_1 s_2 \rho \\ s_1 s_2 \rho & s_2^2 \end{bmatrix}$ , where  $s_1^2$  and  $s_2^2$  are the sample variances of, respectively,  $Y_1$  and  $Y_2$  and  $\rho = -0.8$  is a correlation coefficient reflecting our prior belief of their dependence structure; and Zellner’s g-prior for the variance-covariance matrix of  $\beta$ .

To make our simulation close to our empirical application, we generated 1,000 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 5 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 2,000 estimated models in the pre-intervention period (one for each data set and model type), each having 1,000 draws from the resulting posterior distribution. Finally, we predicted the counterfactual series in the absence of intervention for three-time horizons, namely, after 1 month, 3 months, and 6 months from the intervention, for a total of 30,000 estimated effects (one for each data set, model type, impact size and time horizon).

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

$$\frac{|\hat{\tau}_t((1, 0); (0, 0)) - \bar{\tau}_t((1, 0); (0, 0))|}{\bar{\tau}_t((1, 0); (0, 0))},$$

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 without normalizing the bias different effect sizes are not immediately comparable. To see this, consider that a small bias for estimating a substantial effect is better than that same bias when trying to estimate a small effect.

#### 4.4.2 Results

Tables 6 reports the average interval length under M1 and M2 for all 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 and assumes a certain level of stability. Figure 10 shows the absolute

percentage errors for the first time horizon. We see, unsurprisingly, that it decreases as the effect size increases. This suggests that small effects are more difficult to detect. To confirm this claim, in Figure 11, we report the percentage of times we detect a causal effect over the 1,000 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 easily detected. This phenomenon occurs primarily 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.

Finally, Table 7 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  $Y_2$ .

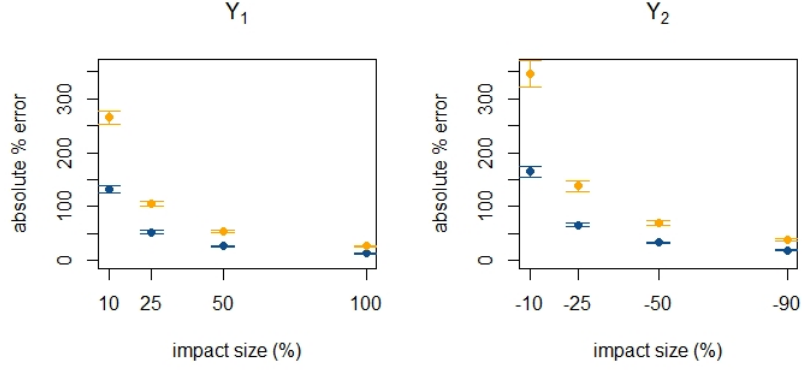
Overall, the simulation results suggest that when the model is correctly specified, the proposed approach performs well in estimating the causal effect of an intervention. Conversely, when the model is misspecified, the estimation error increases and the credible intervals do not achieve the required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

In practice, we recommend testing the adequacy of our model before performing substantive analysis by using posterior predictive checks. Figures 12 and 13 provide examples results obtained under M1 and Figures 14 and 15 show the posterior predictive checks under both M1 and M2. From their observation we can immediately see that M1 yields a better approximation of the empirical density of the simulated data and lower residual autocorrelation than M2.

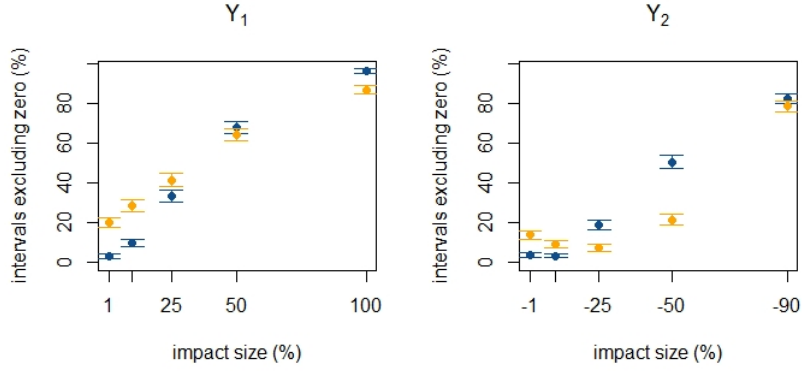
**Table 6:** Length of credible intervals around the temporal average general effect,  $\bar{\tau}_t((1,0);(0,0))$  estimated under M1 and M2 for each effect size and time horizon.

		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%, -1%)	20.93	21.10	27.62	27.80	46.58	46.28
	(+10%, -10%)	21.34	21.37	28.09	28.15	46.98	46.89
	(+25%, -25%)	21.33	21.30	28.18	28.09	47.11	46.97
	(+50%, -50%)	21.30	21.31	28.11	28.11	47.02	46.91
	(+100%, -99%)	21.38	21.25	28.24	28.06	47.12	46.90
M2	(+1%, -1%)	30.39	30.39	30.40	30.41	30.48	30.47
	(+10%, -10%)	30.48	30.48	30.50	30.50	30.57	30.58
	(+25%, -25%)	30.48	30.46	30.51	30.49	30.60	30.58
	(+50%, -50%)	30.45	30.43	30.47	30.46	30.55	30.54
	(+100%, -99%)	30.49	30.49	30.52	30.51	30.60	30.57

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



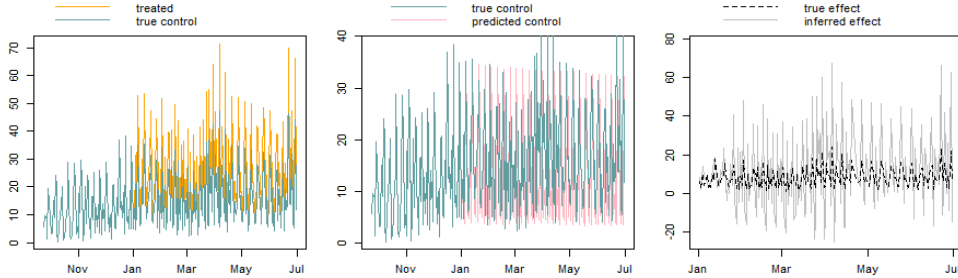
**Figure 11:** Average proportion of credible intervals excluding zero ( $\pm 2$  s.e.m) at the first time horizon under M1 (blue) and M2 (orange) for all impact sizes.



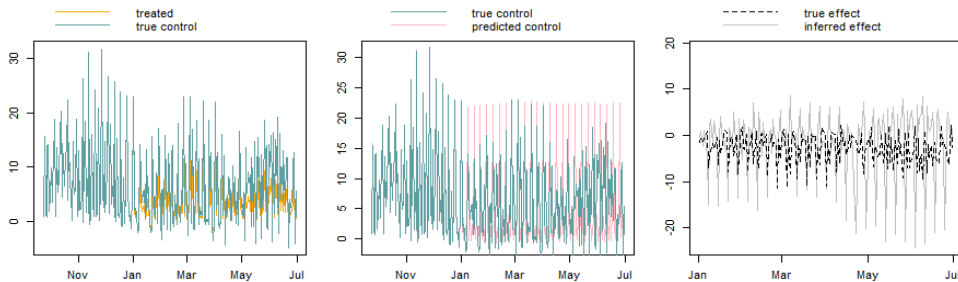
**Table 7:** Interval coverage under M1 and M2 for each effect size and time horizon.

		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%, -1%)	96.0	95.0	96.1	95.3	96.0	96.3
	(+10%, -10%)	95.9	94.9	96.0	95.2	95.9	96.3
	(+25%, -25%)	96.0	95.0	96.0	95.3	96.0	96.2
	(+50%, -50%)	96.1	94.9	96.1	95.2	96.1	96.2
	(+100%, -99%)	95.9	95.0	96.1	95.3	96.0	96.3
M2	(+1%, -1%)	86.8	88.4	85.5	87.2	82.0	84.6
	(+10%, -10%)	87.0	88.5	85.7	87.3	82.1	84.7
	(+25%, -25%)	87.0	88.6	85.7	87.3	82.1	84.7
	(+50%, -50%)	86.9	88.6	85.6	87.3	82.0	84.7
	(+100%, -99%)	86.9	88.6	85.7	87.3	82.1	84.6

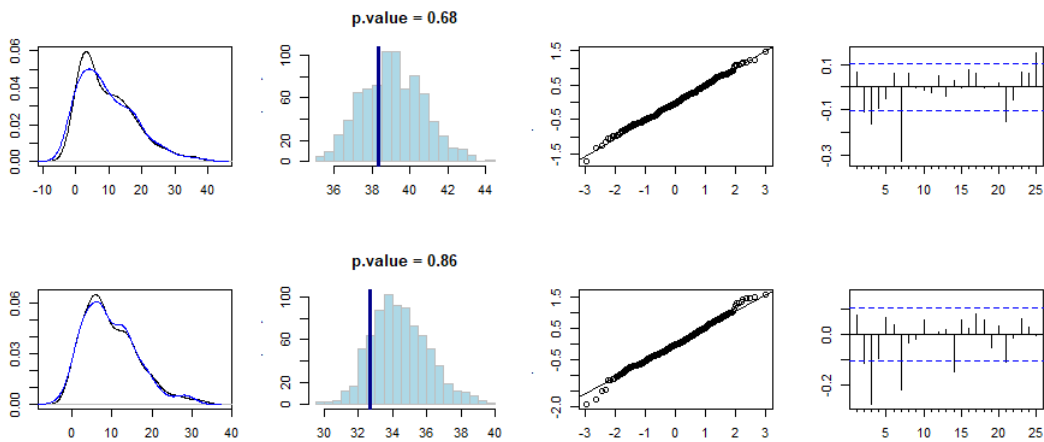
**Figure 12:** For one of the simulated data sets at 6-month horizon, the figure plots: (a) simulated time series assuming an effect size of +50% (orange) vs true counterfactual series generated under model (33) (blue); (b) true counterfactual vs predicted counterfactual series under M1; (c) true effect (black dashed line) vs the inferred effect under M1.



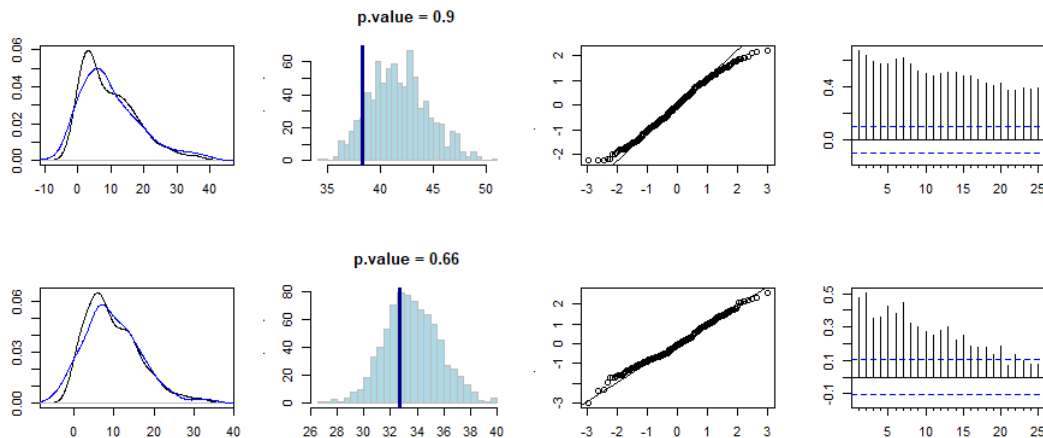
**Figure 13:** For one of the simulated data sets at 6-month horizon, the figure plots: (a) simulated time series assuming an effect size of -50% (orange) vs true counterfactual series generated under model (33) (blue); (b) true counterfactual vs predicted counterfactual series under M1; (c) true effect (black dashed line) vs the inferred effect under M1.



**Figure 14:** Posterior predictive checks under M1 for  $Y_1$  (first row) and  $Y_2$  (second row) for one of the simulated data sets. Starting from the left: i) density of observed data (black) plotted against the posterior predictive mean (blue); ii) observed maximum compared to the distribution of the maximum from the posterior draws; iii) Normal QQ-Plot of standardized residuals; iv) autocorrelation function of standardized residuals.



**Figure 15:** Posterior predictive checks under M2 for  $Y_1$  (first row) and  $Y_2$  (second row) for one of the simulated data sets. Starting from the left: i) density of observed data plotted against the posterior predictive mean; ii) observed maximum compared to the distribution of the maximum from the posterior draws; iii) Normal QQ-Plot of standardized residuals; iv) autocorrelation function of standardized residuals.



## 4.5 Empirical analysis

We now describe the results of our empirical application where we analyze a marketing campaign run by an Italian supermarket chain in its Florence’s stores. The campaign consisted of introducing a permanent price reduction on a selected subset of store brands. The main goal of the policy change 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.

### 4.5.1 Data & methodology

Our data consists of daily sales data 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. Unlike the empirical application in Section 3.5, to further reduce Sundays’ effects, this time we focus on hourly average sales.

As an example, Figure 24 shows the time series of the average number units sold per hour by one pair of cookies, their price, and the autocorrelation function. The plots show a strong weekly seasonal pattern<sup>13</sup>. 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

<sup>13</sup>The same plots for all the remaining store and competitor brands are provided in Appendix A.2.

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 from the data that the intervention on the store brands has influenced the competitor cookies’ prices policy. Indeed, all competitor brands (with the exception of brand 10) received a temporary promotion matching the time of the intervention, suggesting that competitors may have reacted to the new policy.<sup>14</sup>

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 regressors include: 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; a set of synthetic controls selected among one category that did not receive active treatment (e.g., wine sales). 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 6. As a result, the unit prices can not be part of our models; nevertheless, they are important drivers of sales, especially during promotions (Neslin et al., 1985; Blattberg et al., 1995; Pauwels et al., 2002). We solved this issue by using a modified 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 what would have happened in the absence of intervention).

Finally, to speed up computations, the set of synthetic controls is selected in two steps: first, we select the best ten matches among the 260 possible control series in the “wines” category by dynamic time warping;<sup>15</sup> 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 4.3.4, we predicted 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.

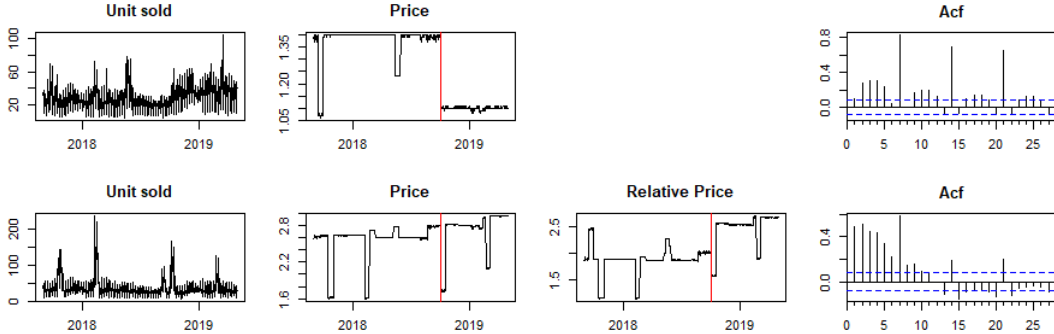
---

<sup>14</sup>See Figure 33 in Appendix A.2.

<sup>15</sup>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, contiguous set of matrix elements satisfying some conditions. 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).



**Figure 16:** Time series of unit sold daily, price per unit and autocorrelation function for the 6<sup>th</sup> pair of cookies (i.e., store brand 6 and the corresponding competitor brand). For the competitor, the relative price plot shows the ratio between its unit price and the price of the store brand.



#### 4.5.2 Results

We now present the results for the best MBSTS model with both a trend and seasonality component. Our posterior predictive checks selected this model, see Appendix A.2 for the details and for a description of the other models tried. Convergence diagnostics are provided in Appendix B.7.

The estimates of the temporal average general effect are reported in Table 8, which reveals 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 17 plots the general effect  $\hat{\tau}_t((1,0);(0,0))$  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.2 for additional plots.

Overall, these results suggest that the policy change had a minor impact on the store brands' sales. Furthermore, since we do not detect an effect after the first month, it 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 company 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 not effective. This result is robust to different prior assumptions, see Appendix B.6 for a detailed sensitivity analysis. In

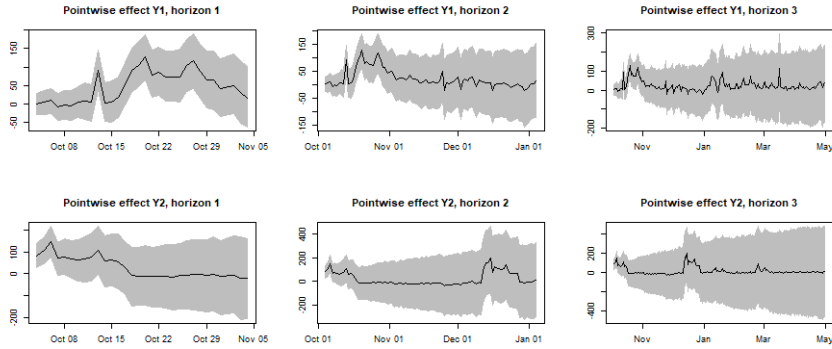
particular, we obtain similar results when, instead of using the individual prices, we include among the predictors the difference in price or the price ratio between the store and competitor brand (see Tables 17 and 18 in Appendix A.1).

As discussed in the introduction, we could have analyzed the data by aggregating the sales of store and competitor brands and treating each aggregate as a univariate time series. This procedure, however, 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 9 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 9 reports the summary temporal average effect on all cookie pairs obtained by combining the individual estimates with a meta-analysis, as described in Section 4.3.6. 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 a univariate analysis we are not able to 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, whereas the meta-analysis shows that the effect on competitor brands is not significant. Overall, despite a similar result for the tenth pair, however, we would have reached wrong conclusions for pairs 4,7 and 8, and we would have reported the misleading finding of an overall positive impact on the sales of store-competitor aggregate.

To further illustrate all the different types of effects that it is possible to estimate in a multivariate setting, we estimated the mean marginal effect and the conditional effect  $\hat{\tau}_t((1, 1); (0, 1))$ . Ultimately, we found three significant marginal effect on the sales of store brands but there is no evidence that the new price policy has had an effect in a scenario where both cookies in pairs are treated compared to the scenario where only the competitor brand is treated. The results are given in Table 15 and Table 16 in Appendix A.1.

**Figure 17:** General causal effect of the permanent price reduction on the fourth store-competitor pair at 1 month, 3 months and 6 months after the intervention.



**Table 8:** Posterior mean and 95% credible intervals of the temporal average general causal effects of the new price policy on the ten store (s) - competitor (c) pairs computed at three time horizon. In this table,  $\hat{\tau}_t$  stands for the general effect  $\hat{\tau}_t((1, 0), (0, 0))$ . There is evidence of a causal effect when the credible intervals do not include zero.

		<i>Time horizon:</i>								
Pair		1 month			3 months			6 months		
		$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	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

**Table 9:** Univariate temporal average causal effect ( $\hat{\tau}_t$ ) 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. There is evidence of a causal effect when the credible intervals do not include zero.

	<i>Time horizon:</i>								
	1 month			3 months			6 months		
	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	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

## 4.6 Discussion

In this section we presented the novel approach CausalMBSTS to estimate causal effects in panel settings with interference and multiple treated units.

Our approach starts from the discussion of the assumptions, where we addressed the issue of interference between units by relying on the partial temporal no-interference assumption. Then, we introduced three classes of estimands focusing on the heterogeneous causal effect and proposed to estimate them by using Multivariate Bayesian Structural Time Series to forecast the group outcome in the absence of intervention. Finally, we tested our approach on a simulation study, and then we used it to re-analyze the permanent price reduction run by the Italian supermarket chain.

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, it allows variable selection (via the addition of a spike-and-slab prior) and the underlying distributional assumptions can be tested in a very natural way by posterior inference.

## 5 Causal effect of multiple interventions

The estimation of causal effects in time series settings under the RCM is notoriously a challenging task, due to added complications arising from serial dependence and seasonality. Furthermore, it becomes even more complex in the context of observational studies where the analyst has no control on the assignment mechanism, since the tools commonly used in randomization-based inference are not available. To complicate things even further, we might be asked to evaluate the effect of an intervention occurring on multiple units (panel setting) or the effect of multiple interventions (multi-intervention setting).

In previous Sections 3 and 4 we addressed the issue of an intervention occurring in a panel setting. First, we conducted inference with a novel approach, C-ARIMA, to estimate the heterogeneous effect on each statistical unit under the temporal no-interference assumption; then, we grouped the units on the basis of shared characteristics and modeled their dependence structure with MBSTS models, thereby explicitly accounting for the interference within each group.

We now address the multi-intervention setting: to the best of our knowledge, this is the first attempt to estimate the heterogeneous casual effect of multiple interventions occurring on a time series in the context of observational studies.

We begin by introducing new classes of causal estimands. Then, to perform inference, we use the methodology derived in Section 3.3, thereby extending C-ARIMA to estimate causal effects in a multi-intervention setting. This work is motivated by the analysis of the first two regulated Bitcoin futures launched in December 2017 with the aim of investigating their possible impact on Bitcoin volatility.

This section is structured as follows: in Section 5.1 we present the background of the empirical application; the new causal estimands and their estimators are described, respectively, in Sections 5.2 and 5.3; finally, Section 5.4 illustrates the results of our empirical study.

### 5.1 Background

Bitcoin is a peer-to-peer payment system created in 2009 under the pseudonym of Satoshi Nakamoto (Nakamoto, 2008). In contrast to fiat money relying on central banks and intermediaries, Bitcoin is decentralized, meaning that its value is not backed by any central bank. In short, transaction data are recorded in *blocks*, each containing a reference (the hash) to the previous ones, thereby forming a chain know as the *blockchain*. Transactions are validated through the so-called *mining* process, which requires to find the hash by solving a time consuming cryptographic problem. Every time this happens, the new validated block is added to the chain and miners are rewarded with new Bitcoins. Thus, essentially the blockchain constitutes an electronic public ledger of validated transactions, which is stored and updated on miner's

computers (the network nodes) to avoid double spending problems and frauds. Mining is the only way new Bitcoins are introduced in the system and their supply is limited by design to a maximum of 21 million units that will be reached by 2140. More specifically, an algorithm adjusts the difficulty of the numerical problem based on the network performance, so that new blocks are generated every 10 minutes; in addition, the number of generated Bitcoins is halved by design every 210,000 blocks (approximately 4 years). Thus, it can be estimated that by 2140 miners' reward will be roughly zero.

Bitcoins are traded on multiple exchanges, the top A-rated ones by daily volume being Binance, Coinbase and Liquid ([CryptoCompare, 2020b](#)).<sup>16</sup> At the time of writing (December 2020, timestamp 1608940800) Bitcoin is trading at \$25,503 beating its record-high of \$19,345 reached on December 16, 2017 (in 2010 1BTC was valued \$0.05).

During this time of extreme price fluctuations, a major event that affected the Bitcoin network is the failure of Mt. Gox, the leading Bitcoin exchange until February 28, 2014 when it filed for bankruptcy protection after announcing a theft of about 850,000 BTC (the equivalent of \$21.7B as of today price) following a security breach ([McLannahan, 2014](#); [Cermak, 2017](#)). Nonetheless, in 2017 Bitcoin started to rally and it is now breaking all previous records. The growing interest from investors toward cryptocurrencies, probably contributed to the decision by two major derivative exchanges of starting a regulated derivative market for Bitcoin futures. Specifically, on August 2, 2017 CBOE announced its partnership with Gemini Trust Company to use Gemini's Bitcoin market data in the creation of derivatives products for trading; the future was then released on December 10. Meanwhile, CME announced the launch of their contract becoming effective starting from December 18. Since then, having seen its market share quickly eroded by the new future ([Baydakova, 2019](#)), CBOE has stopped listing additional Bitcoin futures for trading and after nine months from the release of the first contract, also the Intercontinental Exchange (ICE) started offering Bitcoin derivatives. Table 10 summarizes the major events occurring in Bitcoin futures' history up to December 2017.

In our empirical analysis we focus on the first two contracts launched by CBOE and CME with the aim of estimating their effect on Bitcoin volatility before expiration.<sup>17</sup> Moreover, they are sufficiently close in time so that we can safely rely in the assumed model specification. Conversely, if two or more interventions are distant in time, some variables originally unrelated to the outcome might become relevant after the first intervention and the model would not fit well to the new observations. Thus, including independent variables can improve the fit and

---

<sup>16</sup>In their most recent Exchange Benchmark Report released in July 2020, CryptoCompare rated 165 exchanges according to 68 qualitative and quantitative metrics. They show that existing metrics such as volume or liquidity can be easily manipulated (e.g. volumes can be inflated through strategies such as trading competitions, airdrops and transaction fee mining) and thus are inadequate to reflect the reliability of the trading venue. For further details regarding these strategies and the rating methodology see [CryptoCompare \(2020a\)](#)

<sup>17</sup>The first CBOE contract expired on January 17, 2018, while the first CME contract expired on January 26, 2018. Further details in [Zuckerman \(2018\)](#).

the predictive performance of the assumed model, provided that they are linked to the outcome and not influenced by the intervention.

In our empirical analysis the outcome variable is Bitcoin volatility; more specifically, based on available data, we used the Garman-Klass volatility proxy, which is calculated from intraday Bitcoin prices. Thus, to select suitable covariates, we follow the existing literature on Bitcoin price drivers. Early works evidence association between Bitcoin price and variables such as the exchange-trade ratio (i.e. ratio between trade volume and transaction volume), hash rate, Google and Wikipedia queries and the Shanghai index (Kristoufek, 2013, 2015; Bouoiyour and Selmi, 2015). In a later study, Li and Wang (2017) find a strong association in the later market (after Mt. Gox failure) between the BTC-USD rate and key economic fundamentals (US interest rate, USD money supply, Bitcoin money supply and transaction volume); they also find that technology factors such as mining difficulty and public recognition influence Bitcoin price only in the early market (before Mt. Gox failure). Similarly, Ciaian et al. (2016) evidence that market forces of Bitcoin supply and demand are strongly related to Bitcoin price changes while proxies of public recognition (i.e. views on Wikipedia, new posts and new members on Bitcoin forums) are mainly associated with BTC-USD rate in the early market.

**Table 10:** Major events in Bitcoin futures' history up to December 2017

Date	Event
2017/08/02	CBOE announces launch of Bitcoin futures by 2017-Q4
2017/10/31	CME announces launch of Bitcoin futures by 2017-Q4
2017/12/01	CME announces launch of futures on Dec 18
2017/12/04	CBOE announces launch of futures on Dec 10
2017/12/10	Launch of CBOE Bitcoin futures
2017/12/18	Launch of CME Bitcoin futures

We now review the set of assumptions in the multi-intervention setting and we discuss them in our empirical context.

**Assumption 2** In our application, the statistical unit is Bitcoin cryptocurrency and the main goal is investigating the effect on volatility generated by the launch of Bitcoin futures by two major regulated exchanges. Since both exchanges disclosed their plans to develop Bitcoin futures, we have two types of interventions: i) the announcements made by the exchanges about the upcoming futures; ii) the actual introduction of the two contracts. The latter qualify as persistent interventions. Indeed, even though the exchange can withdraw its future from the market at any time, the future would trade until its expiration date, which is standardized and set up in advance. Furthermore, since the two contracts have different characteristics (underlying spot price, expiration dates, contract units) the main goal is estimating their heterogeneous effects on volatility.

**Assumption 3** Today volatility may be impacted by previous announcements and by the trading activity of existing futures, but we exclude any influence arising from new contracts and market news that have yet to be announced. In other words, we are ruling out the possibility that market participants have access to privileged information.

**Assumption 6** Bitcoin price might be related to several economic and technology factors such as USD money supply and mining difficulty. Including them as independent variables in the analysis may enhance prediction accuracy and, as a result, the reliability of the estimated causal effect. We select the set of covariates based on a survey of relevant literature and including only those predictors that we can safely believe to be untouched by the interventions.

**Assumption 7** Since we are conditioning on past outcomes and covariates up to the day of the launch, this assumption ensures that the estimated effect of the second contract is not confounded by the first future. Another implication of this assumption is that the effect of the first future is correctly identified up to the launch of the second future. The same applies to the announcements: conditioning on the past, the effect that we observe (if any) at the announcement date is due to that specific announcement and not to past or future ones.

## 5.2 Causal estimands

We can now define the causal estimands of interest. In particular, we provide definitions for three classes of causal effects: the general, the contemporaneous and the pointwise causal effects. However, as detailed below, in the context of observational studies the estimation of the general effect is troubled; thus, in Section 5.3.2 we present estimators for the contemporaneous and the pointwise effects alone.

**Definition 6 (General effects)** *For the two treatment paths  $w_{1:t}$  and  $w'_{1:t}$  the general causal effect at time  $t$  is,*

$$\tau_t(w_{1:t}; w'_{1:t}) = Y_t(w_{1:t}) - Y_t(w'_{1:t}). \quad (34)$$

*The cumulative effect of the paths  $w_{1:t}$  and  $w'_{1:t}$  is the sum of the general causal effects up to time  $t$ ,*

$$\Delta_t(w_{1:t}; w'_{1:t}) = \sum_{s=1}^t \tau_s(w_{1:s}; w'_{1:s}). \quad (35)$$

*The temporal average effect of the paths  $w_{1:t}$  and  $w'_{1:t}$  at time  $t$  is,*



$$\bar{\tau}_t(\mathbf{w}_{1:t}; \mathbf{w}'_{1:t}) = \frac{1}{t} \sum_{s=1}^t \tau_s(\mathbf{w}_{1:s}; \mathbf{w}'_{1:s}) = \frac{\Delta_t(\mathbf{w}_{1:t}; \mathbf{w}'_{1:t})}{t}. \quad (36)$$

**Example 6** *Considering the two paths  $\mathbf{w}_{1:t} = (1, 0)$  and  $\mathbf{w}'_{1:t} = (0, 1)$  in Figure 1 and assuming that they describe the announcements made by the exchanges, the general effect  $\tau_1(1; 0) = Y_1(1) - Y_1(0)$  compares the (log) volatility in case of announcement to the one in the absence of announcement; similarly,  $\tau_2((1, 0); (0, 1)) = Y_2(1, 0) - Y_2(0, 1)$  measures the effect on (log) volatility when the announcement is followed by control compared to the opposite scenario.*

This class of effects is analogous to the the general effect defined in Section 4.2, with the only difference that this one is referred to a univariate multi-intervention setting. Notice that if the treatment was randomly allocated at any point in time, under a sharp null hypothesis of temporal no interference we could have derived the exact distribution of the cumulative and temporal average effects (Bojinov and Shephard, 2019). However, since in our application the treatment is not random, the only effects that we could estimate are  $\tau_1(1; 0)$ ,  $\tau_2((1, 0); (0, 0))$  and  $\tau_3((1, 0, 1); (0, 0, 0))$ . Indeed, the remaining effects require the estimation of the outcomes in presence of active treatment (e.g.,  $Y_2(0, 1)$  in the above example).

To solve this issue, we define a second class of estimands that computes the instant effect of a treatment conditioning on the observed treatment path up to that time point.

**Definition 7 (Contemporaneous effects)** *Indicating with  $\Lambda = \{t_1, \dots, t_M\}$  the subset of time points at which the active treatment is administered, the contemporaneous causal effect of the  $m$ -th treatment at time  $t_m \in \Lambda$  conditioning on the observed treatment path  $\mathbf{w}_{1:t_m-1}^{obs}$  is,*

$$\tau^{(m)}(1; 0) = Y_{t_m}(\mathbf{w}_{1:t_m-1}^{obs}, 1) - Y_{t_m}(\mathbf{w}_{1:t_m-1}^{obs}, 0). \quad (37)$$

*Then, the aggregate contemporaneous effect of all interventions is,*

$$\Delta(1; 0) = \sum_{m=1}^M \tau^{(m)}(1; 0) \quad (38)$$

*and the average contemporaneous effect is,*

$$\bar{\tau}(1; 0) = \frac{1}{M} \sum_{m=1}^M \tau^{(m)}(1; 0) = \frac{\Delta(1; 0)}{M}. \quad (39)$$

In words, every time an intervention occurs we can estimate its contemporaneous effect by conditioning on past outcomes and comparing the observed with the counterfactual (missing) outcome at the same time. Indeed, under Assumption 7 the effect of the  $m$ -th intervention is identifiable and thus we can use the superscript  $m$  to denote its effect conditioning on past

outcomes. The sum of the  $M$  effects is the aggregate contemporaneous effect, and dividing it by the total number of interventions we obtain the average contemporaneous effect. The example below illustrate the estimands by connecting to our empirical application.

**Example 7** Assume that the solid line in Figure 1 represents the observed path of the announcements, i.e.,  $w_{1:t}^{obs} = (1, 0, 1)$ . At  $t_1 = 1$  the contemporaneous effect matches the general effect  $\tau^{(1)}(1; 0) = Y_1(1) - Y_1(0)$  and at  $t_1 + 1 = 2$  we observe control. Then, at  $t_2 = 3$  we have a second announcement and by conditioning on the observed path, we can define the contemporaneous causal effect as  $\tau^{(2)}(1; 0) = Y_3(1, 0, 1) - Y_3(1, 0, 0)$ . Thus, the aggregate contemporaneous effect is  $\Delta(1; 0) = \tau^{(1)}(1; 0) + \tau^{(2)}(1; 0)$  and the average contemporaneous effect is  $\bar{\tau}(1; 0) = \frac{1}{2}\Delta(1; 0)$ .

Finally, we can define a class of causal estimands that measures the heterogeneous effect of the  $n$ -th persistent intervention.

**Definition 8 (Pointwise effects)** Indicating with  $\Lambda = \{t_1, \dots, t_M\}$  the subset of time points at which  $M$  persistent interventions take place, the point effect at time  $t \in \{t_m, \dots, t_{m+1} - 1\}$  of the  $m$ -th persistent intervention conditioning on the observed path  $w_{1:t_m-1}^{obs}$  is

$$\tau_t^{(m)}(w; w') = Y_t(w_{1:t_m-1}^{obs}, w) - Y_t(w_{1:t_m-1}^{obs}, w'). \quad (40)$$

Thus, the cumulative pointwise effect of the  $m$ -th intervention up to time  $t$  is,

$$\Delta_t^{(m)}(w; w') = \sum_{s=t_m}^t \tau_s^{(m)}(w; w') \quad (41)$$

and the temporal average pointwise effect of the  $m$ -th intervention is,

$$\bar{\tau}_t^{(m)}(w; w') = \frac{\Delta_t^{(m)}(w; w')}{t - t_m + 1}. \quad (42)$$

Regarding Equation (40), we recall that, conditioning on previous treatments, there are only two possible potential paths in the interval  $\{t_m, \dots, t_{m+1} - 1\}$ : this is why we can ease the notation by removing the subscript on the treatment assignment, i.e.,  $w_{t_m:t} = w$  and  $w'_{t_m:t} = w'$ . In addition, notice that for  $t = t_m$ , the full class of pointwise effects collapses to the contemporaneous effects and in case of one single persistent intervention ( $M = 1$ ) it matches the pointwise effects as defined in Section 3.2.

**Example 8** Assume that CBOE introduces Bitcoin futures at time  $t_1 = 3$ , as outlined in Figure 2. The point effect at  $t_1$  is the contemporaneous effect  $\tau_{t_1}^{(1)}(1; 0) = Y_3(0, 0, 1) - Y_3(0, 0, 0)$  and the point effect at time  $t_1 + 1 = 4$  is  $\tau_{t_1+1}^{(1)}((1, 1); (0, 0)) = Y_4(0, 0, 1, 1) - Y_4(0, 0, 0, 0)$ . Then, the cumulative and the temporal average pointwise effects are, respectively,  $\Delta_4^{(1)}((1, 1); (0, 0)) = \tau_{t_1}^{(1)}(1; 0) + \tau_{t_1+1}^{(1)}((1, 1); (0, 0))$  and  $\bar{\tau}_4^{(1)}((1, 1); (0, 0)) = \frac{1}{2}\Delta_4^{(1)}((1, 1); (0, 0))$ .

In many practical situations we may be interested in the aggregate cumulative effects. For example, if a promotion is run twice on the same item, we might want to compute the total additional sales attributable to that initiative. For this purpose, we can sum the persistent effects of the  $M$  interventions.

**Definition 9 (Aggregate effects)** *For an integer  $k < \min\{t_m - t_{m-1}\}$ , the aggregate cumulative effect of  $M$  interventions is,*

$$\Delta_k(\mathbf{w}; \mathbf{w}') = \sum_{m=1}^M \sum_{h=1}^k \tau_{t_{m-1}+h}^{(m)}(\mathbf{w}; \mathbf{w}') \quad (43)$$

and the aggregate temporal average effect is,

$$\bar{\tau}_k(\mathbf{w}; \mathbf{w}') = \frac{\Delta_k(\mathbf{w}; \mathbf{w}')}{Mk} \quad (44)$$

Furthermore, by setting  $k = 1$ , we can show that the aggregate cumulative and the aggregate temporal average effects collapse, respectively, to the aggregate contemporaneous effect and the average contemporaneous effect. Thus, the contemporaneous effect can also be interpreted as a limiting case of the pointwise effect, namely, the effect of a “pulse” intervention that lasts only for one time point. As a result, in the next section, the estimators of the contemporaneous effects will be derived as limiting cases of the pointwise effects.

