Using ARIMA Models to make causal statements

Fiammetta Menchetti† & Eugenio Palmieri

November 10, 2021

Berlin Time Series Analysis Meetup

*Postdoctoral researcher, University of Florence

Outline

1. Empirical problem & research goal
2. C-ARIMA approach
   - Causal framework
   - Inference
3. Practical Session

Empirical problem

- On October 4th, 2018, the Florence branch of an Italian supermarket chain store permanently lowered the price of 707 store brands in several categories

Empirical problem

- On October 4th, 2018, the Florence branch of an Italian supermarket chain store permanently lowered the price of 707 store brands in several categories
- Among the 284 items in the "cookies" category, 11 store brands were selected for a permanent price reduction ranging from -5.7% to -23.2%

Empirical problem

- On October 4th, 2018, the Florence branch of an Italian supermarket chain store permanently lowered the price of 707 store brands in several categories
- Among the 284 items in the "cookies" category, 11 store brands were selected for a permanent price reduction ranging from -5.7% to -23.2%
- For each of them, the supermarket chain identified 11 direct competitors.

Empirical problem

- On October 4th, 2018, the Florence branch of an Italian supermarket chain store permanently lowered the price of 707 store brands in several categories
- Among the 284 items in the "cookies" category, 11 store brands were selected for a permanent price reduction ranging from −5.7% to −23.2%
- For each of them, the supermarket chain identified 11 direct competitors.

Figure: Store brands (first row) and direct competitor brands (second row).

Research goal

Task:

Estimating the overall effect of the price policy change on cookies' sales with a special focus on the indirect impact generated on the competitor products

Research goal

Task:

Estimating the overall effect of the price policy change on cookies' sales with a special focus on the indirect impact generated on the competitor products

Not so easy! Inference of a proper "causal" effect is complicated by several issues, including:

- Autocorrelation
- Seasonality
- Holiday effects
- Interference

→

Infer the effect attributable to the policy netting out the portion due to other factors
e.g., increasing trends started in the past
Christmas rush purchases, Sunday effect

"Estimating the causal effect of an intervention in a time series setting: the C-ARIMA approach"

Fiammetta Menchetti, Fabrizio Cipollini, Fabrizia Mealli [2]

1. Causal framework under the Rubin Causal Model, comprising: the **assumptions** allowing the postulation of potential outcomes and related causal contrasts (causal estimands); the assumptions for the identification and estimation of these causal estimands

"Estimating the causal effect of an intervention in a time series setting: the C-ARIMA approach"

Fiammetta Menchetti, Fabrizio Cipollini, Fabrizia Mealli [2]

1. Causal framework under the Rubin Causal Model, comprising: the **assumptions** allowing the postulation of potential outcomes and related causal contrasts (causal estimands); the assumptions for the identification and estimation of these causal estimands
2. An inferential methodology based on **ARIMA models** that involves predicting the outcome in the absence of intervention and contrasting it with the observed outcome

"Estimating the causal effect of an intervention in a time series setting: the C-ARIMA approach"

Fiammetta Menchetti, Fabrizio Cipollini, Fabrizia Mealli [2]

1. Causal framework under the Rubin Causal Model, comprising: the **assumptions** allowing the postulation of potential outcomes and related causal contrasts (causal estimands); the assumptions for the identification and estimation of these causal estimands
2. An inferential methodology based on **ARIMA models** that involves predicting the outcome in the absence of intervention and contrasting it with the observed outcome

C-ARIMA shares many features with "CausalImpact" [1], but it is based on ARIMA models — frequentist alternative to BSTS models, easy to implement, used for a wide class of processes.

Assumptions

Let $\mathrm{W}_{i,t} \in \{0,1\}$ be a random variable describing the treatment assignment of unit $i \in \{1,\ldots,N\}$ at time $t \in \{1,\ldots,T\}$, where 1 denotes that a "treatment" (or "intervention") has taken place and 0 denotes control. We maintain the following assumptions:

Assumptions

Let $W_{i,t} \in \{0,1\}$ be a random variable describing the treatment assignment of unit $i \in \{1,\dots,N\}$ at time $t \in \{1,\dots,T\}$, where 1 denotes that a "treatment" (or "intervention") has taken place and 0 denotes control. We maintain the following assumptions:

(A1: Single persistent intervention)

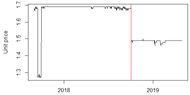
$$\exists t^* \in \{1,\dots,T\}\ s.t\ w_{i,t} = 0\ \ \forall t \le t^*\ and\ \forall t > t^*, w_{i,t} \in \{(1,\dots,1),(0,\dots,0)\}$$

