AdKDD 2021

Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling

Edwin Ng, Zhishi Wang, Athena Dai
Marketing Data Science

Uber

Agenda

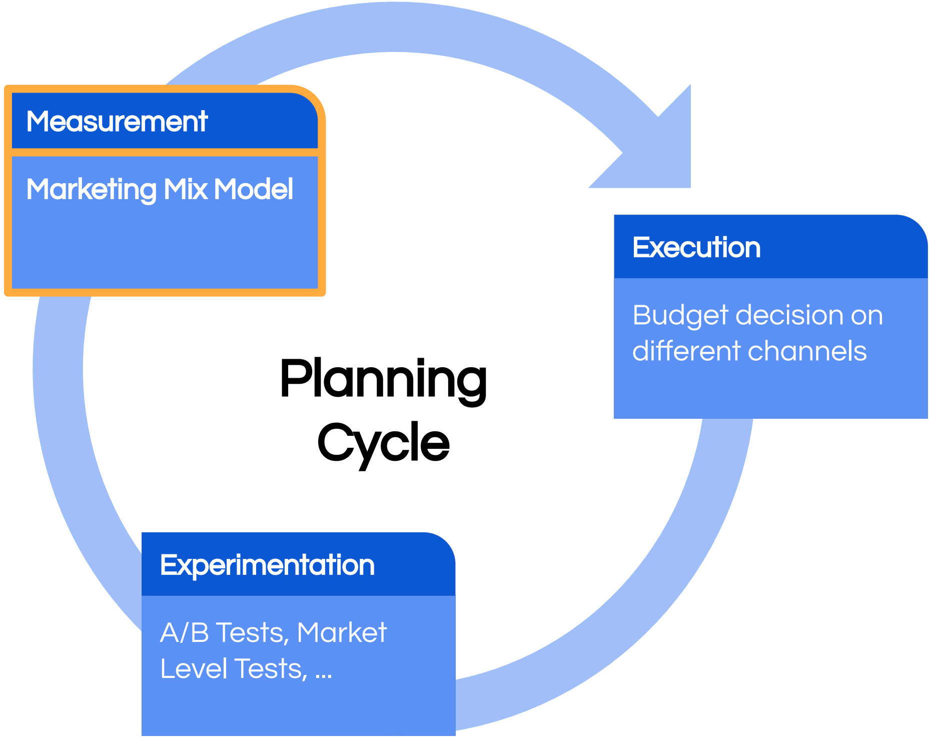
Introduction

Model Overview

Performance

Resources

The Role of Marketing Mix Model (MMM)



Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling.*

Problem Formulation

Basic Marketing Mix Model

$$\hat{y}_t = g(t) \cdot \prod_{p=1}^{P} f_{t,p}(x_{t,p}), \;\; t = 1, \cdots, T,$$

where

- $x_{t,p}$ are the ads spending variables
- $\hat{y}_t$ is the marketing response
- g is a time-series process
- f is the cost curve function
- P is the number of regressors
- T is the number of time points