To give a practical interpretation of the aggregate effects, consider the above example where the effect is measured in terms of the additional sales due to a promotion run twice on the same item: the aggregate cumulative effect is the partial sum of sales up to a specific time point (e.g., total additional sales in the first week); the aggregate temporal average effect gives the sales that, on average, the promotion produced in all the replicas. Naturally, the interventions need to be comparable, e.g., if the promotions have different durations, we can still compute the aggregate cumulative effect but the aggregate temporal average effect is not defined. As a result, since in our application the effect of the CBOE future can only be computed for one week, we only estimate the aggregate contemporaneous effect of the announcements.

### 5.3 Multi-intervention C-ARIMA

In this section, we extend the C-ARIMA model to the multi-intervention setting and we use it to derive estimators for the causal effects defined in Section 5.2. To ease notation, we drop the dependence on the treatment path, i.e., we use  $Y_t$  to indicate  $Y_t(\mathbf{w}_{1:t})$  and  $\tau_t^{(m)}$  to indicate  $\tau_t^{(m)}(\mathbf{w}, \mathbf{w}')$ . We resume the usual notation in Section 5.3.2.

### 5.3.1 The model

Recall from Section 3.3.2 that the general formulation of the C-ARIMA model in a setting with a single intervention is a linear regression with seasonal ARIMA errors and the addition of an effect  $\tau_t$ ,

$$(1 - L^s)^D(1 - L)^d Y_t = \frac{\Theta_Q(L^s)\theta_q(L)}{\Phi_P(L^s)\phi_p(L)}\varepsilon_t + (1 - L^s)^D(1 - L)^d X_t' \beta + \tau_t \quad (45)$$

where  $(1 - L^s)^D$  and  $(1 - L)^d$  are the differencing operators;  $\Theta_Q(L^s)$ ,  $\Phi_P(L^s)$  are the lag polynomials of the seasonal part of the model with period  $s$  and  $\Phi_P(L^s)$  having roots all outside the unit circle. To ease notation, we can indicate with  $T(\cdot)$  the transformation of  $Y_t$  needed to achieve stationarity, i.e.  $T(Y_t) = (1 - L^s)^D(1 - L)^d Y_t$ . Notice that the same transformation is also applied to  $X_t$ . Then, we can define

$$z_t = \frac{\Theta_Q(L^s)\theta_q(L)}{\Phi_P(L^s)\phi_p(L)}\varepsilon_t$$

so that model (45) becomes,

$$S_t = T(Y_t) - T(X_t)' \beta = z_t + \tau_t.$$

Similarly, in a multi-intervention setting where  $M$  interventions take place at time points  $\Lambda \in \{t_1, \dots, t_M\}$ , under Assumption 7 the effect of the  $m$ -th intervention is identifiable for any  $t \in \{t_m, \dots, t_{m+1} - 1\}$ . Thus, at time  $t_m - 1 + k < t_{m+1}$  we have,

$$S_{t_m - 1 + k} = z_{t_m - 1 + k} + \tau_{t_m - 1 + k}^{(m)}. \quad (46)$$

Now, assume we observe  $S_t$  up to time  $t_m - 1$  and indicate with  $H_0$  the situation where  $\tau_t = 0$  for all  $t \in \{t_m, \dots, t_{m+1} - 1\}$ , namely, the  $m$ -th intervention has no effect. Then, the  $k$ -step ahead forecast of  $S_t$  under  $H_0$ , given all the information up to time  $t_m - 1$  is,

$$\hat{S}_{t_m - 1 + k} = E[S_{t_m - 1 + k} | \mathcal{I}_{t_m - 1}, H_0] = \hat{z}_{t_m - 1 + k | t_m - 1}. \quad (47)$$

Notice that  $\hat{S}_{t_m - 1 + k}$  is, by definition, the expectation of the outcome series under the null hypothesis that the  $m$ -th intervention has no effect. Thus, it can be considered an estimate of the missing potential outcomes at time  $t_m - 1 + k$  for a persistent intervention occurring at time  $t_m$ . Furthermore, the limiting case  $k = 1$  leads to the 1-step ahead forecast  $\hat{S}_{t_m}$ , which is an estimate of the missing potential outcome in the absence of intervention for the contemporaneous effect.<sup>18</sup>

---

<sup>18</sup>An alternative formulation for Equation (46) is  $S_{t_m - 1 + k} = z_{t_m - 1 + k} + \sum_{j=1}^m \tau_{t_m - 1 + k}^{(j)}$ , highlighting that the outcome at time  $t_m - 1 + k$  embeds the effect of all previous interventions. This notation still allows the  $m$ -

### 5.3.2 Causal effect estimation

Based on the C-ARIMA model for the multi-intervention setting, we can now derive estimators for the pointwise and the contemporaneous causal effects.

**Definition 10 (Pointwise effects estimators)** Denote with  $\Lambda = \{t_1, \dots, t_M\}$  the subset of time points at which  $M$  persistent interventions take place. For a positive integer  $k \in \{1, \dots, K\}$ , let  $S_{t_m-1+k}(\mathbf{w})$  be the observed time series and let  $\hat{S}_{t_m-1+k}(\mathbf{w}')$  be the  $k$ -step ahead forecast as defined in (47). Then, estimators of the point, cumulative and temporal average effects of the  $m$ -th intervention at time  $t_m - 1 + k$  are, respectively,

$$\hat{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') = S_{t_m-1+k}(\mathbf{w}) - \hat{S}_{t_m-1+k}(\mathbf{w}'). \quad (48)$$

$$\hat{\Delta}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') = \sum_{h=1}^k \hat{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') \quad (49)$$

$$\hat{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') = \frac{\hat{\Delta}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}')}{k}. \quad (50)$$

Furthermore, by setting  $k = 1$  in Equation (48) we get an estimator of the contemporaneous effect of the  $m$ -th intervention at time  $t_m$ ,

$$\hat{\tau}^{(m)}(1; 0) = S_{t_m}(1) - \hat{S}_{t_m}(0).$$

Under the assumption that, conditioning on past information, the interventions are independent, we can sum the pointwise estimators of the  $M$  interventions to get the estimators of the aggregate pointwise effects.

**Definition 11 (Aggregate effects estimators)** An estimator of the aggregate cumulative effect up to  $k$  steps from each intervention is,

$$\hat{\Delta}_k(\mathbf{w}; \mathbf{w}') = \sum_{m=1}^M \sum_{h=1}^k \hat{\tau}_{t_m-1+h}^{(m)}(\mathbf{w}; \mathbf{w}') \quad (51)$$

and an estimator of the aggregate temporal average effect is,

$$\hat{\tau}_k(\mathbf{w}; \mathbf{w}') = \frac{\hat{\Delta}_k(\mathbf{w}; \mathbf{w}')}{Mk}. \quad (52)$$

th intervention to be identifiable, since the  $k$ -step ahead forecast of  $S_{t_m-1+k}$  under  $H_0$  and conditioning on the past information set is,  $S_{t_m-1+k} - \hat{S}_{t_m-1+k} = z_{t_m-1+k} - \hat{z}_{t_m-1+k|t_m-1} + \sum_{j=1}^m \tau_{t_m-1+k}^{(j)} - \sum_{j=1}^{m-1} \tau_{t_m-1+k}^{(j)} = z_{t_m-1+k} - \hat{z}_{t_m-1+k|t_m-1} + \tau_{t_m-1+k}^{(m)}$ .

Notice that by setting  $k = 1$  in Equations (51) and (52) we obtain estimators for the total contemporaneous effect and the average contemporaneous effect.

Finally, we can show that the pointwise estimators and, by extension, the contemporaneous and the aggregate effects estimators, are unbiased. Indeed, as previously outlined, they can be derived from the pointwise estimators.<sup>19</sup>

$$\hat{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}) \sim \left[ \bar{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}'), \sigma_{\varepsilon_m}^2 \sum_{i=0}^{k-1} \psi_{i,m}^2 \right] \quad (53)$$

$$\hat{\Delta}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') \sim \left[ \Delta_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}'), \sigma_{\varepsilon_m}^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_{i,m} \right)^2 \right] \quad (54)$$

$$\hat{\bar{\tau}}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') \sim \left[ \bar{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}'), \frac{1}{k} \sigma_{\varepsilon_m}^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_{i,m} \right)^2 \right] \quad (55)$$

where  $\sigma_{\varepsilon_m}^2$  is the variance of the error terms of the C-ARIMA model estimated on the observations up to time  $t_m - 1$  and the  $\psi_{i,m}$ 's are the coefficients of a MA( $k - 1$ ), i.e. the process of the  $k$ -step ahead prediction error. The values of  $\psi_{i,m}$  are functions of the ARMA parameters ruling  $z_{t_m-1}$ . The full derivation of (53), (54) and (55) is given in Appendix B.8.

Thus, the estimation of the pointwise effect is performed in two steps. First, we estimate a C-ARIMA model up to the day preceding the persistent intervention, i.e., the launch of the two futures by CBOE and CME. Then, we use the covariates and the estimated coefficients to forecast the Garman-Klass volatility proxy up to a pre-specified time horizon; for example, we may be interested in estimating the effect on volatility after few days or few months from the launch of Bitcoin futures. The difference between the observed and the predicted outcome is the estimated pointwise effect. Similarly, we can estimate the contemporaneous effects of each announcement by conditioning to the information set up to the day before the announcement and forecasting the Garman-Klass proxy one-step ahead. If the announcements or the launch of futures had an impact on Bitcoin volatility, we would find a significant deviation from the forecasted outcome.

---

<sup>19</sup>See Appendix B.8 for the details and the full derivation of the variances.

## 5.4 Empirical analysis

### 5.4.1 Data & Methodology

Economic data for this analysis have been gathered from Bloomberg, while Bitcoin daily prices have been collected from CryptoCompare. Notice that since Bitcoin is a (crypto)currency, its price is recorded in terms of another currency, USD in our case, meaning that the Bitcoin price is the BTC-USD exchange rate (i.e. number of USD needed to buy 1 BTC). The other Bitcoin related data (i.e. the hash rate and total number of Bitcoins in circulation) have been gathered from Blockchain.com.

Since Bitcoins are traded on multiple exchanges, the quotation of the BTC-USD rate is not unique, i.e. different exchanges trade Bitcoins at different prices. As explained in [Cermak \(2017\)](#), the price varies across exchanges mainly because of different fee policies and cashout methods but those divergent standards and the slow verification process make arbitrage opportunities difficult to exploit.<sup>20</sup> Aside from the obvious economic implications, this means that the source we take our data from matters a lot: exchange-based data providers present their own quotes, whereas external data providers (e.g. Bloomberg, CryptoCompare) compute their own index, usually a weighted average of all prices across major exchanges.

The main goal of our analysis is to estimate the effect of futures trading on Bitcoin volatility, and, since the upcoming futures were announced by several press releases, we also investigate possible announcements effects. We focus on the first two contracts introduced by CBOE and CME; the analysis period spans from May 1, 2014 to January 8, 2018. The dependent variable is the natural logarithm of the BTC-USD daily volatility as proxied by the unbiased Garman-Klass estimator ([Garman and Klass, 1980](#); [Molnár, 2012](#)), computed as,

$$\begin{aligned}\hat{\sigma}_{GK}^2 &= 0.5(\ln(H) - \ln(L))^2 - (2 \ln 2 - 1)(\ln(C) - \ln(O))^2 \\ \hat{\sigma}_{GK} &= \sqrt{\hat{\sigma}_{GK}^2} \cdot 1.034\end{aligned}$$

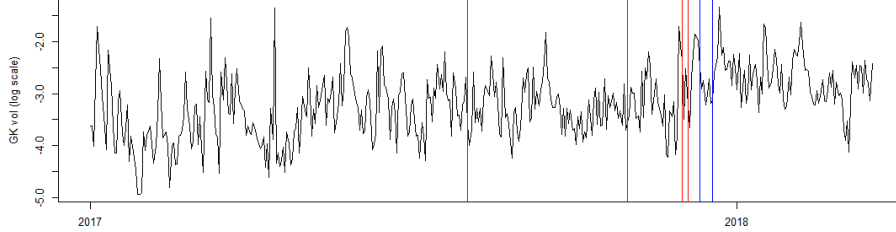
where  $H, L, C$  and  $O$  indicate, respectively, the high, low, close and open BTC-USD rate for the day. Figure 18 shows the evolution of the Garman-Klass estimator (in log scale), while Figure 19 plots the (partial) autocorrelation function and the Normal QQ-Plot.<sup>21</sup>

---

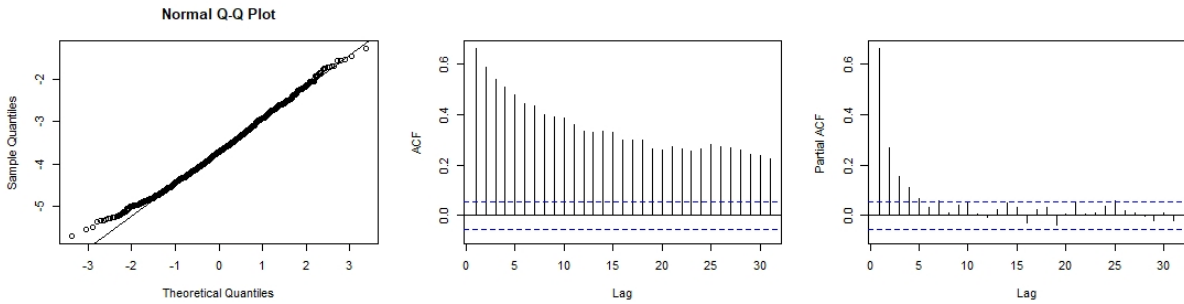
<sup>20</sup>For example, on November 20, 2019 at 4.50 p.m the price for 1 BTC was \$8,104.39 on Coinbase and \$8,146.90 on Bitfinex, with a difference of \$42.51 (Source: CryptoCompare.com)

<sup>21</sup>The acf of the Bitcoin Garman-Klass volatility indicator shows a very slow decay, a behavior which is sometimes labeled as *long-memory*. To capture this pattern, specific models have been developed, like the HAR (Heterogeneous Autoregressive) and its log-counterpart, the log-HAR ([Corsi, 2009](#)). On the other hand, [Cipollini et al. \(2020, Sect. 5.3\)](#) demonstrate that some (2,1)-specifications of different ARMA-like models replicate the ability to approximate the long memory pattern observed in the autocorrelation of realized variances, a feature which has made the HAR model popular. Such an ability is due to the high estimated *persistence* and the presence of a second order parameter, usually significantly negative, in the AR part.

**Figure 18:** Evolution of the Garman-Klass volatility estimator in log scale starting from January 2017. The red vertical bars represent the announcement dates, whereas the blue bars represent the launch of CBOE and CME futures. See Table 10 for the details and the exact dates.



**Figure 19:** Normal QQ-Plot, autocorrelation and partial autocorrelation functions of the Garman-Klass volatility proxy in log scale.



As detailed in Section 5.3.2, to estimate the effect on Bitcoin volatility of the two futures and each related announcement, we need to fit a C-ARIMA model up to the day before each intervention, which, in our case, leads to the estimation of six different models: based on the first four, we predict the Garman-Klass proxy 1-step ahead to compute the contemporaneous effect of each announcement, whereas based on the last two models we can compute the effect of futures at different time horizons. In particular, since the CME future was launched 1 week after the CBOE future, for the former we estimate the temporal average pointwise effect at 1-week horizon, whereas for the CME futures we can also consider the 2-week and the 1-month horizons.

Notice that since the Garman-Klass proxy is in log scale, the difference between the observed and the predicted volatility is equal to the log of their ratio. This means that we are assuming a multiplicative effect on Bitcoin volatility and, as a result, we focus on the average contemporaneous effect and on the temporal average pointwise effects. Indeed, they both have a financial interpretation after re-exponentiating, being, respectively, the geometric average of the contemporaneous effects and of the point effects.

Volatility is typically stationary and non-seasonal, hence, our six independent models are all built from an  $ARMA(p, q)$ . The models also include covariates, so as to improve the forecast of Bitcoin volatility in the absence of intervention. We selected the set of covariates based



on a survey of relevant literature. In particular: we included the daily log returns and the Garman-Klass proxies of the EUR-USD exchange rate, the MSCI Emerging Market Index, the Shanghai Stock Exchange Composite Index and the MSCI World Index; among the economic factors, we selected the Federal Reserve money supply M1 aggregate, the US monthly inflation rate, the Federal funds Target Rate and the US GDP; technology factors are represented by the hash rate (estimated number of tera hashes per second the Bitcoin network is performing) and Bitcoin supply (total number of Bitcoins in circulation). All covariates are in log scale and have been made stationary in case they are not.<sup>22</sup> Finally, since the second Bitcoin halving took place on July 9, 2016, we also included a dummy variable taking value 1 after the halving. Table 20 in Appendix A.1 summarizes the correlation between the covariates and the Garman-Klass volatility. Overall, the Garman-Klass proxy seems to have a small correlation with the volatility of the EUR-USD exchange rate.

The selection of the six independent models was based on the Bayesian Information Criterion (BIC). Since the announcement dates are close to each other and to the actual launch of the futures, the characteristics of the data are in all cases well described by the same C-ARMA(2, 1) model, i.e.,

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t + X_t' \beta + \tau_t.$$

In order to shed light on the driving forces underneath Bitcoin volatility, we also performed a further analysis by investigating the impact of Bitcoin futures on the daily transaction volumes. Indeed, the possible volatility surge (decline) following the launch of the two contracts might be driven by increased (decreased) volumes. Figure 20 shows the evolution of daily volumes throughout the analysis period and from the log scale plot we can immediately notice that Bitcoin transactions experienced a sharp increase during 2014. However, as shown in Figure 21, by focusing on a restricted time period, the time series of daily volumes in log scale becomes more stable. Thus, we select the logarithmic transform of Bitcoin daily volumes as the outcome variable and perform the analysis in the period starting from February 2015.

Unlike the Garman-Klass proxy, log-volumes are not stationary (the KPSS test rejects the null hypothesis of stationarity at the 1% level) and show a weekly seasonal pattern; thus, the model selected with the BIC criterion is a C-ARIMA(1, 1, 1)(0, 0, 2)<sub>7</sub>, which is written as follows,

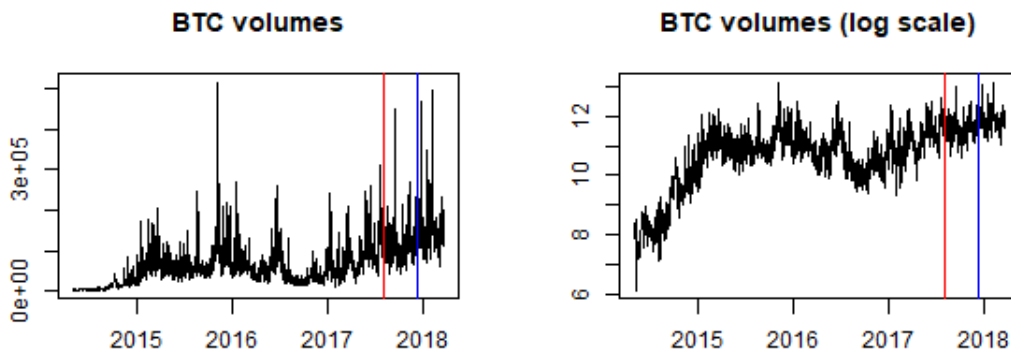
$$(1 - L)Y_t = \frac{\Theta_2(L^7)\theta_1(L)}{\phi_1(L)}\varepsilon_t + X_t'\beta + \tau_t$$

The next Section reports the result of our empirical analysis on the Garman-Klass volatility proxy and Bitcoin daily volumes.

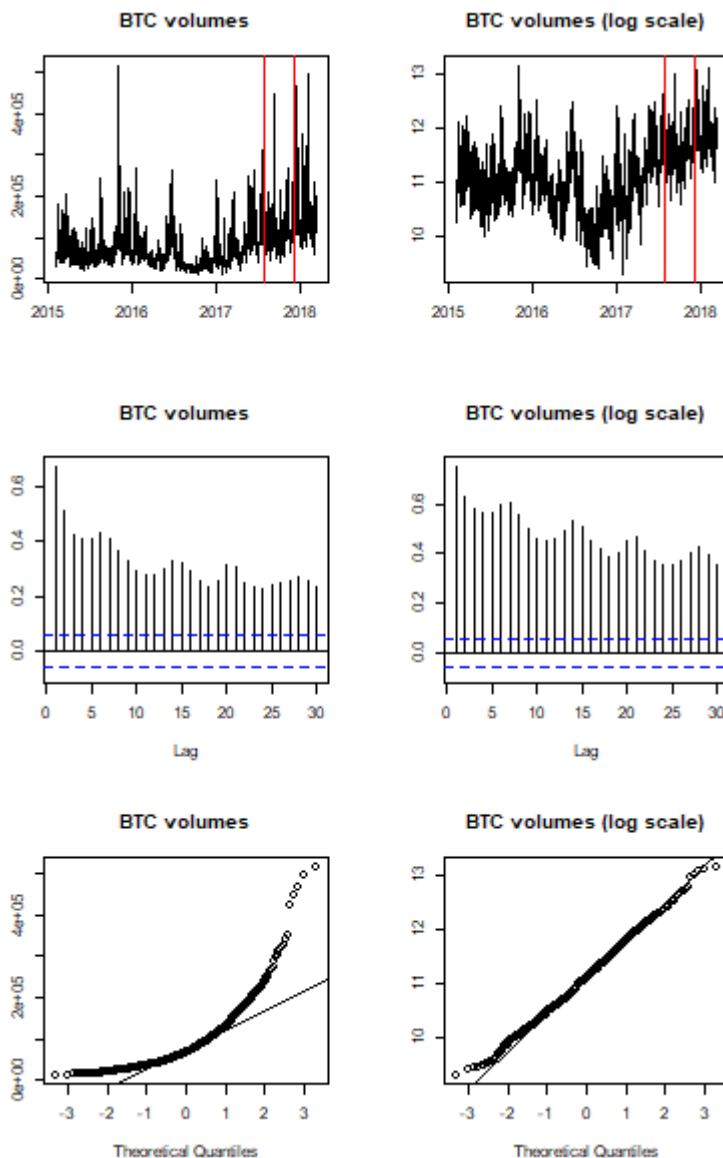
---

<sup>22</sup>First differences have been taken for `m1`, `midrate`, `gdp` and `hash`, whereas `btcsupply` has been differenced twice.

**Figure 20:** Evolution of Bitcoin daily transaction volumes (raw data, log scale) throughout the analysis period. The red vertical bar indicates the date of first announcement and the blue vertical bar denoted the launch date of the first future.



**Figure 21:** Evolution of Bitcoin daily transaction volumes (raw data, log scales), autocorrelation functions and Normal QQ-Plots for the period spanning from February 2015 to March 2018.



### 5.4.2 Results

The parameter estimates of the six C-ARMA(2,1) models, the estimated contemporaneous effect of the announcements and the estimated temporal average pointwise effects of the two futures are reported in Table 11.<sup>23</sup> Even though the Garman-Klass volatility proxy is computed from the intraday prices and could have been related to the main Bitcoin price drivers found in the literature, our findings do not support this idea: lagged values of the volatility seem to be sufficient to describe the dynamics of Bitcoin volatility. We find no evidence of significant contemporaneous effects of the announcements, meaning that the simple news of upcoming futures was not sufficient to spark investors' interests in the (crypto)currency. We, however, observe an interesting result for the actual introduction of the two futures. Indeed, there is no evidence of a significant effect related to CBOE futures, whereas the effects related to CME futures are significant at all time horizons. For example, based on the observed volatility up to 1 week from the launch of CME futures, we find that that the Garman-Klass proxy was 2.18 times higher (on average) than what would have been observed in the absence of futures. Figure 22 shows the forecasted series and the pointwise causal effects. Additional plots are given in Appendix A.2.

Having found no evidence of association between the covariates and the Garman-Klass proxy, as a robustness check to our results we also estimated six alternative models with no regressors. The estimated causal effects based on the alternative models are in line with the results reported in this section (see Table 19 in Appendix A.1).

Finally, Table 12 shows the results of the analysis performed on Bitcoin daily transaction volumes (in log scale). This time we found significant effects for both the CBOE and the CME futures at all time horizons; again, there is no evidence of a causal effect contemporaneous to the announcements. Interestingly, the effects of the two futures have opposite sign: transaction volumes decreased of approximately 16% due to the the launch of the CBOE contract, which might indicate that for a short period of time investors privileged the derivative over of the underlying; conversely, the launch of the CME contract increased volumes and this might have mediated the effect on volatility. In other words, this suggests that, benefiting from the increased transparency of the market, investors' interest toward Bitcoin rose, boosting Bitcoin trading and, in turn, its volatility. Figure 23 provides a graphical representation of these results.

---

<sup>23</sup>The log-ARMA used in the dissertation coincides right with the MLOG(2,1) in Cipollini et al. (2020), the best performing model of that paper. Moreover, the estimated parameters are in line with those reported there (for example, Table 11, Ann. 1 has  $\alpha_1 = 0.452$ ,  $\alpha_2 = -0.264$ ,  $\beta_1 = 0.781$ , to be compared with Figure 1, LNLS-MLOG(2,1) panel in Cipollini et al. (2020)). Finally, the acf of residuals in Figure 43 evidences clearly that they are substantially uncorrelated, supporting the idea that memory in the data is well captured from the model.

**Table 11:** Estimated coefficients (standard errors in parentheses) of the C-ARIMA models fitted on **Garman-Klass volatility proxy (in log scale)** up to the day before each intervention. In this table,  $\hat{\tau}^{(m)}$  indicates the estimated contemporaneous effect of each announcement;  $\hat{\tau}$  is the average contemporaneous effect;  $\hat{\tau}^{(CBOE)}$  is the temporal average pointwise effect of the CBOE future; and,  $\hat{\tau}_t^{(CME)}$  is the temporal average pointwise effect of the CME futures at 1-week, 2-weeks and 3-weeks horizons (indicated with  $t = 7$ ,  $t = 14$  and  $t = 21$ ).

	<i>Dependent variable:</i>					
	Ann.1	Ann.2	Ann.3	Ann.4	CBOE	CME
$\phi_1$	1.233*** (0.060)	1.220*** (0.058)	1.229*** (0.057)	1.226*** (0.057)	1.232*** (0.056)	1.233*** (0.056)
$\phi_2$	-0.264*** (0.051)	-0.252*** (0.050)	-0.259*** (0.049)	-0.256*** (0.049)	-0.261*** (0.049)	-0.262*** (0.049)
$\theta_1$	-0.781*** (0.047)	-0.776*** (0.045)	-0.783*** (0.044)	-0.781*** (0.045)	-0.782*** (0.044)	-0.784*** (0.044)
$c$	-3.790*** (0.121)	-3.777*** (0.122)	-3.770*** (0.126)	-3.770*** (0.126)	-3.765*** (0.132)	-3.768*** (0.129)
eurusd_vol	0.006 (0.021)	0.004 (0.020)	0.002 (0.020)	0.001 (0.020)	-0.001 (0.020)	-0.001 (0.020)
mxwo_vol	0.008 (0.024)	0.002 (0.023)	0.003 (0.023)	0.005 (0.023)	0.002 (0.023)	0.003 (0.023)
mxef_vol	-0.001 (0.021)	0.002 (0.021)	0.003 (0.020)	0.003 (0.020)	0.006 (0.020)	0.006 (0.020)
shc_vol	-0.009 (0.029)	-0.018 (0.028)	-0.018 (0.027)	-0.018 (0.027)	-0.015 (0.027)	-0.014 (0.027)
eurusd	0.222 (0.191)	0.198 (0.184)	0.202 (0.184)	0.200 (0.184)	0.208 (0.184)	0.215 (0.184)
mxwo	-0.075 (0.239)	-0.076 (0.234)	-0.037 (0.234)	-0.046 (0.234)	-0.039 (0.234)	-0.039 (0.233)
mxef	-0.098 (0.309)	-0.068 (0.300)	-0.112 (0.299)	-0.105 (0.299)	-0.104 (0.298)	-0.102 (0.297)
shc	0.056 (0.189)	0.056 (0.187)	0.071 (0.187)	0.071 (0.188)	0.076 (0.187)	0.081 (0.187)
m1	-0.240 (0.363)	-0.260 (0.351)	-0.248 (0.351)	-0.249 (0.351)	-0.261 (0.351)	-0.261 (0.351)
inflation	-0.005 (0.064)	-0.018 (0.063)	-0.006 (0.062)	-0.007 (0.062)	-0.009 (0.062)	-0.009 (0.062)
gdp	0.456 (0.467)	0.285 (0.444)	0.293 (0.446)	0.290 (0.446)	0.300 (0.446)	0.303 (0.445)
midrate	0.154 (0.325)	0.138 (0.323)	0.137 (0.324)	0.138 (0.325)	0.138 (0.325)	0.154 (0.320)
hash	0.047 (0.180)	-0.040 (0.162)	-0.039 (0.159)	-0.035 (0.159)	-0.035 (0.158)	-0.029 (0.158)
tot.btc	-0.505 (2.998)	-0.614 (2.957)	-0.715 (2.963)	-0.645 (2.966)	-0.688 (2.962)	-0.688 (2.953)
halv	-0.033 (0.193)	0.024 (0.186)	0.064 (0.188)	0.069 (0.188)	0.096 (0.194)	0.089 (0.191)
$\hat{\tau}^{(m)}$	-0.54 (0.50)	-0.07 (0.50)	0.41 (0.50)	-0.31 (0.50)		
$\hat{\tau}$	-0.13 (0.25)					
$\hat{\tau}^{(CBOE)}$					-0.26 (0.38)	
$\hat{\tau}_{t=7}^{(CME)}$						0.78** (0.38)
$\hat{\tau}_{t=14}^{(CME)}$						0.76** (0.37)
$\hat{\tau}_{t=21}^{(CME)}$						0.69* (0.37)
Observations	1,153	1,243	1,274	1,277	1,284	1,291
$\sigma^2$	0.259	0.254	0.257	0.257	0.258	0.257
Akaike Inf. Crit.	1,736.402	1,846.683	1,904.189	1,910.221	1,924.778	1,929.746

Note:

·p<0.1; \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

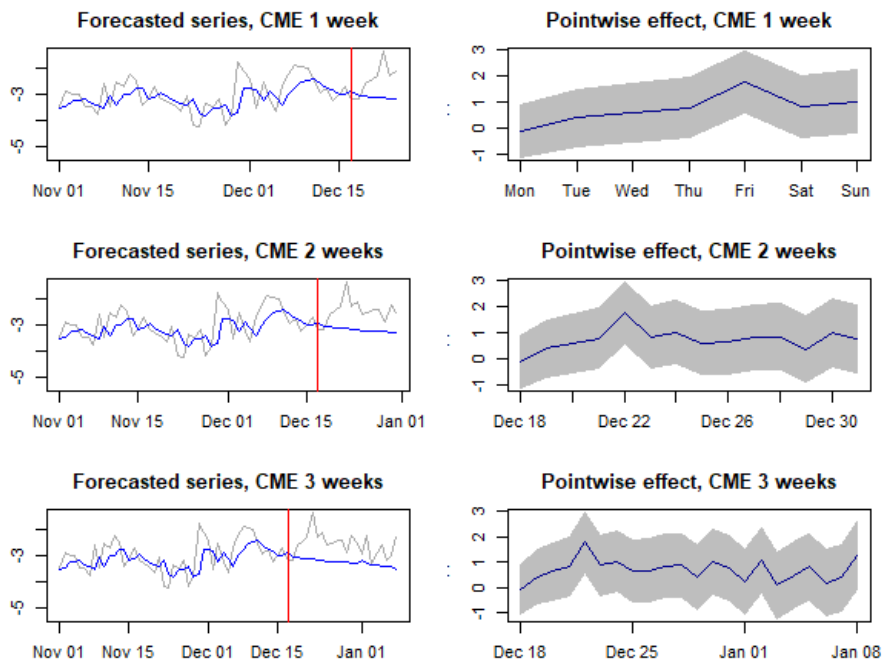
**Table 12:** Estimated coefficients (standard errors in parentheses) of the C-ARIMA models fitted on **Bitcoin daily volumes (in log scale)** up to the day before each intervention. In this table,  $\hat{\tau}^{(m)}$  indicates the estimated contemporaneous effect of each announcement;  $\hat{\tau}$  is the average contemporaneous effect;  $\hat{\tau}^{(CBOE)}$  is the temporal average pointwise effect of the CBOE future; and,  $\hat{\tau}_t^{(CME)}$  is the temporal average pointwise effect of the CME futures at 1-week, 2-weeks and 3-weeks horizons (indicated with  $t = 7$ ,  $t = 14$  and  $t = 21$ ).

	<i>Dependent variable:</i>					
	Ann.1	Ann.2	Ann.3	Ann.4	CBOE	CME
$\phi_1$	0.454*** (0.044)	0.460*** (0.042)	0.461*** (0.041)	0.460*** (0.041)	0.463*** (0.041)	0.465*** (0.040)
$\theta_1$	-0.925*** (0.025)	-0.932*** (0.023)	-0.931*** (0.024)	-0.931*** (0.024)	-0.931*** (0.023)	-0.933*** (0.023)
$\Theta_1$	0.121*** (0.035)	0.122*** (0.034)	0.112*** (0.033)	0.112*** (0.033)	0.118*** (0.033)	0.117*** (0.033)
$\Theta_2$	0.122*** (0.034)	0.117*** (0.031)	0.111*** (0.031)	0.111*** (0.031)	0.107*** (0.031)	0.106*** (0.030)
eurusd_vol	0.013 (0.019)	0.008 (0.018)	0.003 (0.018)	0.003 (0.018)	0.001 (0.018)	0.001 (0.018)
mxwo_vol	0.008 (0.021)	0.001 (0.021)	0.001 (0.020)	0.0005 (0.020)	-0.001 (0.020)	-0.001 (0.020)
mxef_vol	0.019 (0.019)	0.026 (0.018)	0.028 (0.018)	0.028 (0.018)	0.030 (0.018)	0.031 (0.018)
shc_vol	0.004 (0.024)	-0.005 (0.023)	-0.007 (0.023)	-0.006 (0.023)	-0.005 (0.023)	-0.005 (0.023)
eurusd	0.232 (0.156)	0.226 (0.151)	0.223 (0.151)	0.223 (0.150)	0.224 (0.150)	0.225 (0.150)
mxwo	-0.600** (0.198)	-0.585** (0.195)	-0.554** (0.194)	-0.553** (0.194)	-0.545** (0.194)	-0.530** (0.193)
mxef	0.403 (0.255)	0.367 (0.249)	0.324 (0.247)	0.322 (0.246)	0.322 (0.245)	0.305 (0.244)
shc	-0.065 (0.156)	-0.048 (0.155)	-0.043 (0.155)	-0.043 (0.155)	-0.042 (0.155)	-0.041 (0.154)
m1	0.124 (0.294)	0.097 (0.284)	0.100 (0.283)	0.100 (0.283)	0.090 (0.283)	0.091 (0.282)
inflation	0.033 (0.049)	0.027 (0.048)	0.025 (0.047)	0.025 (0.047)	0.024 (0.047)	0.024 (0.047)
gdp	0.635 (0.477)	0.529 (0.443)	0.546 (0.443)	0.546 (0.443)	0.548 (0.442)	0.551 (0.441)
midrate	0.261 (0.246)	0.250 (0.246)	0.262 (0.246)	0.262 (0.245)	0.257 (0.245)	0.259 (0.242)
hash	0.121 (0.155)	0.064 (0.135)	0.088 (0.131)	0.088 (0.131)	0.083 (0.130)	0.080 (0.130)
tot.btc	7.241* (3.565)	7.208* (3.499)	7.369* (3.442)	7.400* (3.439)	7.377* (3.437)	7.377* (3.416)
halv	-0.383 (0.279)	-0.416 (0.272)	-0.423 (0.272)	-0.423 (0.272)	-0.420 (0.273)	-0.427 (0.270)
$\hat{\tau}^{(m)}$	-0.41 (0.39)	0.22 (0.39)	-0.07 (0.39)	-0.30 (0.39)		
$\hat{\tau}$	-0.14 (0.20)					
$\hat{\tau}^{(CBOE)}$					-0.17** (0.07)	
$\hat{\tau}_{t=7}^{(CME)}$						0.42*** (0.07)
$\hat{\tau}_{t=14}^{(CME)}$						0.24*** (0.04)
$\hat{\tau}_{t=21}^{(CME)}$						0.07*** (0.03)
Observations	912	1,002	1,033	1,036	1,043	1,050
$\sigma^2$	0.159	0.159	0.159	0.158	0.159	0.158
Akaike Inf. Crit.	934.949	1,021.979	1,052.144	1,053.061	1,060.798	1,062.398

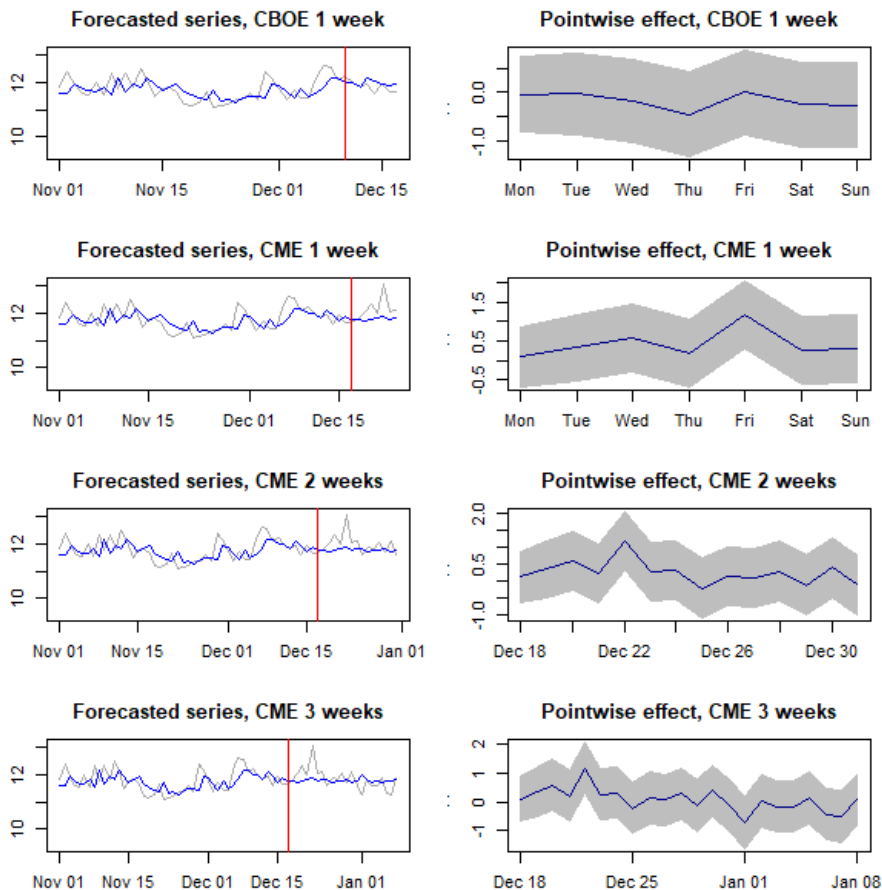
Note:

·p<0.1; \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

**Figure 22:** Observed (grey) and forecasted (blue) **Garman-Klass volatility proxy** (in log scale) at 1-week, 2-weeks and 3-weeks horizons. The vertical red bar indicates the date of CME futures launch. The right charts show the resulting pointwise effect within its 95% confidence bounds.



