Manuscript submitted to *The Econometrics Journal*, pp. S1–S18.

Combining counterfactual outcomes and ARIMA models for policy evaluation: Online supplement

Fiammetta Menchetti[†], Fabrizio Cipollini[†] and Fabrizia Mealli[†,‡]

[†] *DiSIA, University of Florence, Viale Morgagni 59, 50134 Florence, Italy.*
E-mail: fiammetta.menchetti@unifi.it, fabrizio.cipollini@unifi.it
[‡] *Department of Economics, European University Institute, Via delle Fontanelle, 18, 50014 Fiesole, Italy.*
E-mail: fabrizia.mealli@unifi.it

Summary
The Rubin Causal Model (RCM) is a framework that allows to define the causal effect of an intervention as a contrast of potential outcomes. In recent years, several methods have been developed under the RCM to estimate causal effects in time series settings. None of these makes use of ARIMA models, which are instead very common in the econometrics literature. In this paper, we propose a novel approach, C-ARIMA, to define and estimate the causal effect of an intervention in observational time series settings under the RCM. We first formalize the assumptions enabling the definition, the estimation and the attribution of the effect to the intervention. We then check the validity of the proposed method with a simulation study. In the empirical application, we use C-ARIMA to assess the causal effect of a permanent price reduction on supermarket sales. The CausalArima R package provides an implementation of our proposed approach.

Keywords: *Causal Inference, Econometrics, Potential Outcomes, Time Series*

S1. BOOTSTRAP ALGORITHM

Inference on the causal effects of interest can be performed in two ways: if one trusts the Normality assumption of the error term, it is possible to derive confidence intervals and/or hypothesis tests based on Equation (4.14); otherwise, we can resort to a bootstrap strategy. We list below the steps needed to perform bootstrap-based inference.

Algorithm S1.1. Let $\widehat{\varepsilon}_1$ be the vector of model residuals of dimension $t^* \times 1$. Indicate with $\widehat{\vartheta}$ the vector of estimated ARMA parameters and with N the number of bootstrap replications.

Step *1.* Compute A (once for all replications) according to Equation (4.13).
Step *2.* From $\widehat{\varepsilon}_1$, sample with reimmission a vector of residuals $\widehat{\varepsilon} = (\widehat{\varepsilon}_1, \widehat{\varepsilon}_2)$ of dimension $(t^* + K)$.
Step *3.* Based on the entire vector of resampled residuals, simulate $z_t^\dagger$ based on the entire vector of resampled residuals and the estimated ARMA parameters $\widehat{\vartheta}$.
Step *4.* Substitute the simulated ARMA-errors $z^\dagger = (z_1^\dagger, z_2^\dagger)$ and the A matrix computed in Step 1 in Equation (4.12).
Step *5.* Repeat N times from Step 2 to Step 4 to derive the sampling distribution of the causal effect at each point in time.
Step *6.* For each $t \in \{t^* + 1, \ldots, t^* + K\}$ compare the estimated effect $\widehat{\tau}_t^\dagger$ with its sampling distribution, given by the corresponding row of the $K \times N$ matrix derived from previous step.

S2. 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 S2.1 and S2.2 illustrate, respectively, the simulations design and the results.

S2.1. Design

We generate 2000 replications from the following ARIMA$(1,0,1)(1,0,1)_7$ model,[1]

$$\mathrm{Y}_t = \beta_1 \, \mathrm{X}_{1,t} + \beta_2 \, \mathrm{X}_{2,t} + z_t$$
$$z_t = \frac{\theta_q(L)\Theta_Q(L^s)}{\phi_p(L)\Phi_P(L^s)} \varepsilon_t.$$

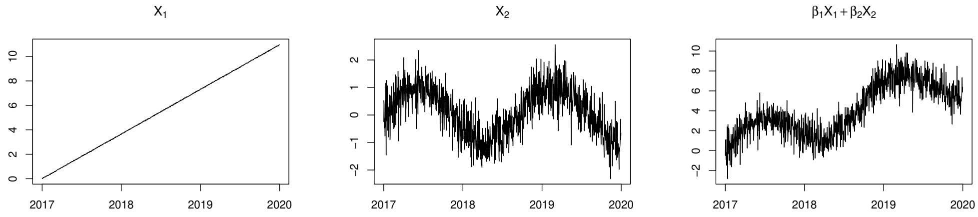
The two covariates of the regression equation are generated as $\mathrm{X}_{1,t} = \alpha_1 t + u_{1,t}$ and $\mathrm{X}_{2,t} = \sin(\alpha_2 t) + u_{2,t}$, with $\alpha_1 = \alpha_2 = 0.01$, $u_{1,t} \sim N(0, 0.02)$, $u_{2,t} \sim N(0, 0.5)$ and coefficients $\beta_1 = 0.7$ and $\beta_2 = 2$, respectively; 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 $\sigma_\varepsilon = 2$. Figure S1 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 interventions: i) a level shift with 4 different magnitudes, i.e., +0% (absence of effect), +10%, +25%, +50%; ii) two time-varying, sinusoidal-shaped interventions resembling an effect that fades after a while to increase again near the end of the analysis period (indicated with IRR 1 and IRR 2). Figure S2 provides a graphical representation of the two types of interventions for one of the simulated time series as well as the pattern of the irregular effects. Notice that IRR 1 is designed such that the effect is negative after three months from the intervention and zero at the end of the analysis period. Instead, the effect under IRR 2 is always positive except when it is exactly zero at the 3-month horizon.

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 Section 4.5. Recall from Section 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 and the best fitting model selected by BIC minimization. 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 8000 and the total number of estimated causal effects is $168,000$ (one for each

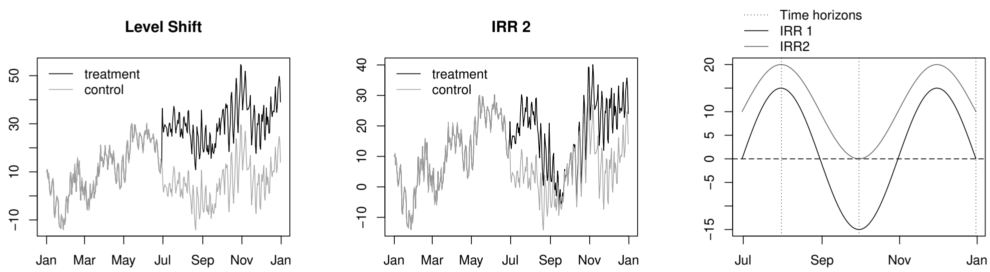
[1] Notice that $D = d = 0$ implies $\tau_t \equiv \tau_t^\dagger$.

Figure S1. Evolution of the generated covariates, $X_{1,t}$ and $X_{2,t}$ and their combination according to the simulated model, $\beta_1 X_{1,t} + \beta_2 X_{2,t}$.



(a) (b) (c)

Figure S2. (a) Level shift of +25% for one simulated series; (b) irregular effect (IRR 2); (c) pattern of the irregular effects.



(a) (b) (c)

time series, model, tested intervention and time horizon). We measure the performance of both approaches in terms of the following indicators:

(a) the probability of rejecting the null hypothesis of absence of effect when it is true (*type I error probability*);
(b) the probability of correctly rejecting the null hypothesis when it is false (*power*);
(c) computational time.

We can also compare the two approaches based on the interval coverage (i.e., how many times the confidence interval at the 95% level contains the true effect) and bias (i.e., deviation of the estimated effects from the corresponding true values).

S2.2. Results

Table S1 reports the type I error probability for the point, cumulative and temporal average estimands under C-ARIMA and for the coefficient of the dummy variable under REG-ARIMA. As expected, both approaches result in type I error probabilities at the desired threshold ($\alpha = 0.05$) at each time horizon. Furthermore, there are no major differences between using the correct model or selecting it by BIC.

