POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Bayesian Learning and Monte Carlo Simulation final project

Willaey Kalisto, Leonard Nader, Barbui Andrea, Croci Federico

Student IDs: 10987637, 11030544, 10728562, 10731153
Advisor: Bassetti federico
Co-advisor: Viselli Andrea
Academic Year: 2023-2024

Contents

1 Data Analysis

1.1 Introduction

The project aims to analyze the relationship between CO2 emissions and various socio-economic indicators across different countries and years. In particular, we investigate how variables such as energy use, GDP per capita, population size, urbanization rate, internet user's number, and the percentage of low-carbon energy sources influence CO2 emissions. The goal is to build a regression model that accurately explains CO2 emissions based on these covariates while exploring transformations for enhanced model performance.

1.2 Dataset description

The dataset "CO2.csv" contains the following indicators relevant to CO2 emissions:

1. **Country**: Name of the country.
2. **y**: Year of the data entry.
3. **EnergyUse**: Energy use measured in kilograms of oil equivalent per capita.
4. **GDP**: Gross Domestic Product per capita, adjusted for purchasing power parity (PPP) in constant 2017 international dollars.
5. **pop**: Population estimates, an integer value.
6. **co2**: Annual CO2 emissions per capita (target variable).
7. **lowcarbon**: Percentage of energy derived from low-carbon sources, excluding traditional biofuels.
8. **urb**: Urban population percentage.
9. **internet**: Number of internet users based on World Bank and United Nations data, an integer value.

The dataset aims to explore how these variables interact and contribute to CO2 emissions trends across different contexts.

1.3 Data exploration

The dataset "CO2.csv" comprises socio-economic indicators from various countries across different years, focusing on factors influencing CO2 emissions. Initial exploration involves examining summary statistics, correlations and scatter plots to understand relationships among the variables.
When first looking at the dataset, it is noticeable that it contains variables with different units of measurement and very different orders of magnitude.
Furthermore, in Table 1, some of the main information are reported, such as the minimum and maximum values, median, mean, and the values of the first and third quartile for each of the 8 numerical variables.

Variable	y	EnergyUse	GDP	pop
Min.	2005	1907	2280	294976
1st Qu.	2006	17801	11588	6642848
Median	2007	31659	23021	20733758
Mean	2007	37438	27896	167472770
3rd Qu.	2008	47448	42820	62614317
Max.	2009	196676	114889	6873077808
Variable	co2percap	Lowcarbon_energy	internet	urb
Min.	0,271	0,001	47416	18,24
1st Qu.	3,946	3,046	2193462	55,08
Median	6,758	12,259	5253921	68,45
Mean	7,576	15,331	36371561	67,18
3rd Qu.	10,055	18,710	15170024	80,45
Max.	34,571	81,746	1742929484	100

Table 1 Summary of the dataset

Summary statistics highlight different features among the variables: while **EnergyUse** and **GDP** per capita exhibit skewed distributions with notable outliers (due to the presence of data concerning the world), **pop** and **internet** usage show more evenly distributed data.
The pairs plot is useful to detect patterns and relationships between variables and to identify potential correlations, additionally it contains a kernel density estimation of the distribution along the diagonal. For this purpose, the plots have been reported in Figure 1.

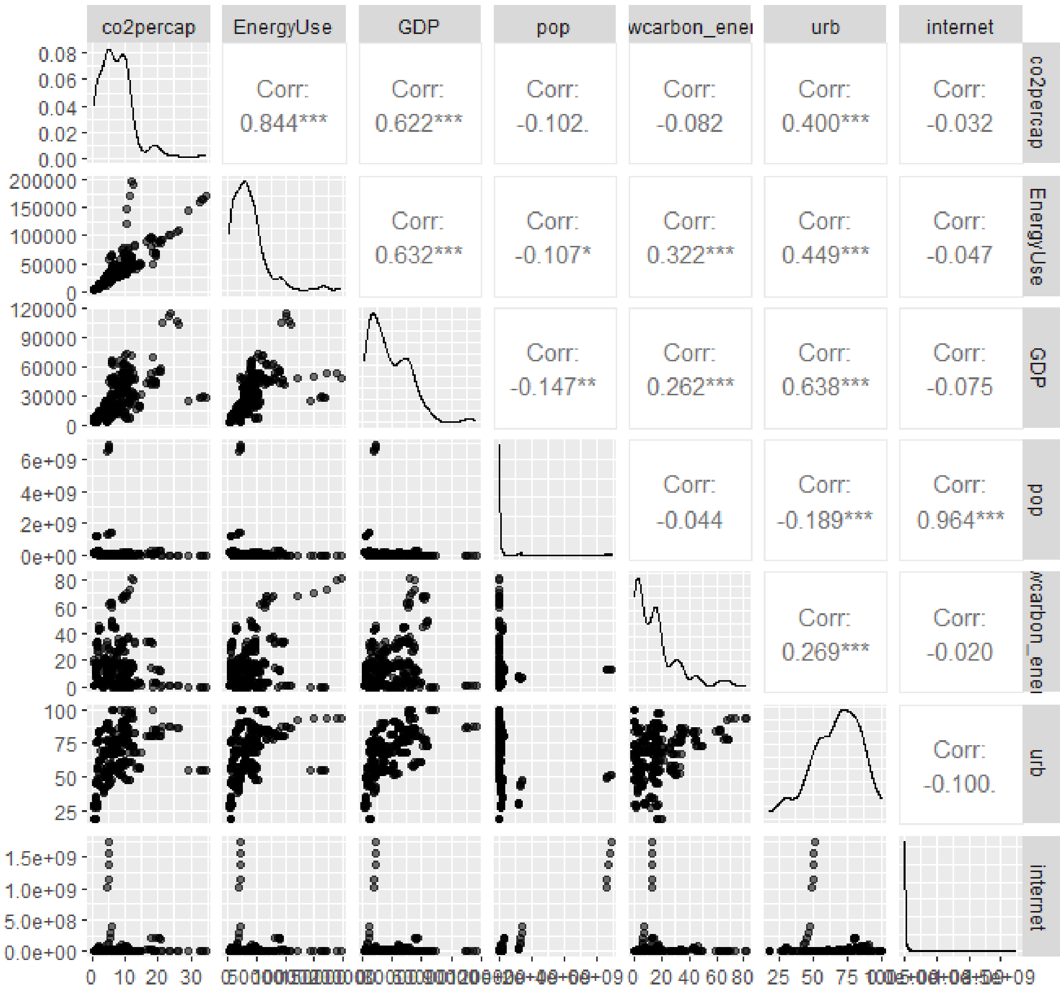

Figure 1 Paris plot between the variables

From Figure 1, it can be seen that there is an almost linear relationship between **co2percap** and **EnergyUse**. It is also interesting to note that as **urb** and **GDP** increase, **co2percap** tends to rise, hence **GDP** positively correlates with **co2percap**, indicating higher emissions in wealthier nations. Meanwhile, for the other variables, there is no strong relationship with the target variable.
The scatter plots of **Internet** and **pop** suggest the presence of outliers.
In Figure 2, the full correlation matrix is reported.