**Figure 23:** Observed (grey) and forecasted (blue) **Bitcoin daily volumes** (log scale) at 1-week, 2-weeks and 3-weeks horizons. The vertical red bar indicates the date of futures launch. The right charts show the resulting pointwise effect within its 95% confidence bounds.



## 5.5 Discussion

In this section we extended the C-ARIMA approach to estimate the heterogeneous effects of multiple interventions. First, we formalized the assumptions to estimate the causal effects in a multi-intervention setting. We then defined and estimated two classes of causal estimands: the contemporaneous effect (i.e., the instantaneous effect of an intervention conditioning on past outcomes) and the pointwise effect of a persistent intervention (i.e., the effect at each point in time between two subsequent persistent interventions).

Our motivating example is the introduction of the first two regulated futures by the CBOE and CME, two leading derivatives exchanges, with the goal of estimating the effect on Bitcoin volatility due to the launch of the two contracts. We also investigated the presence of announcement effects as well as the impact on Bitcoin daily volumes.

Our results indicate that the CME contract produced an increase of Bitcoin volatility. This effect could have been mediated by the rise in daily transactions, as we found evidence of a positive causal effect on volumes. Conversely, despite a small negative impact on transaction volumes due to the CBOE contract, we found no evidence that the first regulated future impacted Bitcoin volatility. Finally, we did not find evidence of announcement effects.

## 6 Conclusions

Causal inference in time series settings is a challenging task, partly because a simple association may be easily mistaken for a causal nexus and partly due to serial dependence, which brings additional difficulties during the estimation process. Moreover, in a panel setting where a treatment assigned to some units affects the others (a situation known as “interference”), inferring a causal effect is particularly demanding.

Having its roots in the potential outcomes framework, we believe that the RCM can aid the estimation of causal effects in such settings. Indeed, the RCM allows the construction of “what if” scenarios and sets the theoretical foundations underneath the attribution of the uncovered effect to a specific intervention. In particular, in this research we analyzed three situations that researchers and practitioners commonly encounter in a time series analysis: i) a single intervention occurring simultaneously on multiple non-interfering series; ii) multiple time series subject to a simultaneous treatment that, due to cross-unit interactions, may affect other series that were not intervened on; iii) multiple interventions on a single time series.

We first introduced a common causal framework building the theoretical foundations for a causal analysis under the RCM; we then defined new classes of causal estimands in each of the three settings and we proposed to estimate them using two novel methodologies: C-ARIMA and CausalMBSTS. The C-ARIMA approach can successfully estimate the effect of an intervention on a single time series as well as on multiple non-interfering series. Indeed, with a simulation study we showed that it performs well compared to a standard intervention analysis method in a situation where the effect takes the form of a fixed change in the level of the outcome; it also outperforms the latter when the effects are irregular and time-varying. Instead, CausalMBSTS can estimate the heterogeneous causal effect of an intervention in a panel setting where the time series interact with one another. Based on multivariate Bayesian models, it is a flexible methodology that allows to model the dependence structure between the time series in a very natural way, whilst enabling variable selection (via the addition of a spike-and-slab prior) and validation (by posterior predictive checks). Finally, we showed how to extend the C-ARIMA approach for the estimation of the heterogeneous causal effects of multiple interventions occurring on a single time series. We also applied the proposed C-ARIMA and CausalMBSTS approaches to evaluate the effect of a permanent price reduction introduced by a supermarket chain in Italy on a selected subset of store brands. Then, we used C-ARIMA in a multi-intervention setting to investigate the impact of the first two regulated Bitcoin futures on Bitcoin volatility and daily transaction volumes.

This research brings both methodological and empirical contributions, introducing two novel approaches to infer causal effects in complex time series settings and showing that the proposed methodologies can be employed in several fields of research, including marketing and finance. We also believe that this research provides several advances to the existing inferential methods



under the RCM. Indeed, recent extensions of the RCM to observational panel studies (i.e., synthetic control methods) mostly focus on a situation where the time series do not interact with one another and they also involve inferential tools that are not usually employed in standard time series analysis. Conversely, the C-ARIMA approach allows the estimation of causal effects under the RCM with standard tools that are extensively used by econometricians and practitioners in many fields. Furthermore, with CausalMBSTS we extended synthetic control methods to a setting with interference and, to foster the adoption of this method by a broad range of researchers, we also released an R package that handles both the definition and the estimation of the multivariate model. By making causal inference tools accessible to a vast audience, we hope we can facilitate the understanding of causal relationships and, in turn, decision making based on solid foundations.

## References

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.
- Abadie, A., Diamond, A., and Hainmueller, J. (2010). 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.
- Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.
- Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132.
- Anger, S., Kvasnicka, M., and Siedler, T. (2011). One last puff? public smoking bans and smoking behavior. *Journal of health economics*, 30(3):591–601.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist’s companion*. Princeton university press.
- Antoniou, A. and Holmes, P. (1995). Futures trading, information and spot price volatility: evidence for the FTSE-100 stock index futures contract using GARCH. *Journal of Banking & Finance*, 19(1):117–129.
- Antoniou, A., Holmes, P., and Priestley, R. (1998). The effects of stock index futures trading on stock index volatility: An analysis of the asymmetric response of volatility to news. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 18(2):151–166.
- Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.
- Baklaci, H. and Tutek, H. (2006). The impact of the futures market on spot volatility: An analysis in Turkish derivatives markets. *WIT Transactions on Modelling and Simulation*, 43:237–246.
- Balke, N. S. and Fomby, T. B. (1994). Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics*, 9(2):181–200.
- Basse, G. W., Feller, A., and Toulis, P. (2019). Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494.
- Baydakova, A. (2019). No change to bitcoin futures plans, CME says, as CBOE pulls back. Coindesk, available at <https://www.coindesk.com/cme-cboe-bitcoin-futures>.

- Ben-Michael, E., Feller, A., and Rothstein, J. (2018). The augmented synthetic control method. Preprint. Available at [arXiv:1811.04170](https://arxiv.org/abs/1811.04170).
- Ben-Michael, E., Feller, A., and Rothstein, J. (2019). Synthetic controls and weighted event studies with staggered adoption. *Preprint. Available at [arXiv: 1912. 03290](https://arxiv.org/abs/1912.03290)*.
- Bessembinder, H. and Seguin, P. J. (1992). Futures-trading activity and stock price volatility. *the Journal of Finance*, 47(5):2015–2034.
- Bhattacharyya, M. and Layton, A. P. (1979). Effectiveness of seat belt legislation on the queensland road toll—an Australian case study in intervention analysis. *Journal of the American Statistical Association*, 74(367):596–603.
- Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001.
- Blattberg, R. C., Briesch, R., and Fox, E. J. (1995). How promotions work. *Marketing science*, 14:G122–G132.
- Bohl, M. T., Diesteldorf, J., and Siklos, P. L. (2015). The effect of index futures trading on volatility: Three markets for chinese stocks. *China Economic Review*, 34:207–224.
- Bojinov, I., Rambachan, A., and Shephard, N. (2020). Panel experiments and dynamic causal effects: A finite population perspective. Preprint. Available at [arXiv:2003.09915](https://arxiv.org/abs/2003.09915).
- Bojinov, I. and Shephard, N. (2019). Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*, 114(528):1665–1682.
- Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.
- Bouoiyour, J. and Selmi, R. (2015). What does Bitcoin look like? *Annals of Economics & Finance*, 16(2):449–492.
- Box, G. E. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, 70(349):70–79.
- Box, G. E. and Tiao, G. C. (1976). Comparison of forecast and actuality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 25(3):195–200.
- Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274.

- Brown, P. J., Vannucci, M., and Fearn, T. (1998). Multivariate bayesian variable selection and prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):627–641.
- Cao, J. and Dowd, C. (2019). Estimation and inference for synthetic control methods with spillover effects. Preprint. Available at [arXiv:1902.07343](https://arxiv.org/abs/1902.07343).
- Card, D. and Krueger, A. B. (1993). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. Technical report, National Bureau of Economic Research.
- Cauley, J. and Im, E. I. (1988). Intervention policy analysis of skyjackings and other terrorist incidents. *The American Economic Review*, 78(2):27–31.
- Cermak, V. (2017). Can Bitcoin become a viable alternative to fiat currencies? An empirical analysis of bitcoin’s volatility based on a GARCH model. Available at <https://ssrn.com/abstract=2961405>.
- Chamberlain, G. (1982). The general equivalence of granger and sims causality. *Econometrica: Journal of the Econometric Society*, 50(3):569–581.
- Chen, C. and Liu, L.-M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88.
- Chen, H., Han, Q., Li, Y., and Wu, K. (2013). Does index futures trading reduce volatility in the Chinese stock market? A panel data evaluation approach. *Journal of Futures Markets*, 33(12):1167–1190.
- Ciaian, P., Rajcaniova, M., and Kancs, A. (2016). The economics of Bitcoin price formation. *Applied Economics*, 48(19):1799–1815.
- Cipollini, F., Gallo, G. M., and Palandri, A. (2020). Realized variance modeling: decoupling forecasting from estimation. *Journal of Financial Econometrics*, 18(3):532–555.
- Corbet, S., Lucey, B., Peat, M., and Vigne, S. (2018). Bitcoin futures—what use are they? *Economics Letters*, 172:23–27.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.
- Cox, D. R. (1958). *Planning of experiments*. Wiley.
- CryptoCompare (2020a). *Exchange Benchmark Report July 2020*. CryptoCompare Research, Available at: <https://www.cryptocompare.com/external/research/exchange-review/>.

- CryptoCompare (2020b). *Exchange Review 2020*. CryptoCompare Research, Available at: <https://www.cryptocompare.com/external/research/exchange-review/>.
- Danthine, J.-P. (1978). Information, futures prices, and stabilizing speculation. *Journal of Economic Theory*, 17(1):79–98.
- Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a bayesian application. *Biometrika*, 68(1):265–274.
- Dube, A. and Zipperer, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies. *IZA Discussion Paper 8944*.
- Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.
- Edwards, F. R. (1988). Does futures trading increase stock market volatility? *Financial Analysts Journal*, 44(1):63–69.
- Enders, W. and Sandler, T. (1993). The effectiveness of antiterrorism policies: A Vector-AutoRegression-intervention analysis. *The American Political Science Review*, 87(4):829–844.
- Figlewski, S. (1981). Futures trading and volatility in the GNMA market. *The Journal of Finance*, 36(2):445–456.
- Forastiere, L., Airoidi, E. M., and Mealli, F. (2020). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*.
- Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of business*, 53(1):67–78.
- Garvey, G. T. and Hanka, G. (1999). Capital structure and corporate control: The effect of antitakeover statutes on firm leverage. *The Journal of Finance*, 54(2):519–546.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 4*, Oxford, UK. Clarendon Press.
- Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 37(3):424–438.
- 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](https://arxiv.org/abs/2004.05027).
- Hale, G., Krishnamurthy, A., Kudlyak, M., Shultz, P., et al. (2018). How futures trading changed Bitcoin prices. *FRBSF Economic Letter*, 12:1–5.
- Harris, L. (1989). S&P 500 cash stock price volatilities. *The Journal of Finance*, 44(5):1155–1175.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.
- Helske, J. (2018). *KFAS: Kalman filter and smoothers for exponential family state space models*. R package version 1.3.3.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.
- Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Jochum, C. and Kodres, L. (1998). Does the introduction of futures on emerging market currencies destabilize the underlying currencies? *IMF Staff Papers*, 45(3):486–521.
- Keogh, E. and Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3):358–386.
- Kim, W., Lee, J., and Kang, K. (2020). The effects of the introduction of bitcoin futures on the volatility of bitcoin returns. *Finance Research Letters*, 33.
- 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 economics*, 25(12):1514–1528.
- Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3(1):1–7.

- Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one*, 10(4):1–15.
- Larcker, D. F., Gordon, L. A., and Pinches, G. E. (1980). Testing for market efficiency: a comparison of the cumulative average residual methodology and intervention analysis. *Journal of Financial and Quantitative Analysis*, 15(2):267–287.
- Larsen, K. (2019). *MarketMatching Package Vignette*. R package version 1.1.2.
- Li, K. T. (2019). Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*.
- Li, X. and Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. *Decision Support Systems*, 95:49–60.
- McKenzie, M. D., Brailsford, T. J., and Faff, R. W. (2001). New insights into the impact of the introduction of futures trading on stock price volatility. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 21(3):237–255.
- McLannahan, B. (2014). Bitcoin exchange Mt Gox files for bankruptcy protection. *The Financial Times*.
- Meyer, B. D., Viscusi, W. K., and Durbin, D. L. (1995). Workers’ compensation and injury duration: evidence from a natural experiment. *The American economic review*, 85(3):322–340.
- Molnár, P. (2012). Properties of range-based volatility estimators. *International Review of Financial Analysis*, 23:20–29.
- Moriarty, E. J. and Tosini, P. A. (1985). Futures trading and the price volatility of gnm certificates further evidence. *The Journal of Futures Markets (pre-1986)*, 5(4):633–641.
- Murry, J. P., Stam, A., and Lastovicka, J. L. (1993). Evaluating an anti-drinking and driving advertising campaign with a sample survey and time series intervention analysis. *Journal of the American Statistical Association*, 88(421):50–56.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
- Neslin, S. A., Henderson, C., and Quelch, J. (1985). Consumer promotions and the acceleration of product purchases. *Marketing science*, 4(2):147–165.
- Noirjean, S., Mariani, M., Mattei, A., and Mealli, F. (2020). Exploiting network information to disentangle spillover effects in a field experiment on teens’ museum attendance. *arXiv preprint arXiv:2011.11023*.

- 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 Services and Outcomes Research Methodology*, 16(1-2):1–21.
- 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](https://arxiv.org/abs/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. *Journal of marketing research*, 39(4):421–439.
- Rambachan, A. and Shephard, N. (2019). A nonparametric dynamic causal model for macroeconomics. Preprint. Available at [arXivpreprintarXiv: 1903. 01637](https://arxiv.org/abs/1903.01637).
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512.
- 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. *Journal of the American Statistical Association*, 94(447):687–700.
- Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200.
- Rubin, D. (1980). Discussion of” randomization analysis of experimental data in the fisher randomization test” by d. basu. *Journal of the American statistical association*, 75(371):591–593.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688–701.
- Rubin, D. B. (1975). Bayesian inference for causality: The importance of randomization. In *The Proceedings of the Social Statistics Section of the American Statistical Association*, pages 233–239.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 6(1):34–58.
- Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, 6(4):377–401.



- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12(4):1151–1172.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331.
- Ryan, A. M., Burgess Jr, J. F., and Dimick, J. B. (2015). Why we should not be indifferent to specification choices for difference-in-differences. *Health services research*, 50(4):1211–1235.
- Salvador, S. and Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580.
- Sävje, F., Aronow, P. M., and Hudgens, M. G. (2020). Average treatment effects in the presence of unknown interference. *Annals of Statistics*. In print.
- Shi, S. (2017). The impact of futures trading on intraday spot volatility and liquidity: Evidence from Bitcoin market. *Preprint*. Available at <https://ssrn.com/abstract=3094647>.
- Sims, C. A. (1972). Money, income, and causality. *The American economic review*, 62(4):540–552.
- Smith, T. C., Spiegelhalter, D. J., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in medicine*, 14(24):2685–2699.
- Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407.
- Stein, J. C. (1987). Informational externalities and welfare-reducing speculation. *Journal of political economy*, 95(6):1123–1145.
- Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical methods in medical research*, 10(4):277–303.
- Sutton, A. J. and Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in medicine*, 27(5):625–650.
- Tchetgen, E. J. T. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75.
- VanderWeele, T. J. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological methods & research*, 38(4):515–544.

- Viviano, D. and Bradic, J. (2019). Synthetic learner: model-free inference on treatments over time. Preprint. Available at [arXiv:1904.01490](https://arxiv.org/abs/1904.01490).
- West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.
- Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. *Trabajos de estadística y de investigación operativa*, 31(1):585–603.
- Zuckerman, M. J. (2018). First Bitcoin futures contract expires at \$10,900, win for bears. COINTELEGRAPH, available at <https://cointelegraph.com/news/first-bitcoin-futures-contract-expires-at-10900-win-for-bears>.

# Appendices

## A

### A.1 Additional tables

Tables 13 and 14 report the results of the estimated C-ARIMA models in the pre-intervention period for, respectively, store and competitor brands.

Tables 15 and 16 report the estimated marginal effect as defined in (19) and the estimated conditional effect  $\hat{\tau}_t((1, 1), (0, 1))$ . Tables 17 and 18 display the results of the trend and seasonal model estimated using a set of covariates with, respectively, the price difference and the price ratio.

Table 19 shows the results of alternative C-ARIMA models fitted on the Garman-Klass volatility proxy without regressors. Finally, Table 20 reports the Pearson's linear correlation coefficients and the Spearman's rho of the covariates included in the Bitcoin application.

**Table 13: Store brands.** Estimated coefficients and (standard errors in parentheses) of the C-ARIMA models fitted to the 11 store brands in the pre-intervention period. The dependent variable is the daily sales counts of each product in log scale.

	<i>Dependent variable:</i>										
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11
$\phi_1$	0.863*** (0.035)	0.865*** (0.037)	0.833*** (0.047)	0.496*** (0.053) 0.221*** (0.052)	0.867*** (0.037)	0.890*** (0.032)	0.828*** (0.042)	0.949*** (0.034)	0.839*** (0.050)	0.364*** (0.053) 0.177** (0.055) 0.216*** (0.054)	0.951*** (0.039)
$\phi_2$											
$\phi_3$											
$\theta_1$	-0.323*** (0.065)	-0.340*** (0.074)	-0.348*** (0.086)		-0.251** (0.082)	-0.458*** (0.063)	-0.191* (0.080)	-0.509*** (0.076) -0.212** (0.074)	-0.476*** (0.085)		-0.841*** (0.072)
$\theta_2$											
$\Phi_1$	0.120* (0.056)	0.113* (0.057)	0.203*** (0.055)	0.260*** (0.053)	0.235*** (0.056)		0.234*** (0.057)	0.240*** (0.059)	0.124* (0.059)	0.774*** (0.204) -0.682** (0.231)	0.688*** (0.146) -0.503** (0.172)
$\Theta_1$						0.181** (0.058)					
$c$	7.131*** (0.167)	6.767*** (0.167)	6.995*** (0.159)	7.912*** (0.152)	7.113*** (0.072)	6.535*** (0.107)	7.740*** (0.099)	7.800*** (0.063)	6.876*** (0.025)	6.312*** (0.110)	6.711*** (0.044)
price	-1.721*** (0.356)	-1.441*** (0.355)	-1.656*** (0.339)	-2.218*** (0.288)	-1.433*** (0.287)	-1.841*** (0.304)	-1.423*** (0.281)	-2.316*** (0.116)	-1.727*** (0.117)	-2.118*** (0.176)	-1.175*** (0.208)
hol	0.176*** (0.033)	0.165*** (0.033)	0.168*** (0.034)	0.160*** (0.029)	0.174*** (0.027)	0.162*** (0.033)	0.162*** (0.027)	0.164*** (0.023)	0.163*** (0.026)	0.076* (0.035)	0.208*** (0.028)
Dec.Sun	0.323*** (0.063)	0.369*** (0.062)	0.306*** (0.069)	0.292*** (0.063)	0.379*** (0.056)	0.411*** (0.067)	0.367*** (0.054)	0.327*** (0.053)	0.390*** (0.051)	0.191* (0.075)	0.374*** (0.072)
Sat	0.265*** (0.024)	0.243*** (0.024)	0.248*** (0.028)	0.276*** (0.026)	0.293*** (0.022)	0.277*** (0.026)	0.292*** (0.021)	0.223*** (0.021)	0.214*** (0.020)	0.135*** (0.035)	0.162*** (0.036)
Sun	-1.291*** (0.028)	-1.355*** (0.027)	-1.321*** (0.031)	-1.213*** (0.029)	-1.140*** (0.026)	-1.261*** (0.029)	-1.173*** (0.025)	-1.203*** (0.025)	-1.291*** (0.022)	-1.398*** (0.037)	-1.514*** (0.036)
Mon	-0.135*** (0.028)	-0.188*** (0.028)	-0.163*** (0.032)	-0.175*** (0.030)	-0.010 (0.026)	-0.062* (0.029)	-0.037 (0.026)	-0.093*** (0.025)	-0.123*** (0.022)	-0.139*** (0.035)	-0.222*** (0.036)
Tue	-0.225*** (0.028)	-0.263*** (0.028)	-0.271*** (0.032)	-0.251*** (0.030)	-0.170*** (0.026)	-0.233*** (0.029)	-0.207*** (0.026)	-0.200*** (0.025)	-0.226*** (0.022)	-0.229*** (0.035)	-0.305*** (0.036)
Wed	-0.247*** (0.027)	-0.250*** (0.027)	-0.266*** (0.030)	-0.271*** (0.028)	-0.209*** (0.025)	-0.243*** (0.028)	-0.232*** (0.025)	-0.249*** (0.025)	-0.266*** (0.021)	-0.262*** (0.037)	-0.312*** (0.036)
Thr	-0.218*** (0.024)	-0.218*** (0.024)	-0.207*** (0.028)	-0.239*** (0.026)	-0.201*** (0.022)	-0.211*** (0.026)	-0.210*** (0.021)	-0.234*** (0.021)	-0.230*** (0.020)	-0.249*** (0.035)	-0.258*** (0.036)
Observations	386	386	386	386	386	386	386	386	386	386	386
$\sigma^2$	0.022	0.022	0.022	0.017	0.014	0.022	0.014	0.011	0.013	0.027	0.017
Alkaike Inf. Crit.	-355.744	-358.769	-352.505	-453.681	-519.897	-366.579	-539.814	-620.166	-577.969	-282.397	-460.103

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01; \*\*\*p<0.001

**Table 14: Competitor brands.** Estimated coefficients and (standard errors in parentheses) of the C-ARIMA models fitted to the 10 competitor brands in the pre-intervention period. The dependent variable is the daily sales counts of each product in log scale.

<i>Dependent variable:</i>										
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
$\phi_1$	-0.442*** (0.026)	0.959*** (0.014)	0.949*** (0.016)	-0.090* (0.043)	0.876*** (0.053)	0.973*** (0.157)	1.800*** (0.088)	0.939*** (0.028)	0.855*** (0.034)	0.786*** (0.083)
$\phi_2$	0.349*** (0.026)	0.850*** (0.038)	-0.161** (0.054)				-0.821*** (0.081)			
$\phi_3$	0.915*** (0.023)									
$\theta_1$	1.402*** (0.013)			0.940*** (0.047)		-0.408** (0.142)	-0.768*** (0.109)	-0.566*** (0.068)	-0.239*** (0.068)	-0.514*** (0.120)
$\theta_2$	1.000*** (0.014)							-0.117 (0.070)		
$\Phi_1$	0.108* (0.054)		0.907*** (0.030)	0.224*** (0.054)	0.162** (0.053)	0.852*** (0.088)			0.281*** (0.053)	
$\Theta_1$			-1.000*** (0.021)			-0.779*** (0.103)				
$c$	8.292*** (0.299)	7.862*** (0.297)	7.828*** (0.150)	8.335*** (0.162)	8.873*** (0.137)	8.380*** (0.127)	10.189*** (0.292)	7.029*** (0.048)	6.773*** (0.050)	6.327*** (0.108)
price	-2.459*** (0.163)	-2.350*** (0.124)	-2.074*** (0.135)	-2.193*** (0.130)	-2.210*** (0.139)	-2.257*** (0.121)	-3.742*** (0.286)	-1.490*** (0.089)	-1.308*** (0.127)	-2.499*** (0.220)
hol	0.145* (0.059)	0.161*** (0.044)	0.099* (0.048)	0.154*** (0.045)	0.107** (0.033)	0.155** (0.029)	0.138* (0.062)	0.178*** (0.025)	0.131*** (0.029)	0.165*** (0.038)
Dec.Sun	0.411*** (0.093)	0.538*** (0.076)	0.446*** (0.068)	0.515*** (0.085)	0.431*** (0.057)	0.379*** (0.056)	0.363*** (0.095)	0.413*** (0.046)	0.338*** (0.063)	0.437*** (0.069)
Sat	0.323*** (0.036)	0.262*** (0.030)	0.272*** (0.013)	0.244*** (0.034)	0.322*** (0.024)	0.357*** (0.027)	0.315*** (0.037)	0.216*** (0.018)	0.263*** (0.025)	0.215*** (0.027)
Sun	-0.984*** (0.047)	-1.143*** (0.039)	-1.083*** (0.018)	-1.351*** (0.045)	-1.011*** (0.034)	-0.992*** (0.031)	-1.057*** (0.050)	-1.303*** (0.020)	-1.288*** (0.030)	-1.303*** (0.029)
Mon	-0.106* (0.051)	-0.182*** (0.042)	-0.172*** (0.018)	-0.199*** (0.048)	-0.050 (0.037)	-0.085** (0.032)	0.017 (0.054)	-0.117*** (0.020)	-0.119*** (0.031)	-0.119*** (0.029)
Tue	-0.254*** (0.051)	-0.240*** (0.042)	-0.266*** (0.018)	-0.267*** (0.048)	-0.206*** (0.037)	-0.241*** (0.032)	-0.166** (0.054)	-0.208*** (0.020)	-0.231*** (0.031)	-0.229*** (0.029)
Wed	-0.237*** (0.046)	-0.222*** (0.039)	-0.284*** (0.016)	-0.285*** (0.044)	-0.197*** (0.033)	-0.254*** (0.030)	-0.218*** (0.049)	-0.237*** (0.020)	-0.242*** (0.029)	-0.270*** (0.028)
Thr	-0.208*** (0.036)	-0.182*** (0.030)	-0.220*** (0.012)	-0.237*** (0.034)	-0.187*** (0.024)	-0.232*** (0.027)	-0.200*** (0.037)	-0.196*** (0.018)	-0.219*** (0.025)	-0.234*** (0.027)
Observations	386	386	385	386	386	386	386	386	386	386
$\sigma^2$	0.080	0.046	0.049	0.044	0.021	0.017	0.094	0.013	0.017	0.027
Alkaike Inf. Crit.	150.457	-66.040	-33.990	-85.448	-372.686	-451.392	207.605	-575.002	-449.421	-281.026

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01; \*\*\*\*p<0.001

**Table 15:** Posterior mean and 95% credible intervals of the temporal average **mean marginal causal effect** of the new price policy on the ten store brands computed at three time horizons. In this table,  $\hat{\tau}_t$  stands for the mean marginal effect  $\hat{\tau}_t(s, 4)$ . There is evidence of a causal effect when the credible intervals do not include zero.

<i>Time horizon:</i>									
	1 month			3 months			6 months		
	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%
1	3.53	-12.27	19.33	2.39	-21.98	26.88	3.55	-32.92	39.98
2	3.55	-7.39	14.51	2.51	-15.09	19.39	3.34	-22.05	29.27
3	4.02	-7.00	16.14	2.71	-15.91	20.70	3.97	-24.13	31.25
4	24.06	2.07	49.43	11.58	-26.51	50.12	12.22	-46.22	68.82
5	1.98	-24.69	28.29	3.73	-40.92	48.08	5.74	-60.05	73.52
6	4.85	-7.97	17.53	5.94	-14.73	26.51	6.78	-24.61	38.53
7	39.19	0.04	77.11	17.33	-40.86	76.01	14.84	-74.46	102.94
8	12.67	-14.32	39.30	11.71	-34.58	54.99	8.63	-57.56	73.01
9	20.46	-9.44	50.67	8.19	-39.87	57.46	6.44	-65.70	82.99
10	6.26	0.52	11.98	4.86	-4.22	14.22	2.72	-11.39	16.84

**Table 16:** Posterior mean and 95% credible intervals of the temporal average **conditional causal effect** of the new price policy on the ten store brands computed at three time horizons. In this table,  $\hat{\tau}_t$  stands for the conditional effect  $\hat{\tau}_t((1, 1), (0, 1))$ . There is evidence of a causal effect when the credible intervals do not include zero.

<i>Time horizon:</i>										
		1 month			3 months			6 months		
	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	
(1)	s	0.09	-0.14	0.39	0.11	-0.14	0.38	0.12	-0.12	0.37
	c	-0.34	-1.68	0.75	-0.63	-1.64	0.78	-0.78	-1.55	0.64
(2)	s	0.08	-0.12	0.30	0.10	-0.14	0.31	0.11	-0.10	0.29
	c	-0.30	-0.86	0.34	-0.54	-0.93	0.34	-0.65	-0.74	0.24
(3)	s	0.11	-0.15	0.36	0.12	-0.12	0.36	0.11	-0.09	0.35
	c	-0.22	-0.79	0.67	-0.37	-0.94	0.58	-0.21	-0.74	0.23
(4)	s	0.28	-0.97	1.50	0.50	-0.92	1.62	0.71	-0.47	4.17
	c	-1.04	-3.92	2.64	-2.13	-4.18	2.36	-3.22	-19.10	1.11
(5)	s	-0.15	-2.69	4.01	-0.12	-7.83	1.30	-0.27	-23.15	1.39
	c	-0.08	-2.48	2.53	-0.08	-2.73	2.67	-0.07	-2.23	3.66
(6)	s	0.17	-0.28	0.57	0.12	-0.31	0.67	-0.02	-0.27	0.55
	c	-0.34	-1.62	0.84	-0.31	-1.90	0.73	-0.29	-1.50	0.70
(7)	s	0.20	-1.16	1.60	0.21	-1.14	1.62	0.20	-1.15	1.63
	c	-1.09	-21.89	18.58	-1.46	-21.83	18.40	-1.26	-22.02	18.63
(8)	s	0.12	-2.75	2.79	0.09	-4.46	4.22	0.18	-6.69	7.86
	c	-0.02	-12.99	14.38	0.15	-19.31	23.83	-0.31	-39.60	32.96
(9)	s	0.64	-42.52	43.54	1.00	-70.38	78.18	0.81	-106.58	119.18
	c	-0.29	-45.17	44.19	-0.25	-74.86	72.15	0.28	-112.62	115.76
(10)	s	0.09	-2.76	3.08	0.08	-5.16	4.39	0.13	-7.60	7.00
	c	0.04	-5.83	6.93	0.07	-8.62	10.90	-0.02	-17.00	16.37

**Table 17:** Temporal average general causal effects of the new price policy on the ten store (s) - competitor (c) pairs computed at three time horizons. In this table,  $\hat{\tau}_t$  stands for the general effect  $\hat{\tau}_t((1,0), (0,0))$  and the results are obtained including in the set of covariates the difference in price between the store and competitor brand prior to the intervention (in the post-intervention period the difference in price is computed from the prior price).

		1 month			3 months			6 months		
		$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%
(1)	s	7.86	-22.72	39.39	6.01	-44.36	54.53	8.69	-62.66	81.65
	c	24.76	-101.23	154.16	18.14	-189.20	223.43	7.94	-299.89	322.01
(2)	s	6.32	-15.06	27.87	4.64	-27.51	36.56	5.78	-43.30	55.55
	c	14.36	-65.53	97.56	8.08	-129.50	142.40	-1.55	-206.59	198.41
(3)	s	7.74	-15.37	31.07	5.76	-32.76	40.91	8.98	-45.53	64.71
	c	17.60	-60.32	98.08	12.58	-116.06	142.92	6.48	-182.11	198.27
(4)	s	<b>47.39</b>	<b>0.94</b>	<b>96.95</b>	23.29	-49.15	104.14	24.21	-88.64	136.26
	c	31.44	-74.80	140.15	23.04	-156.67	205.96	14.52	-259.18	280.48
(5)	s	4.51	-46.29	57.07	8.11	-75.41	91.55	13.40	-108.70	136.45
	c	48.56	-55.74	160.97	18.78	-155.55	199.51	11.59	-255.06	276.53
(6)	s	10.05	-14.63	35.36	12.24	-28.79	54.40	14.69	-45.35	76.51
	c	25.66	-39.05	92.53	7.03	-101.58	117.02	5.53	-159.96	167.62
(7)	s	<b>80.83</b>	<b>6.45</b>	<b>158.56</b>	38.12	-82.24	154.90	34.47	-137.44	209.06
	c	184.75	-216.88	596.71	106.78	-553.29	757.07	92.10	-904.77	1086.75
(8)	s	25.29	-25.76	77.12	23.02	-62.62	103.02	14.70	-111.95	135.90
	c	15.27	-14.96	45.95	5.17	-44.71	53.87	3.01	-68.34	73.61
(9)	s	41.09	-8.93	89.23	16.95	-61.21	99.53	13.91	-102.74	132.98
	c	18.71	-30.61	71.21	2.68	-77.27	80.47	3.93	-114.88	122.98
(10)	s	<b>12.16</b>	<b>1.06</b>	<b>23.02</b>	9.42	-8.54	26.50	5.12	-21.80	32.30
	c	-0.21	-8.89	8.87	1.64	-13.12	17.01	3.64	-17.52	24.97

**Table 18:** Temporal average general causal effects of the new price policy on the ten store (s) - competitor (c) pairs computed at three time horizons. In this table,  $\hat{\tau}_t$  stands for the general effect  $\hat{\tau}_t((1,0), (0,0))$  and the results are obtained including in the set of covariates the price ratio between the store and competitor brand prior to the intervention (in the post-intervention period the ratio is computed from the prior price).

		1 month			3 months			6 months		
		$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%
(1)	s	7.86	-23.99	40.25	5.57	-43.61	56.18	7.60	-65.59	81.24
	c	24.24	-103.31	149.08	18.24	-190.07	236.58	9.94	-302.70	321.87
(2)	s	6.29	-15.08	27.85	4.58	-28.01	36.62	5.78	-43.58	55.33
	c	14.43	-65.19	97.88	8.04	-129.58	142.52	-1.94	-206.72	198.81
(3)	s	7.69	-15.61	31.11	5.69	-33.02	41.00	8.94	-45.58	65.00
	c	17.67	-60.31	98.22	12.55	-116.11	142.85	6.40	-182.21	198.30
(4)	s	47.59	-1.43	95.37	23.49	-52.91	99.97	26.11	-85.00	143.55
	c	30.86	-76.21	142.37	21.79	-156.22	203.90	12.56	-247.89	285.21
(5)	s	4.93	-45.95	56.46	8.44	-74.91	93.47	13.63	-107.63	138.26
	c	48.63	-58.86	160.54	18.78	-161.04	203.72	11.66	-267.79	280.47
(6)	s	9.89	-14.74	34.85	12.05	-29.01	54.04	14.37	-46.42	75.06
	c	25.76	-38.76	92.99	7.05	-100.67	117.62	5.59	-155.74	167.47
(7)	s	<b>80.67</b>	<b>1.53</b>	<b>161.11</b>	36.73	-84.22	156.80	31.45	-150.41	207.70
	c	183.01	-222.65	583.47	108.84	-559.14	799.66	102.14	-892.35	1113.15
(8)	s	23.54	-28.05	73.80	22.06	-59.32	103.49	14.64	-113.07	140.54
	c	14.98	-15.50	44.80	4.46	-44.03	53.53	2.35	-69.75	75.51
(9)	s	41.00	-7.02	87.54	16.93	-64.31	97.09	14.35	-106.63	136.62
	c	18.68	-32.60	69.15	2.66	-82.03	83.46	4.81	-113.13	120.65
(10)	s	<b>12.50</b>	<b>1.45</b>	<b>23.71</b>	9.62	-9.64	27.65	5.07	-23.35	31.58
	c	-0.11	-9.77	9.72	1.72	-13.10	16.31	3.77	-18.52	25.31

**Table 19:** Estimated coefficients (standard errors in parentheses) of alternative C-ARIMA models fitted on **Garman-Klass volatility proxy (in log scale)** up to the day before each intervention. In this table,  $\hat{\tau}^{(m)}$  indicates the estimated contemporaneous effect of each announcement;  $\hat{\tau}$  is the average contemporaneous effect;  $\hat{\tau}^{(CBOE)}$  is the temporal average pointwise effect of the CBOE future; and,  $\hat{\tau}_t^{(CME)}$  is the temporal average pointwise effect of the CME futures at 1-week, 2-weeks and 3-weeks horizons (indicated with  $t = 7$ ,  $t = 14$  and  $t = 21$ ).

