Brodersen, Gallusser, Koehler, Remy and Scott(2014)
Discussion Handout

Topic

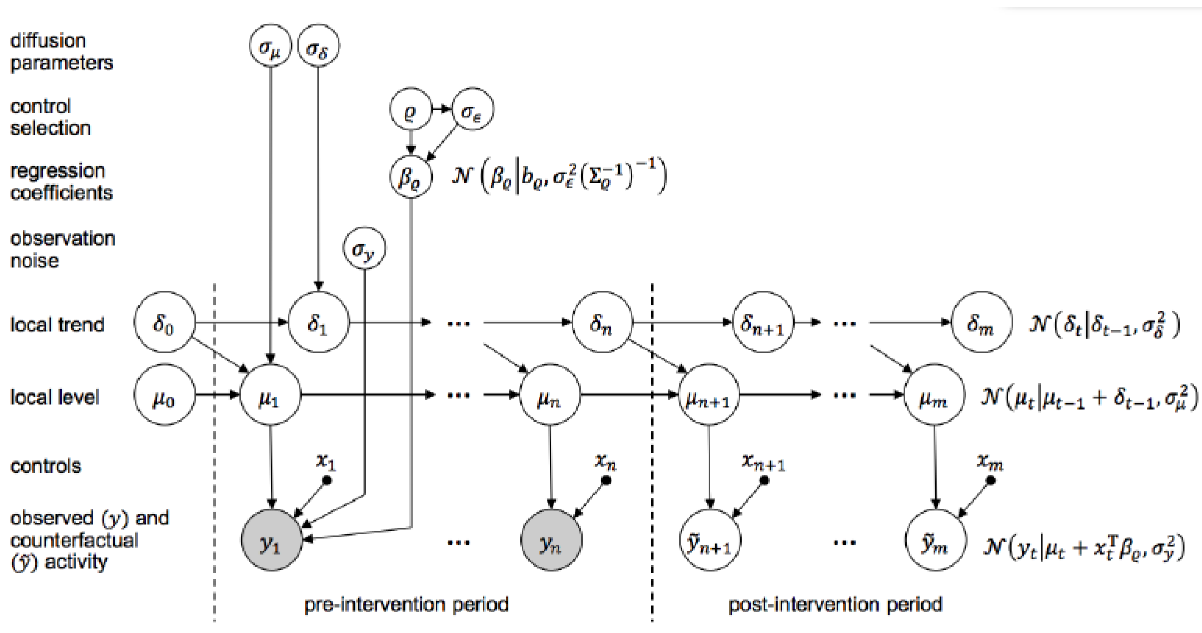
This paper focuses on inferring the causal impact of a market intervention in the time series setting and takes the approach of explicitly modeling the counterfactual of a time series (in state-space model) observed both before and after the intervention.

Contribution

In contrast to the difference in difference(DD hereafter) approach, this paper contributes to better handling the following three problems. First, DD is traditionally based on a static regression model that assumes i.i.d. data despite the fact that the design has a temporal component. Second, most DD analyses only consider two time points: before and after the intervention. Third, when DD analyses are based on time series, previous studies have imposed restrictions on the way in which a synthetic control is constructed from a set of predictor variables, which the authors argue is less desirable than the state-space approach based on the spike and slab prior.

Model

The structure of the model can be summarized in the following graph:



Now we specify the elements of the model in the graph above.

State-Space Representation

The framework of this model is based on the following state-space representation:

$$y_t = Z_t^T \alpha_t + \epsilon_t$$
$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t,$$

where $\epsilon_t \sim N(0, \sigma_t^2)$ and $\eta_t \sim N(0, Q_t)$. R_t is $d \times q$ matrix, and η_t is $q-$dimensional error where $q \leq d$.

There are three components of the state: local linear trend, seasonality, and contemporaneous covariates with static coefficients. This paper focuses on the last one and construct the counterfactual predictions by a synthetic control based on a combination of series that were not treated.

Prior

Inverse Gamma distribution is used as prior for variance parameters governing the distribution of the individual state components. To choose an appropriate set among the many potential controls, this model chooses the spike-and-slab prior over coefficients specified as follows:

Let $\rho = (\rho_1, \ldots, \rho_J)$, where $\rho_j = 1$ if $\beta_j \neq 0$ and $\rho_j = 0$ otherwise. Let β_ρ denote the non-zero elements of the vector β.