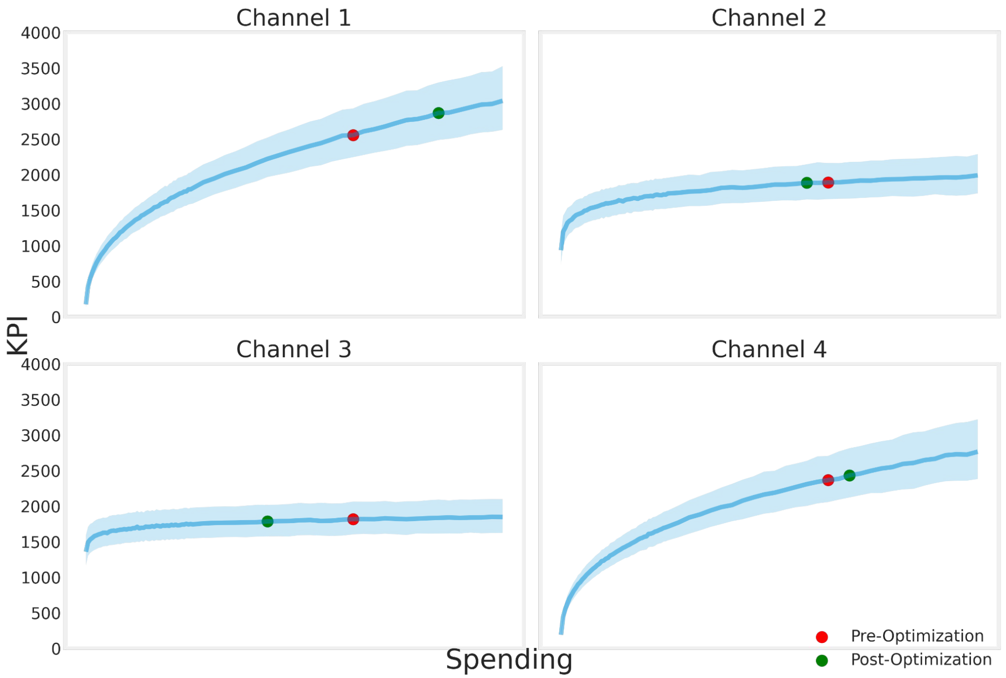


Figure 1. Data-driven cost curves for budget planning with simulated data.

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling*.

Pros and Cons

Benefits

- Low dependency on granular data
- Robust under mobile privacy policy change
- Interpretability
- Capability to forecast

Limitations and Challenges

- Endogeneity
- Multicollinearity
- High-dimensional regression
- Correlation vs. Causality

 Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling*.

Motivation

Motivation

Most of the criticisms about MMM can be addressed by experimentation. However, experimentation is expensive and has other disadvantages that MMM does not have such as the **lack of**

- ➢ forecastability
- ➢ generalizable cost curves
- ➢ insights on organic sales and interaction effects

Objectives

We propose a new solution of MMM which can fully capture all information provided by experimentation with the help of time-varying coefficients and Bayesian framework.

Our Model: Bayesian Time-Varying Coefficient (BTVC)

BTVC Predictive Equation

$$ln(\hat{y}_t) = l_t + s_t + \sum_{p=1}^{P} ln(x_{t,p})\beta_{t,p}$$

$$= l_t + s_t + r_t, \; t = 1, \cdots, T,$$

trend   seasonality   regression

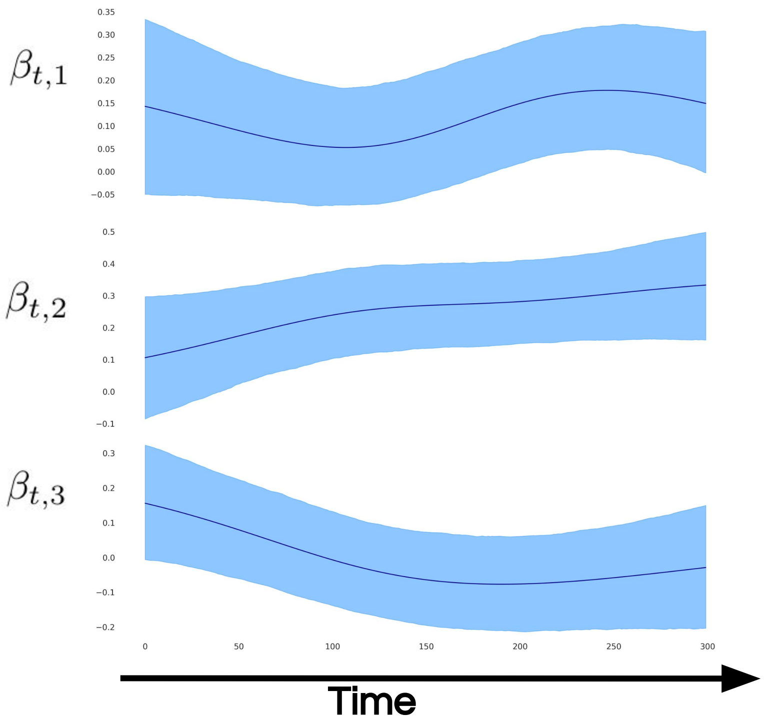


Figure 2. Elasticity is modeled by time-varying coefficients.