	<i>Dependent variable:</i>					
	Ann.1	Ann.2	Ann.3	Ann.4	CBOE	CME
$\phi_1$	1.223*** (0.059)	1.214*** (0.057)	1.224*** (0.056)	1.222*** (0.056)	1.223*** (0.059)	1.214*** (0.057)
$\phi_2$	-0.256*** (0.051)	-0.246*** (0.049)	-0.254*** (0.048)	-0.252*** (0.049)	-0.256*** (0.051)	-0.246*** (0.049)
$\theta_1$	-0.776*** (0.047)	-0.774*** (0.045)	-0.783*** (0.044)	-0.781*** (0.044)	-0.776*** (0.047)	-0.774*** (0.045)
$c$	-3.804*** (0.100)	-3.769*** (0.099)	-3.745*** (0.102)	-3.742*** (0.102)	-3.804*** (0.100)	-3.769*** (0.099)
$\hat{\tau}^{(m)}$	-0.52 (0.51)	-0.05 (0.50)	0.43 (0.50)	-0.32 (0.50)		
$\hat{\tau}$	-0.11 (0.25)					
$\hat{\tau}^{(CBOE)}$					-0.24 (0.38)	
$\hat{\tau}_{t=7}^{(CME)}$						0.80** (0.38)
$\hat{\tau}_{t=14}^{(CME)}$						0.78** (0.37)
$\hat{\tau}_{t=21}^{(CME)}$						0.71* (0.37)
Observations	1,153	1,243	1,274	1,277	1,153	1,243
$\sigma^2$	0.257	0.252	0.254	0.255	0.257	0.252
Akaike Inf. Crit.	1,710.446	1,820.362	1,877.940	1,883.982	1,710.446	1,820.362

Note:

p<0.1; \*p<0.05; \*\*p<0.01; \*\*\*p<0.001



**Table 20:** Pearson's linear correlation coefficient (upper triangular) and Spearman's rho (lower triangular) between Garman-Klass volatility (GK) and the covariates included in the analysis, computed in the period before the first intervention.

	GK	eurusd_vol	mxwo_vol	mxef_vol	shc_vol	eurusd	mxwo	mxef	shc	m1	inflation	gdp	midrate	hash	tot.btc	halv
GK		0.06	-0.04	-0.03	0.02	0.02	-0.01	0.01	0.01	-0.00	0.00	0.01	0.02	0.01	-0.01	-0.00
eurusd_vol	0.04		0.47	0.43	0.35	0.03	-0.03	-0.05	-0.02	-0.03	0.01	-0.00	0.05	-0.04	-0.07	-0.16
mxwo_vol	-0.06	0.44		0.67	0.27	0.02	-0.10	-0.11	-0.07	-0.03	-0.04	-0.00	0.04	-0.04	-0.06	-0.28
mxef_vol	-0.05	0.41	0.63		0.37	0.00	-0.07	-0.09	-0.08	-0.00	-0.02	-0.05	0.06	-0.02	-0.01	-0.22
shc_vol	0.02	0.37	0.30	0.39		-0.03	-0.05	-0.09	-0.12	0.02	-0.01	-0.04	-0.03	-0.01	0.04	-0.55
eurusd	0.01	0.02	0.02	0.00	-0.01		-0.03	-0.04	-0.10	0.00	0.00	-0.05	-0.01	-0.02	0.01	0.05
mxwo	0.04	-0.01	-0.10	-0.04	-0.02	0.01		0.65	0.19	-0.04	0.01	0.01	0.05	-0.01	-0.02	0.04
mxef	0.02	-0.02	-0.08	-0.06	-0.09	-0.04	0.52		0.36	-0.02	-0.01	0.03	0.04	0.03	-0.03	0.06
shc	0.00	0.01	-0.03	-0.01	-0.04	-0.06	0.12	0.26		-0.02	0.04	0.05	-0.00	-0.02	0.01	-0.01
m1	-0.01	-0.04	-0.03	0.01	0.03	0.05	-0.05	-0.01	0.01		-0.03	-0.00	-0.00	-0.02	0.12	0.01
inflation	0.00	0.00	-0.01	-0.04	-0.03	0.02	0.04	-0.02	0.02	-0.04		-0.03	0.00	-0.02	0.02	-0.01
gdp	0.00	0.00	0.00	-0.05	-0.05	-0.08	-0.03	0.03	0.04	-0.01	-0.05		-0.00	-0.01	0.01	-0.01
midrate	0.00	0.06	0.06	0.07	-0.04	-0.00	0.03	0.04	-0.02	-0.00	0.00	-0.01		-0.08	-0.00	0.02
hash	0.02	-0.03	-0.02	-0.01	-0.02	-0.01	-0.00	0.00	-0.04	-0.03	0.01	-0.01	-0.08		0.08	-0.00
tot.btc	-0.00	-0.06	-0.07	-0.02	0.02	0.02	0.02	-0.01	0.04	0.11	-0.00	-0.01	-0.02	0.32		-0.00
halv	-0.01	-0.20	-0.32	-0.25	-0.58	0.04	0.03	0.08	-0.04	-0.02	0.02	-0.01	0.05	-0.01	0.03	

## A.2 Additional plots

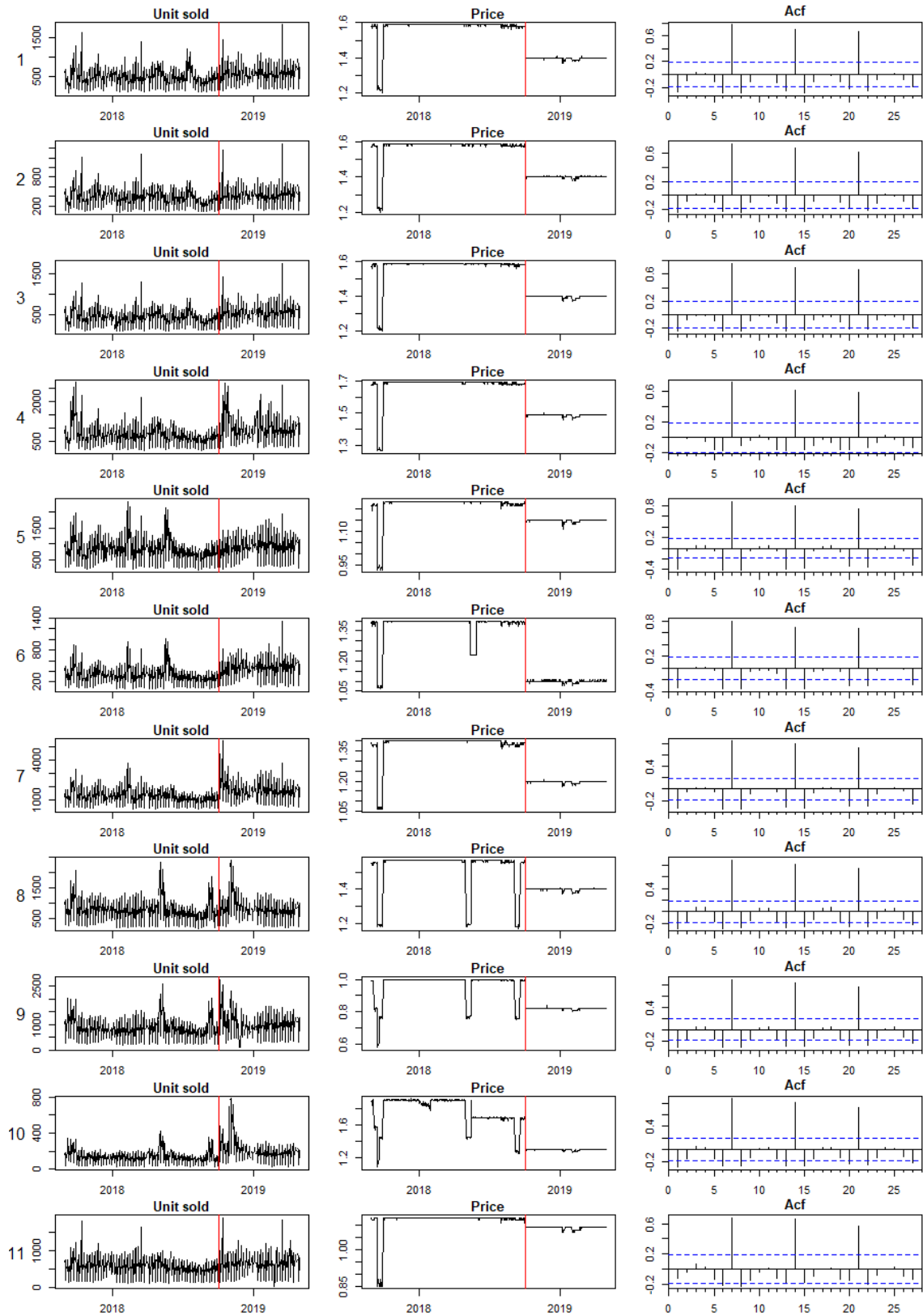
Figures 24 and 25 provide a full illustration of the sales evolution of each store and competitor brand. Figures 32 and 33 provide the same plots in terms of the average sales per hour.

Figure 26 shows the residual diagnostics of the univariate C-ARIMA models fitted to the store brands in the pre-intervention period and Figures 27 and 28 depicts, respectively, the resulting pointwise effects and the forecasted time series in the post-intervention period. Figures 29, 30 and 31 do the same for competitor brands.

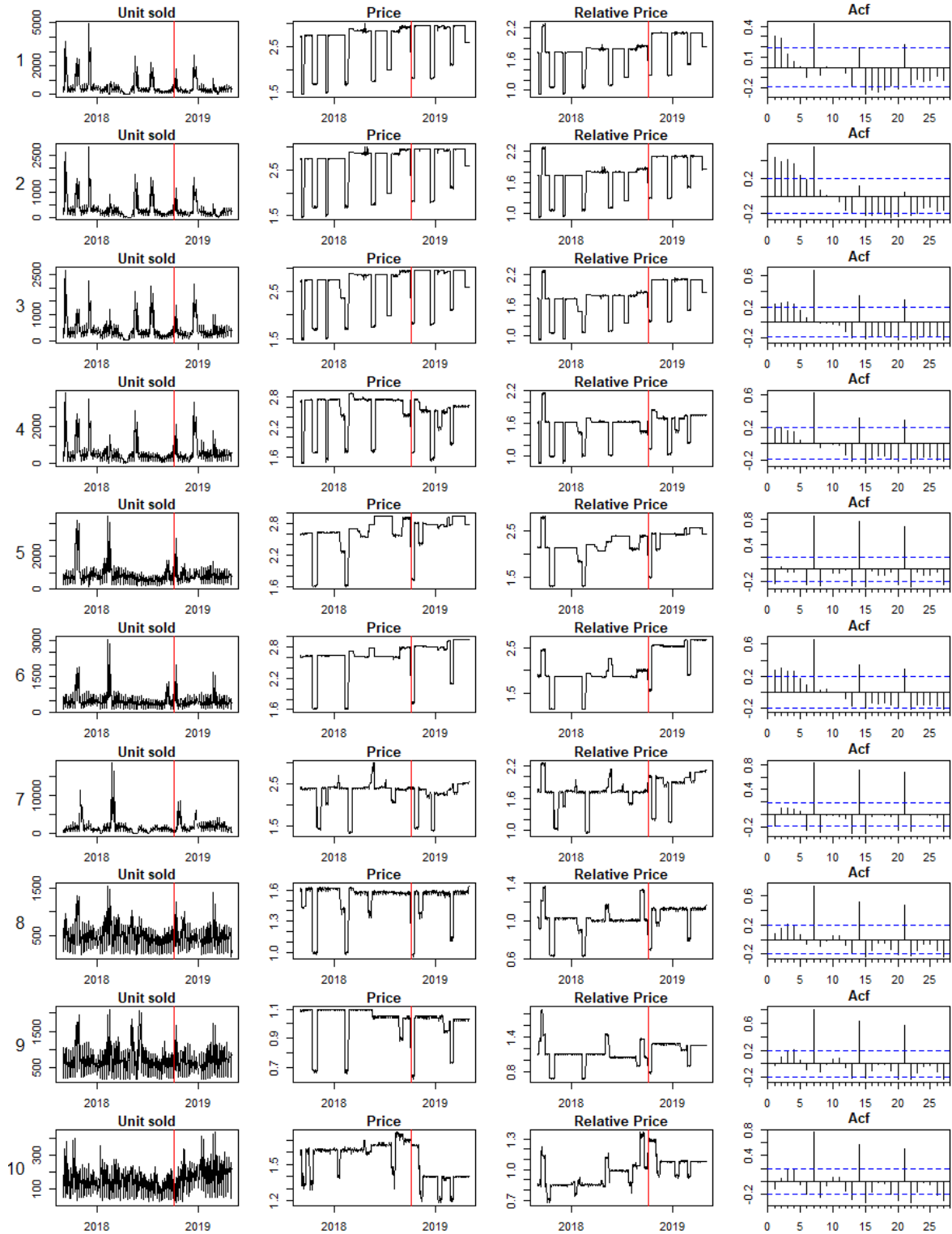
Figures 34 to 37 plot the estimated causal effect, the posterior predictive checks and the inclusion probabilities for the selected MBSTS trend-seasonal model, whereas Figures 38 to 40 show the posterior predictive checks for alternative models.

Figure 41 shows the evolution of the predictors selected for the analysis on Bitcoin volatility. Figures 43 and 44 provide additional plots for the analysis performed on the Garman-Klass volatility proxy, whereas Figure 45 plot the residual diagnostics for the analyses performed on daily transaction volumes.

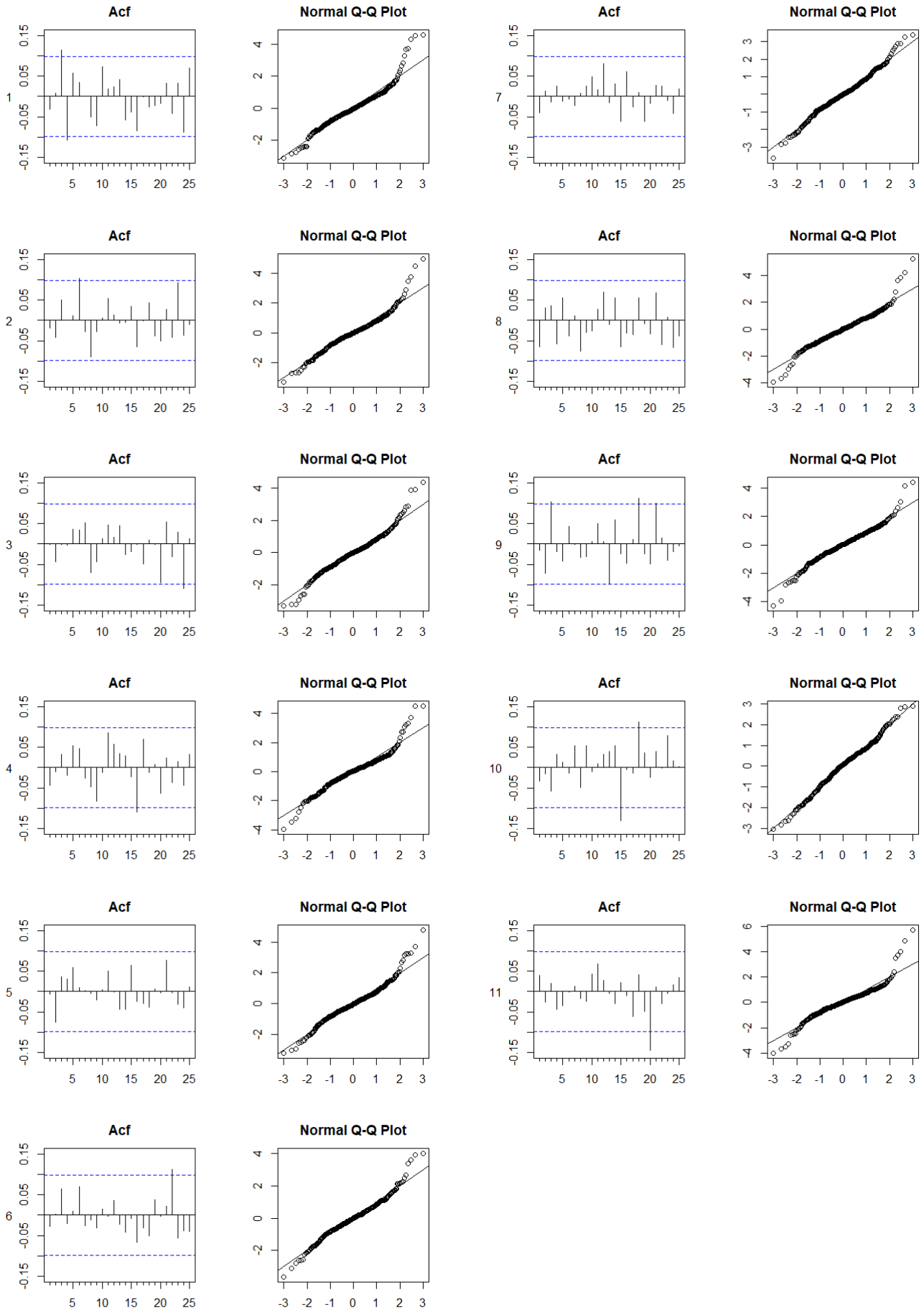
**Figure 24: Store brands.** Time series of unit sold daily, price per unit and autocorrelation function for the 11 store brands. The red vertical bar indicates the intervention date.



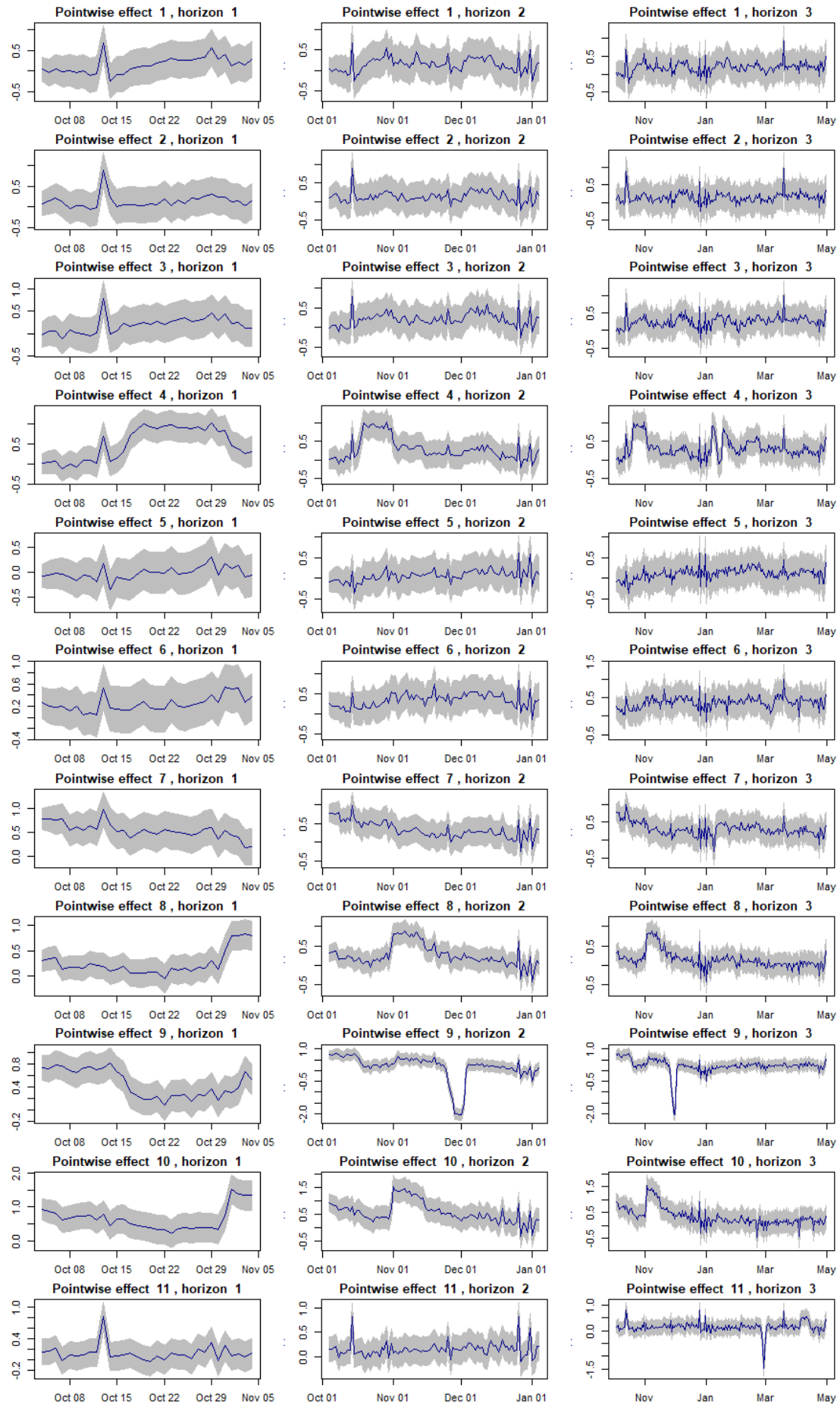
**Figure 25: Competitor brands.** Time series of unit sold daily, price per unit and autocorrelation function for the 10 competitor brands. The red vertical bar indicates the intervention date.



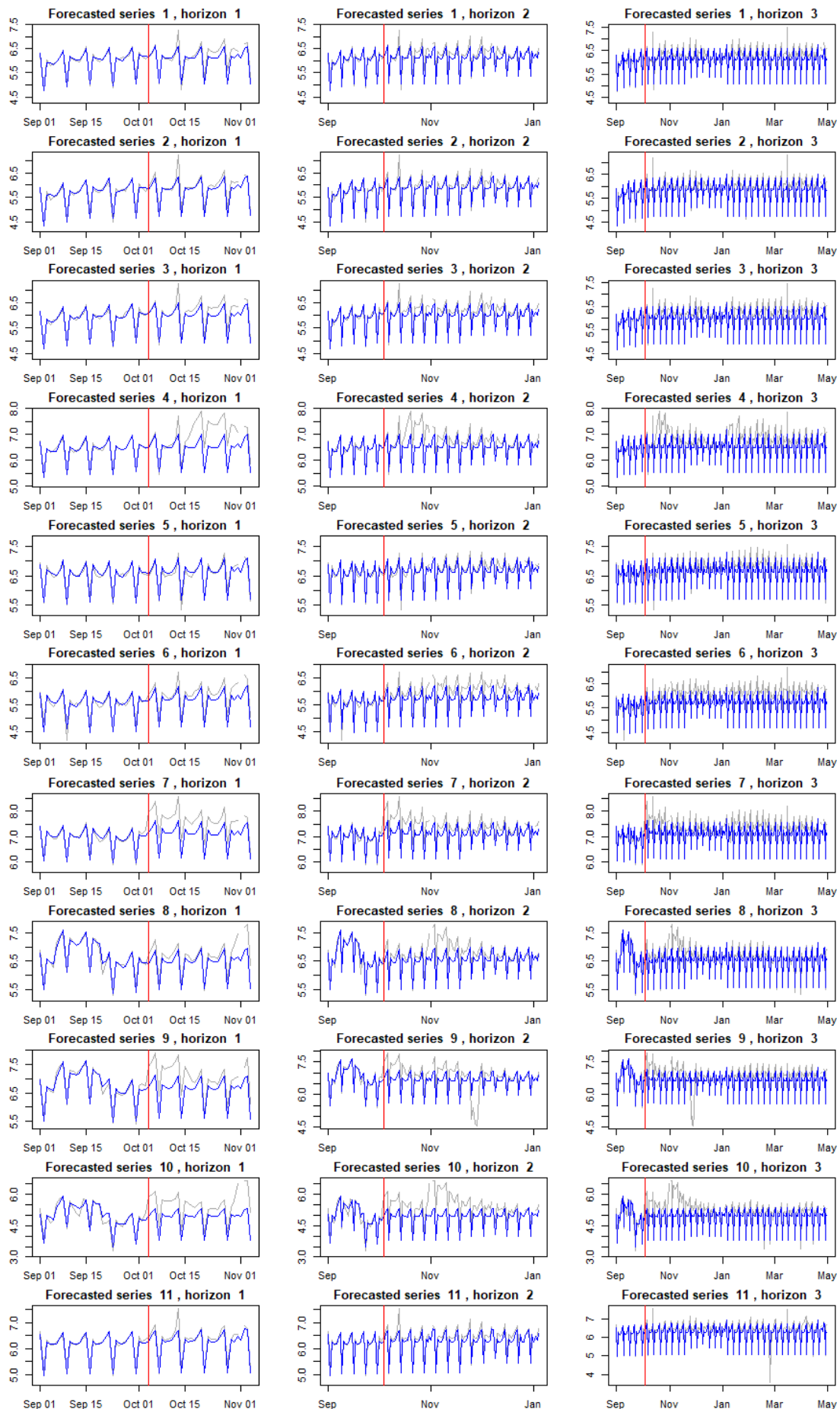
**Figure 26: Store brands.** Residual diagnostics (autocorrelation functions and Normal QQ plots) of the C-ARIMA models fitted to the time series of units sold (in log scale).



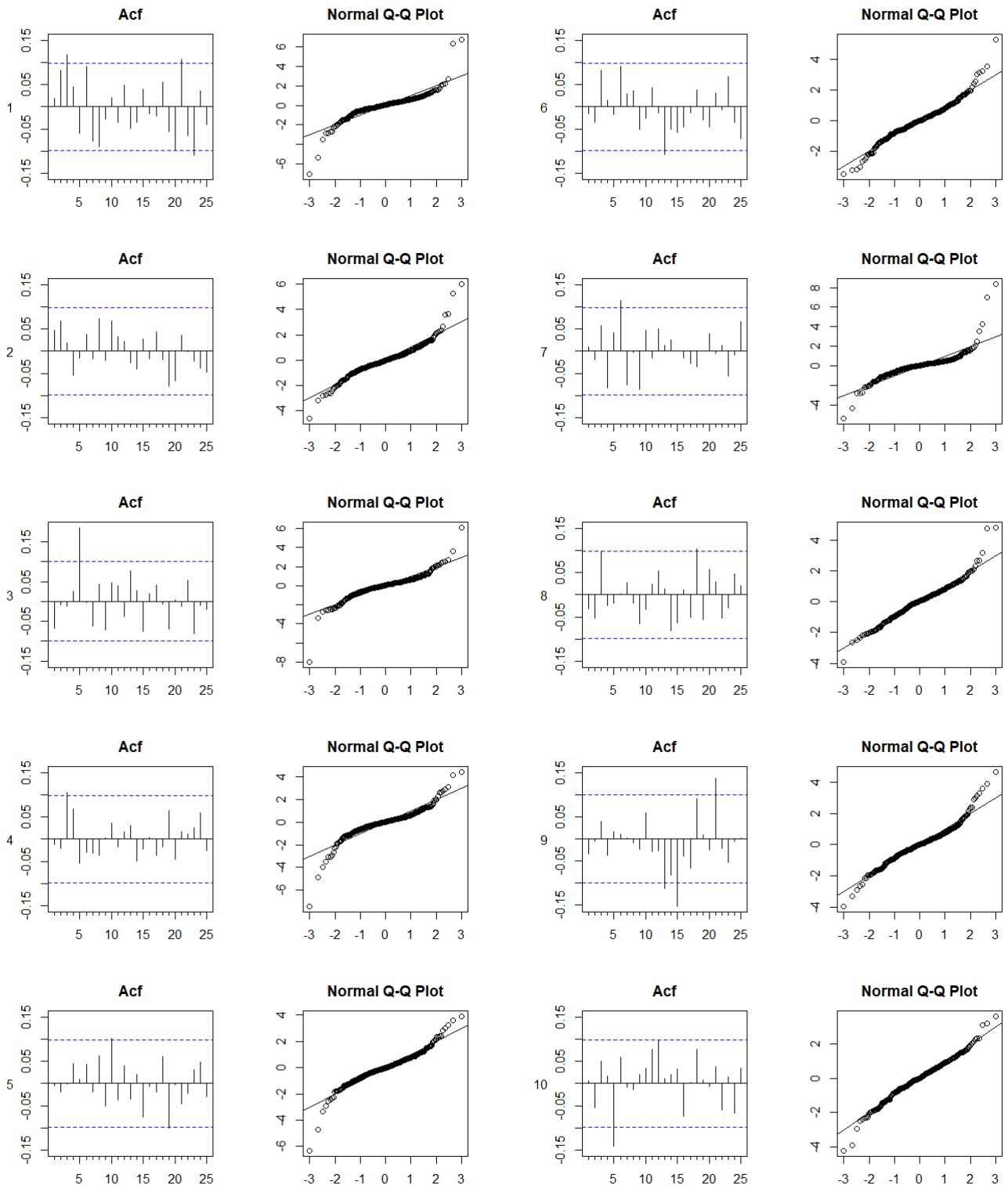
**Figure 27: Store brands.** Pointwise causal effect of the new price policy on the sales of store-brand products for each time horizon (1, 3 and 6 months) estimated via C-ARIMA (the dependent variable is the daily sales counts of each product in log scale).



**Figure 28: Store brands.** Observed sales (grey) and forecasted sales (blue) of each store brand and for each time horizon (1, 3 and 6 months). The red vertical bar indicates the intervention date.

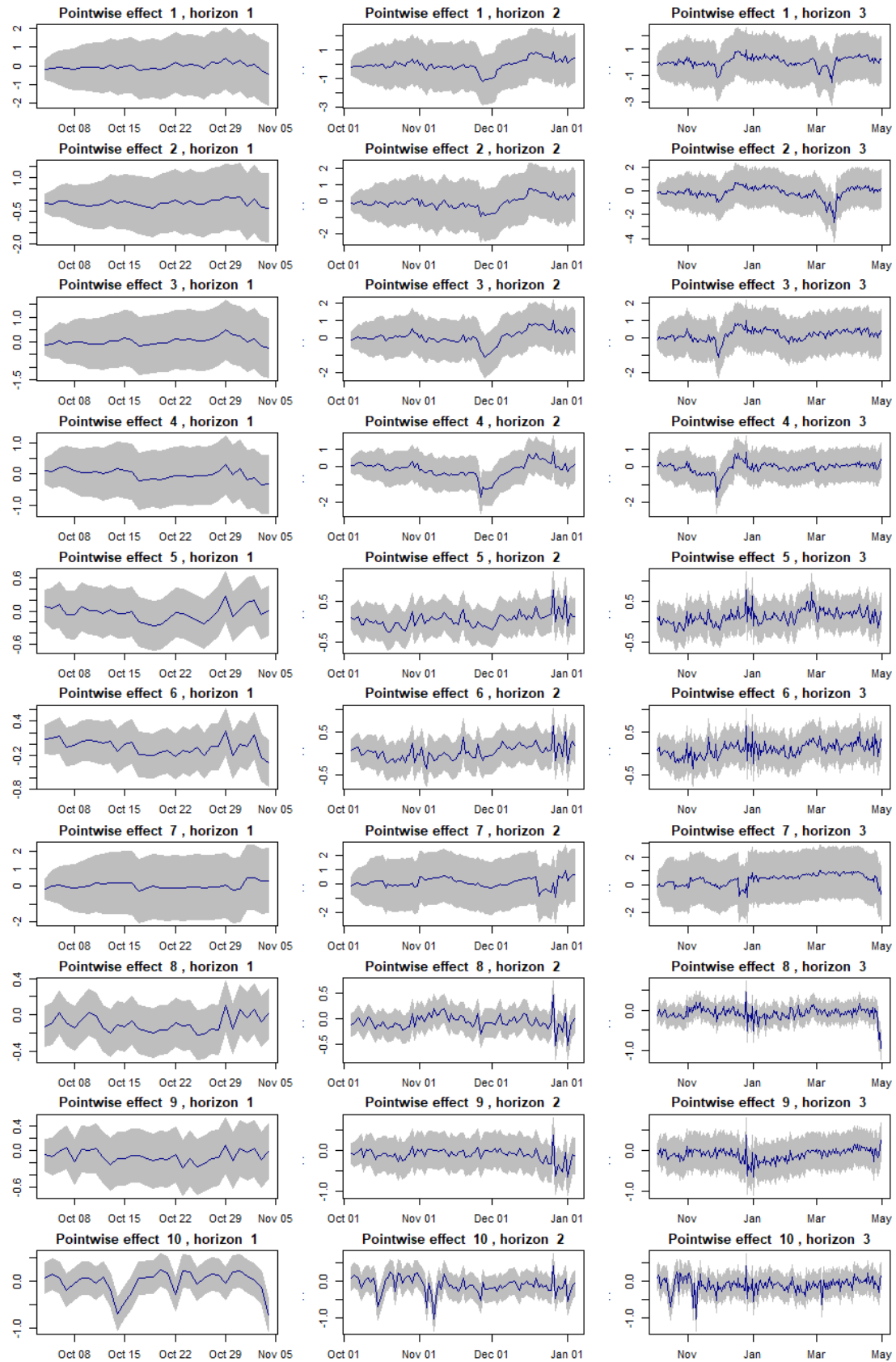


**Figure 29: Competitor brands.** Residual diagnostics (autocorrelation functions and Normal QQ plots) of the C-ARIMA models fitted to the time series of units sold (in log scale).