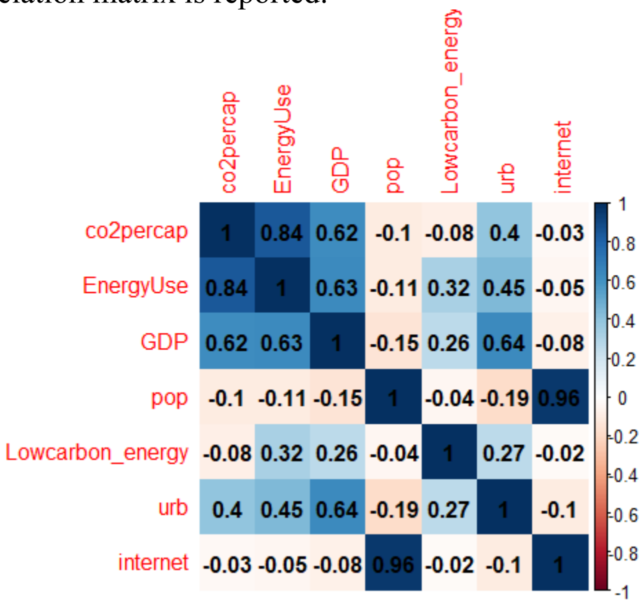

Figure 2 Correlation matrix of the covariates

The correlation matrix reveals significant associations: **GDP** per capita shows a strong positive correlation with **EnergyUse** (0.63), with our target variable **co2percap** (0.62) and with **urb** (0.64), indicating higher economic activity in urbanized regions. While **co2percap** exhibit a strong positive correlation with **EnergyUse** (0.84) and with **GDP** (0.62), moreover it shows a moderate positive correlation with **urb** (0.4).
Before proceeding, we took the logarithm of the variables **pop**, **internet**, **EnergyUse**, and **GDP** because they have much higher values compared to the other variables.
Additionally, categorical variables were treated by creating dummy variables. Assuming we have the variable x which involves K categories $C_1, \dots, C_k$, we construct K-1 indicator variables as follows,

$$I_j(x) = \begin{cases} 1 & \text{if x } = \mathrm{C}_j \\ 0 & otherwise \end{cases} \qquad j = 1, \dots, K-1$$

where K-1 variables are considered to avoid the issue of multicollinearity.

2 Model Specifications

2.1 Bayesian Linear Regression Model

In a linear regression problem we assume that the relationship between the target variable **co2percap** and the other variables is linear with Gaussian noise, while the likelihood is the following:

$$\mathrm{y} = (y_1, \dots, y_n) \sim \mathcal{N}_n(\beta_0 1 + \beta X, \sigma^2 I_n) \quad \beta = (\beta_1, \dots, \beta_k)$$

Where **1** is a vector of ones of length n and I_n is the identity matrix.

Two different priors have been considered in the analysis: the Zellner's informative g-prior and the Jeffreys-Zellner-Siow (JZS) prior.
The Zellner's informative g-prior is defined as follows:

$$\beta \mid \sigma^2 \sim \mathcal{N}_k(0, g\sigma^2(X'X)^{-1})$$
$$(\beta_0, \sigma^2) \sim \pi(\beta_0, \sigma^2) = \sigma^{-2}$$

As can be seen, it assigns a normal distribution prior to the regression coefficients, moreover the BAS package centers the prior for β at zero. The variance of the prior distribution for β is proportional to $g\sigma^2(X'X)^{-1}$, this implies that the variance decreases with the increasing number of predictors.
The prior for σ^2, $\pi(\sigma^2) = \sigma^{-2}$, is a non-informative improper prior since it does not favor any particular value of σ^2 a priori and it does not integrate to a finite value over the entire parameter space.
g is the Hyper-parameter and can be interpreted as a measure of the amount of information available in the prior relative to the sample, meaning that a smaller value of g implies a stronger prior belief, thus it makes the prior more informative. Since the choice of g is a critical aspect of the model, several values for it have been tested, as can be seen in the analyses in the third paragraph.

On the other hand, the Jeffreys-Zellner-Siow (JZS) prior provides an objective approach putting a prior on g, thus having a hierarchical prior.
The JZS prior is based on a Cauchy prior on the coefficients $\beta_j|\sigma^2$, which can be represented as a mixture of g priors with an $\text{Inverse} - \text{Gamma}(1/2, n/2)$ prior on g, that is

$$p(\beta_j|\sigma^2) = \int N(\beta_j|0, g \cdot \sigma^2 \cdot (X'X)^{-1})p(g)dg,$$
$$where \quad p(g) = \frac{(n/2)^{1/2}}{\Gamma(1/2)} g^{-3/2} \exp(-n/(2g)),$$
$$and \quad p(\sigma^2) \propto \sigma^{-2}$$

However it is not possible to compute the posterior in closed form. The bas.lm routine employs MCMC algorithms to approximate the posterior.