Extra Steps:

In BTVC, we can also re-write trend and seasonality into a similar regression framework:

- trend is treated as floating levels
- seasonality is treated as regression on fourier series terms

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling*.

Express Coefficients as Kernel Smoothed Variables

Kernel Smoothing Equations

$$\beta_{t,p} = \sum_{j} w_j(t) \cdot b_{j,p}$$

$$w_j(t) = k(t_j t) / \sum_{j=1}^{J} k(t_j, t)$$

where

- $\beta_{t,p}$ is the coefficient at time point t for p-th regressor
- $b_{j,p}$ is the latent variable (a.k.a knot)
- $k(\cdot, \cdot)$ is the kerenl function
- J is the number of knots

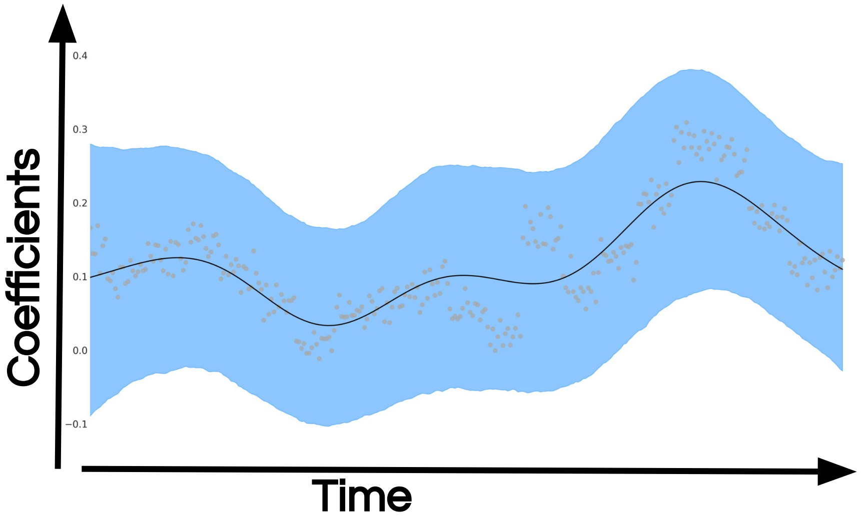


Figure 3. Coefficients derived by kernel smoothing.

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling.*

Choice of Kernel

- We use Triangular kernel for trend / seasonality regression

$$k_{\text{lev}}(t, t_l) = 1 - \frac{|t - t_l|}{t_{j+1} - t_j}$$

$$\text{if } t_j \leq t \leq t_{j+1} \text{ and } l \in \{j, j+1\}$$

- We use Gaussian kernel for regression

$$k_{\text{reg}}(t, t_j; \rho) = \exp\left(-\frac{(t - t_j)^2}{2\rho^2}\right)$$

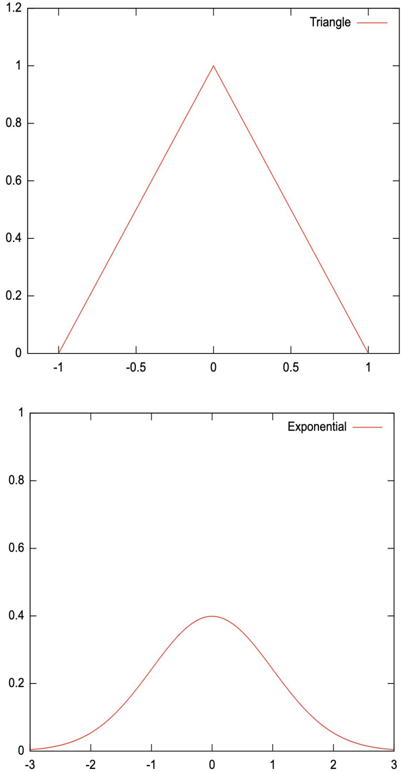


Figure 4. Triangular and Gaussian kernel

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling*.