Table S2 reports the power of the tests based on the point effect (for C-ARIMA) and the dummy coefficient (for REG-ARIMA). When the intervention takes the form of a level shift, unsurprisingly REG-ARIMA yields better results than C-ARIMA. Indeed, the former model is especially suited for interventions in this form. However, when we consider irregular interventions, REG-ARIMA fails to detect the negative effect under IRR 1 and incorrectly rejects the null hypothesis of absence of effect at the second time horizon under IRR2. Conversely, C-ARIMA performs well when the effect is irregular: it is able to detect the negative effect at the second horizon under IRR 1 and it doesn't reject the null when the effect is zero (third horizon under IRR 1, second horizon under IRR 2 as it is shown in Figure S2).

Finally, the computational time needed to perform simulations under each method is displayed in Table S3. In this simulation setup, C-ARIMA is 4 times faster than REG-ARIMA. This is mainly due to the fact that C-ARIMA estimates the model only once and then predicts a counterfactual for any given time point in the post-intervention period. As a result, we can estimate all the three effects defined in Section 3 with no additional computational effort. Instead, under REG-ARIMA we can estimate only one effect (i.e., the dummy coefficient) and model estimation has to be repeated for each time horizon. This is another advantage of C-ARIMA: defining the estimands as contrast of potential outcomes allows to perform model estimation and estimation of causal effects in two different steps. Conversely, under REG-ARIMA and intervention analysis in general, the estimation of the effect overlaps with model estimation. This affects computational speed and also exclude the resulting effect from having a causal connotation.

To add further details, we report the results of the bias and interval coverage in Table S4 and Table S5, respectively. We notice that C-ARIMA always achieves the desired 95% coverage, whereas the intervals computed under REG-ARIMA fails to include the true effect when this is irregular. Similarly, REG-ARIMA is more biased than C-ARIMA for IRR 1 and at the first time horizon for IRR 2. Figure S3 also presents the evolution of the variance of the point, cumulative and temporal average causal effect estimators.

Concluding, REG-ARIMA approach is more computationally demanding and fails to detect irregular interventions (both in terms of power and interval coverage). As expected, REG-ARIMA 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 C-ARIMA 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, BIC has the property to give consistent model selection and, in our case, it is able to select the true model 74% of the times.

Table S1. Type I error probability for C-ARIMA and REG-ARIMA under the null hypothesis of absence of effect.

	TRUE			BIC		
	1m	3m	6m	1m	3m	6m
C-ARIMA $\tau_t^\dagger$	0.0505	0.0575	0.0595	0.0505	0.0565	0.0585
C-ARIMA $\Delta_t^\dagger$	0.0545	0.0545	0.0555	0.0540	0.0555	0.0550
C-ARIMA $\bar{\tau}_t^\dagger$	0.0545	0.0545	0.0555	0.0540	0.0555	0.0550
REG-ARIMA β_D	0.0510	0.0515	0.0480	0.0505	0.0530	0.0480

Note: In this table, type I error probabilities are reported for each test and time horizon: 1 month (1m), 3 months (3m) and 6 months (6m) from the intervention. The table shows the results under the true model (denoted by $TRUE$) and under the best-fitting model selected by BIC (denoted by BIC).

Table S2. Power of the test based on $\widehat{\tau}_t^\dagger$ (for C-ARIMA) and $\widehat{\beta}_D$ (for REG-ARIMA).

	C-ARIMA, $\tau_t^\dagger$						REG-ARIMA, β_D					
	TRUE			BIC			TRUE			BIC		
	1 m	3 m	6 m	1 m	3 m	6 m	1 m	3 m	6 m	1 m	3 m	6 m
IRR 1	0.480	0.477	0.060	0.482	0.478	0.059	**0.052**	**0.058**	0.053	**0.054**	**0.062**	0.051
IRR 2	0.724	0.058	0.242	0.726	0.056	0.242	1.000	**1.000**	1.000	1.000	**1.000**	1.000
+10%	0.243	0.248	0.242	0.243	0.250	0.242	1.000	1.000	1.000	1.000	1.000	1.000
+25%	0.895	0.887	0.879	0.895	0.887	0.879	1.000	1.000	1.000	1.000	1.000	1.000
+50%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: The numbers in bold highlight when the power function significantly deviates from what is expected. The results are reported for both the true model ($TRUE$) and the best-fitting model (BIC) at three time horizons: 1 month (1m), 3 months (3m) and 6 months (6m) from the intervention. The different impact sizes ranging from +10% to +100% in the rows denote estimated effects in the form of level shifts, whereas IRR 1 and IRR 2 indicate the irregular effects.

Table S3. Computational time for simulations under C-ARIMA and REG-ARIMA.

Model	Time horizons	Estimands	Series	Estimated effects	Time (sec.)
C-ARIMA	185	3	2,000	$1,1 \cdot 10^6$	6,641
REG-ARIMA	3	1	2,000	$6 \cdot 10^3$	27,035

Note: Under C-ARIMA, it is possible to define three causal estimands (cf. Section 3), whereas under REG-ARIMA we can define only one effect, i.e., β_D. Model estimation under C-ARIMA is performed only once irrespective of the time horizons for the prediction; therefore, we can simultaneously estimate three causal effects for each day following the intervention, for a total of $185 \cdot 3 \cdot 2000$ estimated effects. Instead, the estimation of β_D under REG-ARIMA needs to be repeated for each time horizon.

Table S4. Absolute difference between the true effect and $\widehat{\bar{\tau}}_t^{\dagger}$ (for C-ARIMA) and $\widehat{\beta}$ (for REG-ARIMA).

	C-ARIMA, $\bar{\tau}_t^{\dagger}$						REG-ARIMA, β_D					
	TRUE			BIC			TRUE			BIC		
	1 m	3 m	6 m	1 m	3 m	6 m	1 m	3 m	6 m	1 m	3 m	6 m
IRR1	3.639	3.259	2.833	3.637	3.259	2.830	9.601	3.390	3.372	9.578	3.393	3.345
IRR2	3.639	3.259	2.833	3.637	3.259	2.830	6.402	2.284	2.258	6.387	2.285	2.234
+10%	3.639	3.259	2.833	3.637	3.259	2.830	0.857	0.853	0.850	0.855	0.848	0.844
+25%	3.639	3.259	2.833	3.637	3.259	2.830	0.857	0.853	0.850	0.855	0.848	0.844
+50%	3.639	3.259	2.833	3.637	3.259	2.830	0.857	0.853	0.850	0.855	0.848	0.844

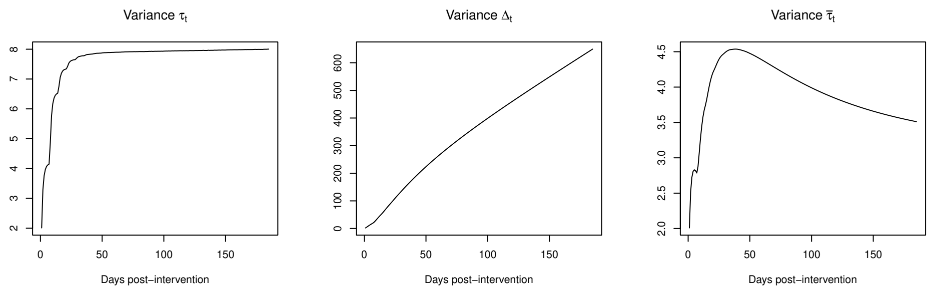
Note: The table reports the bias of the estimated effects from the true effect at each time horizon under the true generative model (denoted by $TRUE$ and under the best-fitting model selected by BIC (denoted by BIC). The different impact sizes ranging from +10% to +100% in the rows denote estimated effects in the form of level shifts, whereas IRR 1 and IRR 2 indicate the irregular effects.