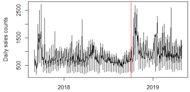
Assumptions

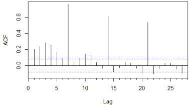
Let $W_{i,t} \in \{0,1\}$ be a random variable describing the treatment assignment of unit $i \in \{1,\dots,N\}$ at time $t \in \{1,\dots,T\}$, where 1 denotes that a "treatment" (or "intervention") has taken place and 0 denotes control. We maintain the following assumptions:

(A1: Single persistent intervention)

$$\exists t^* \in \{1,\dots,T\} \ s.t \ w_{i,t} = 0 \ \ \forall t \leq t^* \ and \ \forall t > t^*, w_{i,t} \in \{(1,\dots,1),(0,\dots,0)\}$$





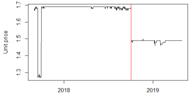


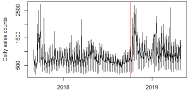
Assumptions

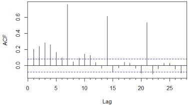
Let $W_{i,t} \in \{0, 1\}$ be a random variable describing the treatment assignment of unit $i \in \{1, \dots, N\}$ at time $t \in \{1, \dots, T\}$, where 1 denotes that a "treatment" (or "intervention") has taken place and 0 denotes control. We maintain the following assumptions:

(A1: Single persistent intervention)

$$\exists t^* \in \{1, \dots, T\} \; s.t \;\; w_{i,t} = 0 \;\; \forall t \leq t^* \; and \; \forall t > t^*, w_{i,t} \in \{(1, \dots, 1), (0, \dots, 0)\}$$







(A2: Temporal no-interference)

$$For \; all \; i \in \{1, \dots, N\}, \; Y_{i,t}(w_{1:N,t^*+1:T}) = Y_{i,t}(w_{i,t^*+1:T})$$

Assumptions

(A3: Covariates-treatment independence)

$$\mathrm{X}_{i,t}(\mathrm{w}_{i,t^*+1:T}) = \mathrm{X}_{i,t}(\mathrm{w}'_{i,t^*+1:T}).$$

(A4: Non-anticipating individualistic treatment)

$$\Pr(\mathrm{W}_{1:N,t^*+1} = \mathrm{w}_{1:N,t^*+1} \,|\, \mathrm{W}_{1:N,1:t^*}, \mathrm{Y}_{1:N,1:T}(\mathrm{w}_{1:N,1:T}), \mathrm{X}_{1:N,1:T}) =$$
$$= \prod_{i=1}^{N} \Pr(\mathrm{W}_{i,t^*+1} = \mathrm{w}_{i,t^*+1} \,|\, \mathrm{Y}_{i,1:t^*}(\mathrm{w}_{i,1:t^*}), \mathrm{X}_{i,1:t^*}).$$

Above assumptions are essential to define, estimate and attribute the causal effect to the intervention. Moreover, they allow us to ease notation: under Assumption 1, for all $t > t^*$ we can write $\mathrm{w}_{i,t} = \mathrm{w}_i$ and if Assumption 2 holds we can also drop the i subscript. From now on, we can use $\mathrm{Y}_t(\mathrm{w})$ to denote the potential outcome of a generic unit at time $t > t^*$.

Causal estimands

Definition

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

$$\tau_t(1;0) = \mathsf{Y}_t(1) - \mathsf{Y}_t(0). \tag{1}$$

Causal estimands

Definition

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

$$\tau_t(1;0) = \mathrm{Y}_t(1) - \mathrm{Y}_t(0). \tag{1}$$

The cumulative causal effect is

$$\Delta_t(1;0) = \sum_{s=t^*+1}^{t} \tau_s(1;0). \tag{2}$$

Causal estimands

Definition

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

$$\tau_t(1;0) = \mathsf{Y}_t(1) - \mathsf{Y}_t(0). \tag{1}$$

The cumulative causal effect is

$$\Delta_t(1;0) = \sum_{s=t^*+1}^{t} \tau_s(1;0). \tag{2}$$

The temporal average causal effect is

$$\bar{\tau}_t(1;0) = \frac{1}{t-t^*} \sum_{s=t^*+1}^{t} \tau_s(1;0) = \frac{\Delta_t(1;0)}{t-t^*}. \tag{3}$$