2.2 Model selection

As a first approach, we applied the Bayesian Information Criterion (BIC). Its primary aim is to select the model that best explains the data with the smallest number of parameters.
The BIC criterion is defined as:

$$BIC = -2\ln[p(\, y \mid \hat{\beta}, \widehat{\sigma^2}, M\,)] + (p+1)\ln(n)$$

where n is the number of observations, p is the number of predictors (without the intercept), and $\mathcal{L}(\, y \mid \beta, \sigma^2, M\,)$ is the data likelihood given the coefficients β and the variance σ^2 for model M. Denote with $(\hat{\beta}, \widehat{\sigma^2})$ the maximum likelihood estimate (MLE), then the BIC is computed using the likelihood evaluated at its maximum.
This approach assign an equal prior probability to each model M and it uses a non-informative prior on the coefficients:

$$\pi(\,\beta \mid \sigma^2\,) \propto 1,$$
$$\pi(\sigma^2) \propto \frac{1}{\sigma^2}$$

By default, the model selection with the BIC follows the Backward Selection method. This process begins with a regression model that includes all n available features. The BIC value for this full model is calculated and then compared with the BIC values of models with one fewer feature ($n - 1$ features). If a simpler model (with $n - 1$ features) has a lower BIC value, it becomes the new model. This procedure is repeated, continually removing one feature at a time and recalculating the BIC, until no further reduction in BIC is possible. The outcome of this method is a model that provides the best balance between goodness of fit and model simplicity, thus preventing overfitting.

However, the best choice is not always the best model because the unknown law which generates the data varies and there may also be sudden changes. In fact, different models may be informative in ways that single models do not and achieve a better fit.
Therefore, as a second approach, Bayesian model averaging (BMA) was considered.
The BMA is based on the following posterior distribution of the parameters:

$$p(\theta \mid y_i) = \sum_{j=1}^{2^p} p(\theta \mid y, M_j)\, p(M_j \mid y)$$

where $p(M_j \mid \boldsymbol{y})$ is the posterior probability of the model M_j and p is the number of covariates.
We can therefore compute the predictions as a weighted average of the individual model predictions, with weights given by the posterior probabilities of each model M_j, that is

$$\hat{y_i} = \sum_{j=1}^{2^p} \widehat{y_i^{(j)}}\, p(M_j \mid \boldsymbol{y})$$

2.3 Variable transformations

The purpose of this section is to define a criterion to adopt for evaluating variable transformations to enhance model performance. First of all, the types of variable transformations adopted are: multiplication between variables, squaring and cubing.
The whole process is shown in the algorithm in Figure 3.

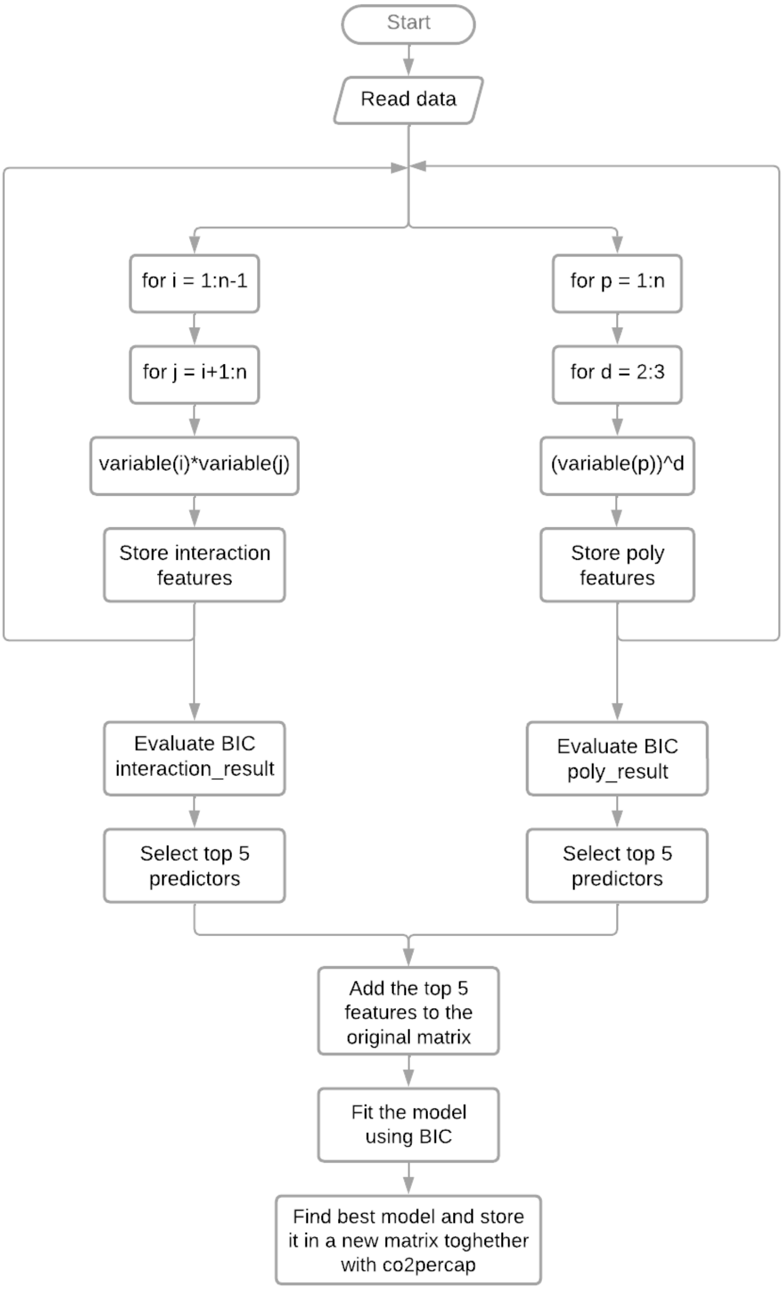

Figure 3 Variable transformations algorithm

In the flow chart, n indicates the number of variables of interest (in our case 6, '**EnergyUse**', '**GDP**', '**pop**', '**urb**', '**internet**', '**Lowcarbon_energy**').
The outcome of this process will then be used in the following paragraphs to fit a regression model to explain CO2 emissions with the other variables and to make predictions.

