Bayesian Forecasting of CO2 Emissions

Mallory Wang, Noah Kochanski

4/16/2022

Section 1

Motivation

Motivation

- Climate affects everyone on Earth
- Considered by many to be one of the biggest threats to humanity
- Increasing demand in the U.S. to understand, track and predict our effect on global warming

Question: Does taking a Bayes approach to forecasting improve our estimates of environmental data?

Section 2

Data Introduction

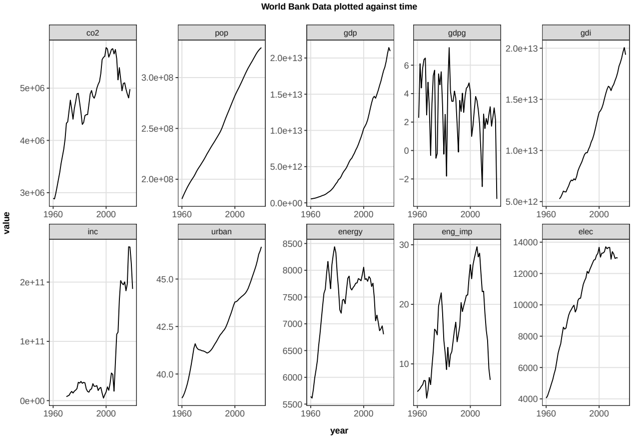
Data Introduction

- Data comes from the World Bank on the United States from 1960 to 2018
- Includes CO2 emissions, economic data (GDP, GDI, primary income), and energy consumption
- Many highly correlated variables

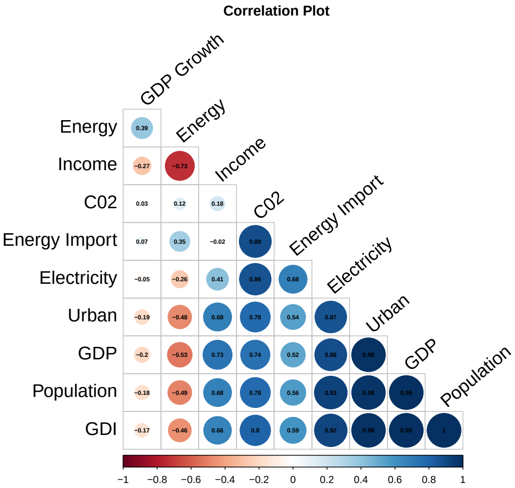
Data Introduction

- Year: annual
- C02 Emissions: measured in kt
- Population: total population
- Population in Urban Agglomeration: population in urban agglomerations of more than one million
- Economic Data:
  - Gross Domestic Product, Gross Domestic Income, Net Primary Income
- Energy Data:
  - Energy Use, Net Energy Import, Electric Power Consumption

Data Introduction



Data Introduction



Section 3

Bayesian Approach

Basic Model

Our first model, only using year to predict CO2 emissions,

$$y_i = \beta + \beta_{year} x_i + \epsilon_i$$

Our conditional:

$$Y_i | x_i, \beta, \sigma^2 \sim N(\beta^T x_i + \epsilon_i, \sigma^2)$$

Likelihood:

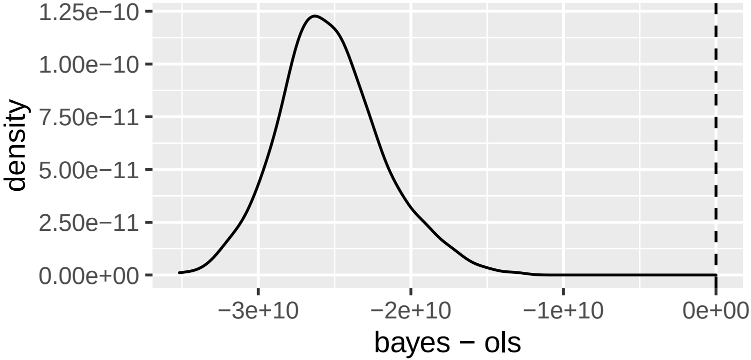
$$p(y_1, ..., y_n | x_i, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(y_i - \beta^T x_i)^2}{2\sigma^2} \right)$$