C-ARIMA

let us assume $\{\mathrm{Y}_t(\mathrm{w})\}$ evolving as

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

where,

- $\phi_p(L)$ and $\theta_q(L)$ are lag polynomials having roots all outside the unit circle
- given this representation, the point causal effect at time $t > t^*$ is $\tau_t \equiv \mathrm{Y}_t(\mathrm{w}=1) - \mathrm{Y}_t(\mathrm{w}=0)$
- $\tau_t = 0 \; \forall t \leq t^*$ and $\mathbb{1}_{\{\mathrm{w}=1\}}$ is an indicator function which is one if $\mathrm{w} = 1$
- ε_t is white noise with mean 0 and variance σ^2_ε

C-ARIMA

let us assume $\{\mathrm{Y}_t(\mathrm{w})\}$ evolving as

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

where,

- $\phi_p(L)$ and $\theta_q(L)$ are lag polynomials having roots all outside the unit circle
- given this representation, the point causal effect at time $t > t^*$ is $\tau_t \equiv \mathrm{Y}_t(\mathrm{w}=1) - \mathrm{Y}_t(\mathrm{w}=0)$
- $\tau_t = 0 \ \forall t \leq t^*$ and $\mathbb{1}_{\{\mathrm{w}=1\}}$ is an indicator function which is one if $\mathrm{w} = 1$
- ε_t is white noise with mean 0 and variance σ^2_ε

C-ARIMA

let us assume $\{Y_t(w)\}$ evolving as

$$(1-L^s)^D(1-L)^d Y_t(w) = \frac{\Theta_Q(L^s)\theta_q(L)}{\Phi_P(L^s)\phi_p(L)}\varepsilon_t + (1-L^s)^D(1-L)^d X_t'\beta + \tau_t \mathbb{1}_{\{w=1\}} \tag{4}$$

where,

- $\phi_p(L)$ and $\theta_q(L)$ are lag polynomials having roots all outside the unit circle
- given this representation, the point causal effect at time $t > t^*$ is $\tau_t \equiv Y_t(w=1) - Y_t(w=0)$
- $\tau_t = 0 \; \forall t \leq t^*$ and $\mathbb{1}_{\{w=1\}}$ is an indicator function which is one if $w = 1$
- ε_t is white noise with mean 0 and variance σ^2_ε
- $\Theta_Q(L^s)$, $\Phi_P(L^s)$ are the lag polynomials of the seasonal part of the model having roots all outside the unit circle
- $(1-L^s)^D$ and $(1-L)^d$ are the differencing operators to ensure stationarity

After some manipulation, Equation (4) becomes

$$S_t = z_t + \tau_t \mathbb{1}_{\{\mathrm{w}=1\}}$$

where, $S_t = T(Y_t) - T(X_t)'\beta$ and $T(\cdot)$ is the transformation of Y_t needed to achieve stationarity, i.e. $T(\mathrm{Y}_t) = (1 - L^s)^D (1 - L)^d \mathrm{Y}_t$; z_t includes the stationary part of the model, namely,

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

If $\mathrm{w} = 0$, absence of intervention, the k-step ahead forecast of S_t conditionally on the information up to time t^* is

$$\hat{S}_{t^*+k}(0) = E[S_{t^*+k}(0)|\mathcal{I}_{t^*}] = E[z_{t^*+k}|\mathcal{I}_{t^*}] = \hat{z}_{t^*+k|t^*}$$

After some manipulation, Equation (4) becomes

$$S_t = z_t + \tau_t \mathbb{1}_{\{\mathrm{w}=1\}}$$

where, $S_t = T(Y_t) - T(X_t)'\beta$ and $T(\cdot)$ is the transformation of Y_t needed to achieve stationarity, i.e. $T(\mathrm{Y}_t) = (1 - L^s)^D (1 - L)^d \mathrm{Y}_t$; z_t includes the stationary part of the model, namely,

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

If $\mathrm{w} = 0$, absence of intervention, the k-step ahead forecast of S_t conditionally on the information up to time t^* is

$$\hat{S}_{t^*+k}(0) = E[S_{t^*+k}(0)|\mathcal{I}_{t^*}] = E[z_{t^*+k}|\mathcal{I}_{t^*}] = \hat{z}_{t^*+k|t^*}$$

$$\begin{aligned} \hat{\tau}_{t^*+k}(1;0) &= S_{t^*+k}(1) - \hat{S}_{t^*+k}(0) \\ &= z_{t^*+k} - \hat{z}_{t^*+k|t^*} + \tau_{t^*+k}. \end{aligned}$$

After some manipulation, Equation (4) becomes

$$S_t = z_t + \tau_t \mathbb{1}_{\{\mathrm{w}=1\}}$$

where, $S_t = T(Y_t) - T(X_t)'\beta$ and $T(\cdot)$ is the transformation of Y_t needed to achieve stationarity, i.e. $T(\mathrm{Y}_t) = (1 - L^s)^D(1 - L)^d\,\mathrm{Y}_t$; z_t includes the stationary part of the model, namely,