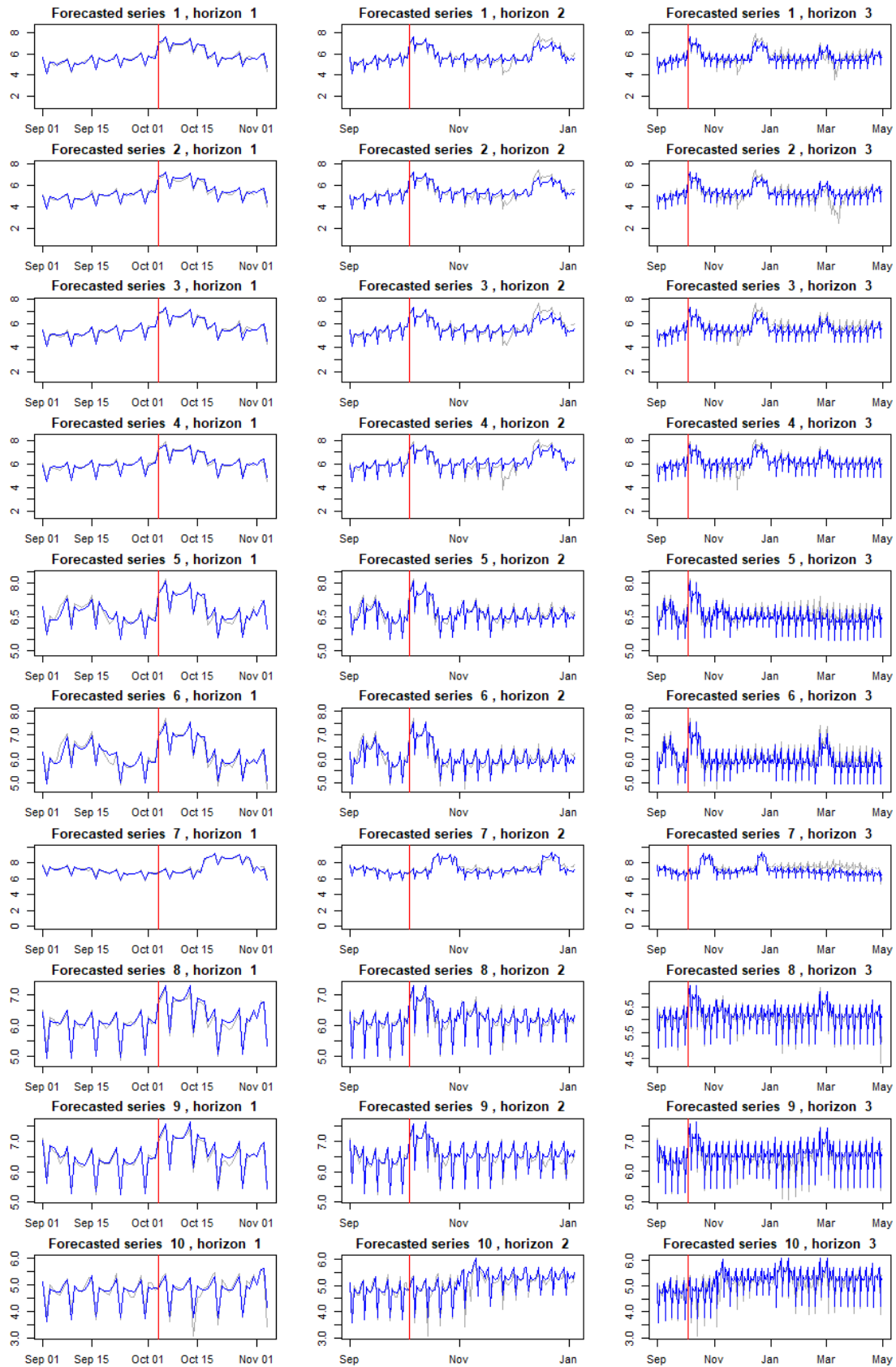




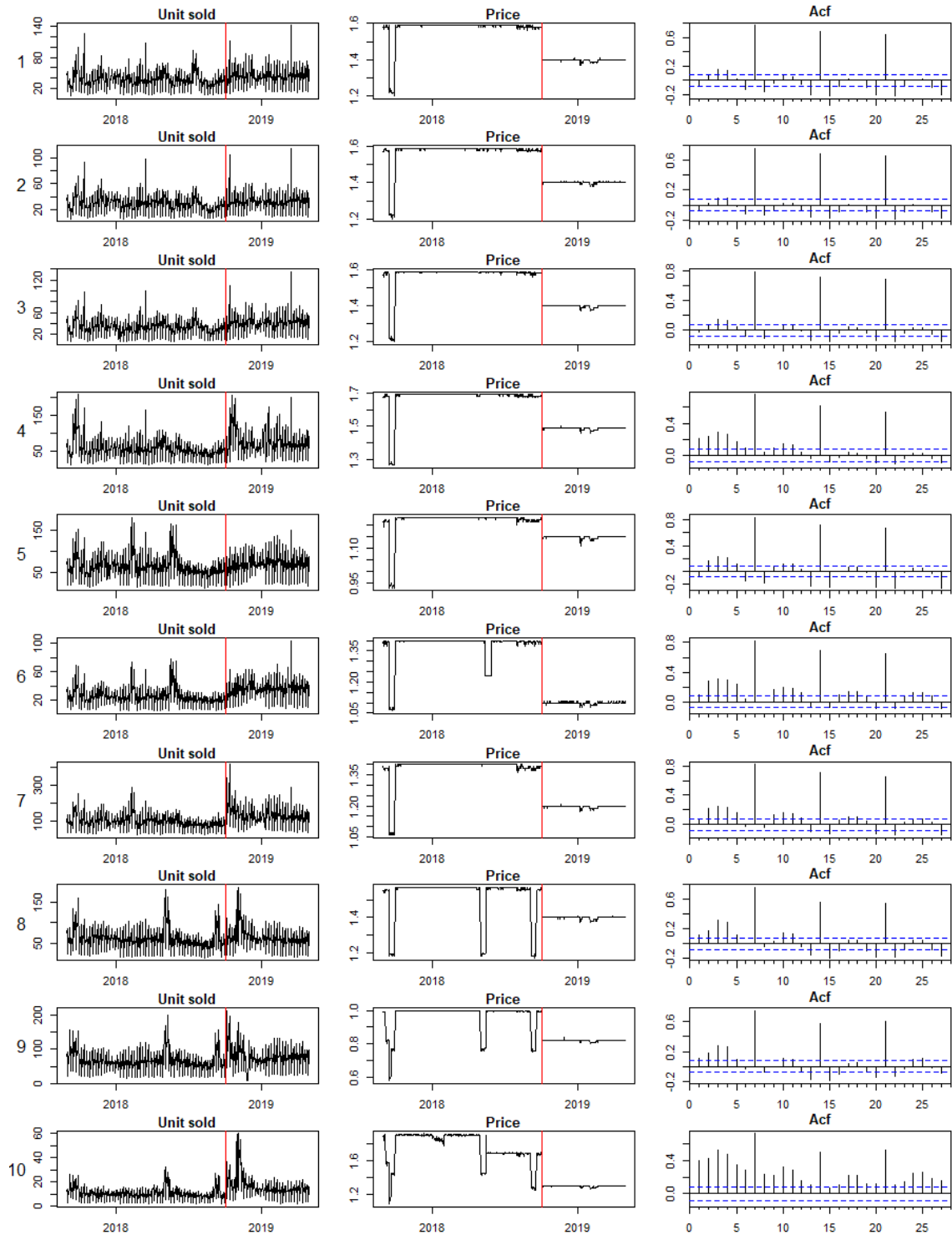
**Figure 30: Competitor brands.** Pointwise causal effect of the new price policy on the sales of competitor-brand products for each time horizon (1, 3 and 6 months) estimated via C-ARIMA (the dependent variable is the daily sales counts of each product in log scale).



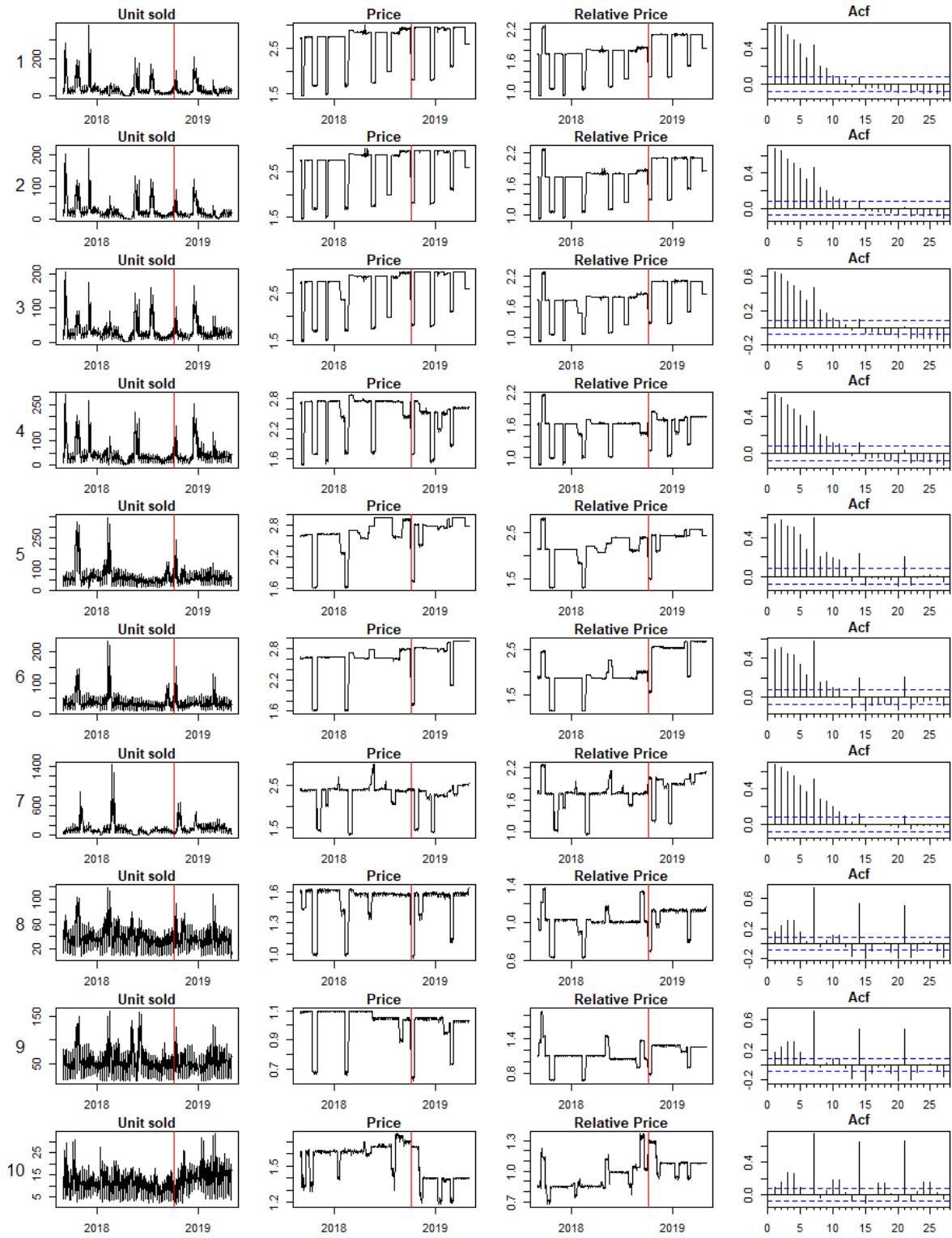
**Figure 31: Competitor brands.** Observed sales (grey) and forecasted sales (blue) of each competitor brand and for each time horizon (1, 3 and 6 months). The red vertical bar indicates the intervention date.



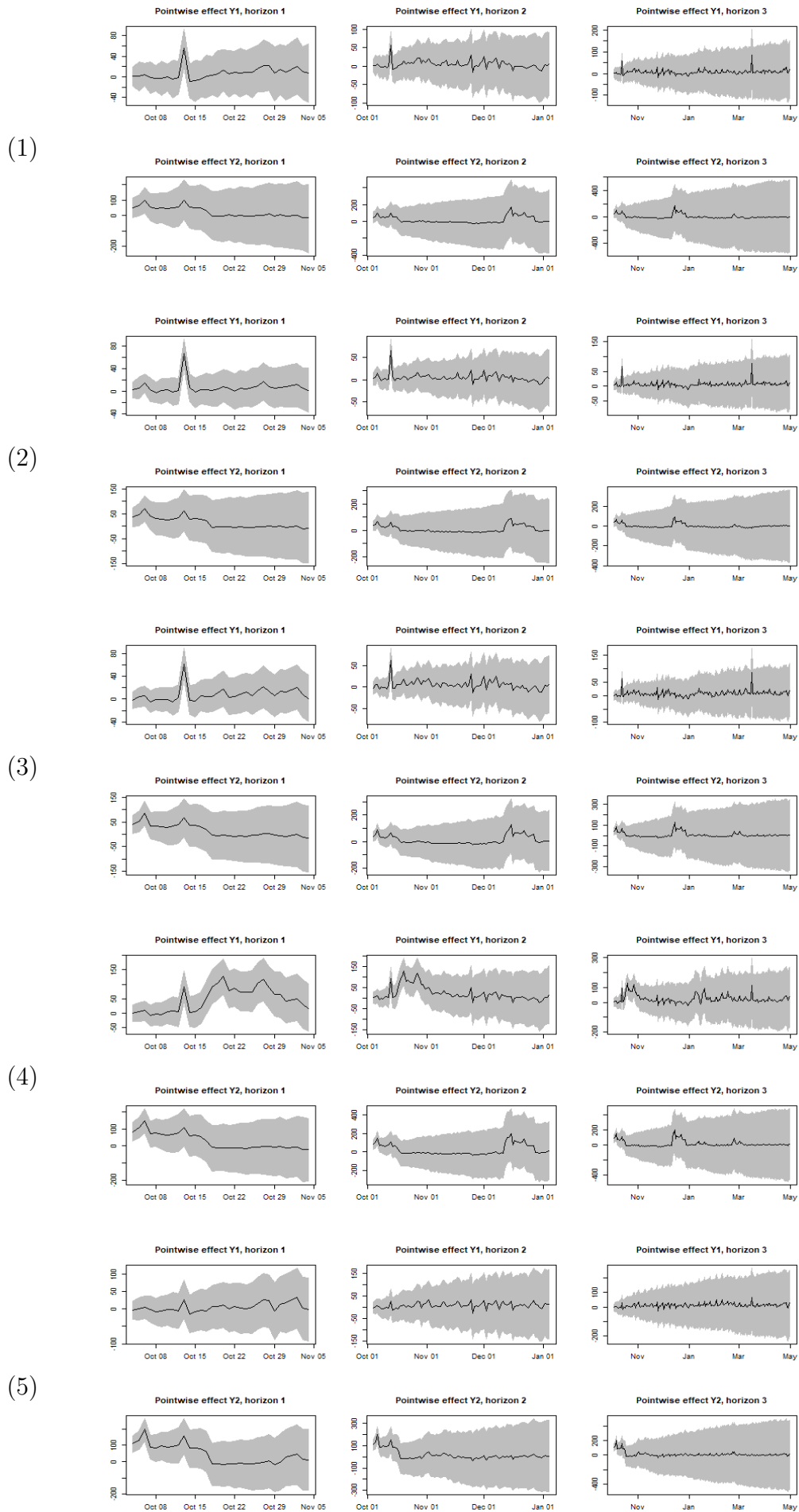
**Figure 32: Store brands.** Evolution of average sales per hour, price per unit and autocorrelation function for the 10 store brands in the pairs. The red vertical bar indicates the intervention date.



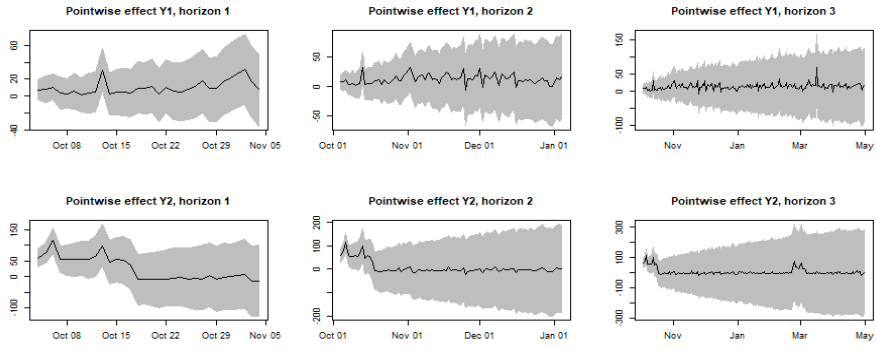
**Figure 33: Competitor brands.** Evolution of average sales per hour, price per unit and auto-correlation function for the 10 competitor brands in the pairs. The red vertical bar indicates the intervention date.



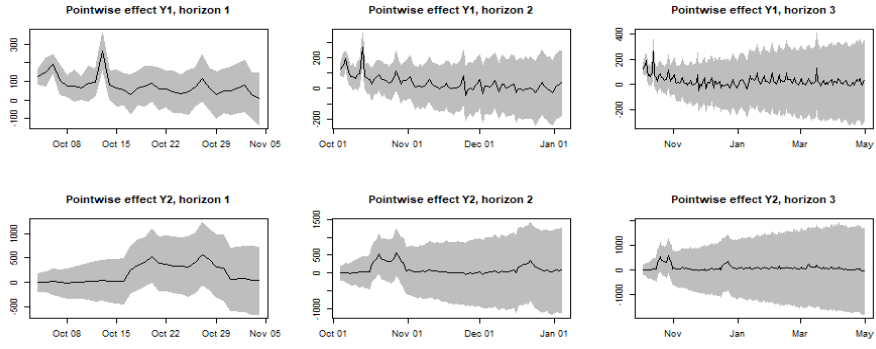
**Figure 34:** Pointwise causal effect of the permanent price reduction on each store-competitor pair at 1 month, 3 months and 6 months after the intervention.



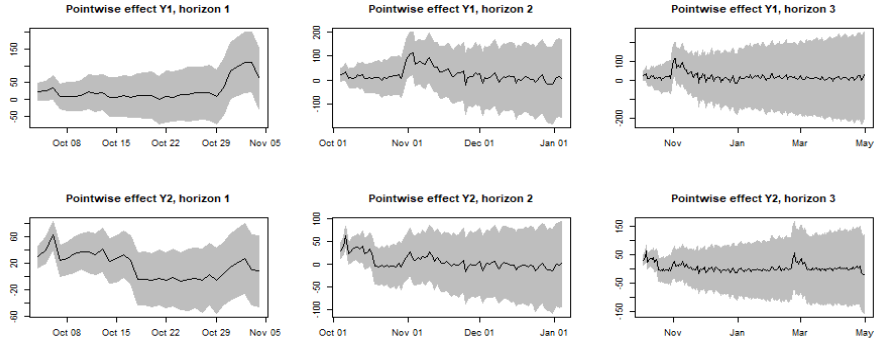
(6)



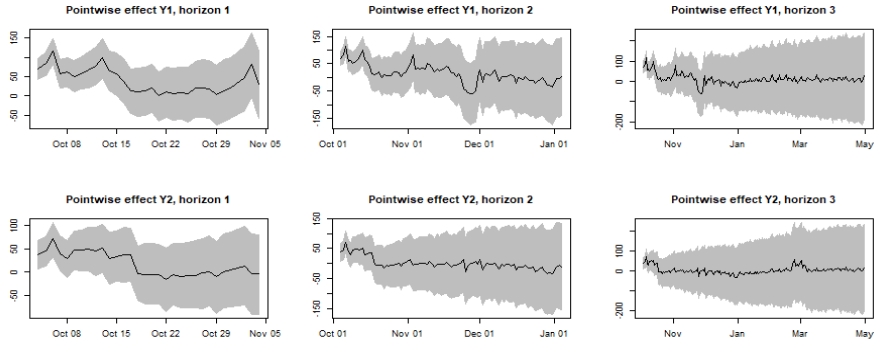
(7)



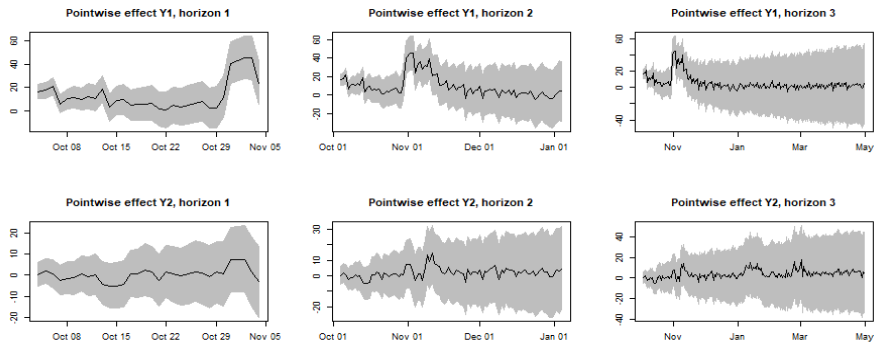
(8)



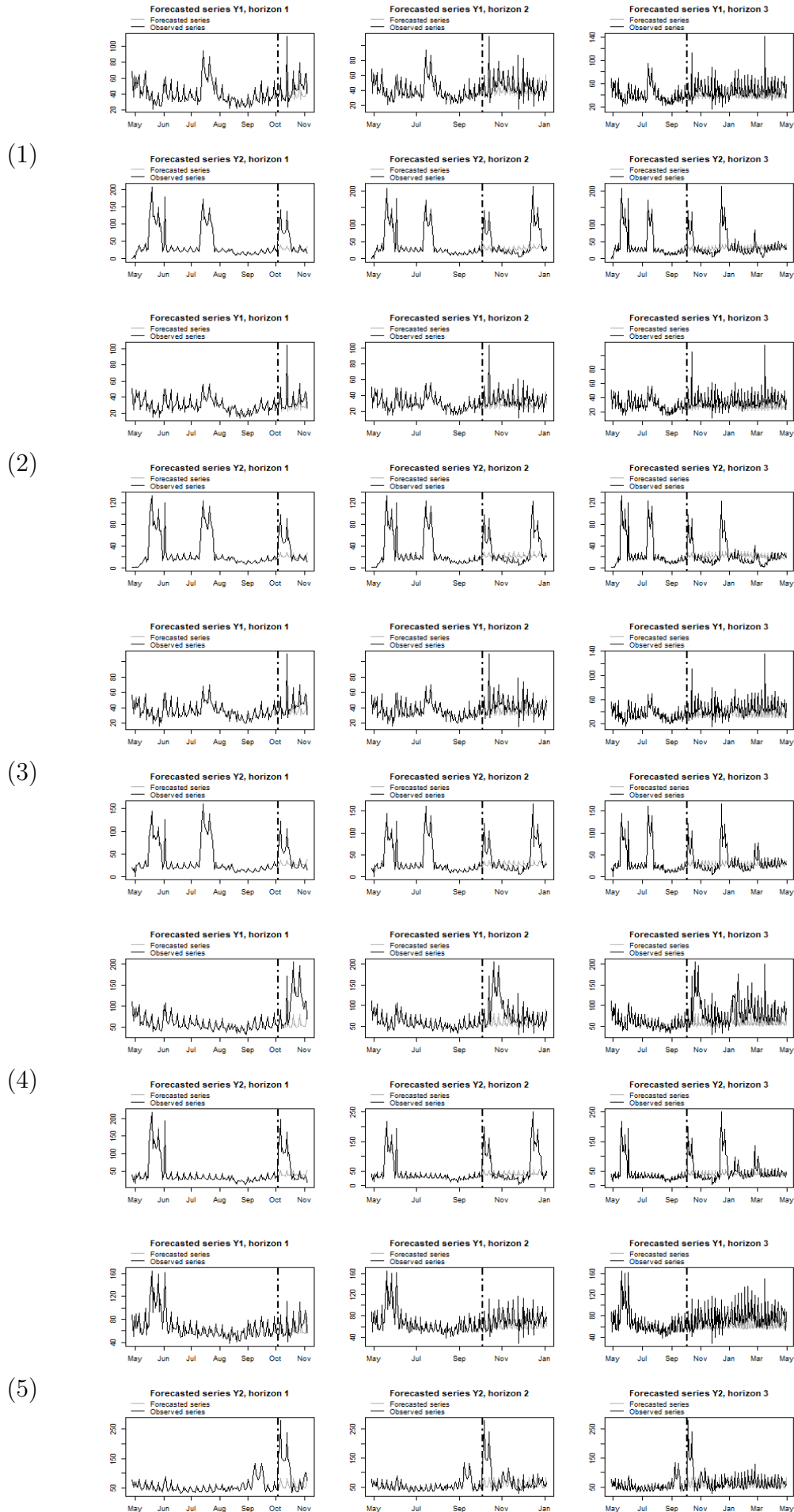
(9)



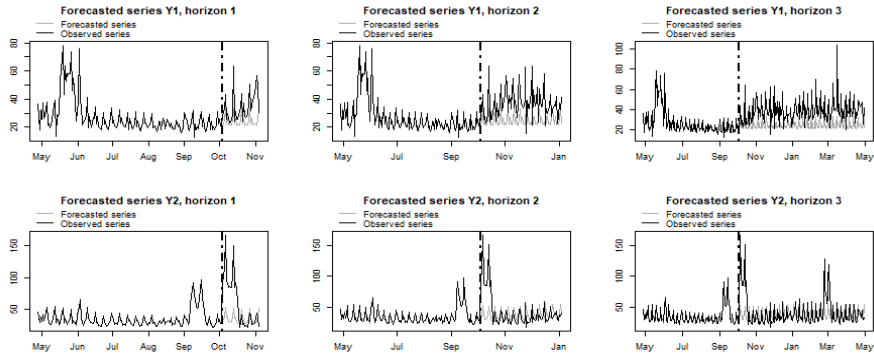
(10)



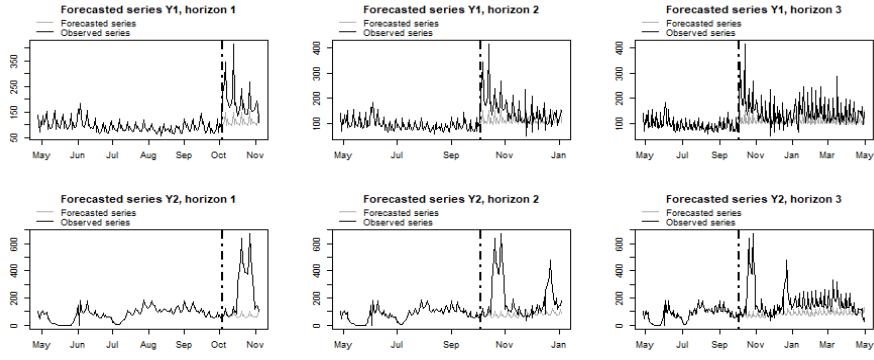
**Figure 35:** For each store-competitor pair, observed outcome (in grey) plotted against the counterfactual outcome in the absence of intervention (in blue) after 1 month, 3 months and 6 months from the intervention, indicated by the red vertical line.



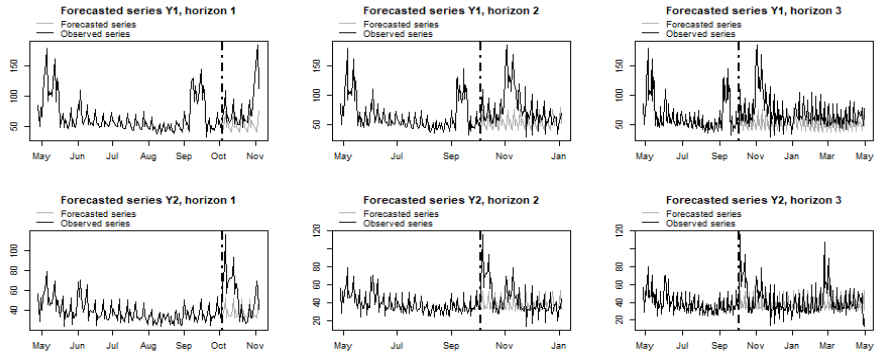
(6)



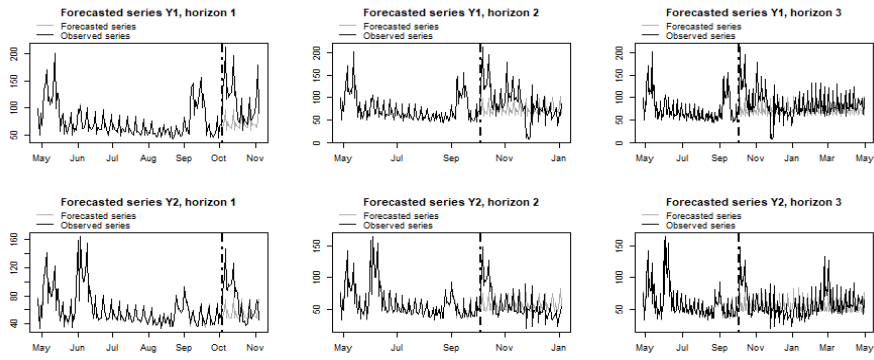
(7)



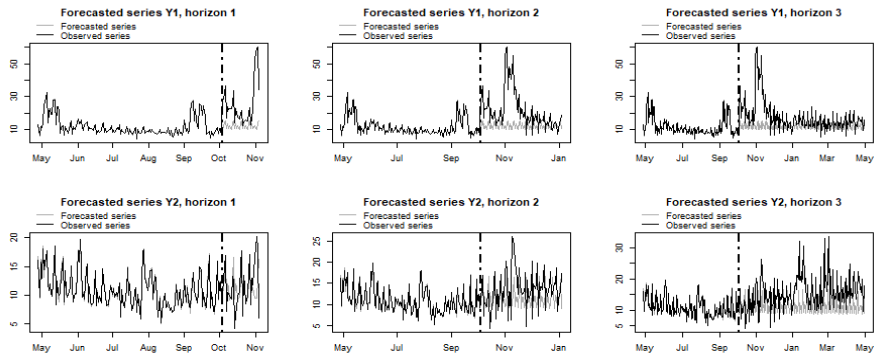
(8)



(9)

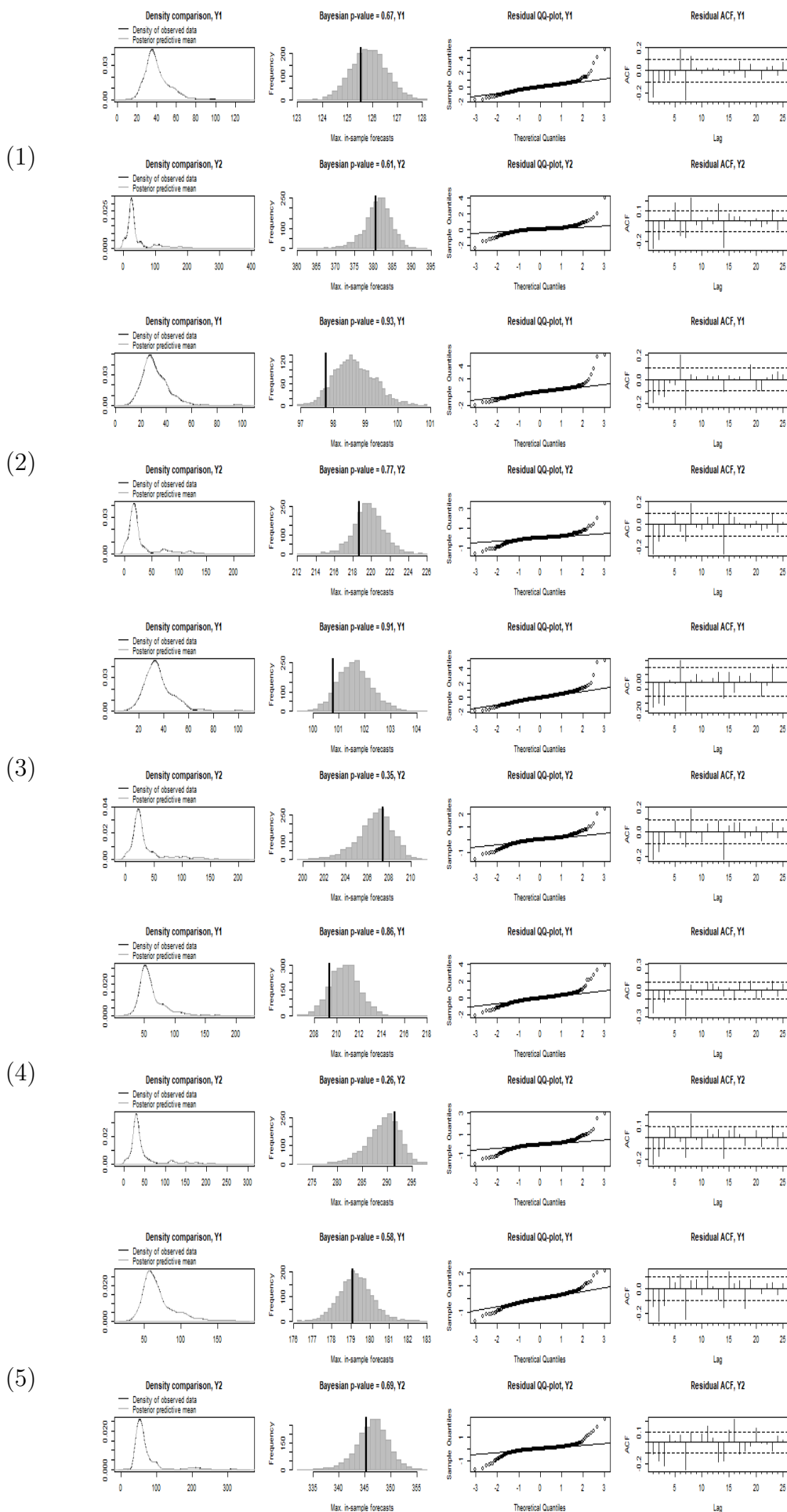


(10)

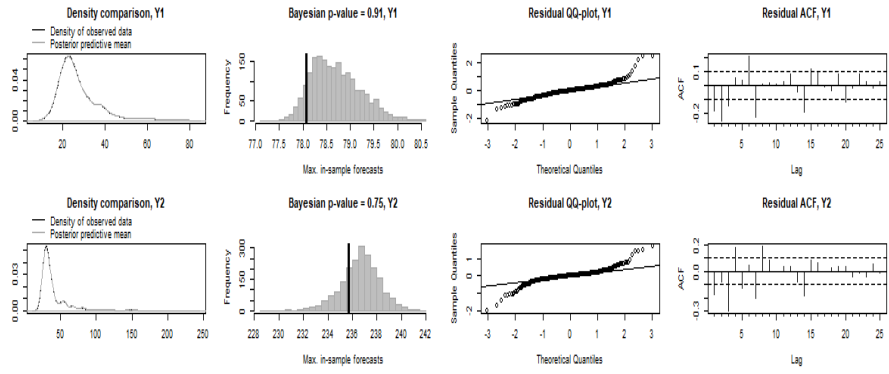




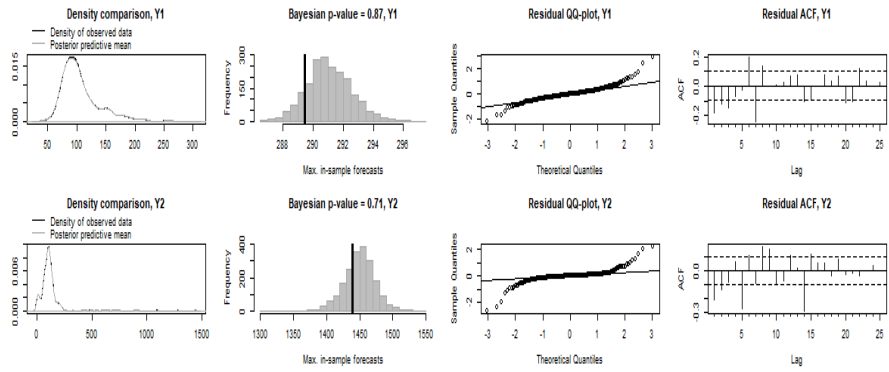
**Figure 36:** Posterior predictive checks for each pair. Starting from the left: i) density of observed data (black) plotted against the posterior predictive mean (blue); ii) observed maximum compared to the distribution of the maximum from the posterior draws; iii) Normal QQ-Plot of standardized residuals; iv) autocorrelation function of standardized residuals.



