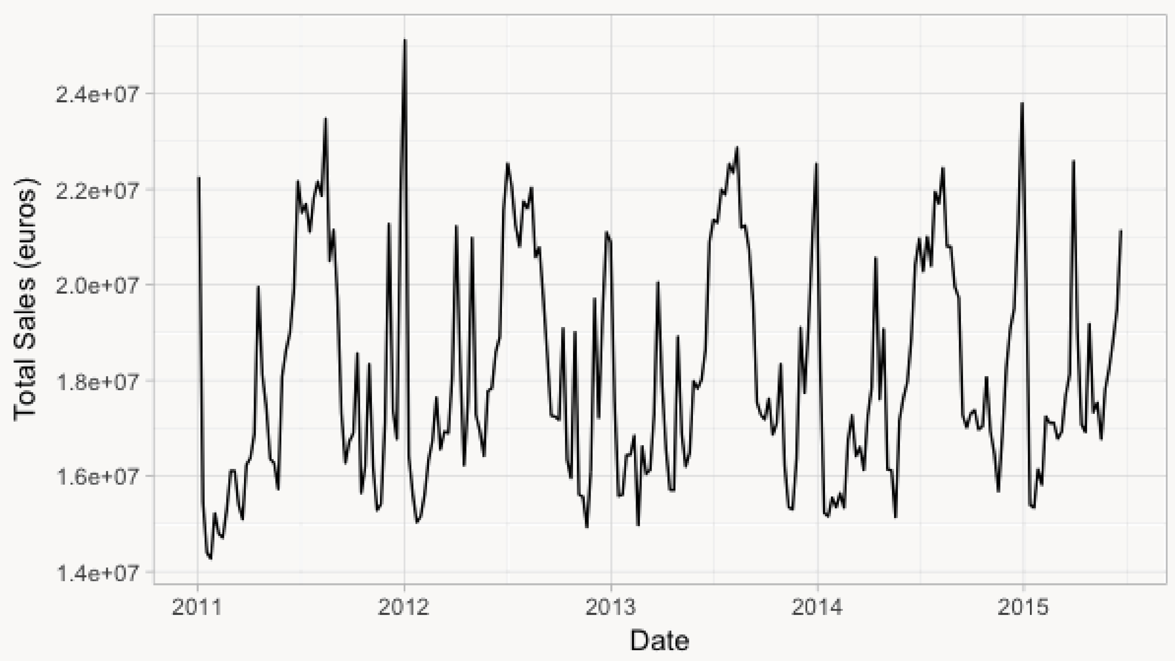
Bayesian structural time series models for advertising expenditures

Víctor Gallego, Pablo Suárez-García, Pablo Angulo and David Gómez-Ullate
ICMAT-CSIC, UCM

Goals and Intro

- Bridge the gap between dynamical systems and ML methods?
- A case study with time series data:



- Predict sales Y_t for the Krusty Burger company week after week.
- Few weekly observations. Predictor variables:
  - Economical: IPC, ICC, unemployment rate...
  - Climate: temperature, precipitation...
  - Special events: holiday, sports...
  - **Investment levels** (advertising channels): Out-of-home, Radio, TV, Online, ...
- **Objective: Help Krusty Burger choose its media plan for the next week**.

Background

- ARIMA models. Traditional tool in econometrics:

$$Y_t = \sum_{p=1}^{P} \alpha_p Y_{t-p} + \sum_{q=0}^{Q} \beta_q \epsilon_{t-q}$$

- After some baselines, results were not as good as we expected. We needed more interpretable models which can take into account experts' beliefs.
- Dynamic Linear Models (DLMs) come in handy: modular design.
  - Observation eq.:
    $Y_t = F_t\theta_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, V_t) \qquad \theta_t \in \mathbb{R}^d$
  - State eq.:
    $\theta_{t+1} = G_t\theta_t + H_t\epsilon'_t \qquad \epsilon'_t \sim \mathcal{N}(0, W_t)$
- Bayesian Structural Time Series (BSTS): slightly more general.
- Nerlove-Arrow model as a linear ODE:

$$\frac{dA}{dt} = qu(t) - \delta A(t),$$

with $A(t)$ being the goodwill and $u(t)$ the advertising spending rate.