3 Posterior Analysis

3.1 g-prior

Starting with the g-prior approach, various values of g were tested, and, in particular, no significant change was observed in the posterior distribution of coefficients (β) and in the 95% credible intervals for the coefficients (β). Additionally, both models were evaluated: the one without variable transformation and the one with it. The values for the coefficients obtained are reported below.

no variable transf.	post mean	post SD	post p(B != 0)	variable transf.	post mean	post SD	post p(B != 0)
Intercept	7,575742	0,147341	1	Intercept	7,575742	0,06715	1
EnergyUse	5,97507	0,305524	1	EnergyUse	-49,023327	1,865589	1
GDP	0,328556	0,466514	1	GDP	-6,739706	2,710999	1
pop	0,176288	0,273224	1	Lowcarbon_energy	0,692225	0,060058	1
Lowcarbon_energy	-0,096977	0,00986	1	urb	0,521796	0,069781	1
internet	-0,29139	0,252421	1	EnergyUse_x_GDP	1,94418	0,394276	1
urb	-0,033567	0,012285	1	EnergyUse_x_urb	-0,054299	0,006971	1
y_2006	0,011058	0,46716	1	EnergyUse_x_Lowcarbon_energy	-0,076512	0,005542	1
y_2007	-0,002403	0,469623	1	EnergyUse_poly_2	2,047199	0,194327	1
y_2008	-0,041912	0,473063	1	GDP_poly_2	-0,647019	0,193394	1
y_2009	0,006694	0,477229	1				

Table 2 g-prior posterior statistics of the coefficients

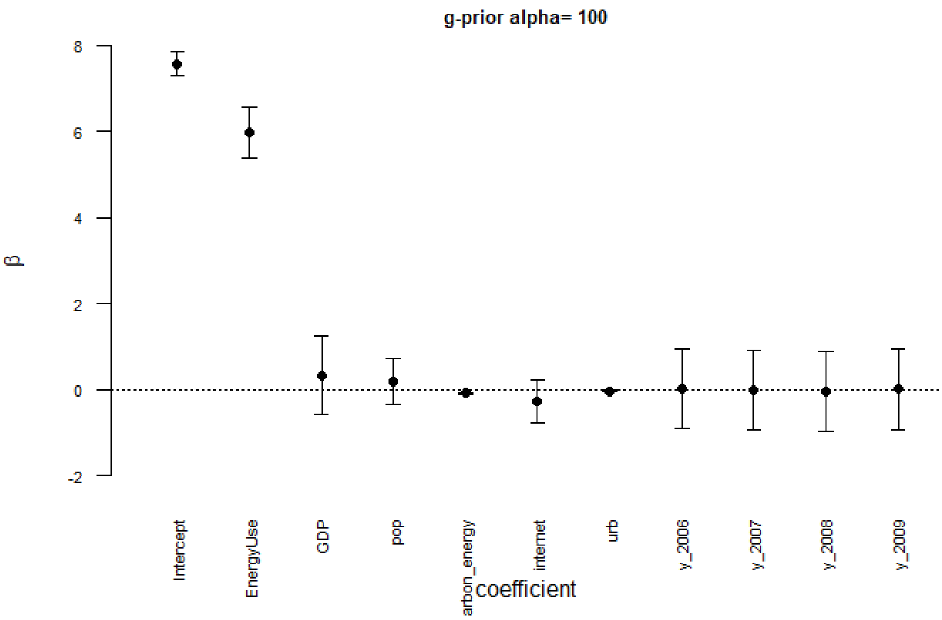

Figure 4 95% credible intervals for the coefficients (without variable transformations)

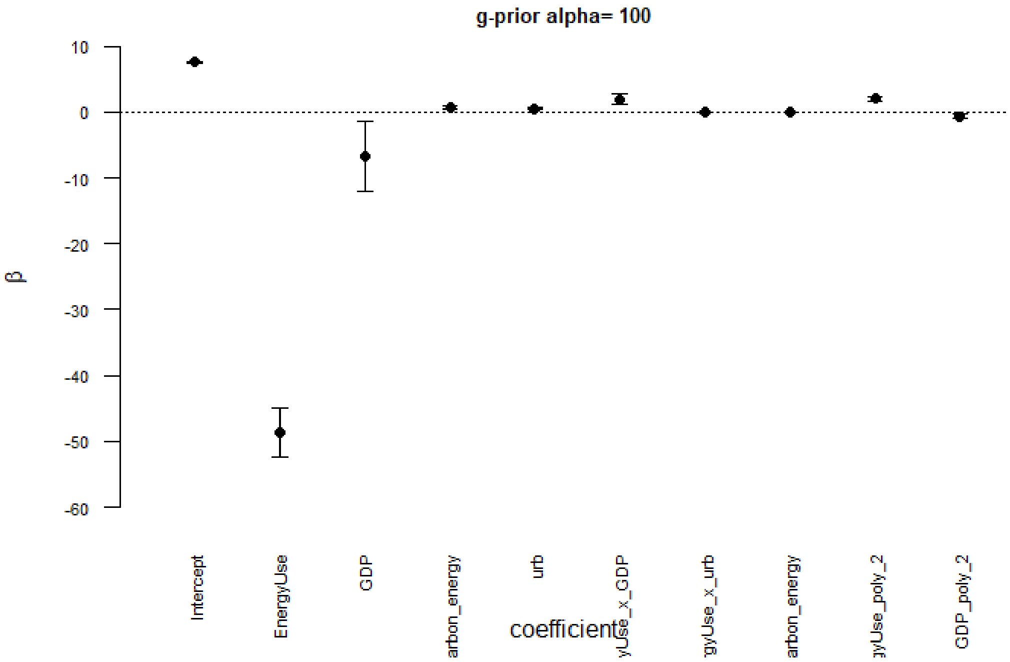

Figure 5 95% credible intervals for the coefficients (with variable transformations)

It is interesting to compare the posterior distribution of the coefficient for **GDP** in both cases.