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

If $\mathrm{w} = 0$, absence of intervention, the k-step ahead forecast of S_t conditionally on the information up to time t^* is

$$\hat{S}_{t^*+k}(0) = E[S_{t^*+k}(0)|\mathcal{I}_{t^*}] = E[z_{t^*+k}|\mathcal{I}_{t^*}] = \hat{z}_{t^*+k|t^*}$$

$$\begin{aligned}\hat{\tau}_{t^*+k}(1;0) &= S_{t^*+k}(1) - \hat{S}_{t^*+k}(0) \\ &= z_{t^*+k} - \hat{z}_{t^*+k|t^*} + \tau_{t^*+k}.\end{aligned}$$

Point causal effect estimator!

Inference

Theorem (Part 1)

Let $\{Y_t\}$ follow the regression model with ARIMA errors defined in Equation (4) and, for any $k > 0$ let $H_0 : \tau_{t^*+k}(1;0) = 0$ the null hypothesis that the intervention has no effect. Then, the estimators of the point, cumulative and temporal average effects under H_0 can be expressed as,

$$\hat{\tau}_{t^*+k}(1;0)|H_0 = \sum_{i=0}^{k-1} \psi_i \varepsilon_{t^*+k-i} \tag{5}$$

$$\hat{\Delta}_{t^*+k}(1;0)|H_0 = \sum_{h=1}^{k} \varepsilon_{t^*+h} \sum_{i=0}^{k-h} \psi_i \tag{6}$$

$$\hat{\bar{\tau}}_{t^*+k}(1;0)|H_0 = \frac{1}{k} \sum_{h=1}^{k} \varepsilon_{t^*+h} \sum_{i=0}^{k-h} \psi_i, \tag{7}$$

where, the ψ_i's are the coefficients of a moving average of order $k-1$ whose values are functions of the ARMA parameters in Equation (4).

Inference

Theorem (Part 2)

In case the ε_t error term is assumed normally distributed, Equations (5)–(7) become

$$\hat{\tau}_{t^*+k}(1;0)|H_0 \sim N\left[0, \sigma_\varepsilon^2 \sum_{i=0}^{k-1} \psi_i^2\right] \tag{8}$$

$$\widehat{\Delta}_{t^*+k}(1;0)|H_0 \sim N\left[0, \sigma_\varepsilon^2 \sum_{h=1}^{k} \left(\sum_{i=0}^{k-h} \psi_i\right)^2\right] \tag{9}$$

$$\hat{\bar{\tau}}_{t^*+k}(1;0)|H_0 \sim N\left[0, \frac{1}{k^2}\sigma_\varepsilon^2 \sum_{h=1}^{k} \left(\sum_{i=0}^{k-h} \psi_i\right)^2\right] \tag{10}$$

If one relies on the normality of the error term, inference can be based on Equations (8)–(10); otherwise, one can compute empirical critical values from Equations (5)–(7) by bootstrapping the errors from the model residuals.

Inference

Summarizing, to estimate the effect of an intervention with C-ARIMA we need to follow these steps:

1. 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
2. based on the process learned in the pre-intervention period, perform a prediction step and obtain an estimate of the counterfactual outcome during the post-intervention period
3. 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

Concluding remarks

- The proposed C-ARIMA approach can be used for policy evaluation (e.g., new regulation, passage of a law) and, in general, to estimate the causal effect of interventions in time series settings where there are no control units available, as in those cases where all units are treated
- In observational panel studies where a group of control units is also present, difference-in-differences and synthetic control methods may be preferred when the temporal dimension is short compared to the cross-sectional dimension

Practical Session

GitHub

FMenchetti/CausalArima

In the past few days, you were given sales data of one of the store brand products analyzed in the paper. We will now show how to use C-ARIMA to answer the following questions:

1. What is the total number of units sold due to the price reduction after 1-week, 1 month and 3 months from the intervention?
2. What is the total number of units sold at the end of the analysis period?
3. Plot the causal effect and residual diagnostics
4. Assume a multiplicative effect, how much more did the product sold due to the premanent price reduction?

Go to the exercise!

References

Thanks for your attention!

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

[2] Menchetti, F., Cipollini, F. and Mealli, F. [2021]. Estimating the causal effect of an intervention in a time series setting: the C-ARIMA approach, *Preprint. Available at* *https://arxiv.org/abs/2103.06740* .