Table S5. Interval coverage in percentage of the true effects within the estimated intervals around $\bar{\tau}_t^{\dagger}$ (for C-ARIMA) and β_D (for REG-ARIMA).

	C-ARIMA, $\bar{\tau}_t(1;0)$						REG-ARIMA, β					
	TRUE			BIC			TRUE			BIC		
	1 m	3 m	6 m	1 m	3 m	6 m	1 m	3 m	6 m	1 m	3 m	6 m
IRR1	94.55	94.55	94.45	94.60	94.45	94.50	0.00	12.50	12.40	0.00	12.50	12.60
IRR2	94.55	94.55	94.45	94.60	94.45	94.50	0.05	45.15	46.80	0.05	44.60	46.55
+10%	94.55	94.55	94.45	94.60	94.45	94.50	94.90	94.85	95.20	94.95	94.70	95.20
+25%	94.55	94.55	94.45	94.60	94.45	94.50	94.90	94.85	95.20	94.95	94.70	95.20
+50%	94.55	94.55	94.45	94.60	94.45	94.50	94.90	94.85	95.20	94.95	94.70	95.20

Note: The different impact sizes ranging from +10% to +100% in the rows denote estimated effects in the form of level shifts. 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

Figure S3. Evolution of the variances of the point, cumulative and temporal average causal effect estimators in the post-intervention period.



S3. ADDITIONAL RESULTS

S3.1. Store brands

Table S6. Causal effect estimates of the permanent price reduction on sales of store-brand cookies after one month, three months and six months from the intervention.

	Time horizon:		
	1 month	3 months	6 months
Item	$\widehat{\overline{\tau}}_t^{\dagger}$	$\widehat{\overline{\tau}}_t^{\dagger}$	$\widehat{\overline{\tau}}_t^{\dagger}$
1	0.14	$0.15^{\cdot}$	0.18^{**}
2	0.14	0.13	0.14^{*}
3	$0.19^{\cdot}$	0.21^{**}	0.25^{***}
4	0.49^{***}	0.30^{***}	0.32^{***}
5	-0.02	0.07	0.11
6	0.24^{*}	0.34^{***}	0.37^{***}
7	0.55^{***}	0.34^{***}	0.30^{***}
8	0.26^{**}	0.25^{***}	0.14^{*}
9	0.47^{***}	0.20^{***}	0.21^{***}
10	0.66^{***}	0.57^{***}	0.33^{***}
11	$0.12^{\cdot}$	0.16^{*}	0.14^{**}
Note:			$^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001

Note: In this table, $\widehat{\overline{\tau}}_t^{\dagger}$ is the estimated temporal average effect (in all the estimated models d = D = 0, therefore this is the effect on the original variable and $\widehat{\overline{\tau}}_t^{\dagger} = 0$ implies no effect). The empirical critical values are computed by bootstrapping the errors from model residuals as described in Algorithm S1.1.

Figure S4. Daily time series of unit sold, price per unit and autocorrelation function for the 11 store brands. The red vertical bar indicates the intervention date.

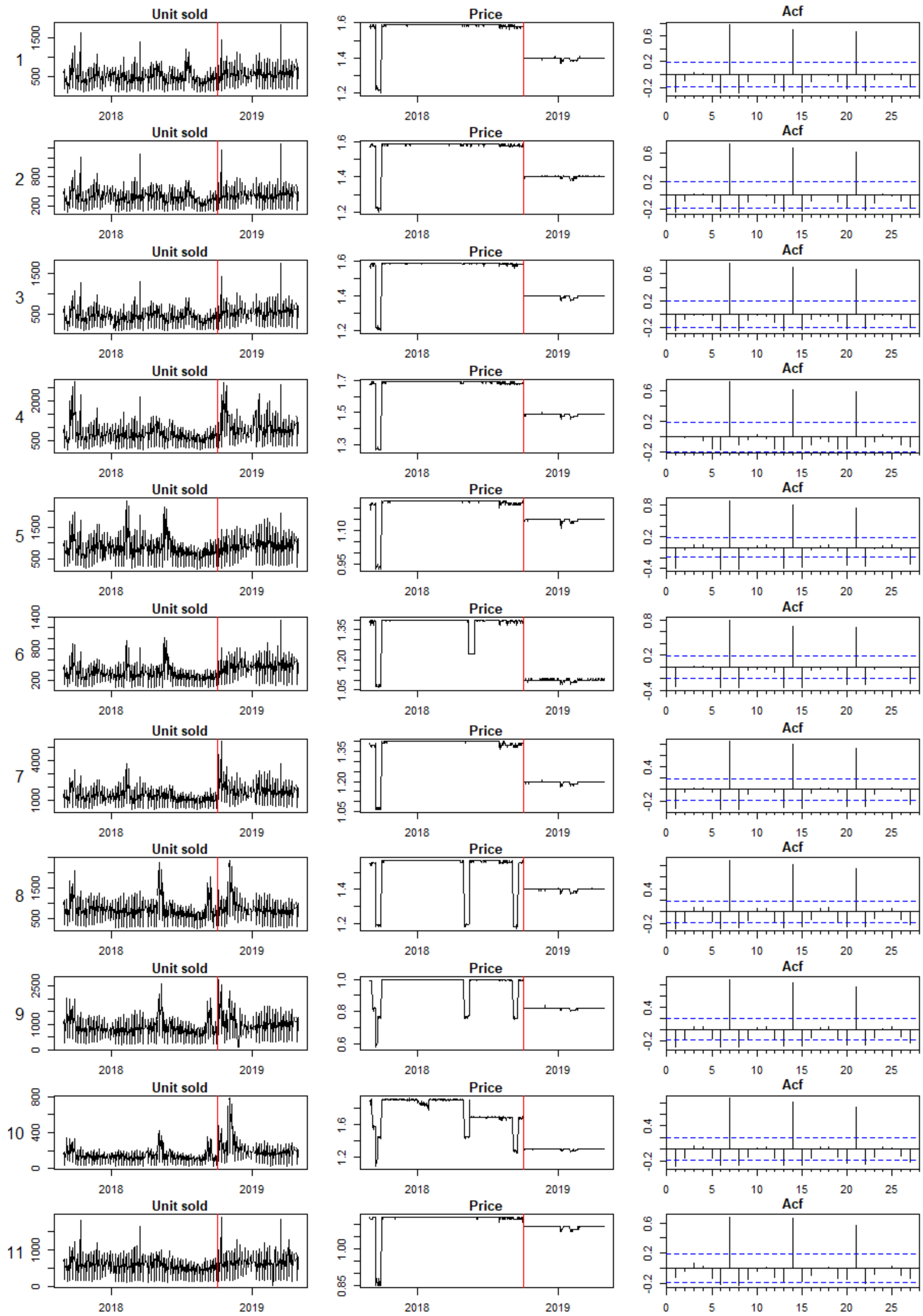


Figure S5. Residual diagnostics (autocorrelation functions and Normal QQ plots) of the C-ARIMA models fitted to the time series of units sold (in log scale).