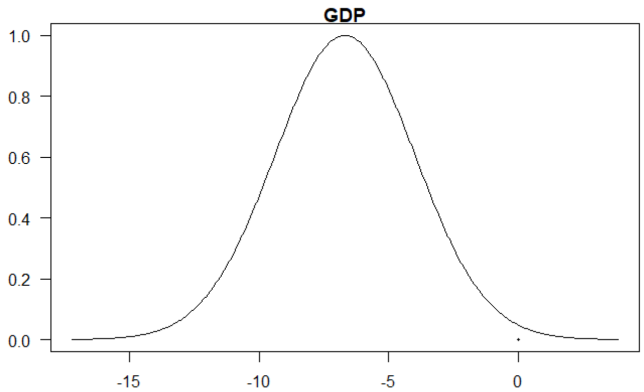

Figure 7 Posterior distribution of GDP's coefficient with variable transformation

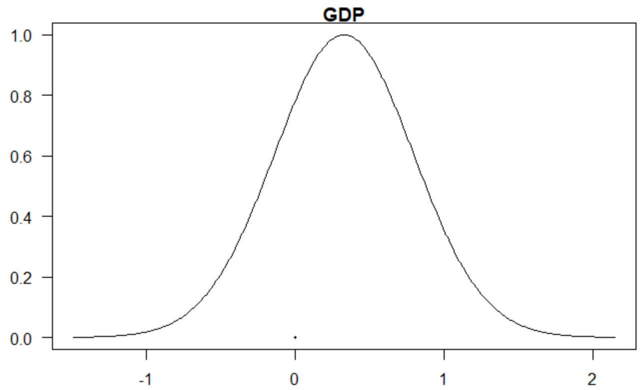

Figure 6 Posterior distribution of GDP's coefficient without variable transformation

From Figures 6 and 7, it can be noted that the covariate **GDP** has a considerable impact on the response in the case of variable transformation, whereas in the case without variable transformation, it has little impact because zero is very close to the center of the posterior distribution of its β.

3.2 JZS prior

We continued analyzing the regression model with JZS prior for both models (with and without variable transformations). The values for the coefficients obtained are reported below.

no variable transf.	post mean	post SD	post p(B != 0)	variable transf.	post mean	post SD	post p(B != 0)
Intercept	7,57574	0,03313	1	Intercept	7,575742	0,06715	1
EnergyUse	4,68693	0,74349	1	EnergyUse	-49,087281	1,866805	1
GDP	2,5382	0,85713	1	GDP	-6,748498	2,712767	1
pop	10,56479	2,70307	1	Lowcarbon_energy	0,693128	0,060097	1
Lowcarbon_energy	-0,03469	0,02586	1	urb	0,522477	0,069826	1
internet	-0,21893	0,20605	1	EnergyUse_x_GDP	1,946716	0,394533	1
urb	-0,08035	0,07376	1	EnergyUse_x_urb	-0,05437	0,006975	1
y_2006	-0,10807	-0,1174	1	EnergyUse_x_Lowcarbon_energy	-0,076611	0,005546	1
y_2007	-0,28236	-0,14523	1	EnergyUse_poly_2	2,04987	0,194453	1
y_2008	-0,46348	0,17255	1	GDP_poly_2	-0,647863	0,19352	1
y_2009	-0,59879	0,19581	1				

Table 3 JZS-prior posterior statistics of the coefficients

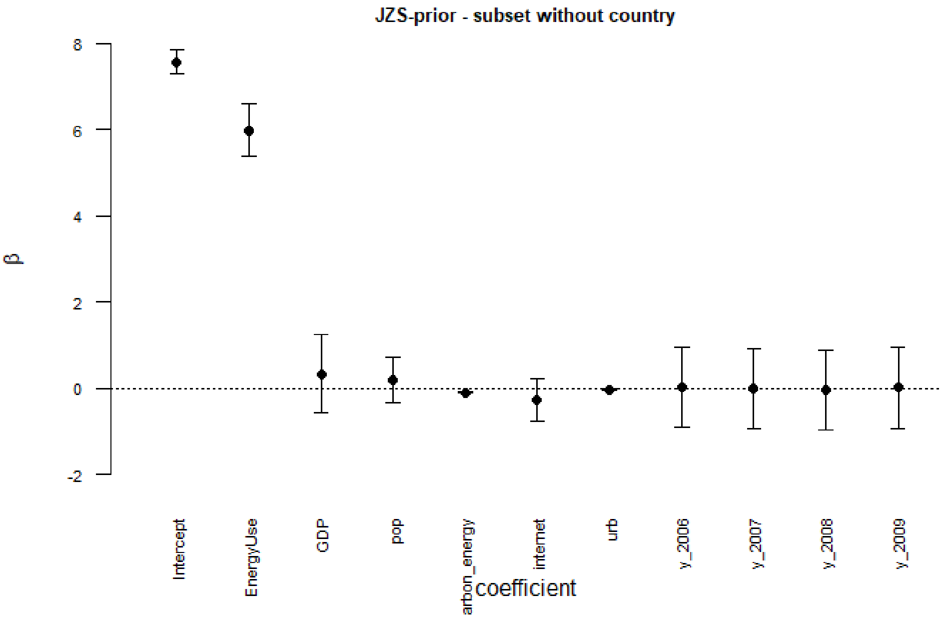

Figure 8 95% credible intervals for the coefficients (without variable transformations)

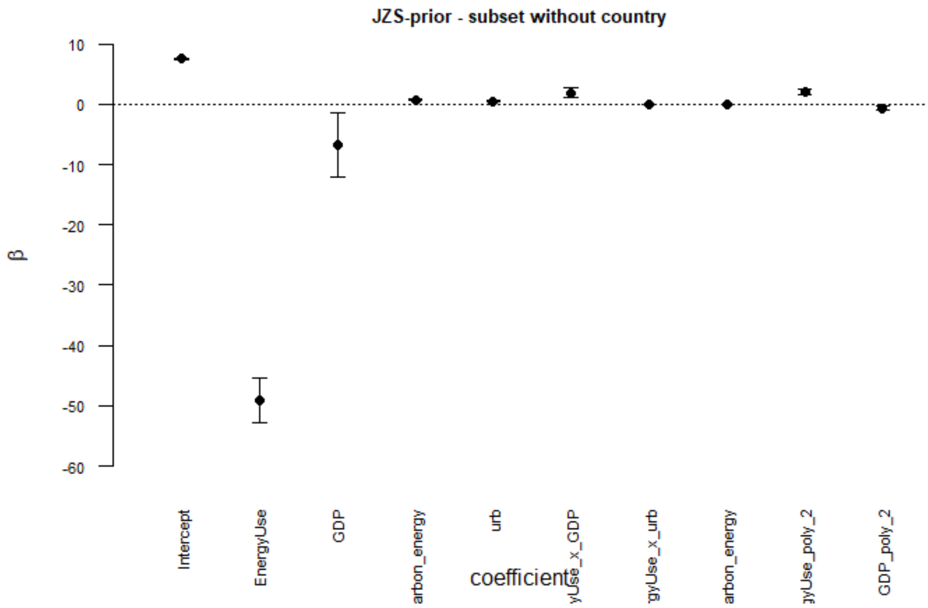

Figure 9 95% credible intervals for the coefficients (with variable transformations)

The posterior distribution of the coefficient for **GDP** in both cases is quite similar to those obtained with the g-prior, where without variable transformation it appears that **GDP** has little impact on the response.