$$p(\rho, \beta, 1/\sigma_\epsilon^2) = p(\rho) p(\sigma_\epsilon^2 \mid \rho) p(\beta_\rho \mid \rho, \sigma_\epsilon^2)$$

using

$$p(\rho) = \prod_{j=1}^{J} \pi_j^{\rho_j} (1 - \pi_j)^{1-\rho_j}$$
$$\beta_\rho \mid \sigma_\epsilon^2 \sim N\left(b_\rho, \sigma_\epsilon^2 (\Sigma_\rho^{-1})^{-1}\right)$$
$$\frac{1}{\sigma_\epsilon^2} \sim g\left(\frac{\nu_\epsilon}{2}, \frac{s_\epsilon}{2}\right)$$

Inference

Posterior inference in this paper can be broken down into three pieces. First, the authors simulate draws of the model parameters and the state vector given the observed data in the training period. Second, the authors use the posterior simulations to simulate from the posterior predictive distribution over the counterfactual time series given the observed pre-intervention activity, and use the posterior simulations to simulate from the posterior predictive distribution over the counterfactual time series given the observed pre-intervention activity. Third, the authors use the posterior predictive samples to compute the posterior distribution of the pointwise impact. Lastly, the posterior simulation is done by a Gibbs sampler.

Application

The authors analyzes an advertising campaign run by one of Google's advertisers in the United States. In particular, they inferred the campaign's causal effect on the number of times a user was directed to the advertiser's website from the Google search results page.

Time series of search-related visits to the advertiser's website (including both organic and paid clicks), pointwise (daily) incremental impact of the campaign on clicks, and cumulative impact of the campaign on clicks are shown in the figure below.

As we can see from the figure, following the onset of the campaign, observations quickly began to diverge from counterfactual predictions: the actual number of clicks was consistently higher than what would have been expected in the absence of the campaign. To sum up, the authors we found a cumulative lift of 85 900 clicks (posterior expectation), or 21%, with a [12%; 30%] interval, and the analysis replicated almost perfectly the original analysis that had access to a randomised set of controls.

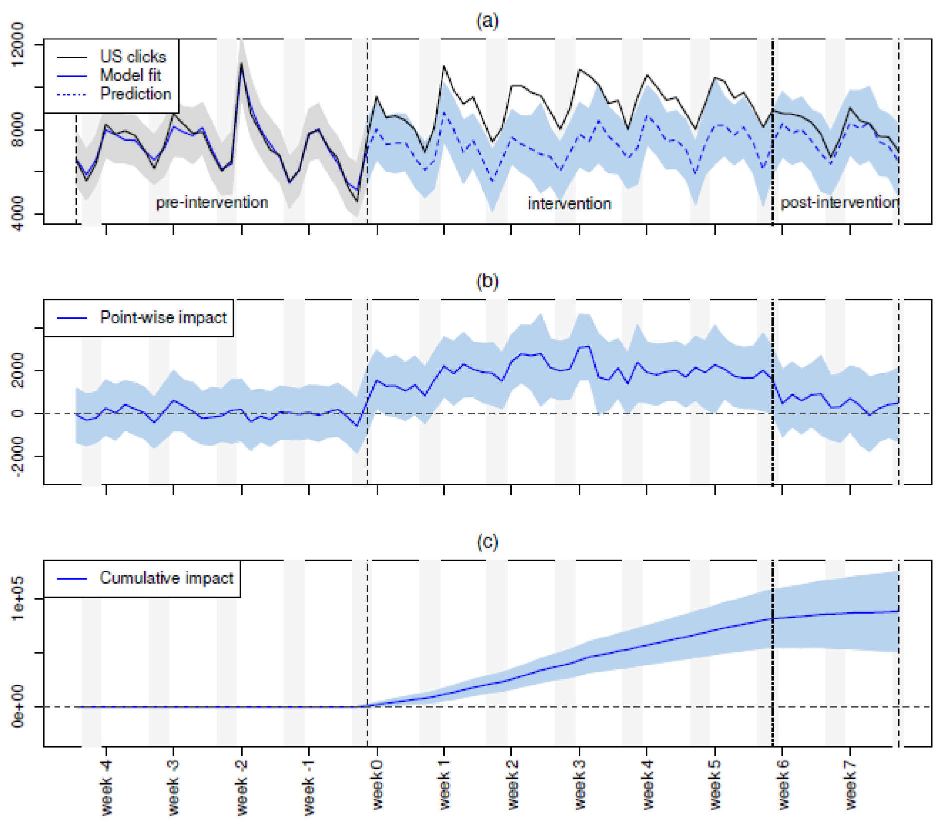


FIGURE 5. *Causal effect of online advertising on clicks in treated regions. (a) Time series of search-related visits to the advertiser's website (including both organic and paid clicks). (b) Pointwise (daily) incremental impact of the campaign on clicks. Shaded vertical bars indicate weekends. (c) Cumulative impact of the campaign on clicks.*