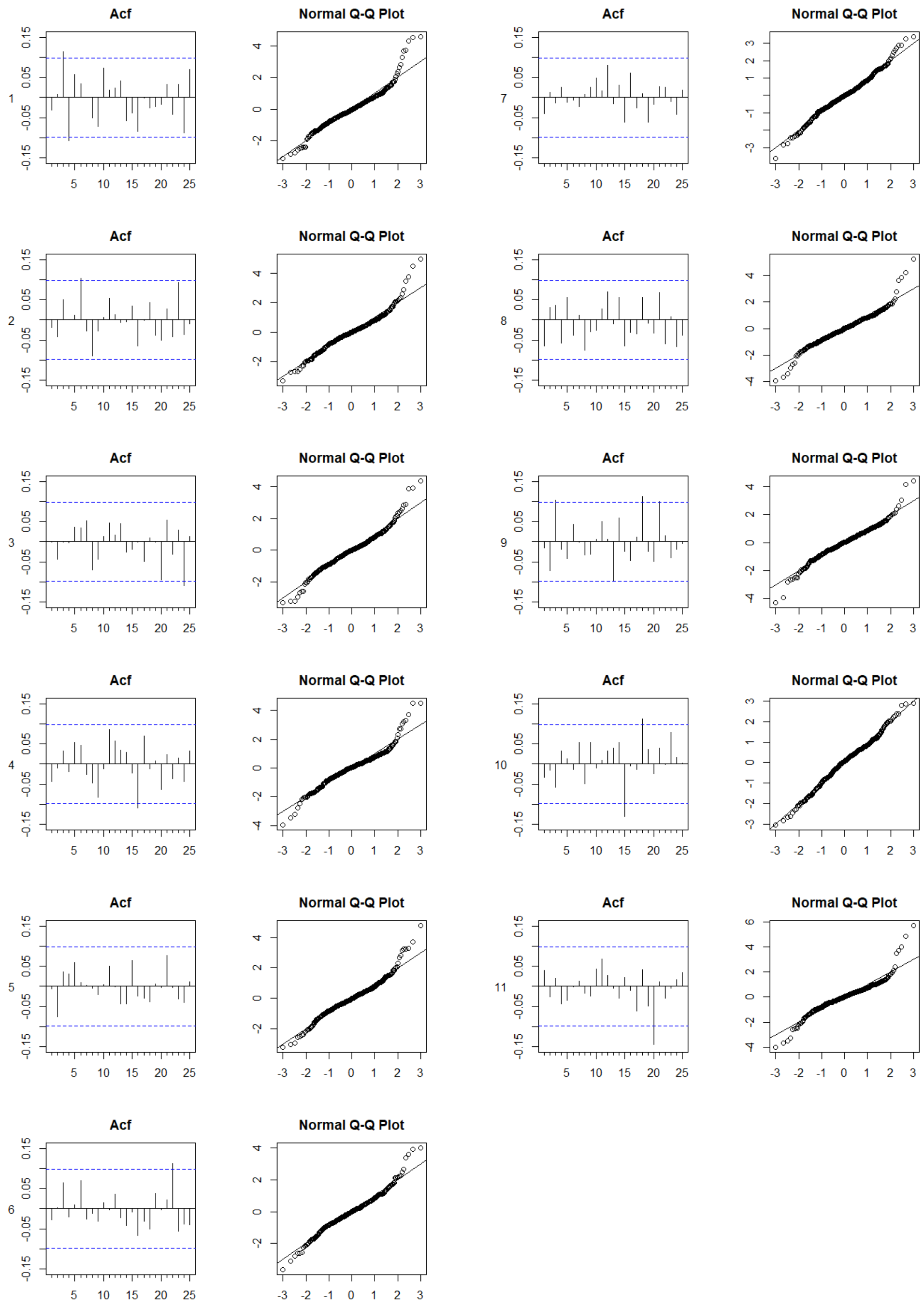


Figure S6. Point causal effect of the new price policy on the sales of store-brand products for each time horizon (1, 3 and 6 months) estimated via C-ARIMA (the dependent variable is the daily sales counts of each product in log scale).

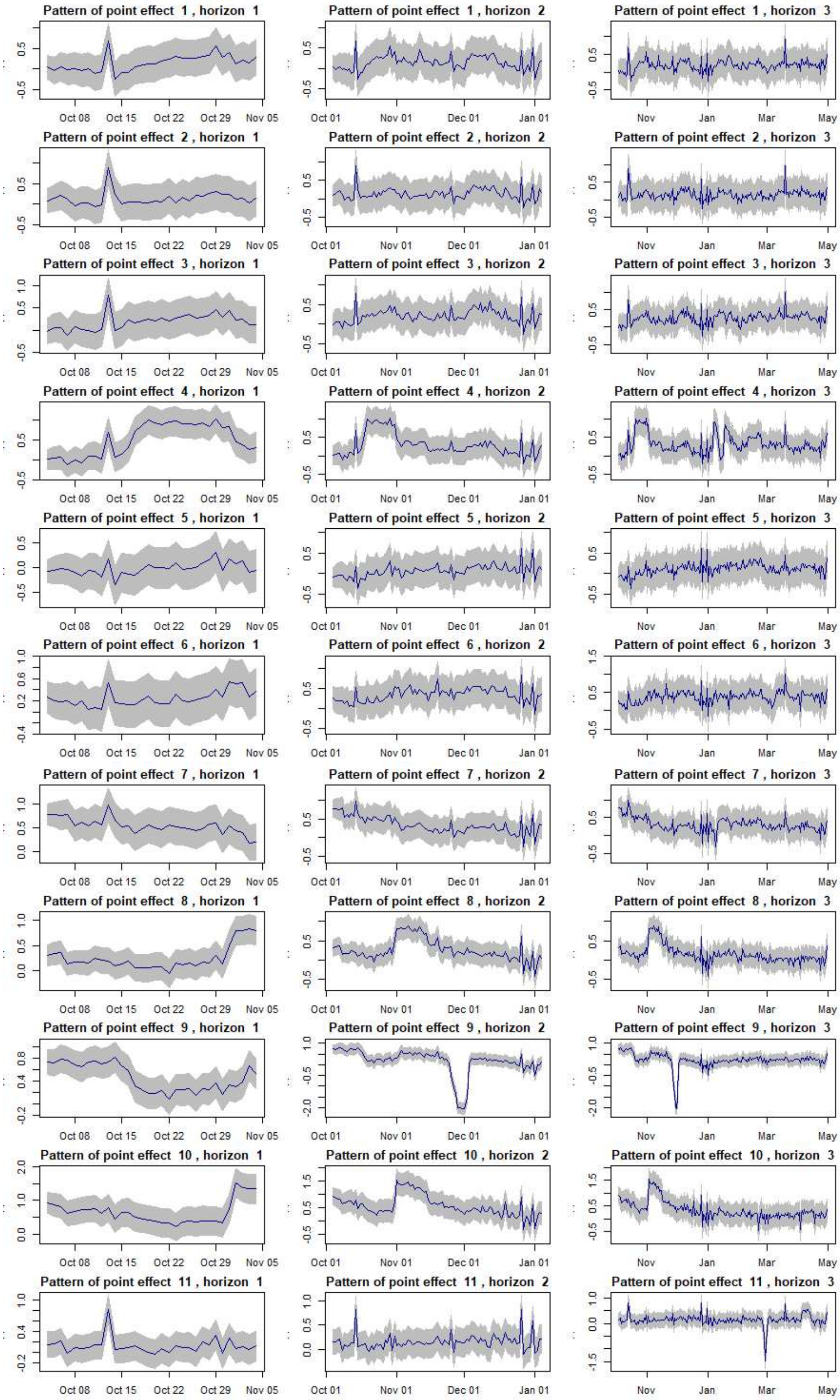


Figure S7. Observed sales (grey) and forecasted sales (blue) of each store brand and for each time horizon (1, 3 and 6 months). The red vertical bar indicates the intervention date.