3.3 Bayesian Information Criterion (BIC)

Applying the BIC criterion in the first case, without considering variable transformations, the outcome obtained is shown in Figure 10.

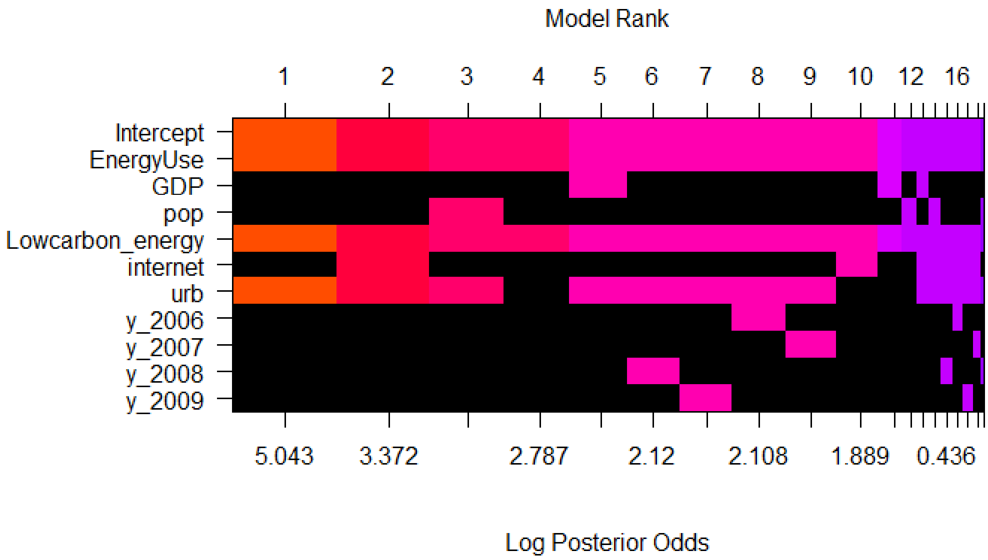

Figure 10 Model ranking heatmap

no variable transf.	P(B != 0 | Y)	model 1
Intercept	1	1
EnergyUse	1	1
GDP	0	0
pop	0	0
Lowcarbon_energy	1	1
internet	0	0
urb	1	1
y_2006	0	0
y_2007	0	0
y_2008	0	0
y_2009	0	0
BF	NA	1
PostProbs	NA	0,516
R2	NA	0,755
dim	NA	4
logmarg	NA	-1.457

Table 4 Best model summary

As can be seen, in the model without variable transformations, only 3 covariates were retained in the best model: **EnergyUse**, **Lowcarbon_energy**, and **urb**. From this analysis, it was also highlighted that the probability of the GDP to belongs to the models is quite low (0.059), so it is coherent with the previous analyses.

While the outcome of the second case, considering variable transformations, is the following.

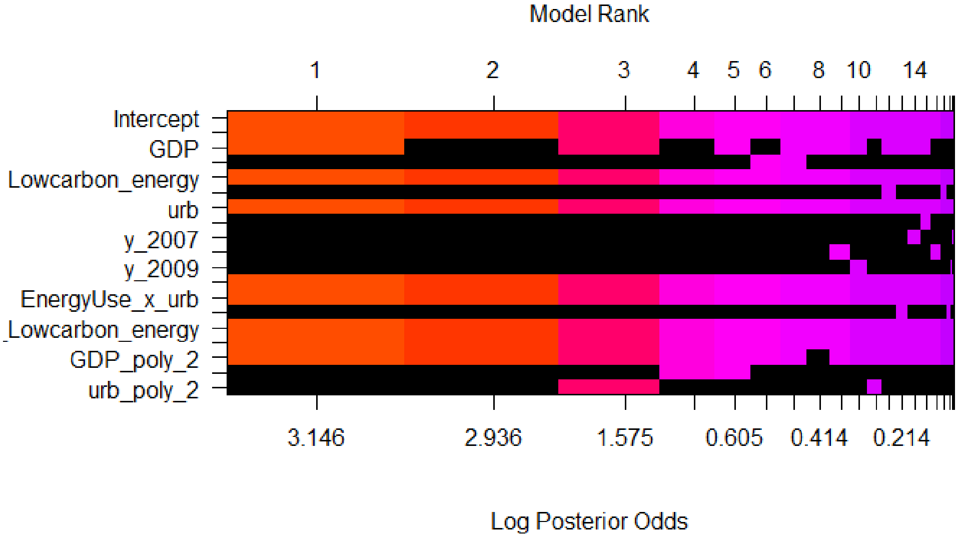

Figure 11 Model ranking heatmap

variable transf.	P(B != 0 | Y)	model 1
Intercept	1	1
EnergyUse	1	1
GDP	1	1
pop	0	0
Lowcarbon_energy	1	1
internet	0	0
urb	1	1
y_2006	0	0
y_2007	0	0
y_2008	0	0
y_2009	0	0
EnergyUse_x_GDP	1	1
EnergyUse_x_urb	1	1
EnergyUse_x_internet	0	0
EnergyUse_x_Lowcarbon_energy	1	1
EnergyUse_poly_2	1	1
GDP_poly_2	1	1
pop_poly_3	0	0
urb_poly_2	0	0
BF	NA	1
PostProbs	NA	0,274
R2	NA	0,949
dim	NA	10
logmarg	NA	-1187,879

Table 5 Best model summary

As it can be seen, the covariates that contribute to the best model are: **EnergyUse**, **GDP**, **Lowcarbon_energy**, **urb**, **EnergyUse_x_GDP**, **EnergyUse_x_urb**, **EnergyUse_x_Lowcarbon_energy**, **EnergyUse_poly_2**, **GDP_poly_2**.

In this case, **GDP** is included in the best model. Moreover, the above mentioned covariates are the ones used in the regression models and predictions (considering variable transformations).