Hyperparameters Tuning for Bias and Variance Tradeoff

Tuning Strategy

- J and ρ control the condition of overfitting vs. underfitting.
- Overfitted models tend to have low bias high variance.
- Underfitted models tend to have high bias low variance.
- We define optimal J and ρ by time-based cross-validation (a.k.a backtest)

Expanding Window Backtest

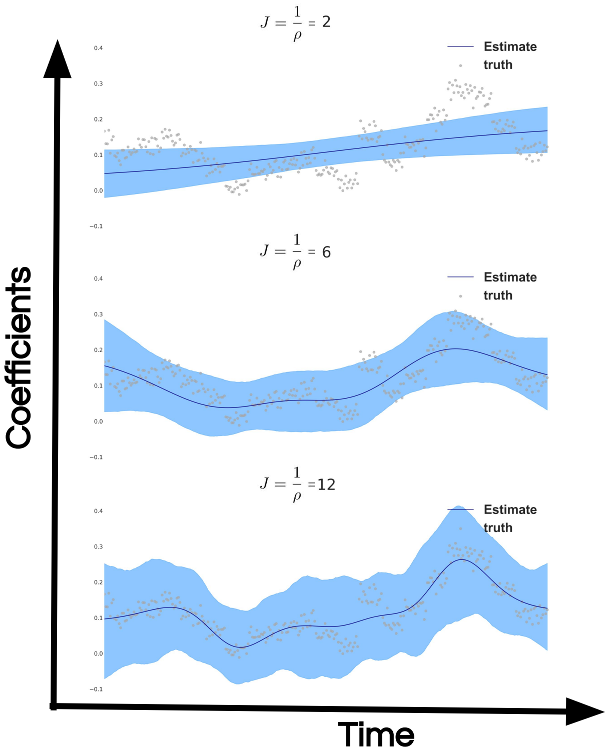


Figure 5. Estimated coefficients on different values of J and ρ fitted

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling*.

Incorporate Experimentation Results under Bayesian Framework

Informative Bayesian

Response Data Likelihood:

$$ln(y_t) \sim \texttt{Student}(\nu, ln(\hat{y}_t), \sigma^2)$$

Optional Likelihood (a.k.a "Time-Point" Priors) from Experimentation:

$$\beta_{t,p} \sim N(\mu_{t,p}, \sigma^2_{t,p})$$

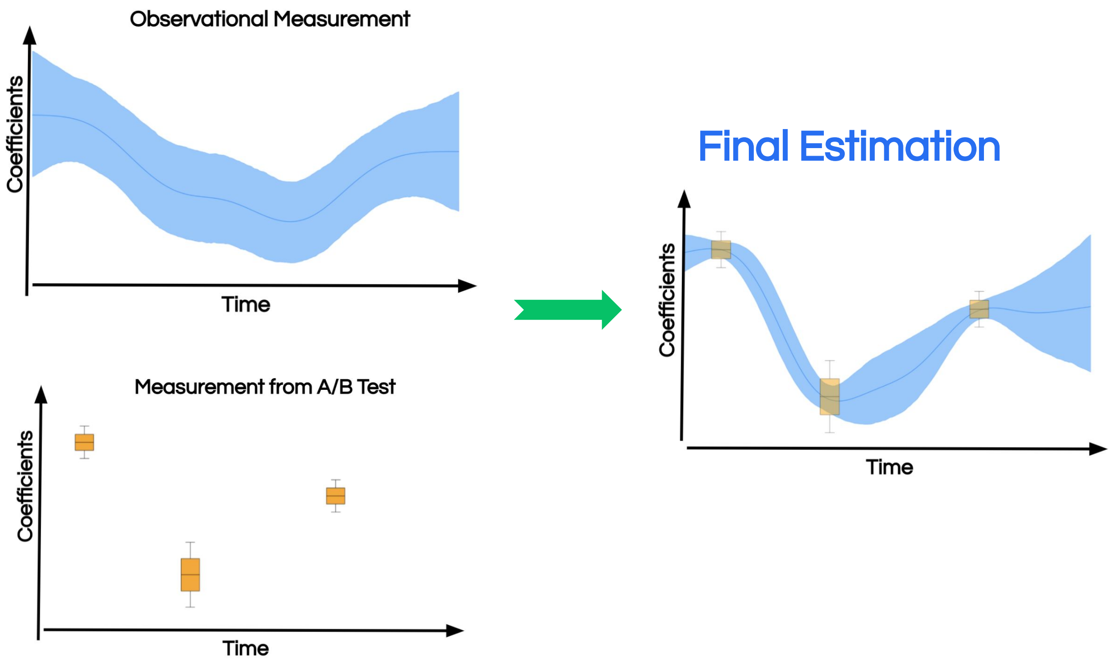


Figure 6. Bayesian Framework in BTVC

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling*.

Coefficients Curve Fitting Benchmark

Random Walk Simulations

We conduct a simulation study based on the following process:

$$y_t = \text{trend} + \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \beta_{3t}x_{3t} + \epsilon_t, \quad t = 1, \cdots T$$

where $\text{trend}, \beta_{1t}, \beta_{2t}, \beta_{3t}$ are all random walks and $T = 300$

Model	$\beta_1(t)$	$\beta_2(t)$	$\beta_3(t)$
BSTS	0.0067	0.0078	0.0080
tvReg	0.0103	0.0103	0.0096
BTVC	0.0030	0.0026	0.0029

Table 1. Average of mean squared errors based on 100 times simulations. We compare Bayesian structural time series (BSTS)[1] and time-varying coefficient for single and multiple equation regression (tvReg)[2]

Informative Priors

We further examine behavior when we provide informative "time-point" priors for some t.

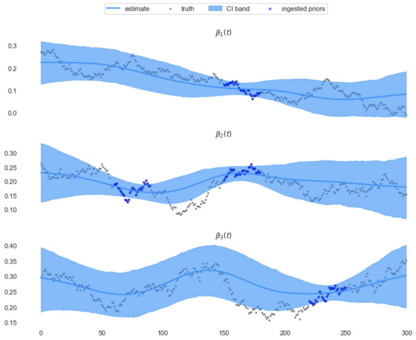


Figure 7. Example of informative priors

	SMAPE		Pinball Loss			
	w/o priors	w priors	w/o priors lower	w/o priors upper	w priors lower	w priors upper
$\beta_1(t)$	0.39	0.21	0.0009	0.0032	0.0005	0.0019
$\beta_2(t)$	1.37	1.25	0.0021	0. 0019	0.0011	0.0014
$\beta_3(t)$	0.30	0.18	0.0017	0.0028	0.0014	0.0013

Table 2. SMAPE and pinball loss of coefficient estimates for models without and with informative priors.

1. Isabel Casas and Ruben Fernandez-Casal. 2021. tvReg R package version 0.5.4
2. Steven L Scott and Hal R Varian. 2014.

Forecast Accuracy Benchmark

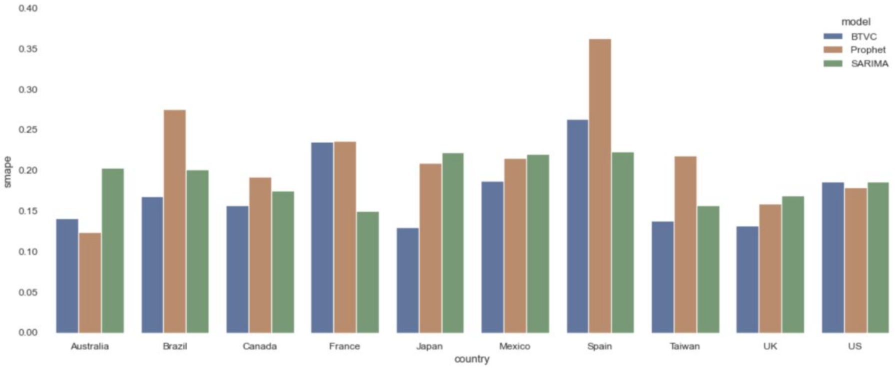


Figure 7. SMAPE comparison across different countries

Real Case Study

As a forecast model, we also compare accuracy with expanding windows backtest with Symmetric Mean Absolute Error (SMAPE) where

$$\text{SMAPE} = \sum_{t=1}^{h} \frac{|F_t - A_t|}{(|F_t| + |A_t|)/2}, h = 28$$

Model	Mean of SMAPE	Std of SMAPE
SARIMA	0.191	0.027
Prophet	0.218	0.067
BTVC	0.174	0.045