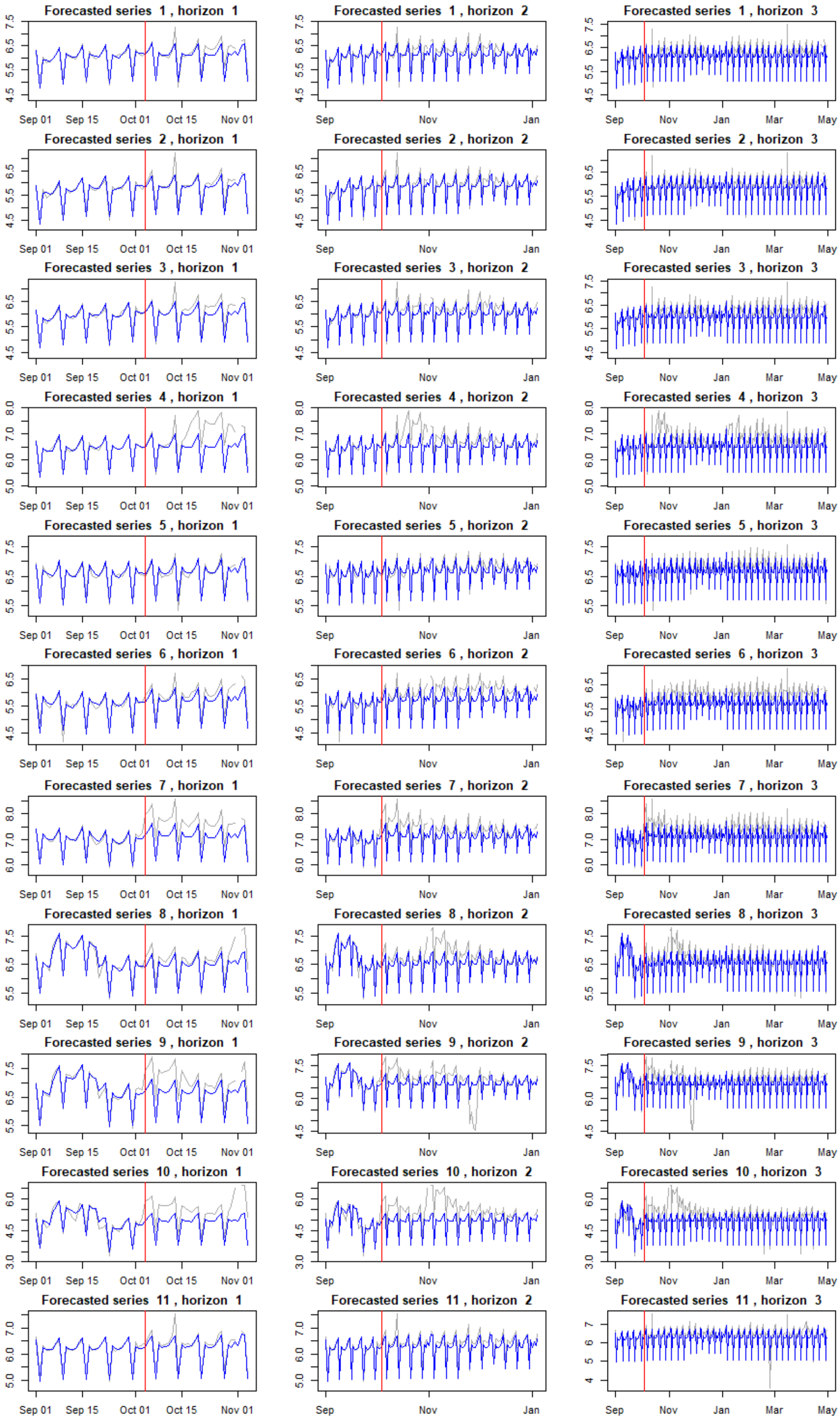


Table S7. Estimated coefficients of the C-ARIMA models fitted to the 11 store brands in the pre-intervention period.

	Dependent variable:										
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11
ϕ_1	0.863^{***}	0.865^{***}	0.833^{***}	0.496^{***}	0.867^{***}	0.890^{***}	0.828^{***}	0.949^{***}	0.839^{***}	0.364^{***}	0.951^{***}
	(0.035)	(0.037)	(0.047)	(0.053)	(0.037)	(0.032)	(0.042)	(0.034)	(0.050)	(0.053)	(0.039)
ϕ_2				0.221^{***}						0.177^{**}	
				(0.052)						(0.055)	
ϕ_3										0.216^{***}	
										(0.054)	
θ_1	-0.323^{***}	-0.340^{***}	-0.348^{***}		-0.251^{**}	-0.458^{***}	-0.191^{*}	-0.509^{***}	-0.476^{***}		-0.841^{***}
	(0.065)	(0.074)	(0.086)		(0.082)	(0.063)	(0.080)	(0.076)	(0.085)		(0.072)
θ_2								-0.212^{**}			
								(0.074)			
Φ_1	0.120^{*}	0.113^{*}	0.203^{***}	0.260^{***}	0.235^{***}		0.234^{***}	0.240^{***}	0.124^{*}	0.774^{***}	0.688^{***}
	(0.056)	(0.057)	(0.055)	(0.053)	(0.056)		(0.057)	(0.059)	(0.059)	(0.204)	(0.146)
Θ_1						0.181^{**}				-0.682^{**}	-0.503^{**}
						(0.058)				(0.231)	(0.172)
c	7.131^{***}	6.767^{***}	6.995^{***}	7.912^{***}	7.113^{***}	6.535^{***}	7.740^{***}	7.800^{***}	6.876^{***}	6.312^{***}	6.711^{***}
	(0.167)	(0.167)	(0.159)	(0.152)	(0.072)	(0.107)	(0.099)	(0.063)	(0.025)	(0.110)	(0.044)
price	-1.721^{***}	-1.441^{***}	-1.656^{***}	-2.218^{***}	-1.433^{***}	-1.841^{***}	-1.423^{***}	-2.316^{***}	-1.727^{***}	-2.118^{***}	-1.175^{***}
	(0.356)	(0.355)	(0.339)	(0.288)	(0.287)	(0.304)	(0.281)	(0.116)	(0.117)	(0.176)	(0.177)
hol	0.176^{***}	0.165^{***}	0.168^{***}	0.160^{***}	0.174^{***}	0.162^{***}	0.162^{***}	0.164^{***}	0.163^{***}	0.076^{*}	0.208^{***}
	(0.033)	(0.033)	(0.034)	(0.029)	(0.027)	(0.033)	(0.027)	(0.023)	(0.026)	(0.035)	(0.028)
Dec.Sun	0.323^{***}	0.369^{***}	0.306^{***}	0.292^{***}	0.379^{***}	0.411^{***}	0.367^{***}	0.327^{***}	0.390^{***}	0.191^{*}	0.374^{***}
	(0.063)	(0.062)	(0.069)	(0.063)	(0.056)	(0.067)	(0.054)	(0.053)	(0.051)	(0.075)	(0.072)
Sat	0.265^{***}	0.243^{***}	0.248^{***}	0.276^{***}	0.293^{***}	0.277^{***}	0.292^{***}	0.223^{***}	0.214^{***}	0.135^{***}	0.162^{***}
	(0.024)	(0.024)	(0.028)	(0.026)	(0.022)	(0.026)	(0.021)	(0.021)	(0.020)	(0.035)	(0.036)
Sun	-1.291^{***}	-1.355^{***}	-1.321^{***}	-1.213^{***}	-1.140^{***}	-1.261^{***}	-1.173^{***}	-1.203^{***}	-1.291^{***}	-1.398^{***}	-1.514^{***}
	(0.028)	(0.027)	(0.031)	(0.029)	(0.026)	(0.029)	(0.025)	(0.025)	(0.022)	(0.037)	(0.036)
Mon	-0.135^{***}	-0.188^{***}	-0.163^{***}	-0.175^{***}	-0.010	-0.062^{*}	-0.037	-0.093^{***}	-0.123^{***}	-0.139^{***}	-0.222^{***}
	(0.028)	(0.028)	(0.032)	(0.030)	(0.026)	(0.029)	(0.026)	(0.025)	(0.022)	(0.035)	(0.036)
Tue	-0.225^{***}	-0.263^{***}	-0.271^{***}	-0.251^{***}	-0.170^{***}	-0.233^{***}	-0.207^{***}	-0.200^{***}	-0.226^{***}	-0.229^{***}	-0.305^{***}
	(0.028)	(0.028)	(0.032)	(0.030)	(0.026)	(0.029)	(0.026)	(0.025)	(0.022)	(0.035)	(0.036)
Wed	-0.247^{***}	-0.250^{***}	-0.266^{***}	-0.271^{***}	-0.209^{***}	-0.243^{***}	-0.232^{***}	-0.249^{***}	-0.266^{***}	-0.262^{***}	-0.312^{***}
	(0.027)	(0.027)	(0.030)	(0.028)	(0.025)	(0.028)	(0.025)	(0.025)	(0.021)	(0.037)	(0.036)
Thr	-0.218^{***}	-0.218^{***}	-0.207^{***}	-0.239^{***}	-0.201^{***}	-0.211^{***}	-0.210^{***}	-0.234^{***}	-0.230^{***}	-0.249^{***}	-0.258^{***}
	(0.024)	(0.024)	(0.028)	(0.026)	(0.022)	(0.026)	(0.021)	(0.021)	(0.020)	(0.035)	(0.036)
Observations	386	386	386	386	386	386	386	386	386	386	386
σ^2	0.022	0.022	0.022	0.017	0.014	0.022	0.014	0.011	0.013	0.027	0.017
Akaike 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:	$^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001										