With semi-conjugate prior for β:

$$p(y | X, \beta, \sigma^2) \propto \exp - \frac{1}{2\sigma^2} SSR(\beta)$$

Basic Model Results

Using 1970 to 2005 as training data and 2006 to 2015 for testing,

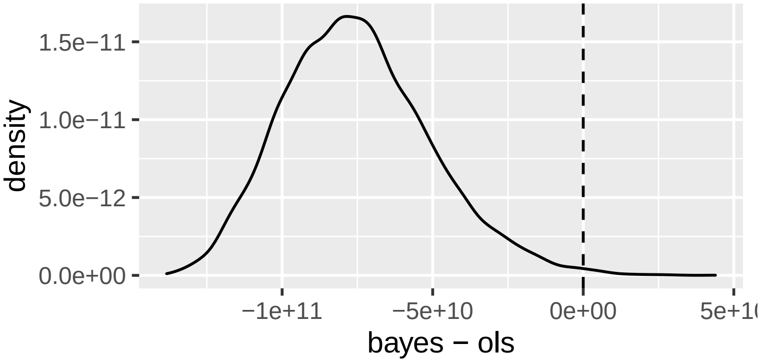


Adding Covariates

After adding lagged data and removing covariates which were highly correlated, the model takes the form,

$$Y_t = \beta' \begin{bmatrix} X_t \\ X_{t-5} \\ Y_{t-5} \end{bmatrix} + \epsilon_t$$

More complex model

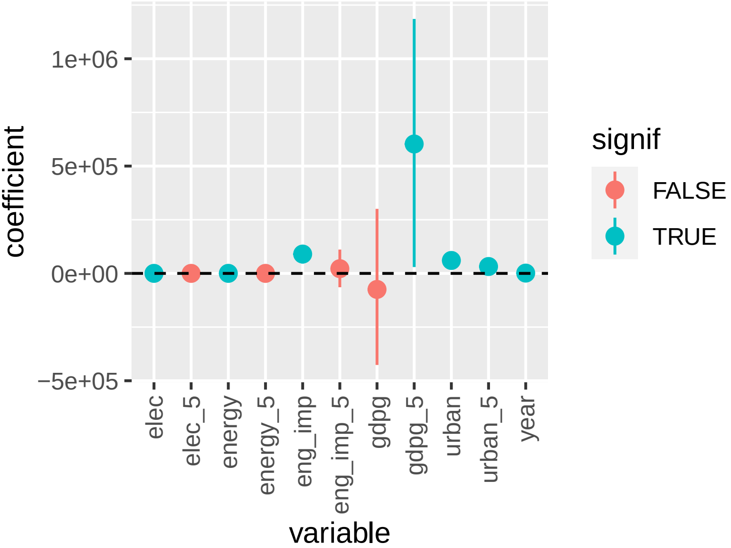


Model Selection with Gibbs Sampler

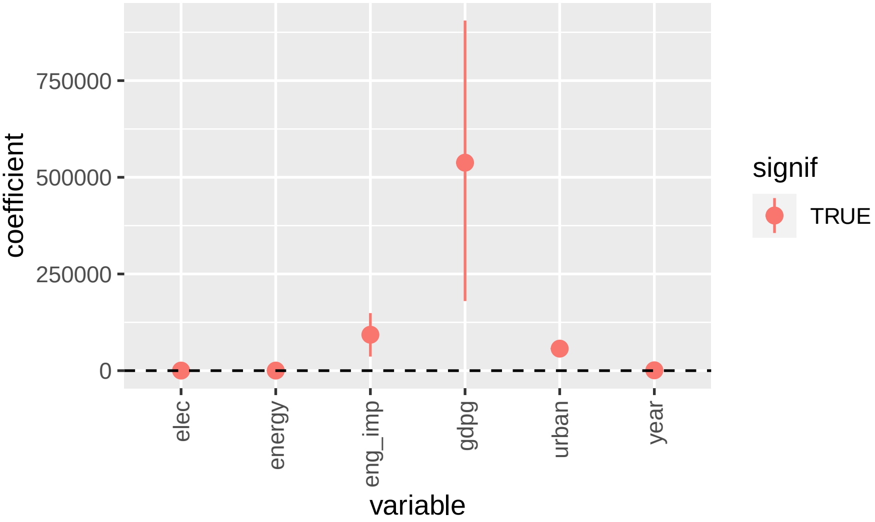
1. Create Z matrix such that $\beta_j = z_j \times b_j$ where $z_j \in \{0, 1\}$
2. Random order sample from $p(z_j | z_{-j}, y, X)$.
3. Update $z^{(s+1)}$.
4. Sample $\sigma^{2(s+1)}$ from $p(\sigma^2 | z^{(s+1)}, y, X)$.

5) Sample $\beta^{(s+1)}$ from $p(\beta | z^{(s+1)}, \sigma^{2(s+1)}, y, X)$.

Model Selection with lagged covariates



Model Selection without lagged covariates



Section 4

Conclusion and Future Work

Conclusion and Future Work

- Bayes approach improved estimates when compared to the frequentist method
- Limited by our knowledge of time series method
- Expect that ARIMA or MA models would improve the frequentist approach