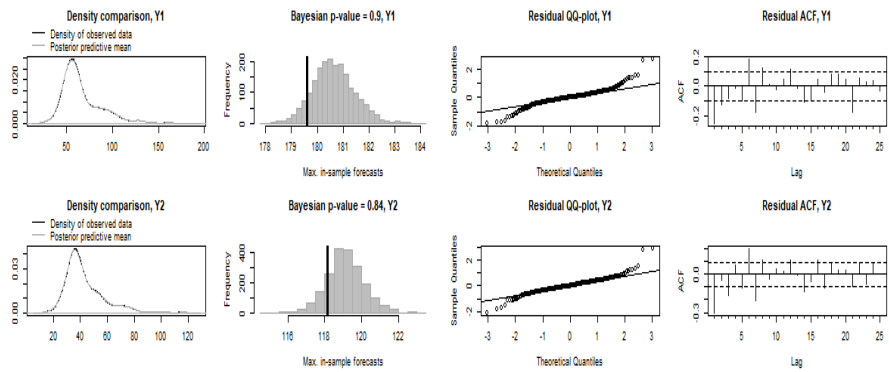
(6)



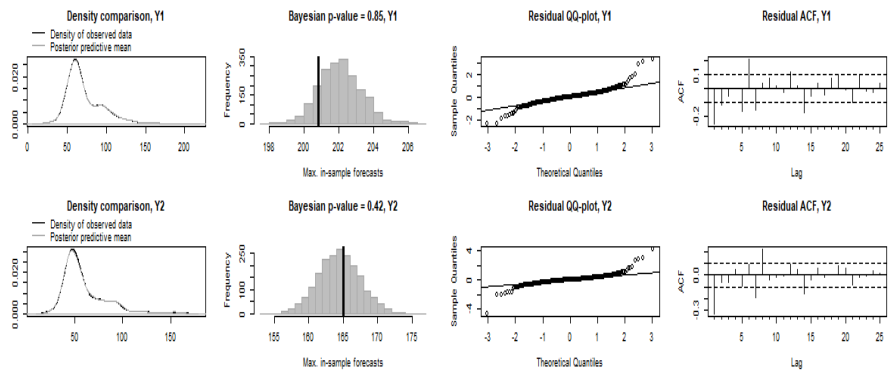
(7)



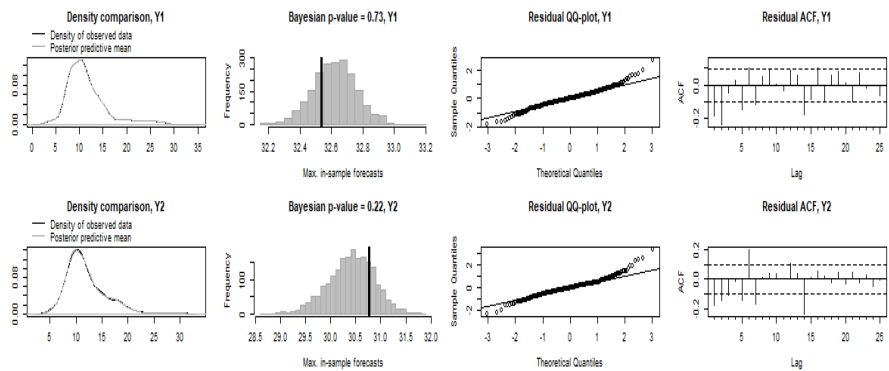
(8)



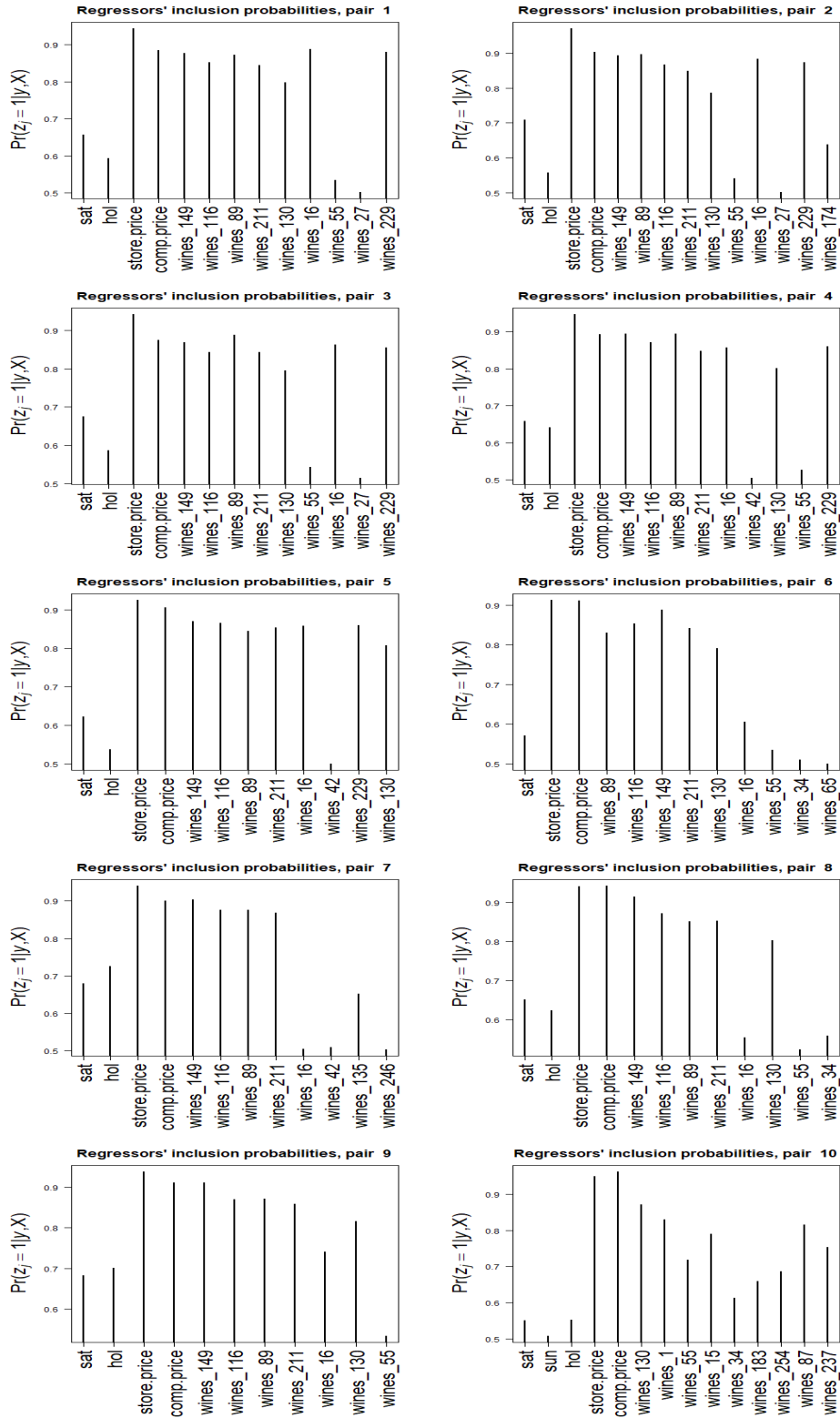
(9)



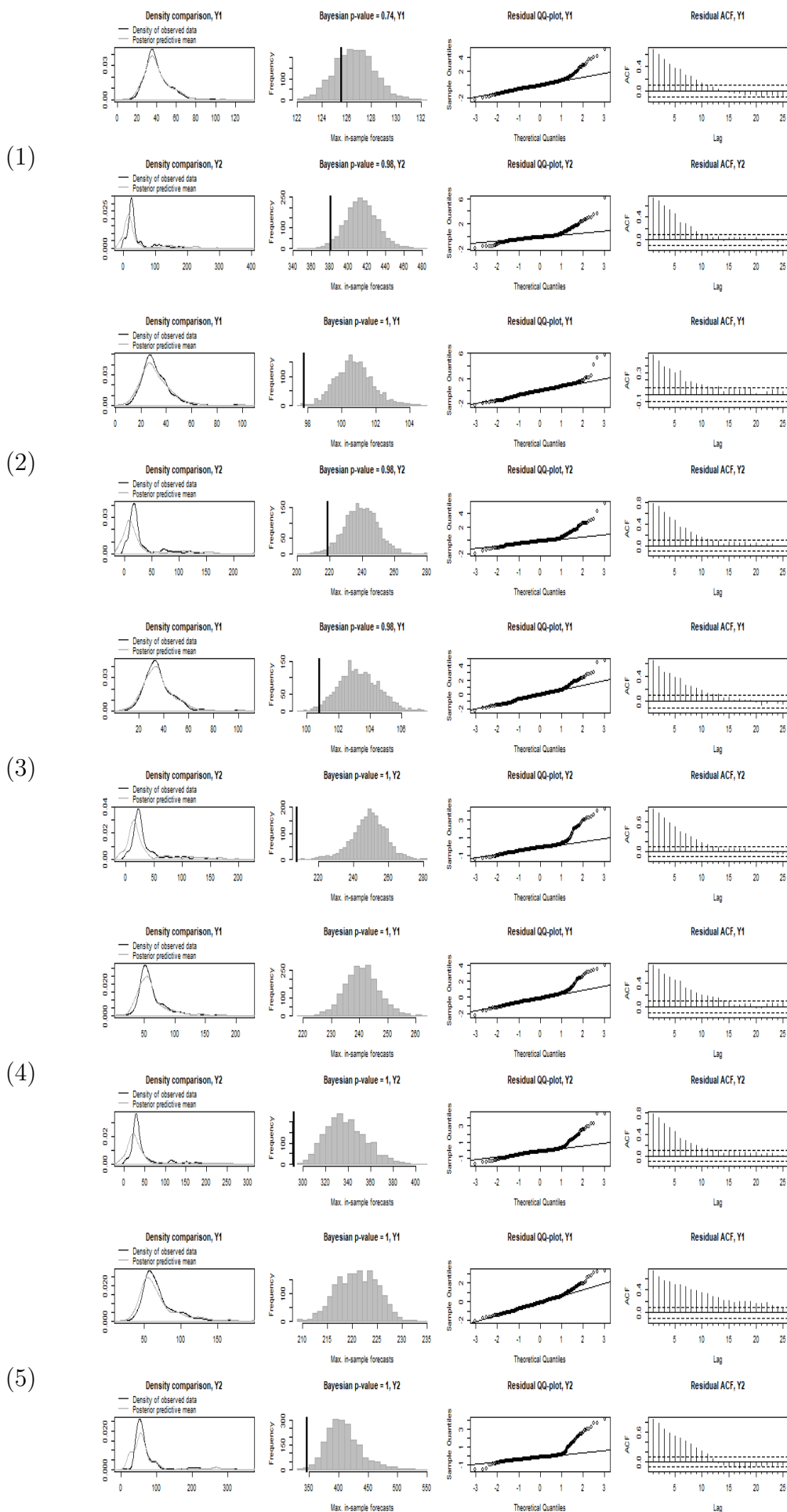
(10)



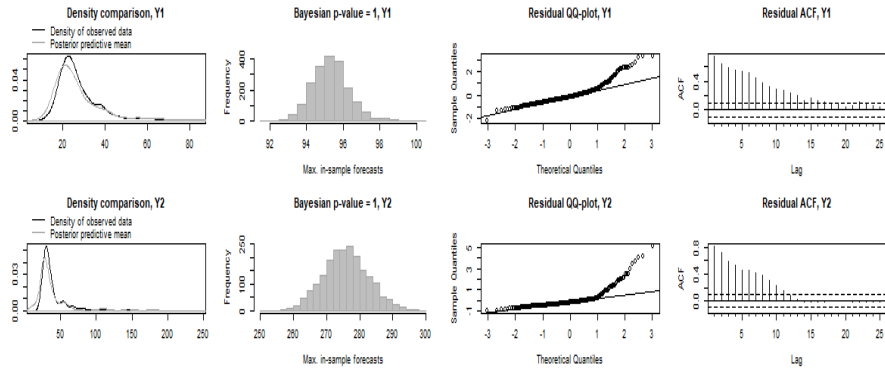
**Figure 37:** Inclusion probabilities above the 0.5 threshold of the regressors included in the MBSTS models estimated on each store-competitor pair.



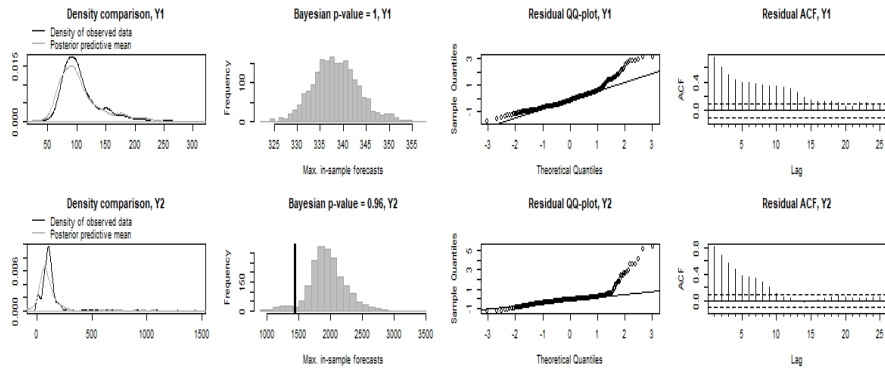
**Figure 38:** Posterior predictive checks for a seasonal MBSTS model. Starting from the left: i) density of observed data (black) vs posterior predictive mean (blue); ii) observed maximum vs distribution of the maximum from the posterior draws; iii) Normal QQ-Plot of standardized residuals; iv) autocorrelation function of standardized residuals.



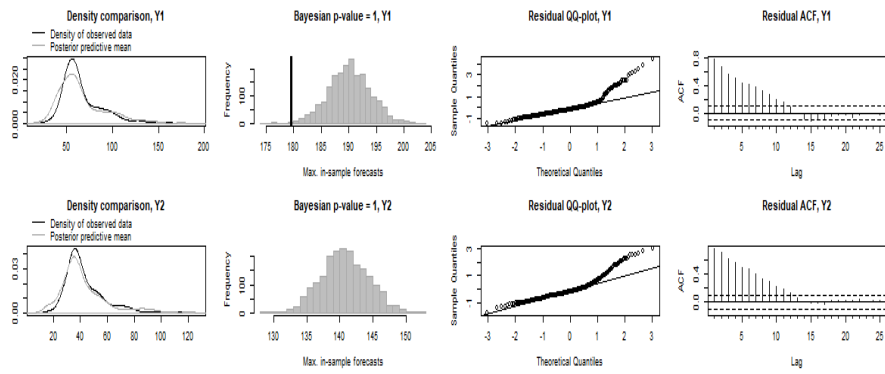
(6)



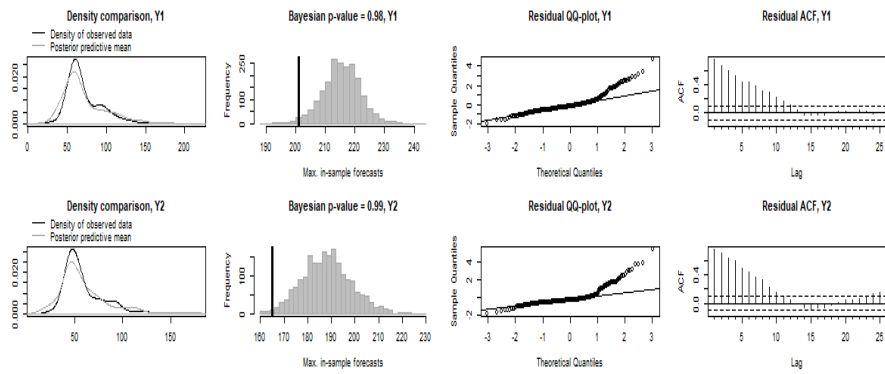
(7)



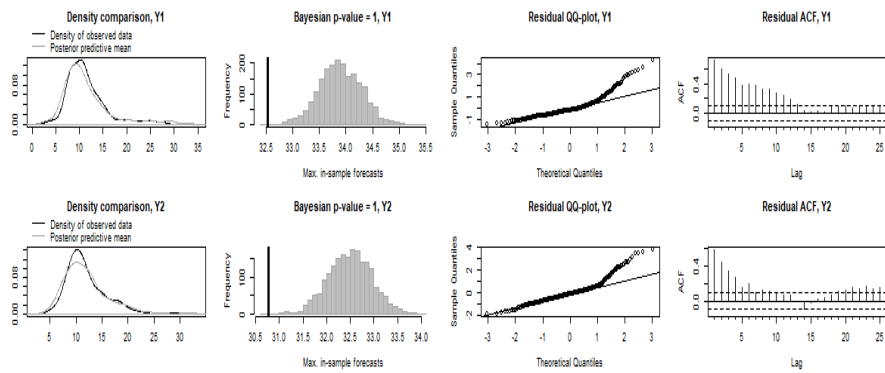
(8)



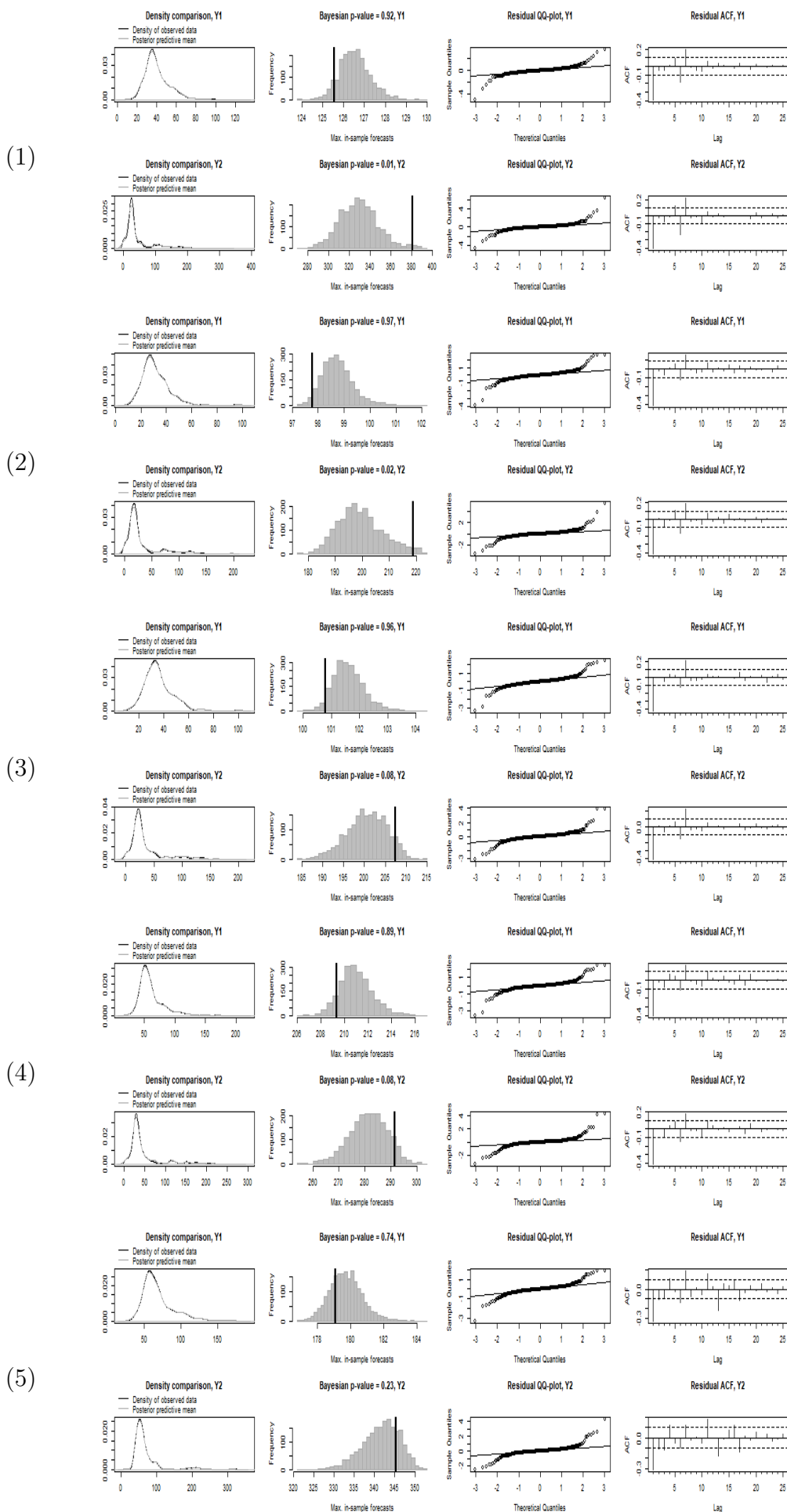
(9)



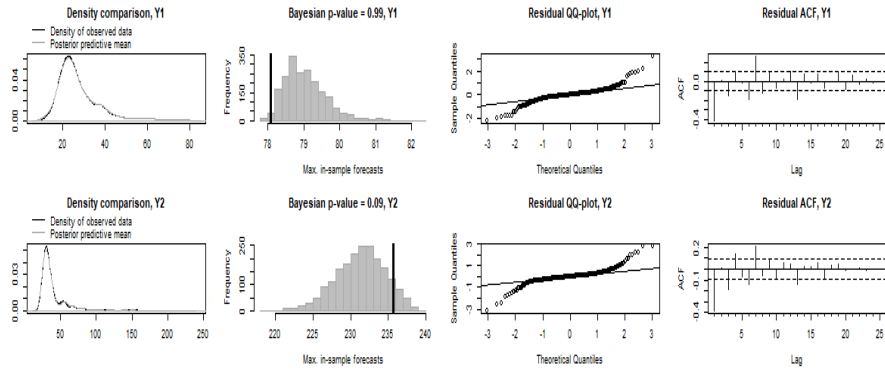
(10)



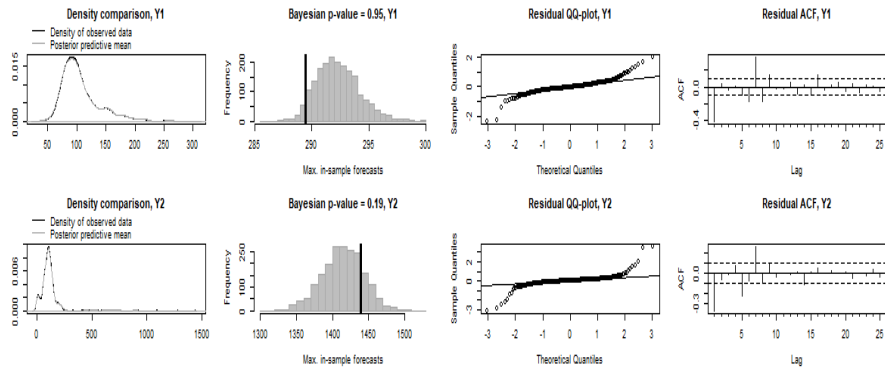
**Figure 39:** Posterior predictive checks for a trend MBSTS model. Starting from the left: i) density of observed data (black) plotted against the posterior predictive mean (blue); ii) observed maximum compared to the distribution of the maximum from the posterior draws; iii) Normal QQ-Plot of standardized residuals; iv) autocorrelation function of standardized residuals.



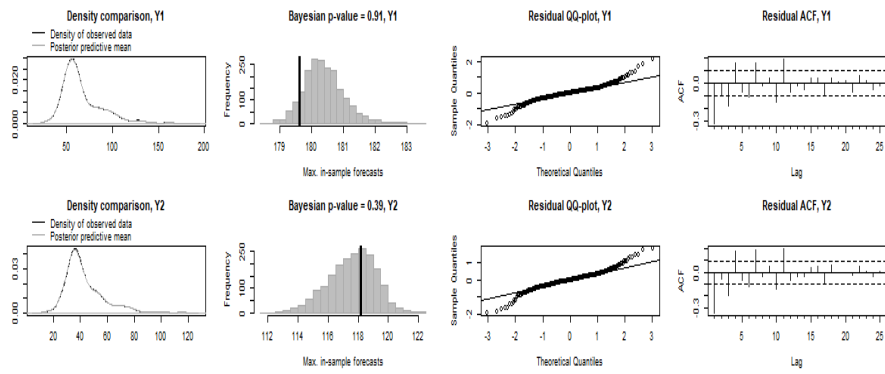
(6)



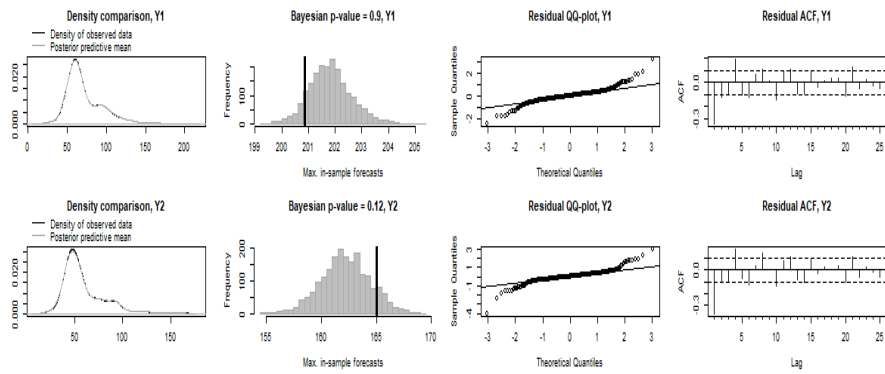
(7)



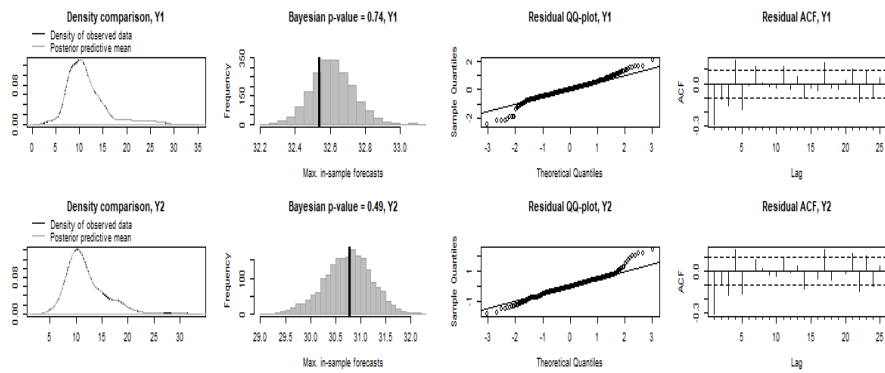
(8)



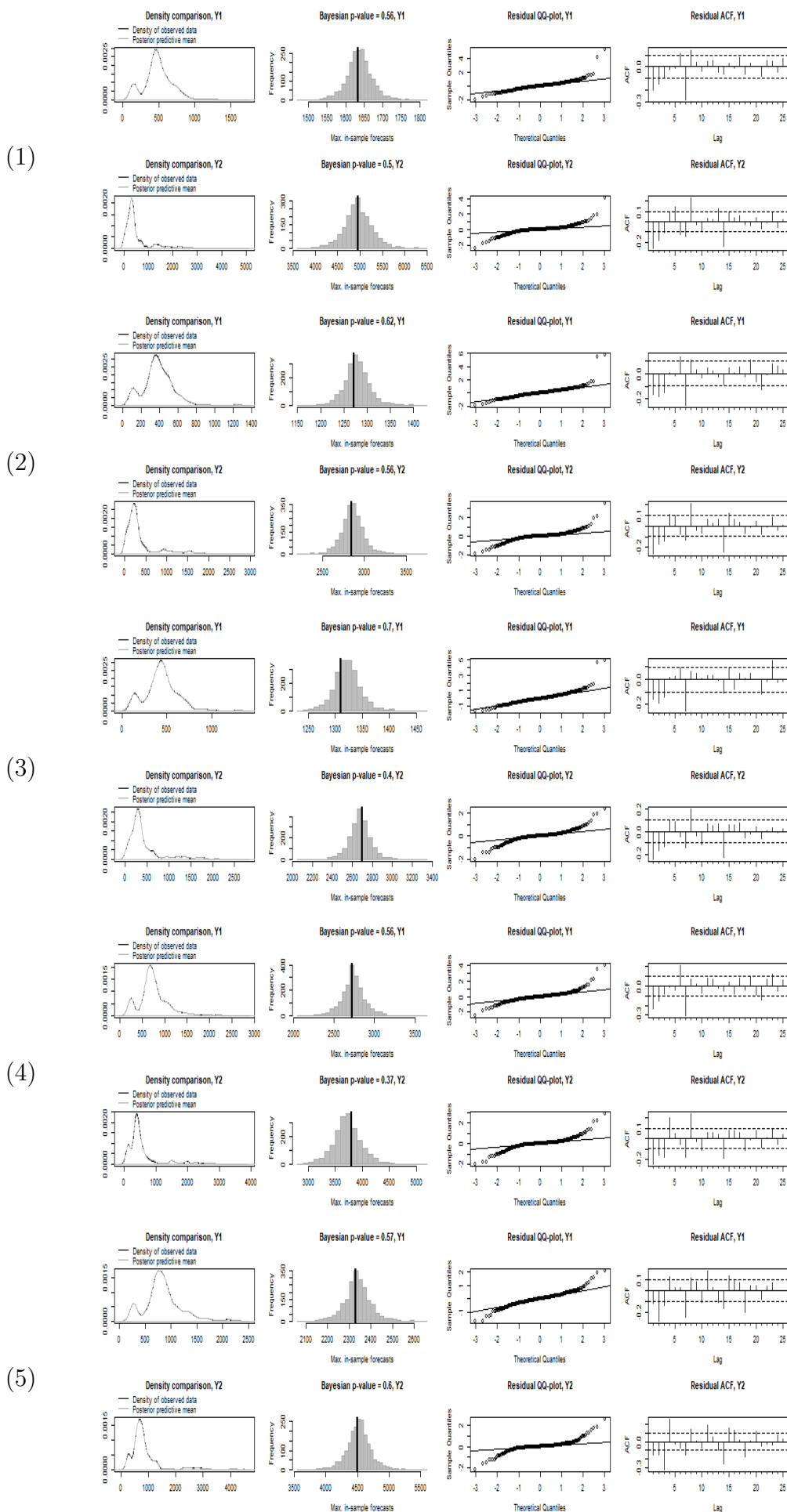
(9)



(10)

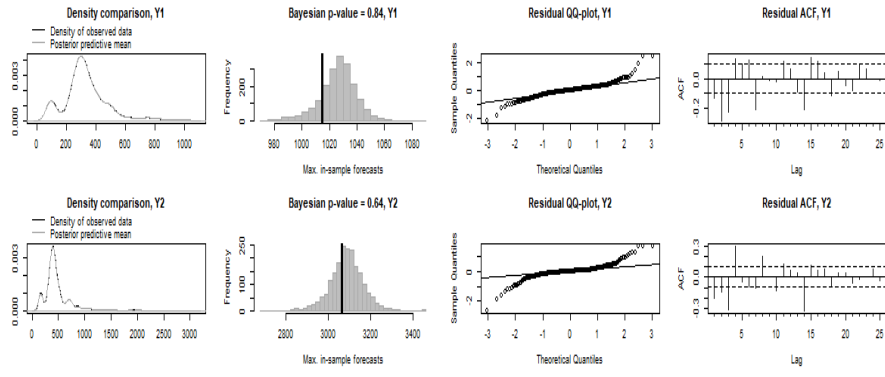


**Figure 40:** Posterior predictive checks for a trend and seasonal MBSTS model estimated on the daily units sold. Starting from the left: i) density of observed data (black) plotted against the posterior predictive mean (blue); ii) observed maximum compared to the distribution of the maximum from the posterior draws; iii) Normal QQ-Plot of standardized residuals; iv) autocorrelation function of standardized residuals.

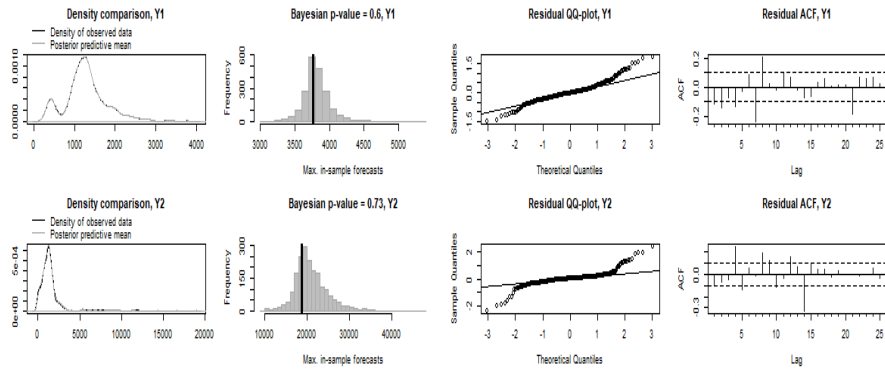




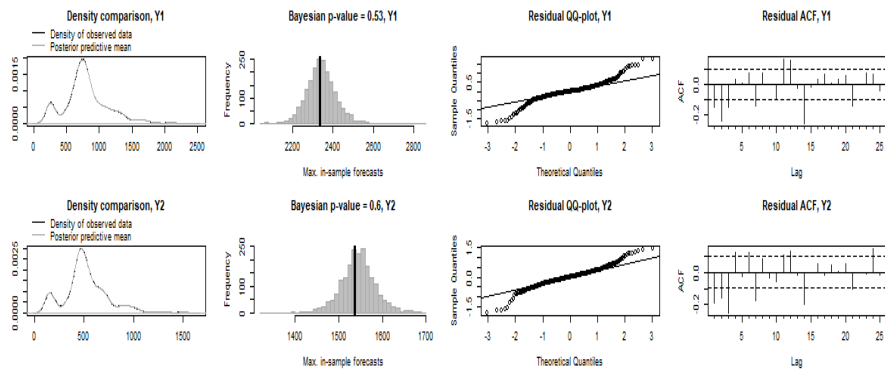
(6)



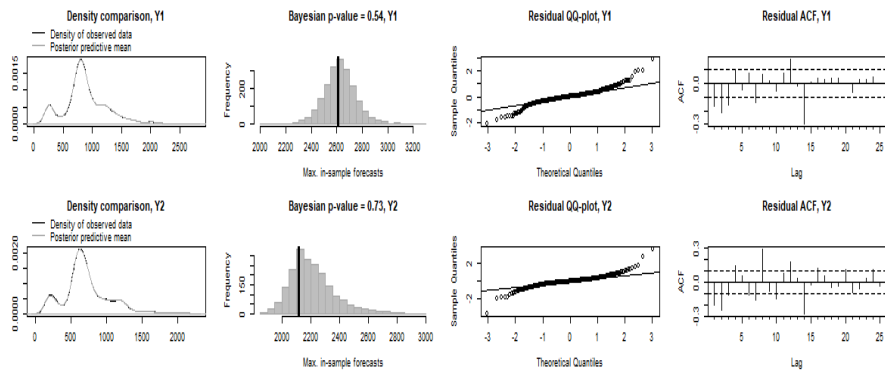
(7)



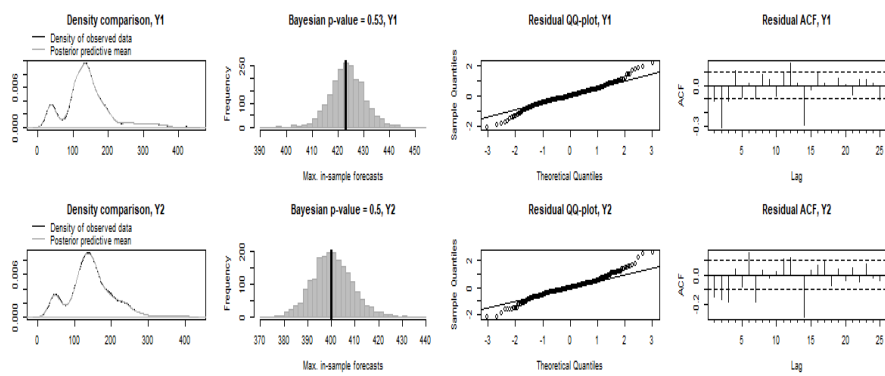
(8)



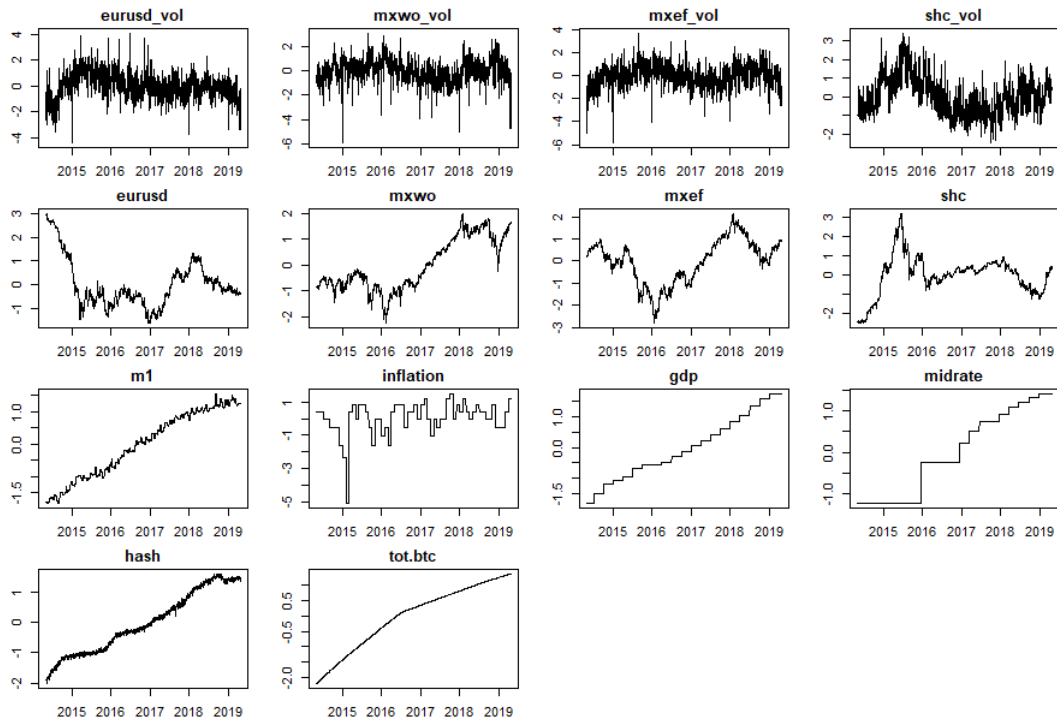
(9)



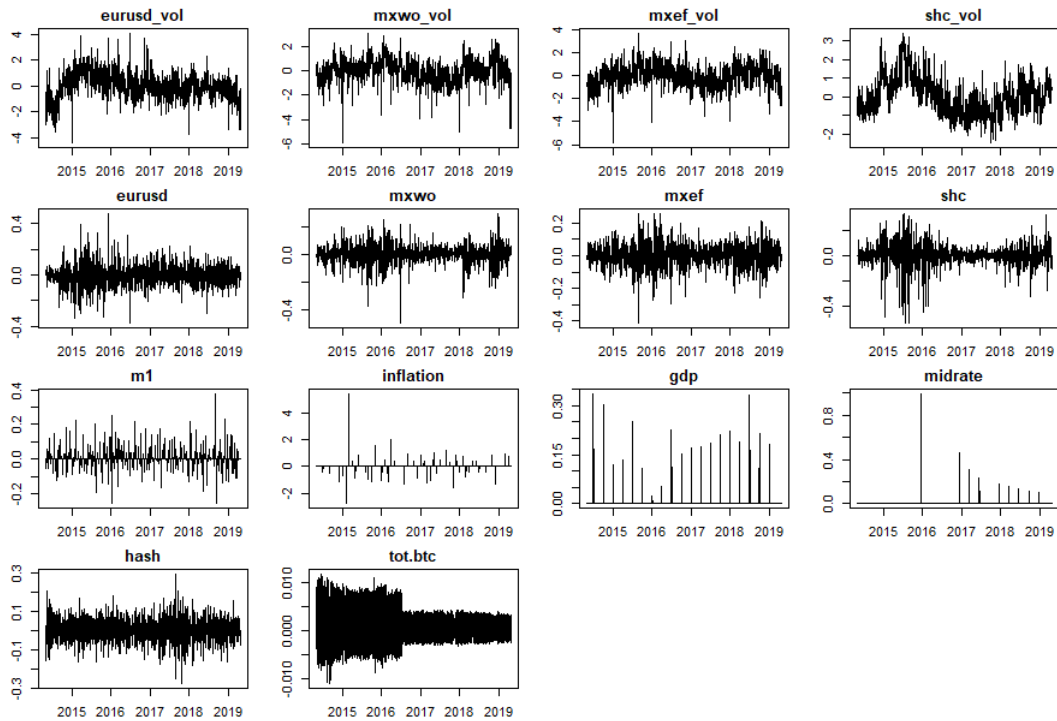
(10)