Note: The dependent variable is the daily sales counts of each product in log scale. Standard errors within parentheses.

S3.2. Competitor brands

Table S8. Causal effect estimates of the permanent price reduction on sales of competitor-brand cookies after one month, three months and six months from the intervention.

	Time horizon:		
	1 month	3 months	6 months
Item	$\widehat{\overline{\tau}}_t^{\dagger}$	$\widehat{\overline{\tau}}_t^{\dagger}$	$\widehat{\overline{\tau}}_t^{\dagger}$
1	−0.04	0.02	0.04
2	−0.13	−0.07	−0.15
3	0.04	0.09	0.17
4	0.00	−0.13	−0.04
5	−0.03	0.05	0.12*
6	−0.05	0.03	0.09
7	0.04	0.11	0.4
8	−0.09	−0.06	−0.08$^{\cdot}$
9	−0.09	−0.11	−0.1
10	−0.03	−0.12*	−0.11*
Note:		$^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001	

Note: In this table, $\widehat{\overline{\tau}}_t^{\dagger}$ is the estimated temporal average effect (in all the estimated models d = D = 0, therefore this is the effect on the original variable and $\widehat{\overline{\tau}}_t^{\dagger} = 0$ implies no effect). The empirical critical values are computed by bootstrapping the errors from model residuals in Algorithm S1.1.

Figure S8. Daily time series of unit sold, price per unit, relative price to store brand and autocorrelation function for the 10 competitor brands. The red vertical bar indicates the intervention date.

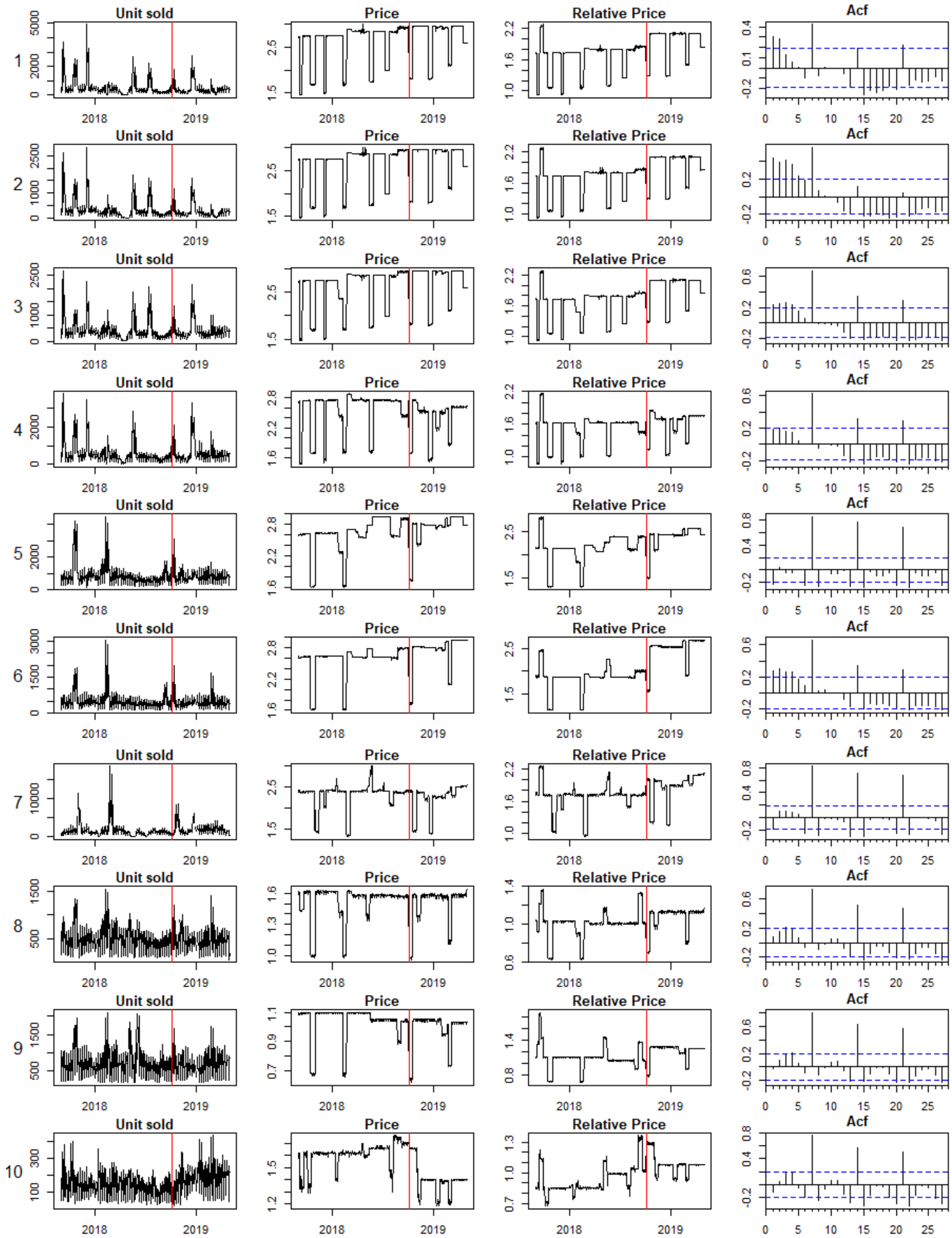


Figure S9. Residual diagnostics (autocorrelation functions and Normal QQ plots) of the C-ARIMA models fitted to the time series of units sold (in log scale).