Model construction

- We discretize the N-A model, allowing for k different channels

$$A_t = (1-\delta)A_{t-1} + \sum_{i=1}^{k} q_i u_{i(t-1)} + \epsilon_t$$

- Now, we may frame it as a DLM!

$$\theta_t = G_t\theta_{t-1} + H_t\epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, W_t),$$

where

$$\theta_t = \begin{bmatrix} A_t \\ q_1 \\ \vdots \\ q_k \end{bmatrix}, \quad G_t = \begin{bmatrix} (1-\delta) & u_{1(t-1)} & \dots & u_{k(t-1)} \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \quad H_t = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$Y_{NA,t} = \left[1, 0, \dots, 0\right] \theta_t + \epsilon'_t \qquad \epsilon'_t \sim \mathcal{N}(0, V_t)$$

Model augmentation and inference

- Via the superposition principle we can specify the model

$$Y_t = Y_{NA,t} + Y_{T,t} + T_{S,t} + Y_{R,t}$$

where
  - $Y_{NA,t}$ is the discretization of the N-A model from before.
  - $Y_{T,t}$ is a trend component (local level model).
  - $Y_{S,t}$ is the seasonal part (period 52).
  - $Y_{R,t}$ are explanatory variables.
- For the regression component

$$Y_{R,t} = X_t\beta + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

, we use a spike and slab prior that is expressed as

$$p(\beta, \gamma, \sigma^2) = p(\beta_\gamma|\gamma, \sigma^2)p(\sigma^2|\gamma)p(\gamma).$$

with $\gamma_i = 1$ iff $\beta_i \neq 0$.
A usual choice for the γ prior is a product of Bernoulli distributions:

$$\gamma \sim \Pi_i \pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}.$$

- Making the model more robust: we can replace the assumption of gaussian errors with student T errors

$$Y_t = F_t\theta_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{T}_\nu(0, \tau^2).$$

- Inference using MCMC (Gibbs sampler). Obtains draws $\rho^{(1)}, \rho^{(2)}, \dots, \rho^{(K)}$ from the posterior distribution, then the usual predictive equation

$$p(\bar{y}|y_{1:t}) = \int p(\bar{y}|\rho)p(\rho|y_{1:t})d\rho.$$

Decision Support System

$$\begin{aligned} \underset{u_{(t+1),1} \cdots u_{(t+1),k}}{\text{maximize}} \quad & E[\bar{y}_{t+1}|y_{1:t}, x_{t+1}, u_{t+1}] \\ \text{subject to} \quad & \sum_{i=1}^{k} u_{(t+1),i} \leq b_{t+1} \\ & Var[\bar{y}_{t+1}|y_{1:t}, x_{t+1}, u_{t+1}] \leq \sigma^2, \end{aligned}$$

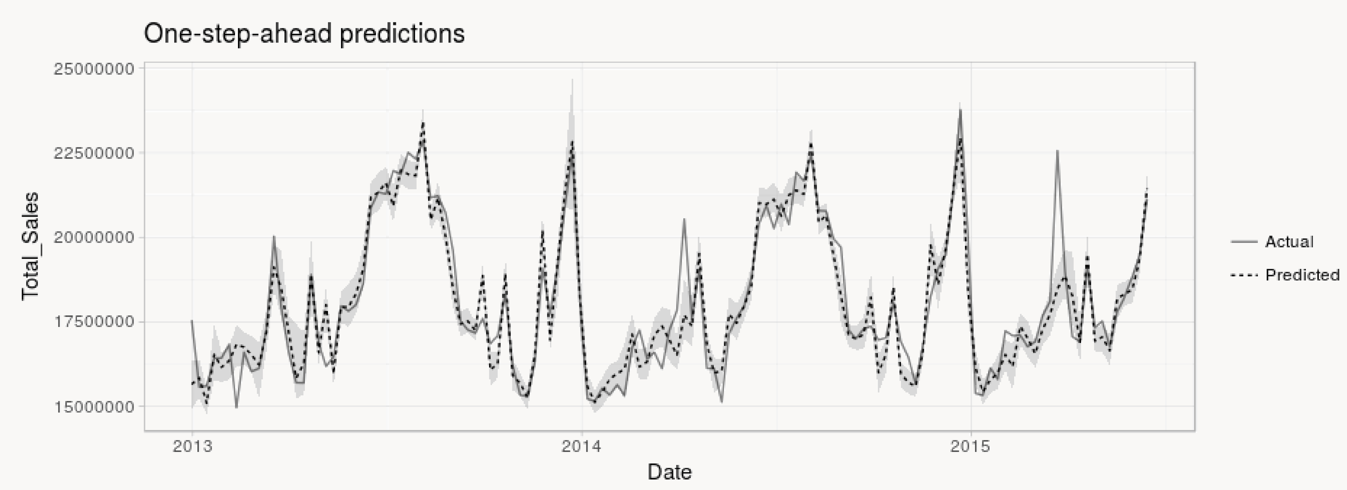
where b_t is the total budget and σ controls the risk.

- Used to evaluate an initial set of investment plans.

Experiments and Results

Our best variant achieved

$$\text{MAPE} := \frac{100\%}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{y_t} \approx 4.59\%$$



Conclusions

- DLM (BSTS) can provide a nice framework to mix dynamical systems and data-driven models.
- The firm can optimize in the investment levels, maximizing expected global sales yet minimizing a risk metric.

victor.gallego@icmat.es