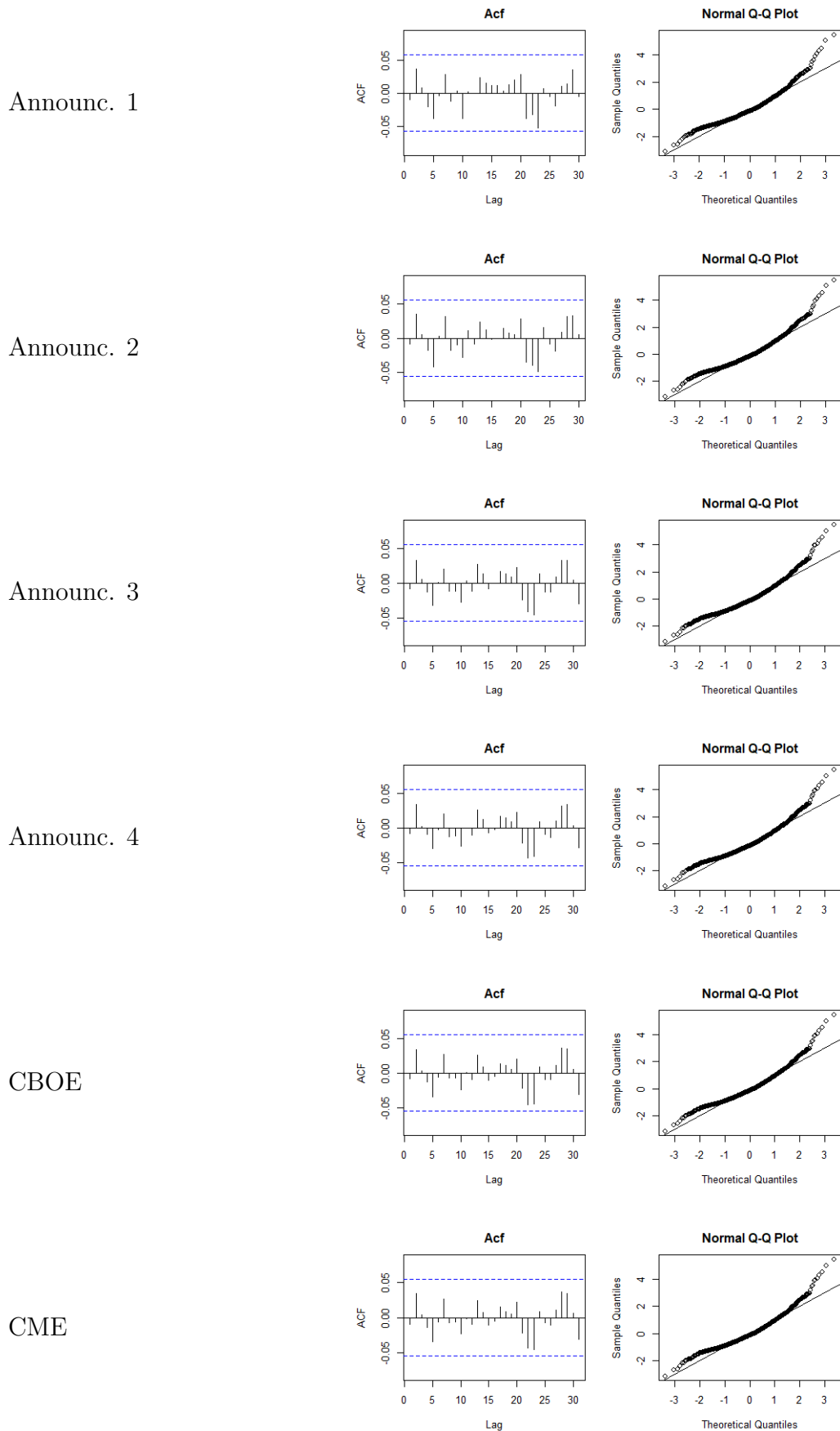
**Figure 41:** Evolution of the covariates included in the analysis before any transformation



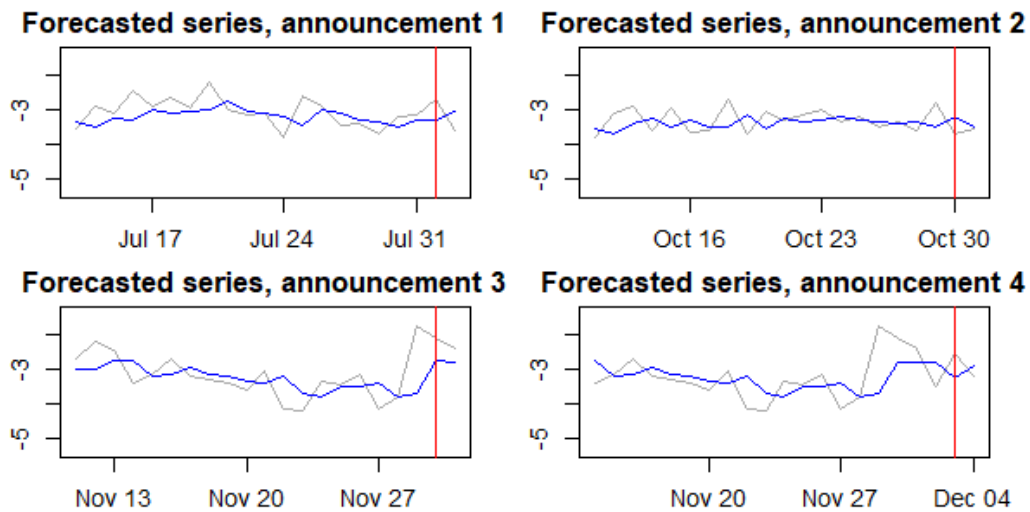
**Figure 42:** Evolution of the covariates included in the analysis, scaled and made stationary



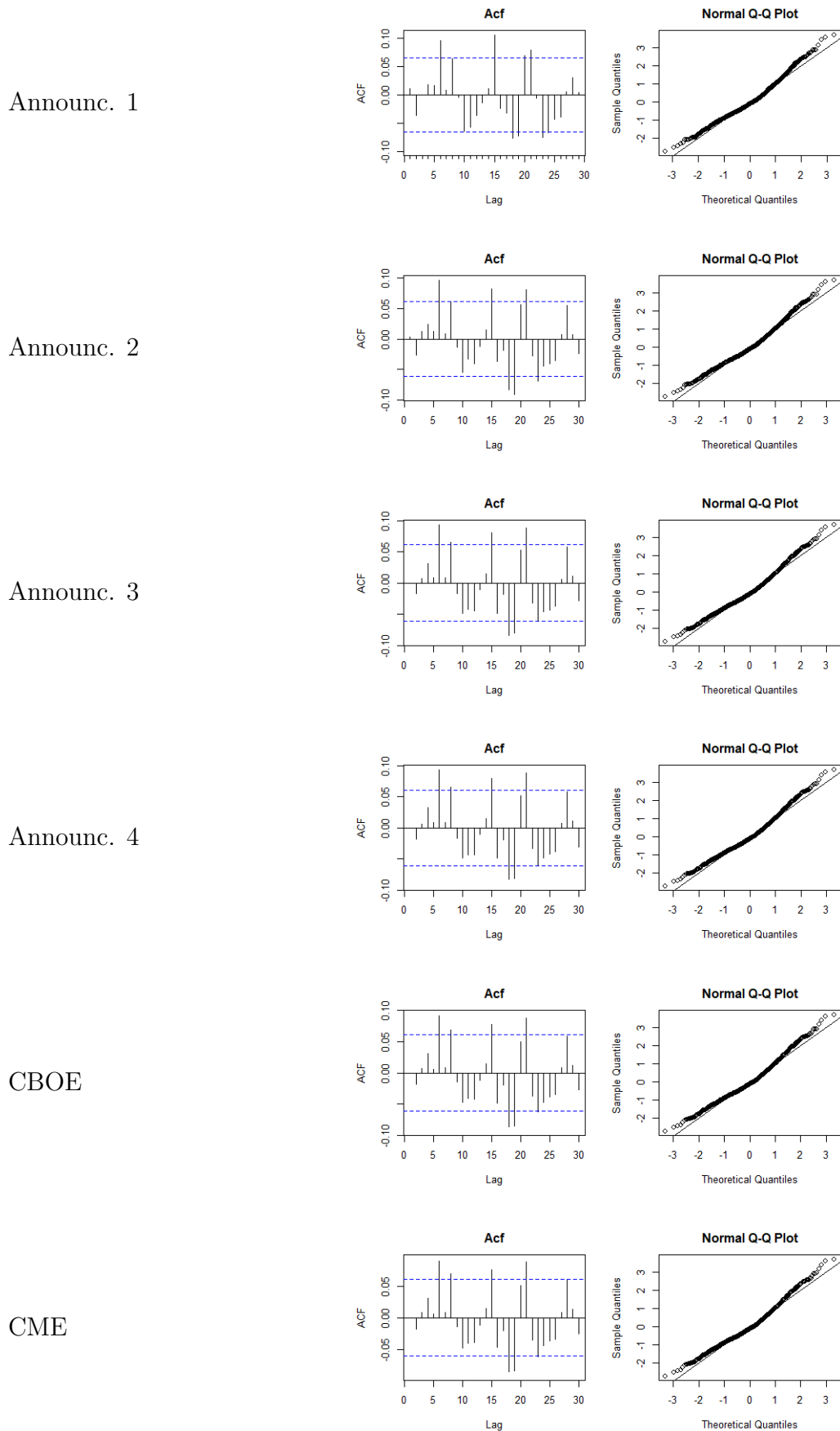
**Figure 43:** Residuals diagnostics (autocorrelation function and Normal Q-Q Plot) of the six independent C-SARIMA models fitted to the **Garman-Klass volatility proxy** (in log scale) up to the day preceding each intervention (four announcements and two futures).



**Figure 44:** Contemporaneous causal effect of the four announcements, computed as the difference, at the intervention date, between the observed **Garman-Klass volatility proxy** (in log scale) (the gray line after the vertical bar) and the 1-step ahead forecast (the blue line after the bar). The red vertical bar indicates the day before the intervention.



**Figure 45:** Residuals diagnostics (autocorrelation function and Normal Q-Q Plot) of the six independent C-SARIMA models fitted to **Bitcoin daily volumes (log scale)** up to the day preceding each intervention (four announcements and two futures).



## B

### B.1 Proof of relations (10), (11) and (12)

Considering the setup formalized in Section 3.3, the  $k$ -step prediction error becomes

$$S_{t^*+k}(\mathbf{w}) - \hat{S}_{t^*+k}(\mathbf{w}') = z_{t^*+k}(\mathbf{w}) + \tau_{t^*+k}(\mathbf{w}; \mathbf{w}') - \hat{z}_{t^*+k|t^*}(\mathbf{w}') = \sum_{i=0}^{k-1} \psi_i \varepsilon_{t^*+k-i} + \tau_{t^*+k}(\mathbf{w}, \mathbf{w}'), \quad (56)$$

Where the last expression comes from the well known relationship that the forecasting error at lag  $k$  is a  $MA(k-1)$ . Thus, for a single  $k$ , (56) implies that

$$S_{t^*+k}(\mathbf{w}) - \hat{S}_{t^*+k}(\mathbf{w}') \sim \left[ \tau_{t^*+k}(\mathbf{w}; \mathbf{w}'), \sigma_\varepsilon^2 \sum_{i=0}^{k-1} \psi_i^2 \right]. \quad (57)$$

Equation (56) allows also to make inference on the cumulative effect and on the temporal average effect:

$$\begin{aligned} \sum_{h=1}^k \left( S_{t^*+h}(\mathbf{w}) - \hat{S}_{t^*+h}(\mathbf{w}') \right) &= \sum_{h=1}^k \sum_{i=0}^{h-1} \psi_i \varepsilon_{t^*+h-i} + \sum_{h=1}^k \tau_{t^*+h}(\mathbf{w}; \mathbf{w}') \\ &= \sum_{h=1}^k \varepsilon_{t^*+h} \sum_{i=0}^{k-h} \psi_i + \sum_{h=1}^k \tau_{t^*+h}(\mathbf{w}; \mathbf{w}'). \end{aligned}$$

Then

$$\sum_{h=1}^k \left( S_{t^*+h}(\mathbf{w}) - \hat{S}_{t^*+h}(\mathbf{w}') \right) \sim \left[ \Delta_{t^*+k}(\mathbf{w}; \mathbf{w}'), \sigma_\varepsilon^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_i \right)^2 \right]$$

and

$$\frac{1}{k} \sum_{h=1}^k \left( S_{t^*+h}(\mathbf{w}) - \hat{S}_{t^*+h}(\mathbf{w}') \right) \sim \left[ \bar{\tau}_{t^*+k}(\mathbf{w}; \mathbf{w}'), \frac{1}{k^2} \sigma_\varepsilon^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_i \right)^2 \right].$$

## B.2 Causal effect estimation on the untransformed variable

So far, we derived estimators for three causal effects defined for the transformed variable  $S_t = T(Y_t) - T(X_t)' \beta$ , but with a further step we can also estimate the effect for the original (untransformed) variable  $Y_t$  (to improve readability, we avoid here the dependency on the treatment paths). For example, if the transformation to achieve stationarity is a seasonal differencing of period  $s$ , we have that,

$$\begin{aligned} \hat{\tau}_{t^*+k} &= S_{t^*+k} - E[S_{t^*+k} | \mathcal{I}_{t^*}, H_0] = T(Y_{t^*+k}) - X'_{t^*+k} \beta - E[T(Y_{t^*+k}) | \mathcal{I}_{t^*}, H_0] + X'_{t^*+k} \beta \\ &= T(Y_{t^*+k}) - E[T(Y_{t^*+k}) | \mathcal{I}_{t^*}, H_0] \\ &= Y_{t^*+k} - E[Y_{t^*+k} | \mathcal{I}_{t^*}, H_0] - Y_{t^*+k-s} + E[Y_{t^*+k-s} | \mathcal{I}_{t^*}, H_0] \end{aligned}$$

Thus, the pointwise causal effect on the original variable is,

$$\hat{\tau}_{t^*+k}^Y = Y_{t^*+k} - E[Y_{t^*+k} | \mathcal{I}_{t^*}, H_0] = \hat{\tau}_{t^*+k} + \delta_k \quad (58)$$

where  $\delta_k = Y_{t^*+k-s} - E[Y_{t^*+k-s} | \mathcal{I}_{t^*}, H_0]$  and  $\delta_k = 0$  when  $k \leq s$ . Finally, from (58) estimating the cumulative and temporal average effects for the original variable it is straightforward,

$$\begin{aligned} \hat{\Delta}_{t^*+k}^Y &= \sum_{h=1}^k \hat{\tau}_{t^*+k}^Y = \hat{\Delta}_{t^*+k} + \sum_{h=1}^k \delta_h \\ \hat{\bar{\tau}}_{t^*+k}^Y &= \frac{1}{k} \sum_{h=1}^k \hat{\tau}_{t^*+k}^Y = \hat{\bar{\tau}}_{t^*+k} + \frac{1}{k} \sum_{h=1}^k \delta_h \end{aligned}$$

### B.3 Proof of relations (23), (24) and (25)

$\beta$  has prior density function given by ,

$$\begin{aligned}\Pr(\beta_\rho | \Sigma_\varepsilon, \boldsymbol{\rho}, \boldsymbol{\theta}) &= (2\pi)^{-p_\rho d/2} \det(\mathbf{H}_\rho)^{-d/2} \det(\Sigma_\varepsilon)^{-p_\rho/2} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{H}_\rho^{-1} \beta_\rho \Sigma_\varepsilon^{-1} \beta_\rho'] \right\} \\ &= (2\pi)^{-p_\rho d/2} \det(\mathbf{H}_\rho)^{-d/2} \det(\Sigma_\varepsilon)^{-p_\rho/2} \exp \left\{ -\frac{1}{2} \text{tr} [\beta_\rho' \mathbf{H}_\rho^{-1} \beta_\rho \Sigma_\varepsilon^{-1}] \right\}\end{aligned}$$

Where  $p_\rho$  is the number of selected regressors. Similarly, the density function  $\Pr(\tilde{\mathbf{Y}}_{1:t^*})$  can be written as,

$$\begin{aligned}\Pr(\tilde{\mathbf{Y}}_{1:t^*} | \beta_\rho, \Sigma_\varepsilon, \boldsymbol{\rho}, \boldsymbol{\theta}) &= (2\pi)^{-dt^*/2} \det(\Sigma_\varepsilon)^{-t^*/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{t^*} (\tilde{\mathbf{Y}}_{1:t^*} - \mathbf{X}_\rho \beta_\rho) \Sigma_\varepsilon^{-1} (\tilde{\mathbf{Y}}_{1:t^*} - \mathbf{X}_\rho \beta_\rho)' \right\} \\ &= (2\pi)^{-dt^*/2} \det(\Sigma_\varepsilon)^{-t^*/2} \exp \left\{ -\frac{1}{2} \text{tr} [(\tilde{\mathbf{Y}}_{1:t^*} - \mathbf{X}_\rho \beta_\rho)' (\tilde{\mathbf{Y}}_{1:t^*} - \mathbf{X}_\rho \beta_\rho) \Sigma_\varepsilon^{-1}] \right\}\end{aligned}$$

Now we can derive the posterior distribution for the regression coefficients as follows,

$$\begin{aligned}\Pr(\beta_\rho | \tilde{\mathbf{Y}}_{1:t^*}, \Sigma_\varepsilon, \boldsymbol{\rho}, \boldsymbol{\theta}) &\propto \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \beta_\rho, \Sigma_\varepsilon, \boldsymbol{\rho}, \boldsymbol{\theta}) \Pr(\beta_\rho | \Sigma_\varepsilon, \boldsymbol{\rho}, \boldsymbol{\theta}) \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [(\tilde{\mathbf{Y}}_{1:t^*} - \mathbf{X}_\rho \beta_\rho)' (\tilde{\mathbf{Y}}_{1:t^*} - \mathbf{X}_\rho \beta_\rho) \Sigma_\varepsilon^{-1}] \right\} \exp \left\{ -\frac{1}{2} \text{tr} [\beta_\rho' \mathbf{H}_\rho^{-1} \beta_\rho \Sigma_\varepsilon^{-1}] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [\beta_\rho' \mathbf{X}'_\rho \mathbf{X}_\rho \beta_\rho \Sigma_\varepsilon^{-1} - 2\beta_\rho' \mathbf{X}'_\rho \tilde{\mathbf{Y}}_{1:t^*} \Sigma_\varepsilon^{-1} + \beta_\rho' \mathbf{H}_\rho^{-1} \beta_\rho \Sigma_\varepsilon^{-1}] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [\beta_\rho' (\mathbf{X}'_\rho \mathbf{X}_\rho + \mathbf{H}_\rho^{-1}) \beta_\rho \Sigma_\varepsilon^{-1} - 2\beta_\rho' \mathbf{X}'_\rho \tilde{\mathbf{Y}}_{1:t^*} \Sigma_\varepsilon^{-1}] \right\}\end{aligned}$$

Which is the kernel of a matrix-normal distribution  $\mathcal{N}(\mathbf{M}, \mathbf{W}, \Sigma_\varepsilon)$ , with  $\mathbf{W} = (\mathbf{X}'_\rho \mathbf{X}_\rho + \mathbf{H}_\rho^{-1})^{-1}$  and  $\mathbf{M} = (\mathbf{X}'_\rho \mathbf{X}_\rho + \mathbf{H}_\rho^{-1})^{-1} \mathbf{X}'_\rho \tilde{\mathbf{Y}}_{1:t^*}$ .

Integration of the above quantity is necessary to derive the posterior distribution of  $\Sigma_\varepsilon$  and yields the inverse of the normalization constant, which is  $\kappa = (2\pi)^{p_\rho d/2} \det(\mathbf{W})^{d/2} \det(\Sigma_\varepsilon)^{p_\rho/2}$ .

However,  $\kappa$  simplifies with the constants singled out from the integral, which are

$$(2\pi)^{-p_\rho d/2} \det(\Sigma_\varepsilon)^{-p_\rho/2} \det(\mathbf{H}_\rho)^{-d/2} \text{ and } (2\pi)^{-dt^*/2} \det(\Sigma_\varepsilon)^{-t^*/2}, \text{ leaving } \\ \det(\mathbf{H}_\rho)^{-d/2} \det(\mathbf{W})^{d/2} (2\pi)^{-dt^*/2} \det(\Sigma_\varepsilon)^{-t^*/2}.$$



$$\begin{aligned}
\Pr(\boldsymbol{\Sigma}_\varepsilon | \tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\varrho}, \boldsymbol{\theta}) &\propto \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta}) \Pr(\boldsymbol{\Sigma}_\varepsilon | \boldsymbol{\varrho}, \boldsymbol{\theta}) \\
&\propto \Pr(\boldsymbol{\Sigma}_\varepsilon | \boldsymbol{\varrho}, \boldsymbol{\theta}) \int \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \boldsymbol{\beta}_\varrho, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta}) \Pr(\boldsymbol{\beta}_\varrho | \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta}) d\boldsymbol{\beta} \\
&\propto \det(\boldsymbol{\Sigma}_\varepsilon)^{-(d+\nu_\varepsilon+t^*+1)/2} \exp\left\{-\frac{1}{2} \text{tr}(\mathbf{S}_\varepsilon \boldsymbol{\Sigma}_\varepsilon^{-1})\right\} \exp\left\{-\frac{1}{2} \text{tr}\left[(\tilde{\mathbf{Y}}'_{1:t^*} \tilde{\mathbf{Y}}_{1:t^*} + \mathbf{M}'\mathbf{W}^{-1}\mathbf{M})\boldsymbol{\Sigma}_\varepsilon^{-1}\right]\right\} \\
&\propto \det(\boldsymbol{\Sigma}_\varepsilon)^{-(d+\nu_\varepsilon+t^*+1)/2} \exp\left\{-\frac{1}{2} \text{tr}\left[(\mathbf{S}_\varepsilon + \tilde{\mathbf{Y}}'_{1:t^*} \tilde{\mathbf{Y}}_{1:t^*} - \mathbf{M}'\mathbf{W}^{-1}\mathbf{M})\boldsymbol{\Sigma}_\varepsilon^{-1}\right]\right\}
\end{aligned}$$

This is the kernel of an Inverse-Wishart distribution with  $\nu = \nu_\varepsilon + t^*$  degrees of freedom and scale matrix  $\mathbf{S}\mathbf{S}_\varepsilon = (\mathbf{S}_\varepsilon + \tilde{\mathbf{Y}}'_{1:t^*} \tilde{\mathbf{Y}}_{1:t^*} - \mathbf{M}'\mathbf{W}^{-1}\mathbf{M})$ . We can also derive the posterior of the latent vector  $\boldsymbol{\varrho}$ ,

$$\Pr(\boldsymbol{\varrho} | \tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\theta}) = \frac{\Pr(\tilde{\mathbf{Y}}_{1:t^*} | \boldsymbol{\varrho}, \boldsymbol{\theta}) \Pr(\boldsymbol{\varrho} | \boldsymbol{\theta})}{\sum_{\boldsymbol{\varrho}} \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \boldsymbol{\varrho}, \boldsymbol{\theta}) \Pr(\boldsymbol{\varrho} | \boldsymbol{\theta})}$$

where,

$$\begin{aligned}
\Pr(\tilde{\mathbf{Y}}_{1:t^*} | \boldsymbol{\varrho}, \boldsymbol{\theta}) &= \int \int \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \boldsymbol{\beta}_\varrho, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta}) \Pr(\boldsymbol{\beta}_\varrho | \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta}) \Pr(\boldsymbol{\Sigma}_\varepsilon | \boldsymbol{\varrho}, \boldsymbol{\theta}) d\boldsymbol{\beta} d\boldsymbol{\Sigma}_\varepsilon \\
&= \int \left( \int \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \boldsymbol{\beta}_\varrho, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta}) \Pr(\boldsymbol{\beta}_\varrho | \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\varrho}, \boldsymbol{\theta}) d\boldsymbol{\beta} \right) \Pr(\boldsymbol{\Sigma}_\varepsilon | \boldsymbol{\varrho}, \boldsymbol{\theta}) d\boldsymbol{\Sigma}_\varepsilon \\
&= \int \det(\mathbf{H}_\varrho)^{-d/2} \det(\mathbf{W})^{d/2} (2\pi)^{-dt^*/2} \det(\boldsymbol{\Sigma}_\varepsilon)^{-t^*/2} \exp\left\{-\frac{1}{2} \text{tr}\left[(\tilde{\mathbf{Y}}'_{1:t^*} \tilde{\mathbf{Y}}_{1:t^*} - \mathbf{M}'\mathbf{W}^{-1}\mathbf{M})\boldsymbol{\Sigma}_\varepsilon^{-1}\right]\right\} \\
&\quad \frac{\det(\mathbf{S}_\varepsilon)^{\nu_\varepsilon/2}}{2^{\nu_\varepsilon d/2} \Gamma_d(\nu_\varepsilon/2)} \det(\boldsymbol{\Sigma}_\varepsilon)^{-(\nu_\varepsilon+d+1)/2} \exp\left\{-\frac{1}{2} \text{tr}(\mathbf{S}_\varepsilon \boldsymbol{\Sigma}_\varepsilon^{-1})\right\} d\boldsymbol{\Sigma}_\varepsilon \\
&= \frac{\det(\mathbf{H}_\varrho)^{-d/2} \det(\mathbf{W})^{d/2} (2\pi)^{-dt^*/2} \det(\mathbf{S}_\varepsilon)^{\nu_\varepsilon/2}}{2^{\nu_\varepsilon d/2} \Gamma_d(\nu_\varepsilon/2)} \cdot \frac{2^{(\nu_\varepsilon+t^*)d/2} \Gamma_d(\nu_\varepsilon + t^*/2)}{\det(\mathbf{S}\mathbf{S}_\varepsilon)^{\nu_\varepsilon+t^*/2}} \\
&= \frac{\det(\mathbf{H}_\varrho)^{-d/2} \det(\mathbf{W})^{d/2} (\pi)^{-dt^*/2} \det(\mathbf{S}_\varepsilon)^{\nu_\varepsilon/2} \Gamma_d(\nu_\varepsilon + t^*/2)}{\Gamma_d(\nu_\varepsilon/2) \det(\mathbf{S}\mathbf{S}_\varepsilon)^{\nu_\varepsilon+t^*/2}}
\end{aligned}$$

Notice that if we set  $\mathbf{H}_\varrho = (\mathbf{X}'_\varrho \mathbf{X}_\varrho)^{-1}$ , the above expressions simplify to  $\mathbf{W} = \frac{1}{2}(\mathbf{X}'_\varrho \mathbf{X}_\varrho)^{-1}$ ,  $\mathbf{M} = \frac{1}{2}(\mathbf{X}'_\varrho \mathbf{X}_\varrho)^{-1} \mathbf{X}'_\varrho \tilde{\mathbf{Y}}_{1:t^*}$  and  $\mathbf{S}\mathbf{S}_\varepsilon = \mathbf{S}_\varepsilon + \tilde{\mathbf{Y}}'_{1:t^*} \tilde{\mathbf{Y}}_{1:t^*} - \frac{1}{2} \tilde{\mathbf{Y}}'_{1:t^*} (\mathbf{X}'_\varrho \mathbf{X}_\varrho)^{-1} \mathbf{X}'_\varrho \tilde{\mathbf{Y}}_{1:t^*}$ .

In order to evaluate the posterior distribution  $\Pr(\boldsymbol{\varrho} | \tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\theta})$  we can resort to the odds and update the elements of the selection vector one component at a time, while the others are held fixed. This ensures that at each step only the most likely model is retained, either the one with  $X_p$  in it or the one without. More formally, let  $\varrho_p = 1$  and indicate with  $\boldsymbol{\varrho}_{-p}$  the vector of all

the elements in  $\boldsymbol{\varrho}$  except  $\varrho_p$ . The full conditional of  $\varrho_p$  is given by,

$$\begin{aligned} \Pr(\varrho_p = 1 | \tilde{\mathbf{Y}}_{1:t^*}, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta}) &= \frac{\Pr(\varrho_p = 1 | \boldsymbol{\theta}) \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \varrho_p = 1, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta})}{\Pr(\varrho_p = 1 | \boldsymbol{\theta}) \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \varrho_p = 1, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta}) + \Pr(\varrho_p = 0 | \boldsymbol{\theta}) \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \varrho_p = 0, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta})} \\ &= \frac{1}{1 + o_p^{-1}} \end{aligned} \quad (59)$$

Where, assuming equal prior probabilities  $\Pr(\varrho_p = 1 | \boldsymbol{\theta}) = \Pr(\varrho_p = 0 | \boldsymbol{\theta})$  we have,

$$o_p = \frac{\Pr(\varrho_p = 1 | \boldsymbol{\theta}) \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \varrho_p = 1, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta})}{\Pr(\varrho_p = 0 | \boldsymbol{\theta}) \Pr(\tilde{\mathbf{Y}}_{1:t^*} | \varrho_p = 0, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta})} = \frac{\Pr(\tilde{\mathbf{Y}}_{1:t^*} | \varrho_p = 1, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta})}{\Pr(\tilde{\mathbf{Y}}_{1:t^*} | \varrho_p = 0, \boldsymbol{\varrho}_{-p}, \boldsymbol{\theta})}$$

Finally, let  $\boldsymbol{\eta}_{1:t^*}^{(r)}$  indicate the disturbances up to time  $t^*$  of the  $r$ -th state. Then,  $\boldsymbol{\eta}_{1:t^*}^{(r)}$  is a  $(t^* \times d)$  matrix independently drawn from a  $\mathcal{N}(0, I_{t^*}, \boldsymbol{\Sigma}_r)$ . Thus we have,

$$\begin{aligned} \Pr(\boldsymbol{\Sigma}_r | \boldsymbol{\eta}_{1:t^*}^{(r)}, \boldsymbol{\theta}) &\propto \Pr(\boldsymbol{\eta}_{1:t^*}^{(r)} | \boldsymbol{\Sigma}_r, \boldsymbol{\theta}) \Pr(\boldsymbol{\Sigma}_r | \boldsymbol{\theta}) \\ &\propto \det(\boldsymbol{\Sigma}_r)^{-t^*/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\eta}_{1:t^*}^{(r)} \boldsymbol{\Sigma}_r^{-1} \boldsymbol{\eta}_{1:t^*}^{\prime(r)}) \right\} \det(\boldsymbol{\Sigma}_r)^{-\frac{\nu_r + d + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{S}_r \boldsymbol{\Sigma}_r^{-1}) \right\} \\ &\propto \det(\boldsymbol{\Sigma}_r)^{-\frac{\nu_r + d + t^* + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (\mathbf{S}_r + \boldsymbol{\eta}_{1:t^*}^{(r)} \boldsymbol{\Sigma}_r^{-1} \boldsymbol{\eta}_{1:t^*}^{\prime(r)}) \right] \right\} \end{aligned}$$

Which is the kernel of an Inverse-Wishart distribution with  $\nu_r + t^*$  degrees of freedom and scale matrix  $\mathbf{S}_r + \boldsymbol{\eta}_{1:t^*}^{\prime(r)} \boldsymbol{\eta}_{1:t^*}^{(r)}$ .

To sample from the joint posterior distribution of the states and model parameters we can employ the following MCMC algorithm.

---

**Algorithm 1** Gibbs sampler to draw from the joint posterior distribution of the states and model parameters

---

**Require:**  $\Sigma_\varepsilon^{(0)}$ ,  $\Sigma_r^{(0)}$ ,  $\theta$ ,  $\mathbf{H}_\theta$ , niter

- 1: **for** s in 1 : niter **do**
  - 2: draw  $\alpha_t^{(s)}$  from  $\Pr(\alpha_t | \mathbf{Y}_{1:t^*}, \Sigma_\varepsilon^{(s-1)}, \Sigma_r^{(s-1)}, \theta)$  using the simulation smoothing by [Durbin and Koopman \(2002\)](#)<sup>24</sup>
  - 3: draw  $\Sigma_r^{(s)}$  from  $\Pr(\Sigma_r | \boldsymbol{\eta}_{1:t^*}^{(r,s)}, \theta)$  according to equation (25)
  - 4: compute  $\tilde{\mathbf{Y}}_{1:t^*}^{(s)}$  and draw  $\boldsymbol{\rho}^{(s)}$  from  $\Pr(\rho_p | \tilde{\mathbf{Y}}_{1:t^*}^{(s)}, \boldsymbol{\rho}_{-p}^{(s)}, \theta)$  by changing  $\rho$  one component at a time and computing its posterior probability (this ensures that every time a component  $\rho_p$  is changed, the most likely model is retained, i.e. either the one with  $X_p$  in or the one without  $X_p$ )
  - 5: draw  $\Sigma_\varepsilon^{(s)}$  from  $\Pr(\Sigma_\varepsilon | \tilde{\mathbf{Y}}_{1:t^*}^{(s)}, \boldsymbol{\rho}^{(s)}, \theta)$  according to equation (24)
  - 6: draw  $\beta_\rho^{(s)}$  from  $\Pr(\beta_\rho | \tilde{\mathbf{Y}}_{1:t^*}^{(s)}, \Sigma_\varepsilon^{(s)}, \boldsymbol{\rho}^{(s)}, \theta)$  according to equation (23)
  - 7: **end for**
- 

## B.4 Unbiased causal effects

**Theorem 1** For a positive integer  $k$ , define  $\hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) = \mathbb{E}[\Pr(\mathbf{Y}_{t^*+k}(\mathbf{w}) | \mathbf{Y}_{t^*}(\mathbf{w}))]$  and  $\hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}') = \mathbb{E}[\Pr(\mathbf{Y}_{t^*+k}(\mathbf{w}') | \mathbf{Y}_{t^*}(\mathbf{w}'))]$ ; under model (21),  $\hat{\mathbf{Y}}_{t^*+k}(\mathbf{w})$  and  $\hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}')$  are the  $k$ -step ahead forecast of  $\mathbf{Y}_{t^*+k}(\mathbf{w})$  and  $\mathbf{Y}_{t^*+k}(\mathbf{w}')$  given the information set up to time  $t^*$ ,  $\mathcal{I}_{t^*} = \{\mathbf{Y}_{1:t^*}, X_{1:t^*}\}$ . Then,  $\hat{\boldsymbol{\tau}}_{t^*+k}(\mathbf{w}; \mathbf{w}') = \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}')$  is the point estimator of the general causal effect and, conditionally on  $\mathcal{I}_{t^*}$  we have,

$$\boldsymbol{\tau}_{t^*+k}(\mathbf{w}; \mathbf{w}') - \hat{\boldsymbol{\tau}}_{t^*+k}(\mathbf{w}; \mathbf{w}') \sim N(\mathbf{0}, \Sigma_{\mathbf{w}} + \Sigma_{\mathbf{w}'}) \quad (60)$$

$$\Delta_{t^*+k}(\mathbf{w}; \mathbf{w}') - \hat{\Delta}_{t^*+k}(\mathbf{w}; \mathbf{w}') \sim \mathcal{N}(\mathbf{0}, \Sigma_{D(\mathbf{w})} + \Sigma_{D(\mathbf{w}')} , \Sigma) \quad (61)$$

$$\bar{\boldsymbol{\tau}}_{t^*+k}(\mathbf{w}; \mathbf{w}') - \hat{\bar{\boldsymbol{\tau}}}_{t^*+k}(\mathbf{w}; \mathbf{w}') \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{k^2}(\Sigma_{D(\mathbf{w})} + \Sigma_{D(\mathbf{w}')}) , \Sigma\right) \quad (62)$$

where,  $\Sigma_{\mathbf{w}} = \text{Var}\left[\mathbf{Y}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) \mid \mathcal{I}_{t^*}\right]$ ,  $\Sigma_{D(\mathbf{w})} = \text{Var}\left[\sum_k (\mathbf{Y}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w})) \mid \mathcal{I}_{t^*}\right]$  with  $\mathbf{w} \in \{\mathbf{w}, \mathbf{w}'\}$  are defined as follows

$$\Sigma_{\mathbf{w}} = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}'_t + \Sigma_\varepsilon \quad (63)$$

$$\Sigma_{D(\mathbf{w})} = \left( \mathbf{D}_{t^*+1} \mathbf{P}_{t^*+1} \mathbf{D}'_{t^*+1} + \sum_k (\mathbf{D}_{t^*+k} \mathbf{R}_{t^*+K-1} \mathbf{C}_{t^*+K-1} \mathbf{R}'_{t^*+K-1} \mathbf{D}'_{t^*+k}) \right) + KH_t \quad (64)$$

and

$$\begin{aligned}\mathbf{D}_{t^*+k} &= \mathbf{Z}_{t^*+k} + \mathbf{D}_{t^*+k+1} \mathbf{T}_{t^*+k}, \quad k = 1, \dots, K-1 \\ \mathbf{D}_{t^*+K} &= \mathbf{Z}_{t^*+K}\end{aligned}$$

**Proof.**

The difference between the general causal effect and its estimator can be written as,

$$\begin{aligned}\tau_{t^*+k}(\mathbf{w}; \mathbf{w}') - \hat{\tau}_{t^*+k}(\mathbf{w}; \mathbf{w}') &= \mathbf{Y}_{t^*+k}(\mathbf{w}) - \mathbf{Y}_{t^*+k}(\mathbf{w}') - \left[ \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}') \right] \\ &= \underbrace{\mathbf{Y}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w})}_A - \underbrace{\left[ \mathbf{Y}_{t^*+k}(\mathbf{w}') - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}') \right]}_B\end{aligned}$$

Let's focus our attention on A and define  $\mathbf{a}_{t^*+k} = E[\boldsymbol{\alpha}_{t^*+k} | \mathcal{I}_{t^*}]$  and  $\mathbf{P}_{t^*+k} = \text{Var}[\boldsymbol{\alpha}_{t^*+k} | \mathcal{I}_{t^*}]$ . Under model (21) we have,

$$\begin{aligned}\mathbf{Y}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) &= \mathbf{Z}_{t^*+k} \boldsymbol{\alpha}_{t^*+k} + \mathbf{X}_{t^*+k} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t^*+k} - E[\mathbf{Y}_{t^*+k}(\mathbf{w}) | \mathcal{I}_{t^*}] \\ &= \mathbf{Z}_{t^*+k} \boldsymbol{\alpha}_{t^*+k} + \mathbf{X}_{t^*+k} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t^*+k} - \mathbf{Z}_{t^*+k} \mathbf{a}_{t^*+k} - \mathbf{X}_{t^*+k} \boldsymbol{\beta} \\ &= \mathbf{Z}_{t^*+k} \boldsymbol{\alpha}_{t^*+k} - \mathbf{Z}_{t^*+k} \mathbf{a}_{t^*+k} + \boldsymbol{\varepsilon}_{t^*+k}\end{aligned}$$

Then,

$$\begin{aligned}E[\mathbf{Y}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) | \mathcal{I}_{t^*}] &= 0 \\ \text{Var}[\mathbf{Y}_{t^*+k}(\mathbf{w}) - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) | \mathcal{I}_{t^*}] &= \mathbf{Z}_{t^*+k} \mathbf{P}_{t^*+k} \mathbf{Z}'_{t^*+k} + \boldsymbol{\Sigma}_{\varepsilon} = \boldsymbol{\Sigma}_{\mathbf{w}}\end{aligned}$$

Following the exact same steps for B we can show that  $\mathbf{Y}_{t^*+k}(\mathbf{w}') - \hat{\mathbf{Y}}_{t^*+k}(\mathbf{w}') \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{w}'})$ . Since the potential paths are independent of each other, relation (60) follows from the properties of the difference of two independent multivariate Normal random variables.