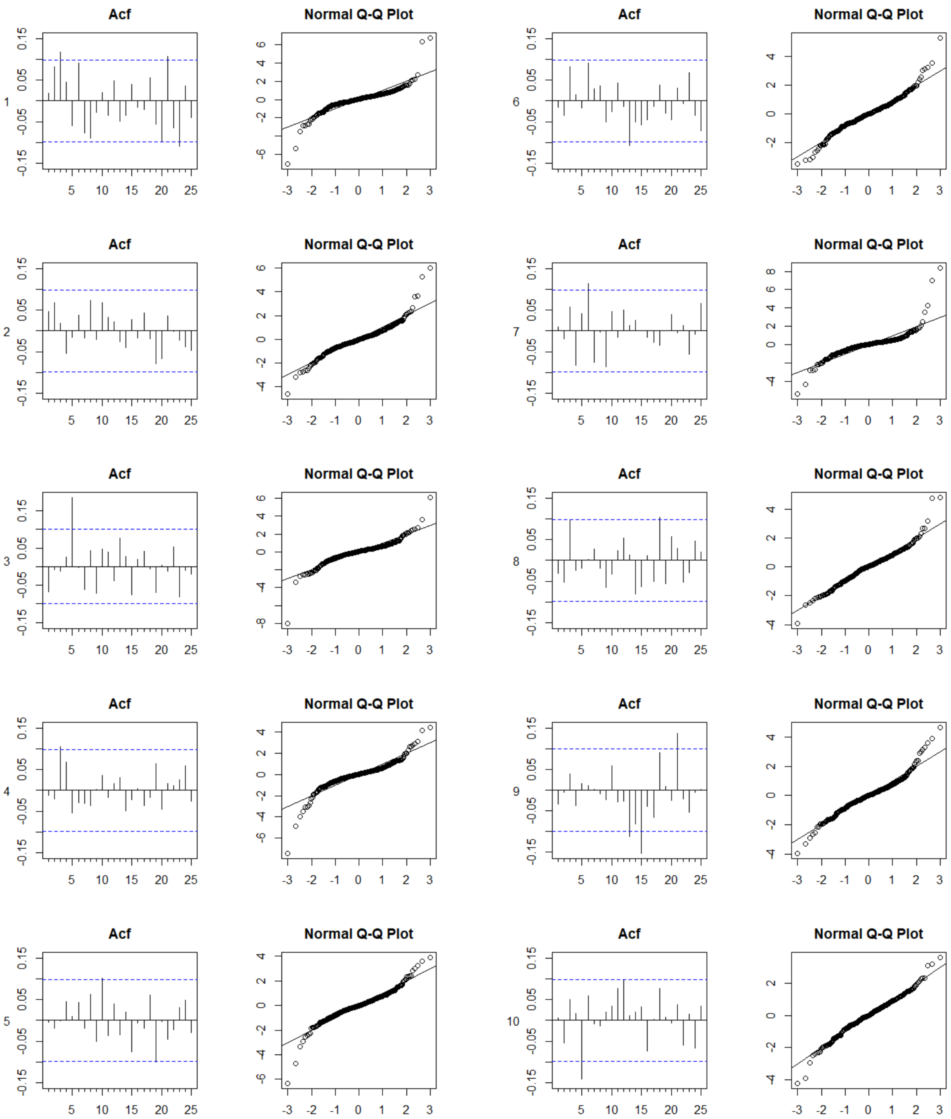


Figure S10. Point causal effect of the new price policy on the sales of competitor-brand products for each time horizon (1, 3 and 6 months) estimated via C-ARIMA (the dependent variable is the daily sales counts of each product in log scale).

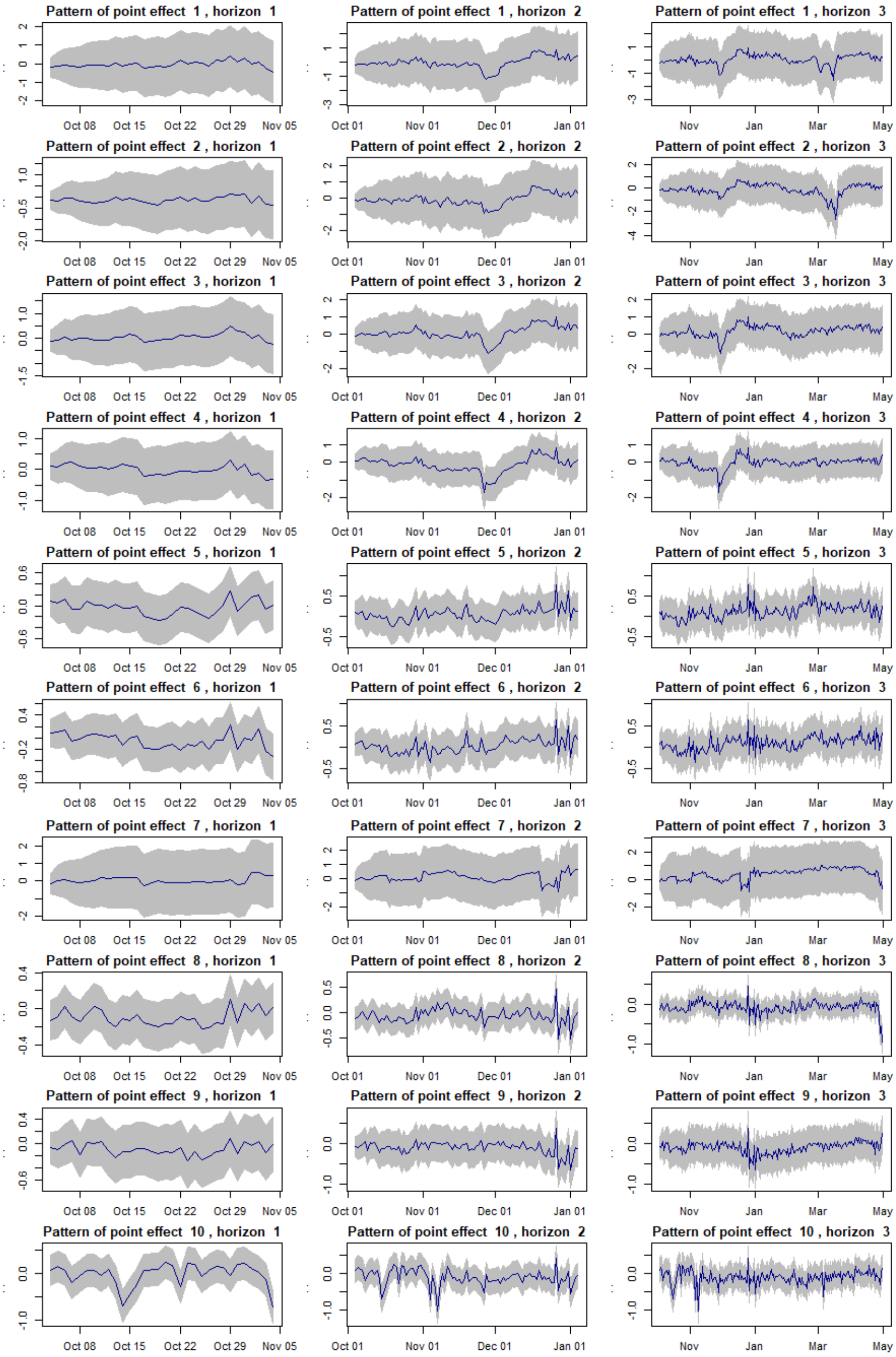


Figure S11. Observed sales (grey) and forecasted sales (blue) of each competitor brand and for each time horizon (1, 3 and 6 months). The red vertical bar indicates the intervention date.