Table 3. SMAPE comparison across models including SARIMA, Prophet and BTVC with 6 splits and 28 days horizon.

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling.*

Implementation in a Scalable Way

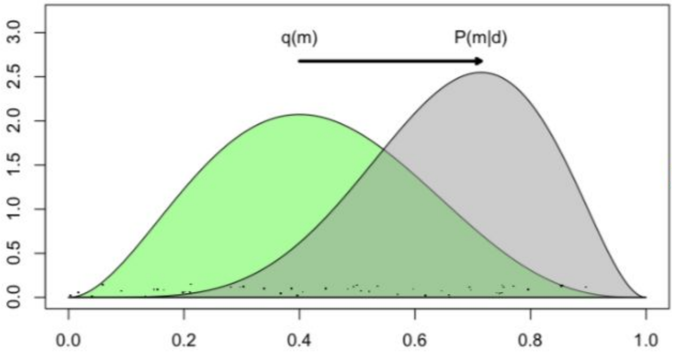
Leveraged Stochastic Variational Inference under Pyro for posteriors sampling.

PYRO

http://pyro.ai/

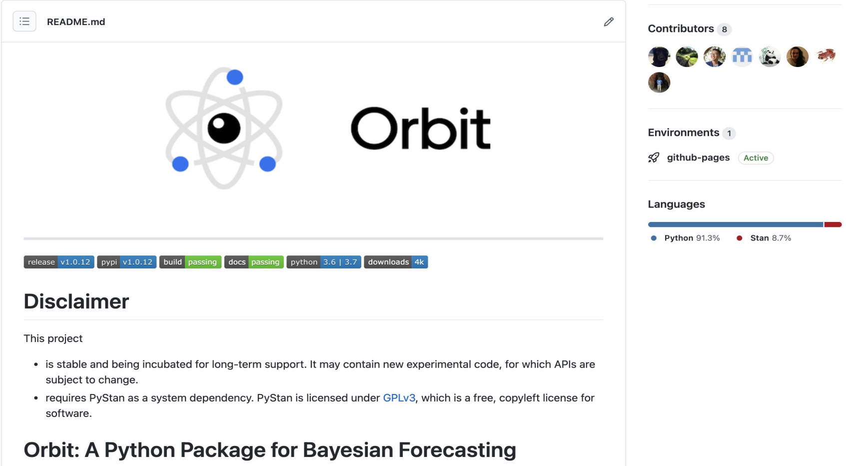
Compared to MCMC, SVI

- is computationally faster
- suitable for large datasets



Implemented under Orbit

https://github.com/uber/orbit



Uber

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling.*

Conclusion

Traditional Marketing Mix Models struggle with endogenous variables, multicollinearity and correlation vs. causality challenge. **Bayesian Time-Varying Coefficient** (**BTVC**) model solves these problems by introducing a natural way to integrate experimentation results through Bayesian framework and time-varying coefficients. Unlike typical Kalman Filter models, **BTVC** provides flexibility for user to customize likelihood functions and priors. Besides, both simulation and real case studies show BTVC has better performance in estimating regression coefficients (compared to **BSTS** and **tvReg**) and forecast accuracy (compared to **SARIMA** and **Prophet**). The methodology is open-sourced under the python package **Orbit** using stochastic variational inference in **Pyro** to perform posteriors sampling.

Ng, Wang and Dai. 2021. *Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling.*

Q & A

We are hiring!

Feel free to contact for any questions or comments

edwinng@uber.com

zhishiw@uber.com

athena.dai@uber.com

Marketing Data Science

Thank you!