A Bayesian time series model of Coca leaf production in Colombia

Sergio H. Garrido M., Emiliano Isaza

DTU MLSM Machine Learning for Smart Mobility

Motivation

The war on drugs is a costly decision process for policy makers in Colombia and other producing countries.

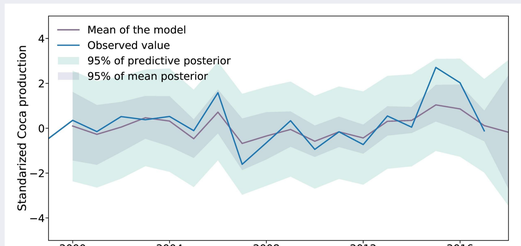
The contributions of this paper are:

- A Bayesian estimation of sq. hectares Coca production in Colombia at a desegregated level.
- The novel Pareto Smoothed Importance Sampling (PSIS) Leave Future Out (LFO) Cross Validation **(PSIS-LFO-CV)** technique is used to compare alternative models and assess thoroughly their validity.

Results

ELPD	
AR(1)	-115.36
VAR(1)	-192.00
HVAR(1) (intercept)	-178.29
HVAR(1) (intercept and coefficients)	-110.94

Expected Log-Predictive Density (ELPD) of our model HVAR(1) (i&c), and the benchmark models. Higher is better.



Example posterior plot of a single departamento.

Model

We propose a **Bayesian** model to predict Coca production in Colombia so that policy makers can **quantify uncertainty** and take it into account in their decision making process. Given the high dimensionality of the problem and the relatively low quantity of data points, we propose a **Hierarchical time series** approach. This is justified by the **"blessing of abstraction"**.

Notation:

- y_t are the sq. hectares produced in time t this is a vector of size 24
- α_c is the intercept of departamento c
- $\beta c, i$ are the effect of departamento c on time on $t-1$ departamento i on time t. $\beta_{c,i}$ is a matrix of shape 24×24
- μ_p and σ_p the hierarchical parameters

Generative process:

$$
\begin{aligned}
\mu_\alpha, \mu_{\beta,c} &\sim \mathcal{N}(0,1) \\
\sigma_\alpha, \sigma_{\beta,c} &\sim \text{HalfCauchy}(0,1) \\
\alpha_c &\sim \mathcal{N}(\mu_\alpha, \sigma_\alpha) \\
\beta_{c,i} &\sim \mathcal{N}(\mu_{\beta,c}, \sigma_{\beta,c}) \\
y_t &\sim \mathcal{N}(\alpha + y_{t-1}\beta, 1)
\end{aligned}
$$

Conclusion and Future work

Even though our model performs better compared to three other specifications using PSIS-LFO-CV. Some future modeling approaches are suggested to improve the model's performance :

- We suggest using other correlated time series as **exogenous variables** -- such as exchange rates.
- Using a **Gaussian Processes** modelling approach.
- Using a **Hidden Markov Model** to model some latent state of the time series.

References:

- Bürkner, P. C., Gabry, J., & Vehtari, A. (2019). Approximate leave-future-out cross-validation for Bayesian time series models.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis. CRC press.
- Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. Psychological review, 118(1), 110.