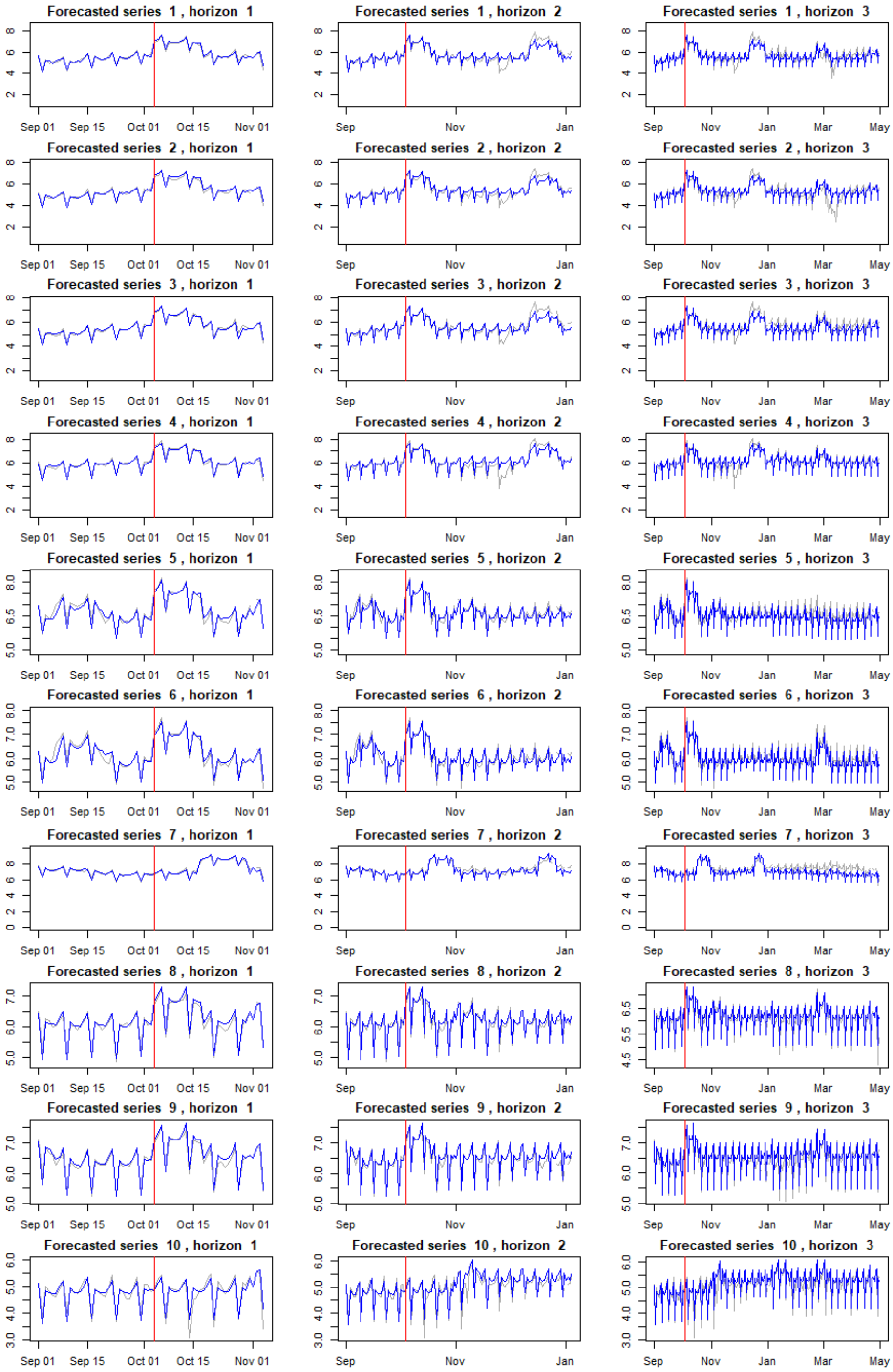


Table S9. Estimated coefficients of the C-ARIMA models fitted to the 10 competitor brands in the pre-intervention period.

	Dependent variable:									
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
ϕ_1	0.947***	0.959***	0.949***	−0.090*	0.876***	0.973***	1.800***	0.939***	0.855***	0.786***
	(0.015)	(0.014)	(0.016)	(0.043)	(0.053)	(0.157)	(0.088)	(0.028)	(0.034)	(0.083)
ϕ_2				0.850***	−0.161**	−0.075	−0.821***			
				(0.038)	(0.054)	(0.130)	(0.081)			
ϕ_3										
θ_1				0.940***		−0.408**	−0.768***	−0.566***	−0.239***	−0.514***
				(0.047)		(0.142)	(0.109)	(0.068)	(0.068)	(0.120)
θ_2								−0.117$^{\cdot}$		
								(0.070)		
Φ_1		0.108*	0.907***	0.224***	0.162**	0.852***			0.281***	
		(0.054)	(0.030)	(0.054)	(0.053)	(0.088)			(0.053)	
Θ_1			−1.000***			−0.779***				
			(0.021)			(0.103)				
c	8.296***	7.862***	7.828***	8.335***	8.873***	8.380***	10.189***	7.029***	6.773***	6.327***
	(0.300)	(0.297)	(0.150)	(0.162)	(0.137)	(0.127)	(0.292)	(0.048)	(0.050)	(0.108)
price	−2.465***	−2.350***	−2.074***	−2.193***	−2.210***	−2.257***	−3.742***	−1.490***	−1.308***	−2.499***
	(0.164)	(0.124)	(0.135)	(0.130)	(0.139)	(0.121)	(0.286)	(0.089)	(0.127)	(0.220)
hol	0.157***	0.161***	0.099*	0.154***	0.107**	0.155***	0.138*	0.178***	0.131***	0.165***
	(0.059)	(0.044)	(0.048)	(0.045)	(0.033)	(0.029)	(0.062)	(0.025)	(0.029)	(0.038)
Dec.Sun	0.404***	0.538***	0.446***	0.515***	0.431***	0.379***	0.363***	0.413***	0.338***	0.437***
	(0.092)	(0.076)	(0.068)	(0.085)	(0.057)	(0.056)	(0.095)	(0.046)	(0.063)	(0.069)
Sat	0.323***	0.262***	0.272***	0.244***	0.322***	0.357***	0.315***	0.216***	0.263***	0.215***
	(0.036)	(0.030)	(0.013)	(0.034)	(0.024)	(0.027)	(0.037)	(0.018)	(0.025)	(0.027)
Sun	−0.982***	−1.143***	−1.083***	−1.351***	−1.011***	−0.992***	−1.057***	−1.303***	−1.288***	−1.303***
	(0.047)	(0.039)	(0.018)	(0.045)	(0.034)	(0.031)	(0.050)	(0.020)	(0.030)	(0.029)
Mon	−0.104*	−0.182***	−0.172***	−0.199***	−0.050	−0.085**	0.017	−0.117***	−0.119***	−0.119***
	(0.051)	(0.042)	(0.018)	(0.048)	(0.037)	(0.032)	(0.054)	(0.020)	(0.031)	(0.029)
Tue	−0.254***	−0.240***	−0.266***	−0.267***	−0.206***	−0.241***	−0.166**	−0.208***	−0.231***	−0.229***
	(0.051)	(0.042)	(0.018)	(0.048)	(0.037)	(0.032)	(0.054)	(0.020)	(0.031)	(0.029)
Wed	−0.237***	−0.222***	−0.284***	−0.285***	−0.197***	−0.254***	−0.218***	−0.237***	−0.242***	−0.270***
	(0.046)	(0.039)	(0.016)	(0.044)	(0.033)	(0.030)	(0.049)	(0.020)	(0.029)	(0.028)
Thr	−0.208***	−0.182***	−0.220***	−0.237***	−0.187***	−0.232***	−0.200***	−0.196***	−0.219***	−0.234***
	(0.036)	(0.030)	(0.012)	(0.034)	(0.024)	(0.027)	(0.037)	(0.018)	(0.025)	(0.027)
Observations	386	386	385	386	386	386	386	386	386	386
σ^2	0.081	0.046	0.049	0.044	0.021	0.017	0.094	0.013	0.017	0.027
Akaike Inf. Crit.	147.413	-66.040	-33.990	-85.448	-372.686	-451.392	207.605	-575.002	-449.421	-281.026
Note:										$^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001

Note: The dependent variable is the daily sales counts of each product in log scale. Standard errors within parentheses.