Based on the above result, we can easily show that the expectation of the difference between the cumulative effect and its estimator is zero. In what follows we derive the proof for  $t' = t^* + K$  but it could be shown for every  $k = 1, \dots, K$ .

$$\begin{aligned}E \left[ \Delta_{t^*+K}(\mathbf{w}; \mathbf{w}') - \hat{\Delta}_{t^*+K}(\mathbf{w}; \mathbf{w}') \mid \mathcal{I}_{t^*} \right] &= E \left[ \sum_{k=1}^K (\tau_{t^*+k}(\mathbf{w}; \mathbf{w}') - \hat{\tau}_{t^*+k}(\mathbf{w}; \mathbf{w}')) \mid \mathcal{I}_{t^*} \right] \\ &= \sum_{k=1}^K E \left[ \tau_{t^*+k}(\mathbf{w}; \mathbf{w}') - \hat{\tau}_{t^*+k}(\mathbf{w}; \mathbf{w}') \mid \mathcal{I}_{t^*} \right] = 0\end{aligned}$$

The derivation of the variance may be somewhat more cumbersome, because the time dependency also come into play. So we have three dependence structures to take into account: the one between the  $d$  series, the one between times and the one between the states. To address this issue it is useful to re-define  $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, H_t, \boldsymbol{\Sigma})$ ; in this way,  $\boldsymbol{\varepsilon}_t$  can be seen as a single-row matrix following a matrix Normal distribution, which is in line with the definition provided in Section 4.3. Thus, we have

$$\begin{aligned} \text{Var} \left[ \Delta_{t^*+K}(\mathbf{w}; \mathbf{w}') - \widehat{\Delta}_{t^*+K}(\mathbf{w}; \mathbf{w}') \mid \mathcal{I}_{t^*} \right] &= \text{Var} \left[ \sum_{k=1}^K \boldsymbol{\tau}_{t^*+k}(\mathbf{w}; \mathbf{w}') - \sum_{k=1}^K \widehat{\boldsymbol{\tau}}_{t^*+k}(\mathbf{w}; \mathbf{w}') \mid \mathcal{I}_{t^*} \right] \\ &= \text{Var} \left[ \sum_{k=1}^K \left( \mathbf{Y}_{t^*+k}(\mathbf{w}) - \widehat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) \right) - \sum_{k=1}^K \left( \mathbf{Y}_{t^*+k}(\mathbf{w}') - \widehat{\mathbf{Y}}_{t^*+k}(\mathbf{w}') \right) \mid \mathcal{I}_{t^*} \right] \end{aligned}$$

Focusing on the first term,

$$\begin{aligned} \text{Var} \left[ \sum_{k=1}^K \left( \mathbf{Y}_{t^*+k}(\mathbf{w}) - \widehat{\mathbf{Y}}_{t^*+k}(\mathbf{w}) \right) \mid \mathcal{I}_{t^*} \right] &= \text{Var} \left[ \sum_{k=1}^K \mathbf{Z}_{t^*+k} \boldsymbol{\alpha}_{t^*+k} - \sum_{k=1}^K \mathbf{Z}_{t^*+k} \mathbf{a}_{t^*+k} + \sum_{k=1}^K \boldsymbol{\varepsilon}_{t^*+k} \mid \mathcal{I}_{t^*} \right] \\ &= \text{Var} \left[ \sum_{k=1}^K \mathbf{Z}_{t^*+k} \boldsymbol{\alpha}_{t^*+k} \mid \mathcal{I}_{t^*} \right] + KH_t \end{aligned}$$

where,

$$\begin{aligned} \text{Var} \left[ \sum_{k=1}^K \mathbf{Z}_{t^*+k} \boldsymbol{\alpha}_{t^*+k} \mid \mathcal{I}_{t^*} \right] &= \text{Var} \left[ \mathbf{Z}_{t^*+1} \boldsymbol{\alpha}_{t^*+1} + \mathbf{Z}_{t^*+2} \boldsymbol{\alpha}_{t^*+2} + \cdots + \mathbf{Z}_{t^*+K} \boldsymbol{\alpha}_{t^*+K} \mid \mathcal{I}_{t^*} \right] \\ &= \text{Var} \left[ \mathbf{Z}_{t^*+1} \boldsymbol{\alpha}_{t^*+1} + \mathbf{Z}_{t^*+2} (\mathbf{T}_{t^*+1} \boldsymbol{\alpha}_{t^*+1} + \mathbf{R}_{t^*+1} \boldsymbol{\eta}_{t^*+1}) + \cdots + \mathbf{Z}_{t^*+K} \boldsymbol{\alpha}_{t^*+K} \mid \mathcal{I}_{t^*} \right] \\ &= \text{Var} \left[ (\mathbf{Z}_{t^*+1} + \mathbf{Z}_{t^*+2} \mathbf{T}_{t^*+1} + \cdots + \mathbf{Z}_{t^*+K} \mathbf{T}_{t^*+K-1} \cdots \mathbf{T}_{t^*+1}) \boldsymbol{\alpha}_{t^*+1} + \right. \\ &\quad \left. + (\mathbf{Z}_{t^*+2} + \mathbf{Z}_{t^*+3} \mathbf{T}_{t^*+2} + \cdots + \mathbf{Z}_{t^*+K} \mathbf{T}_{t^*+K-1} \cdots \mathbf{T}_{t^*+2}) \mathbf{R}_{t^*+1} \boldsymbol{\eta}_{t^*+1} + \right. \\ &\quad \left. + \cdots + \mathbf{Z}_{t^*+K} \mathbf{R}_{t^*+K-1} \boldsymbol{\eta}_{t^*+K-1} \mid \mathcal{I}_{t^*} \right] \end{aligned}$$

Then, defining  $\mathbf{D}_{t^*+1} = \mathbf{Z}_{t^*+1} + \mathbf{Z}_{t^*+2} \mathbf{T}_{t^*+1} + \cdots + \mathbf{Z}_{t^*+K} \mathbf{T}_{t^*+K-1} \cdots \mathbf{T}_{t^*+1}$  we can notice that  $\mathbf{D}_{t^*+1} = \mathbf{Z}_{t^*+1} + (\mathbf{Z}_{t^*+2} + \mathbf{Z}_{t^*+3} \mathbf{T}_{t^*+2} \cdots \mathbf{Z}_{t^*+K} \mathbf{T}_{t^*+K-1} \cdots \mathbf{T}_{t^*+2}) \mathbf{T}_{t^*+1} = \mathbf{Z}_{t^*+1} + \mathbf{D}_{t^*+2} \mathbf{T}_{t^*+1}$ . Thus, in general we have

$$\begin{aligned}\mathbf{D}_{t^*+k} &= \mathbf{Z}_{t^*+k} + \mathbf{D}_{t^*+k+1} \mathbf{T}_{t^*+k}, \quad k = 1, \dots, K-1 \\ \mathbf{D}_{t^*+K} &= \mathbf{Z}_{t^*+K}\end{aligned}$$

and

$$Var \left[ \sum_{k=1}^K \mathbf{Z}_{t^*+k} \boldsymbol{\alpha}_{t^*+k} \mid \mathcal{I}_{t^*} \right] = \left( \mathbf{D}_{t^*+1} \mathbf{P}_{t^*+1} \mathbf{D}'_{t^*+1} + \sum_{k=2}^K (\mathbf{D}_{t^*+k} \mathbf{R}_{t^*+K-1} \mathbf{C}_{t^*+K-1} \mathbf{R}'_{t^*+K-1} \mathbf{D}'_{t^*+k}) \right)$$

This yields to the final result in equation (64). Repeating these steps for the second term we obtain equation (61). Finally, applying the usual properties of variance we obtain relation (62) for the temporal average causal effect,

$$\begin{aligned}Var \left[ \bar{\tau}_{t^*+K}(\mathbf{w}; \mathbf{w}') - \hat{\tau}_{t^*+K}(\mathbf{w}; \mathbf{w}') \mid \mathcal{I}_{t^*} \right] &= Var \left[ \frac{1}{K} \sum_{k=1}^K \tau_{t^*+k}(\mathbf{w}; \mathbf{w}') - \frac{1}{K} \sum_{k=1}^K \hat{\tau}_{t^*+k}(\mathbf{w}; \mathbf{w}') \mid \mathcal{I}_{t^*} \right] \\ &= \frac{1}{K^2} Var \left[ \Delta_{t^*+K}(\mathbf{w}; \mathbf{w}') - \hat{\Delta}_{t^*+K}(\mathbf{w}; \mathbf{w}') \mid \mathcal{I}_{t^*} \right]\end{aligned}$$

Theorem 1, states that the point estimator of the general causal effect and, by extension, the marginal and the conditional causal effect estimators are unbiased. From equation (64) we can infer that the variance of the difference between the cumulative effect and its estimator increases with the variance of both  $\boldsymbol{\varepsilon}_t$  and  $\boldsymbol{\eta}_t$ . Furthermore, the variance is an increasing function of  $\mathbf{D}_t$ , therefore, our uncertainty increases with time, reflecting our intuition that we have less information about potential outcomes that are further from the time of the intervention.

## B.5 Posterior predictive checks

To produce reliable causal effect estimates from the model-based predictions, the assumed model has to adequately describe the data. One way to check the quality of the model fit within a Bayesian framework is to use posterior predictive checks (Rubin, 1981, 1984; Gelman et al., 2013). Intuitively, this entails generating synthetic data sets from the fitted model and comparing them to the observed data.

Typically, we generate replicated data by drawing multiple times from the posterior predictive distribution; then, we compare these draws with the observed data using both numerical and graphical checks (Gelman et al., 2013). More specifically, let  $T(\mathbf{Y}_{1:t^*}, \boldsymbol{\vartheta})$  be a test quantity that depends on the data and the unknown model parameters and denote with  $\mathbf{Y}_{1:t^*}^{new}$  a new vector of observations sampled from the posterior predictive distribution, as outlined in equation (26). To describe the degree of the discrepancy, we use the Bayesian  $p$ -value, which is the probability of observing a test quantity at least as extreme as the observed data,  $T(\mathbf{Y}_{1:t^*}^{new}, \boldsymbol{\vartheta})$ , we denote this by

$$p_B = \Pr(T(\mathbf{Y}_{1:t^*}^{new}, \boldsymbol{\vartheta}) \geq T(\mathbf{Y}_{1:t^*}, \boldsymbol{\vartheta}) | \mathbf{Y}_{1:t^*}). \quad (65)$$

Unlike in frequentist statistics where a  $p$ -value near 0 indicates that the corresponding null hypothesis can be rejected, an extreme Bayesian  $p$ -value denotes that the specific feature of the data captured by the test quantity is inconsistent with the assumed model. For example, if we suspect that our model may not be able to reproduce the large values observed in the data, a suitable test quantity could be the observations' maximum. In this case, a  $p$ -value near 0 indicates that, under the assumed model, it is unlikely to encounter a value larger than the observed maximum; so, if the replicated data were generated under a Normal model, a heavy tail distribution may actually be more appropriate. A Bayesian  $p$ -value can be estimated by computing the proportion of replicated data sets satisfying (65).

We can also provide a graphical representation by plotting the distribution of the test quantity against the observed test quantity; as in a classical setting, the Bayesian  $p$ -value is the right tail-area probability. Another graphical check consists of computing the posterior predictive mean (i.e., the mean of the posterior predictive distribution) and then plotting it against the distribution of the observed data. Generally, graphical model checks are useful for highlighting the systematic discrepancies between the observed and the simulated data.

Finally, for both linear and non-linear regression models, we can also assess the goodness of fit using residual plots. We can think of Bayesian model residuals as a generalization of classical residuals that accounts for the uncertainty in the model parameters.

In Section 4.5, we extensively used posterior predictive checks to select and validate the model used for our empirical analysis.

## B.6 Sensitivity analysis

Model validation performed through posterior predictive checks shows that the structural time series model with a trend and seasonal component adequately describe the data (see [A.2](#) for the details). Nonetheless, posterior inference might still be affected by prior assumptions. Thus, to strengthen our confidence in the assumed model, we performed a sensitivity analysis in order to evaluate to what extent our inferred causal effect changes to different values of the prior hyperparameters.

As described in [Section 4.3.2](#), for the unknown scale matrices of the Inverse-Wishart distributions we chose the following variance-covariance matrix,

$$\mathbf{S}_\varepsilon = \mathbf{S}_r = \begin{bmatrix} hs_1^2 & \sqrt{hk}s_1s_2\rho \\ \sqrt{hk}s_1s_2\rho & ks_2^2 \end{bmatrix},$$

where,  $s_1^2, s_2^2$  are the sample variances, which can be scaled by some positive values  $h$  and  $k$ , and  $\rho$  is the correlation coefficient. Linking the scale matrix to the sample variances is in line with an objective Bayesian approach and can ensure a reasonable scale for the prior ([Brodersen et al., 2015](#)). In our empirical analysis, we set  $h = k = 1$  but we could have used different values. For example, since the sample variance of the competitor brands is, on average, ten times higher than the sample variance of the store brands, another reasonable scaling can be obtained by setting  $h = 0.1, k = 1$ . [Table 21](#) presents the estimated causal effects under different assumptions for the scaling factors.

Another parameter that can influence our posterior inference is the linear correlation coefficient. We set  $\rho = -0.8$  based on our prior belief that the two products in the pair are perfect substitutes, but the correlation might be smaller than what assumed or even positive. [Table 22](#) shows the estimated causal effects under different combinations of the correlation and the scaling factors.

Finally, we assumed  $\mathbf{S}_\varepsilon = \mathbf{S}_r$  but we can also allow the state disturbances to vary more (less) freely than the observation disturbances. The estimated effects under different assumptions for  $\mathbf{S}_r$  are reported in [Table 23](#).

Overall, our estimates seem to be robust to different prior assumptions: even if in some instances we find only one or two significant effects, this still supports our general conclusion that the new price policy was not effective.



**Table 21:** Temporal average general effect estimates at one month horizon under different prior assumptions for the scaling factors  $h$  and  $k$ .

		$h = 1, k = 0.01$			$h = 1, k = 0.1$			$h = 1, k = 1$		
		$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%
(1)	s	7.24	-23.37	36.89	7.25	-23.42	37.53	6.97	-24.25	38.47
	c	23.55	-126.97	178.70	24.34	-118.60	168.55	24.89	-101.30	153.64
(2)	s	7.41	-13.63	30.46	7.15	-13.80	29.48	7.02	-14.79	28.90
	c	12.33	-87.37	114.44	13.33	-77.80	106.46	14.71	-62.26	99.44
(3)	s	7.46	-15.24	29.66	7.68	-15.17	30.33	7.94	-14.08	32.26
	c	13.57	-76.17	100.46	14.39	-70.34	95.71	15.42	-62.17	90.81
(4)	s	<b>47.19</b>	<b>0.25</b>	<b>94.40</b>	<b>47.25</b>	<b>1.05</b>	<b>94.28</b>	<b>47.84</b>	<b>4.71</b>	<b>96.82</b>
	c	26.28	-101.06	150.07	26.93	-93.87	142.41	28.86	-77.93	135.93
(5)	s	3.60	-44.33	52.00	3.46	-45.09	53.48	4.11	-46.65	54.64
	c	41.69	-82.11	159.68	43.67	-72.81	157.08	45.47	-63.13	154.24
(6)	s	9.43	-13.15	33.27	9.48	-13.45	33.57	9.53	-14.45	33.68
	c	22.83	-52.50	95.71	23.33	-47.52	92.92	25.64	-37.88	93.36
(7)	s	<b>79.87</b>	<b>12.19</b>	<b>151.16</b>	<b>78.25</b>	<b>5.65</b>	<b>148.78</b>	<b>78.19</b>	<b>0.15</b>	<b>154.08</b>
	c	165.50	-313.51	621.07	180.33	-262.04	644.09	182.70	-221.16	600.08
(8)	s	24.79	-25.48	78.87	25.20	-28.56	75.59	25.23	-28.60	78.16
	c	14.90	-16.30	47.43	15.83	-15.80	47.50	15.91	-15.15	47.53
(9)	s	40.54	-9.93	91.72	40.34	-10.24	89.36	40.29	-9.84	90.38
	c	15.91	-31.49	64.54	16.63	-31.47	66.75	17.17	-30.76	68.56
(10)	s	<b>12.39</b>	<b>0.81</b>	<b>23.67</b>	<b>12.43</b>	<b>1.00</b>	<b>23.82</b>	<b>12.43</b>	<b>1.35</b>	<b>23.64</b>
	c	0.06	-9.02	9.56	0.16	-8.78	9.37	0.04	-9.36	9.79

		$h = 0.1, k = 0.01$			$h = 0.1, k = 0.1$			$h = 0.1, k = 1$		
		$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%
(1)	s	7.96	-19.49	37.36	7.82	-20.49	38.49	7.56	-21.54	38.87
	c	19.89	-122.41	158.93	19.07	-119.11	151.84	20.61	-104.27	142.84
(2)	s	7.13	-11.29	27.24	6.75	-12.31	26.95	6.71	-12.91	26.21
	c	13.14	-82.89	111.07	13.44	-80.64	102.26	14.72	-62.25	99.20
(3)	s	7.50	-13.08	27.75	7.60	-13.25	28.14	7.83	-11.88	29.91
	c	13.80	-74.26	98.68	14.18	-70.76	92.72	15.30	-61.50	91.00
(4)	s	<b>47.72</b>	<b>2.69</b>	<b>93.21</b>	<b>47.66</b>	<b>2.99</b>	<b>93.63</b>	<b>47.99</b>	<b>4.60</b>	<b>94.78</b>
	c	25.65	-98.05	146.90	26.33	-92.66	139.63	29.72	-78.09	135.79
(5)	s	5.15	-49.86	60.98	4.46	-51.98	62.41	5.46	-53.97	66.87
	c	46.15	-74.29	174.59	43.95	-69.46	154.96	48.31	-55.84	157.87
(6)	s	8.80	-15.48	34.40	9.04	-17.71	36.31	8.90	-15.72	34.29
	c	22.76	-51.75	95.12	23.10	-47.28	91.12	25.19	-37.30	91.33
(7)	s	75.17	-5.10	156.68	77.54	-3.56	160.93	74.41	-9.53	158.00
	c	186.83	-289.26	671.61	177.16	-250.44	586.05	190.17	-201.93	593.67
(8)	s	24.07	-32.97	77.91	24.17	-31.64	77.64	24.35	-31.78	78.01
	c	15.42	-16.28	46.81	15.61	-15.21	46.83	15.80	-15.42	46.50
(9)	s	38.07	-15.31	92.90	38.30	-14.32	91.26	37.89	-13.39	88.85
	c	16.59	-32.28	67.16	16.44	-31.49	64.43	17.27	-30.87	66.77
(10)	s	11.56	-1.61	25.12	11.73	-1.12	24.18	<b>12.00</b>	<b>0.02</b>	<b>23.85</b>
	c	0.25	-8.65	9.83	0.30	-8.71	9.68	0.00	-9.46	9.77

**Table 22:** Temporal average general effect estimates at one month horizon under different prior assumptions for the scaling factors  $h$ ,  $k$  and the linear correlation coefficient  $\rho$ .

		$h = 1, k = 0.1, \rho = -0.3$			$h = 1, k = 1, \rho = -0.3$			$h = 0.1, k = 1, \rho = +0.3$		
		$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%
(1)	s	7.20	-23.85	38.44	7.13	-24.72	39.13	7.88	-21.90	39.20
	c	23.73	-116.98	167.21	24.14	-99.19	152.88	20.89	-101.96	144.23
(2)	s	7.24	-14.58	30.35	7.12	-15.19	29.16	6.85	-13.39	28.50
	c	13.65	-71.93	110.44	14.63	-61.80	99.32	14.61	-66.66	93.83
(3)	s	7.70	-15.19	30.47	7.94	-13.96	32.43	8.03	-12.18	30.54
	c	14.52	-69.13	93.31	15.36	-60.10	91.26	15.07	-60.56	90.46
(4)	s	<b>47.31</b>	<b>0.14</b>	<b>95.55</b>	<b>48.05</b>	<b>4.46</b>	<b>97.81</b>	<b>48.20</b>	<b>3.77</b>	<b>96.39</b>
	c	26.54	-92.09	141.12	28.17	-76.13	134.14	27.36	-80.92	133.40
(5)	s	3.92	-42.58	51.27	4.28	-42.17	53.99	5.26	-48.09	59.97
	c	44.02	-69.30	152.63	48.36	-54.29	155.36	47.98	-52.36	154.19
(6)	s	9.60	-11.69	32.73	9.55	-12.56	31.58	9.38	-13.30	32.70
	c	23.86	-44.24	89.96	25.89	-35.07	89.86	25.68	-35.01	89.41
(7)	s	<b>79.00</b>	<b>6.63</b>	<b>148.42</b>	<b>78.96</b>	<b>1.67</b>	<b>154.79</b>	76.86	-9.14	165.34
	c	187.38	-244.11	635.17	190.74	-198.58	596.36	187.56	-202.75	572.28
(8)	s	25.66	-28.03	76.80	25.65	-26.71	79.36	25.04	-35.08	83.11
	c	16.24	-15.76	48.84	16.09	-15.54	46.68	16.17	-14.73	48.06
(9)	s	40.33	-10.75	90.49	40.49	-8.66	89.99	38.33	-13.22	90.34
	c	17.34	-30.12	65.67	17.64	-31.95	67.90	17.64	-32.55	68.40
(10)	s	<b>12.37</b>	<b>0.88</b>	<b>23.58</b>	<b>12.39</b>	<b>0.71</b>	<b>23.66</b>	11.53	-1.71	24.82
	c	0.28	-8.44	9.53	-0.09	-9.77	9.99	0.05	-9.42	9.85

**Table 23:** Temporal average general effect estimates at one month horizon for  $h = k = 1$ ,  $\rho = 1$  and under different prior assumptions for  $\mathbf{S}_r$ .

		$\mathbf{S}_r = 0.5\mathbf{S}_\varepsilon$			$\mathbf{S}_r = 2\mathbf{S}_\varepsilon$			$\mathbf{S}_r = 10\mathbf{S}_\varepsilon$		
		$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%	$\hat{\tau}_t$	2.5%	97.5%
(1)	s	6.90	-23.91	37.77	7.10	-26.11	39.83	7.09	-34.28	46.37
	c	23.86	-102.42	152.62	24.79	-101.46	154.68	25.29	-115.70	171.43
(2)	s	6.97	-13.72	27.83	7.06	-16.70	31.65	7.26	-24.02	37.26
	c	14.22	-63.03	100.72	14.81	-65.48	97.76	15.80	-72.62	110.79
(3)	s	7.92	-12.94	31.16	7.94	-15.49	34.35	8.05	-22.00	41.74
	c	15.09	-63.38	91.06	15.65	-62.26	91.39	15.68	-73.12	103.23
(4)	s	<b>48.08</b>	<b>4.49</b>	<b>95.57</b>	<b>47.62</b>	<b>2.24</b>	<b>98.14</b>	47.59	-8.74	110.51
	c	27.52	-78.23	131.69	29.77	-78.38	136.75	29.36	-90.11	152.06
(5)	s	4.08	-45.71	55.58	3.32	-47.07	55.74	5.33	-52.97	67.21
	c	45.10	-63.01	152.22	49.28	-58.20	163.80	46.12	-81.66	170.30
(6)	s	9.42	-13.67	32.68	9.60	-14.88	34.67	9.29	-19.14	37.63
	c	23.53	-42.49	87.81	25.84	-38.52	94.63	26.15	-47.89	102.64
(7)	s	<b>77.94</b>	<b>0.62</b>	<b>153.30</b>	78.39	-0.28	156.90	82.02	-9.49	180.16
	c	184.04	-215.19	594.20	181.14	-233.10	602.97	169.11	-307.08	619.20
(8)	s	24.70	-27.11	75.79	25.79	-30.32	81.06	26.76	-37.56	90.71
	c	15.90	-14.53	47.33	16.08	-15.65	49.14	16.38	-21.24	56.59
(9)	s	39.83	-9.11	89.22	41.04	-9.98	92.03	41.00	-20.73	103.17
	c	16.81	-31.86	66.30	16.84	-36.00	69.79	17.26	-45.31	79.88
(10)	s	<b>12.42</b>	<b>1.28</b>	<b>23.61</b>	<b>12.61</b>	<b>1.34</b>	<b>24.15</b>	12.71	-0.72	25.90
	c	0.09	-9.13	9.35	-0.12	-10.10	10.10	-0.50	-12.44	12.13

## B.7 Convergence diagnostics

To make inference with Markov Chain Monte Carlo (MCMC) methods we need to verify that our Markov chain has converged to the stationary distribution. Geweke’s diagnostic test (Geweke, 1992) compares the sample means of two non-overlapping quantiles of the chain (for example, the first 10% and the last 50% of the draws). If the draws are sampled from the same stationary distribution, the sample means are equal and the test statistic is asymptotically Normal.

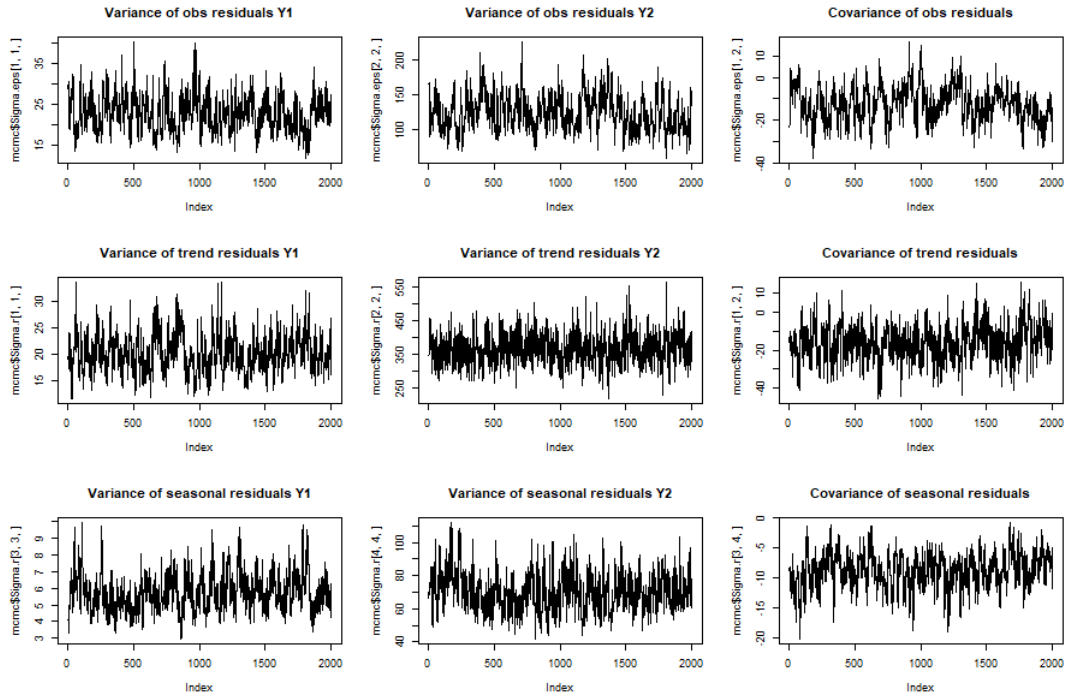
Table 24 shows the resulting p-value for the two-sided test for every parameter of the bivariate models estimated on the 10 store-competitor pairs. The Geweke diagnostic fails to detect non-convergence of the chains to the stationary distribution (at the 5% level, the test fails to reject the null hypothesis of the equality of means in 81 cases out of 90).

Finally, for a visual inspection of the chain convergence, we also include the trace plots for the parameters of the first two models (Figures 46 and 47).

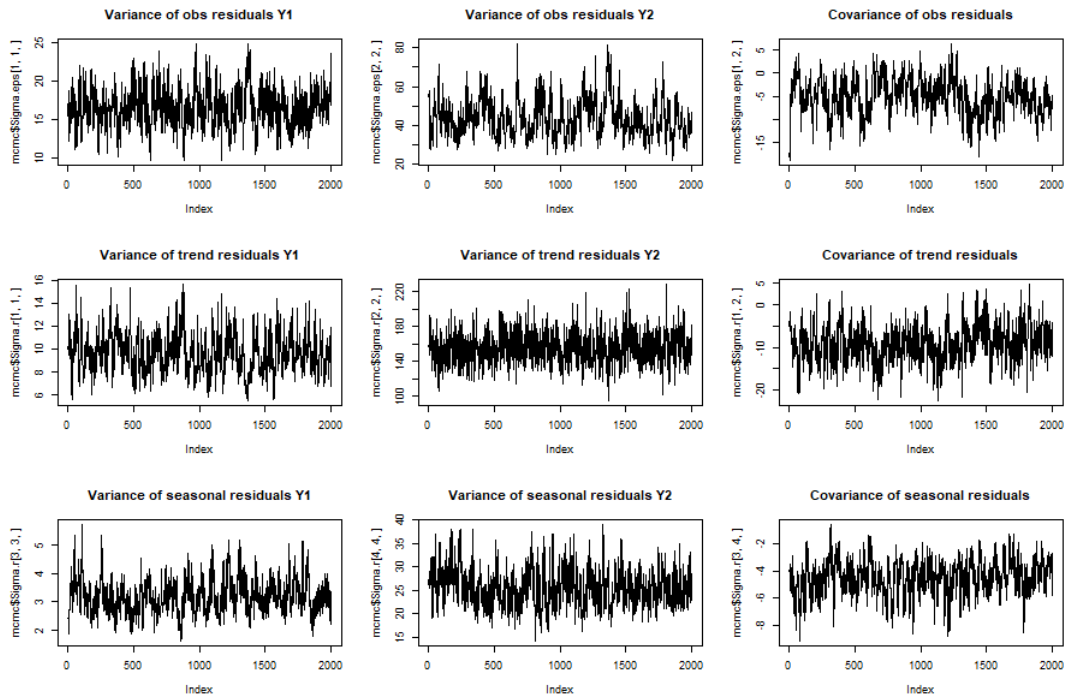
**Table 24:** Geweke’s diagnostics at the lower 10% and upper 50% quantiles. In this table,  $\sigma_i^2$ ,  $i \in \{1, 2\}$  and  $\sigma_{1,2}$  indicate, respectively, the variances and the covariance of the observation disturbances;  $\sigma_{\mu_i}^2$  and  $\sigma_{\mu_{1,2}}$  the variances and the covariance of the trend disturbances;  $\sigma_{\gamma_i}^2$  and  $\sigma_{\gamma_{1,2}}$  the variances and the covariance of the disturbances of the seasonal component.

	$\sigma_1^2$	$\sigma_2^2$	$\sigma_{1,2}$	$\sigma_{\mu_1}^2$	$\sigma_{\mu_2}^2$	$\sigma_{\mu_{1,2}}$	$\sigma_{\gamma_1}^2$	$\sigma_{\gamma_2}^2$	$\sigma_{\gamma_{1,2}}$
1	0.41	0.92	0.75	0.59	0.32	0.71	0.95	0.03	0.03
2	0.94	0.75	0.25	0.71	0.19	0.55	0.88	0.00	0.01
3	0.62	0.99	0.64	0.88	0.52	0.80	0.98	0.01	0.06
4	0.55	0.83	0.96	0.71	0.16	0.67	0.65	0.02	0.07
5	0.59	0.98	0.65	0.97	0.27	0.73	0.88	0.20	0.30
6	0.76	0.83	0.81	0.39	0.17	0.98	0.90	0.17	0.28
7	0.81	0.54	0.16	0.53	0.47	0.04	0.89	0.18	0.10
8	0.29	0.78	0.47	0.40	0.45	0.83	0.30	0.64	0.02
9	0.86	0.65	0.46	0.39	0.24	0.91	0.95	0.87	0.11
10	0.81	0.05	0.91	0.53	0.34	0.79	0.72	0.35	0.04

**Figure 46:** Trace plots of the variance-covariance matrices of the model estimated on the first store-competitor pair.



**Figure 47:** Trace plots of the variance-covariance matrices of the model estimated on the second store-competitor pair.



## B.8 Proof of relations (53), (54) and (55)

Under the setup formalized in Section 5.3, the  $k$ -step ahead forecast  $\hat{S}_{t_m-1+k}$  under  $H_0$ , conditioning to the information set up to time  $t_m - 1$ , can be considered an estimate of the missing potential outcome in the absence of intervention. Thus, an estimator of the point effect is the  $k$ -step ahead prediction error,

$$\begin{aligned} S_{t_m-1+k}(\mathbf{w}) - \hat{S}_{t_m-1+k}(\mathbf{w}') &= z_{t_m-1+k}(\mathbf{w}) + \tau_{t_m-1+k}^{(n)}(\mathbf{w}; \mathbf{w}') - \hat{z}_{t_m-1+k|t_m-1}(\mathbf{w}') \quad (66) \\ &= \sum_{i=0}^{k-1} \psi_{i,m} \varepsilon_{t_m-1+k-i} + \tau_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') \end{aligned}$$

Where the last expression comes from the Wold representation of the covariance stationary process  $z_t$ . Hence,  $\forall t_m \in \Lambda$ ,

$$\hat{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') \sim \left[ \tau_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}'), \sigma_{\varepsilon_m}^2 \sum_{i=0}^{k-1} \psi_{i,m}^2 \right]$$

where  $\sigma_{\varepsilon_m}^2$  is the variance of the error terms of the C-ARIMA model estimated on the observations up to time  $t_m$ . Note that by setting  $k = 1$  we obtain the estimator of the contemporaneous effect. Furthermore, relation (66) allows us to derive the estimators for the cumulative and the temporal average pointwise effects,

$$\begin{aligned} \sum_{h=1}^k (S_{t_m-1+h}(\mathbf{w}) - \hat{S}_{t_m-1+h}(\mathbf{w}')) &= \sum_{h=1}^k \sum_{i=0}^{h-1} \psi_{i,m} \varepsilon_{t_m-1+h-i} + \sum_{h=1}^k \tau_{t_m-1+h}^{(m)}(\mathbf{w}; \mathbf{w}') \\ &= \sum_{h=1}^k \varepsilon_{t_m-1+h-i} \sum_{i=0}^{k-h} \psi_{i,m} + \sum_{h=1}^k \tau_{t_m-1+h}^{(m)}(\mathbf{w}; \mathbf{w}') \end{aligned}$$

Then,

$$\hat{\Delta}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') \sim \left[ \Delta_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}'), \sigma_{\varepsilon_m}^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_{i,m} \right)^2 \right]$$

and

$$\hat{\bar{\tau}}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}') \sim \left[ \bar{\tau}_{t_m-1+k}^{(m)}(\mathbf{w}; \mathbf{w}'), \frac{1}{k^2} \sigma_{\varepsilon_m}^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_{i,m} \right)^2 \right]$$

Finally, to derive the estimators of the aggregate pointwise effects, we need to assume that the individual effects are independent, which is not a stringent assumption in our framework. Indeed, every time we want to estimate an effect, by conditioning to the past information set

we take previous interventions as fixed. In other words, this is an assumption about the effect that previous interventions have when a new intervention take place and we are assuming that this effect is the same in both scenarios. Thus, from the above relations we can derive,

$$\hat{\Delta}_k(w; w') \sim \left[ \Delta_k(w; w'), \sum_{m=1}^M \sigma_{\varepsilon_m}^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_{i,m}^2 \right) \right]$$

$$\hat{\bar{\tau}}_k(w; w') \sim \left[ \bar{\tau}_k(w; w'), \frac{1}{(Mk)^2} \sum_{m=1}^M \sigma_{\varepsilon_m}^2 \sum_{h=1}^k \left( \sum_{i=0}^{k-h} \psi_{i,m}^2 \right) \right]$$

By setting  $k = 1$  we also obtain the estimators of the aggregate contemporaneous effect and the average contemporaneous effect.