4 Prediction Analysis

4.1 Choice of the metrics

To evaluate the quality of our models, we need to define the metrics we will use to compare their performance. We used 4 different metrics that we define in the following. We denote by $\widehat{y_k}$ the output of a model M for the input x_k (hence the predicted value), whose actual observed value is y_k.

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{k=1}^{n} |\widehat{y_k} - y_k|$$

- Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{k=1}^{n} (\widehat{y_k} - y_k)^2$$

- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n} (\widehat{y_k} - y_k)^2}$$

- Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n}\sum_{k=1}^{n} \left|\frac{y_k - \widehat{y_k}}{y_k}\right|$$

The MAE is proportional to the Manhattan norm (l_1 norm) of the prediction error and like for the RMSE, MAE has the advantage of being dimensionally consistent with the target value.
MSE is proportional to the Euclidean norm (l_2 norm) of the error and the RMSE is its square root. Moreover, they penalize large error of prediction with the squaring.
MAPE is less intuitive than MAE and RMSE because it is not symmetric, it treats overestimates and underestimates differently. However, it helps to understand the prediction errors relative to the magnitude of the target values, expressed as a percentage.
In general, lower values of these metrics indicate better model performance.

4.2 Predictions with original features

<u>Predictions with g-prior and JSZ-prior:</u>

To evaluate the models, we must analyze their ability to perform prediction on new data, where the predictive posterior distribution can be computed as:

$$\int \pi(y_{new}|\sigma^2, \beta, X_{new})\pi(\sigma^2, \beta|y, X)d\sigma^2 d\beta$$

To do so, we split the data into 2 sets, the training set and the test set. The training set represents 70% of all the data. First, we considered the original features available in the dataset. The following plots represent the distribution of predictions and confidence bounds when using g-prior with the hyperparameter g set to 100.

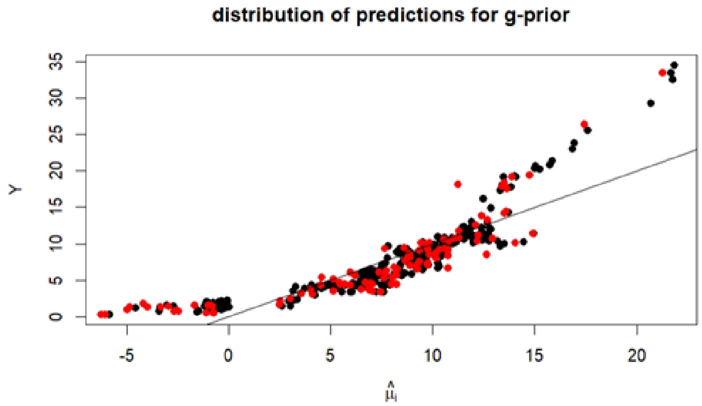

Figure 12 Distribution of predictions

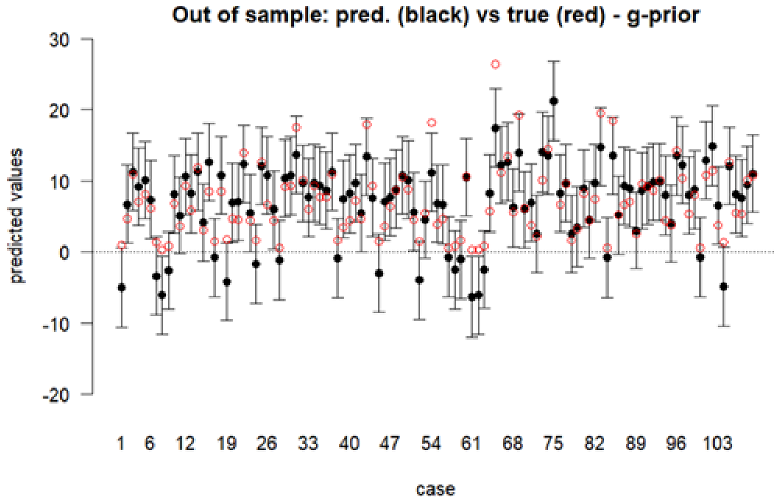

Figure 13 Out of sample predictions

The prediction errors are the following:

- MSE = 10.535
- RMSE = 3.246
- MAE = 2.433
- MAPE = 46.764%

If we use the JSZ-prior instead of g-prior, we observe very similar results, meaning that at least the choice of the prior is not critical for making predictions.

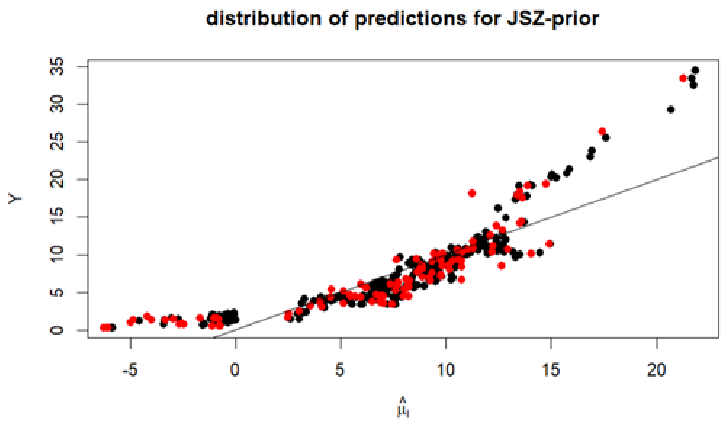

Figure 14 Distribution of predictions

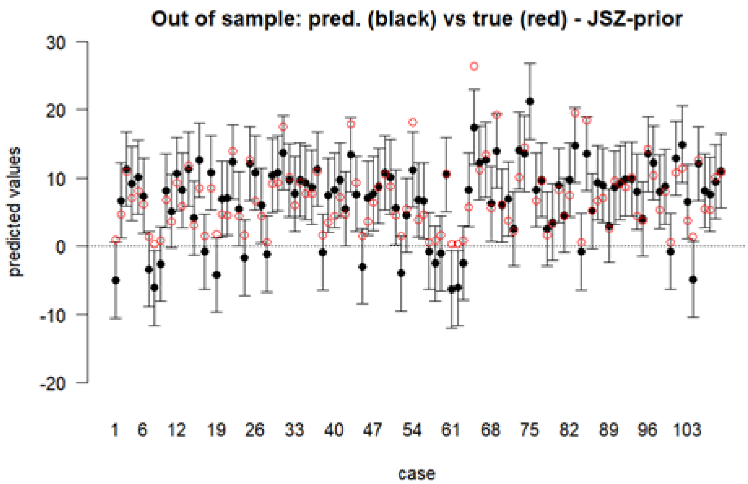

Figure 15 Out of sample predictions

The prediction errors are the same to the nearest thousandth as in the previous case:

- MSE = 10.535
- RMSE = 3.246
- MAE = 2.433
- MAPE = 46.764%

<u>Predictions using Bayesian Information Criterion (BIC):</u>

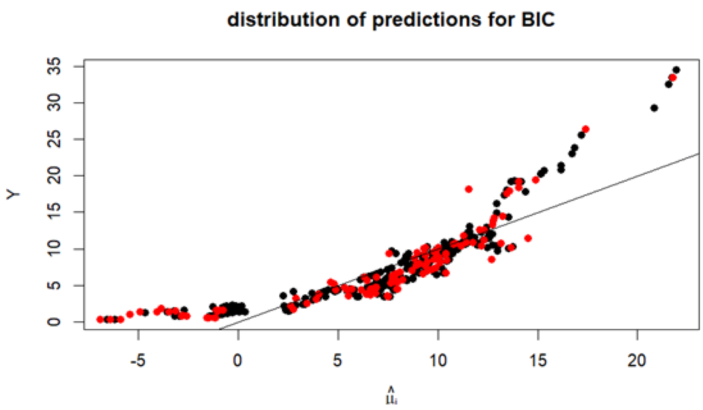

Figure 16 Distribution of predictions

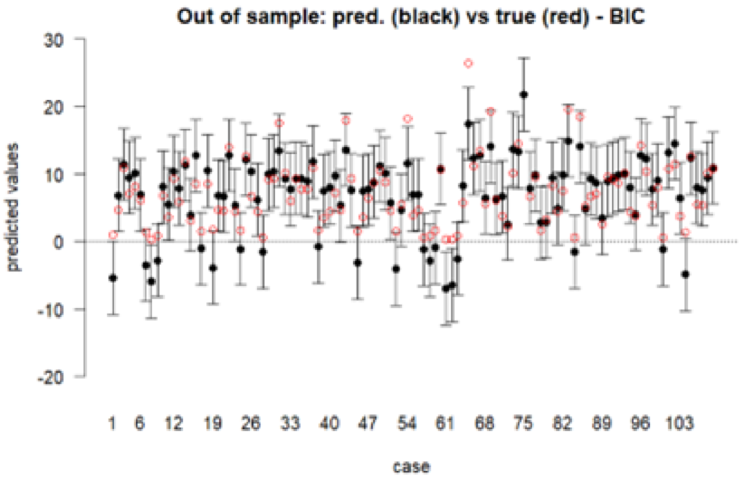

Figure 17 Out of sample predictions

The prediction errors are:

- MSE = 9.863
- RMSE = 3.141
- MAE = 2.358
- MAPE = 68.033%

The prediction performances did not meet our expectations. The prediction distribution plots (Figures 12, 14, 16) show that some points deviate significantly from the diagonal. Similarly, the out-of-sample prediction plots (Figures 13, 15, 17) indicate that in some cases, the true values lie outside the 95% credible interval of the predictive distribution. Also the errors are quite high.
Hence in the following section, we will consider the new features described in the paragraph 2.3, to achieve better prediction performance.

4.3 Predictions with transformed features

<u>Predictions with g-prior and JSZ-prior:</u>

Now we consider the new features we have built from the original ones as seen in the dedicated part. The following plots represent the distribution of predictions and confidence bounds when using g-prior with the hyperparameter g set to 100.

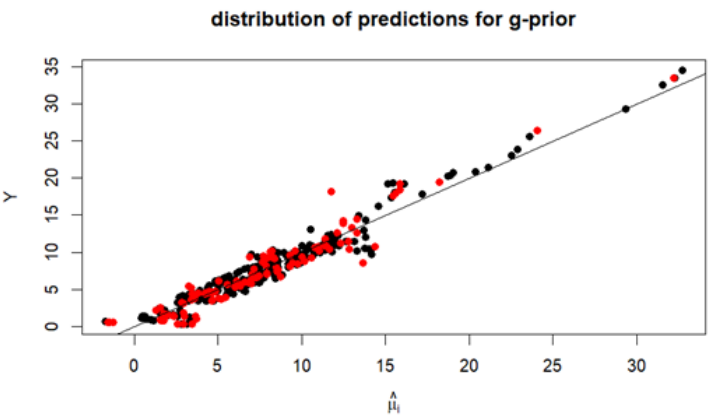

Figure 18 Distribution of predictions

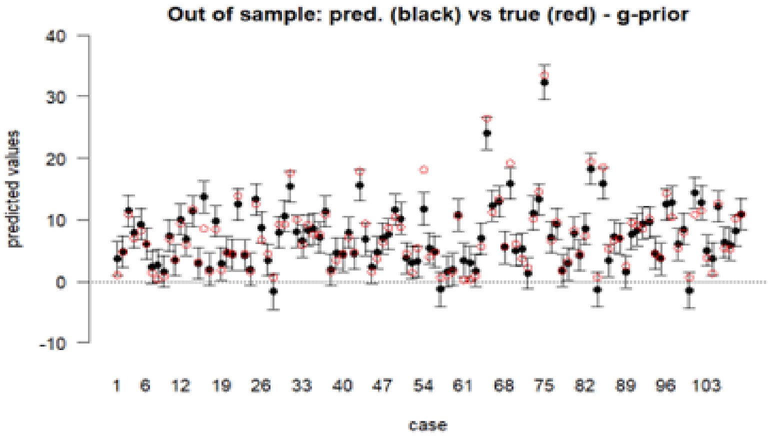

Figure 19 Out of sample predictions

Where the errors are:

- MSE = 2.606
- RMSE = 1.614
- MAE = 1.205
- MAPE = 34.692%

While using the JSZ-prior we ended up with the following outcomes.

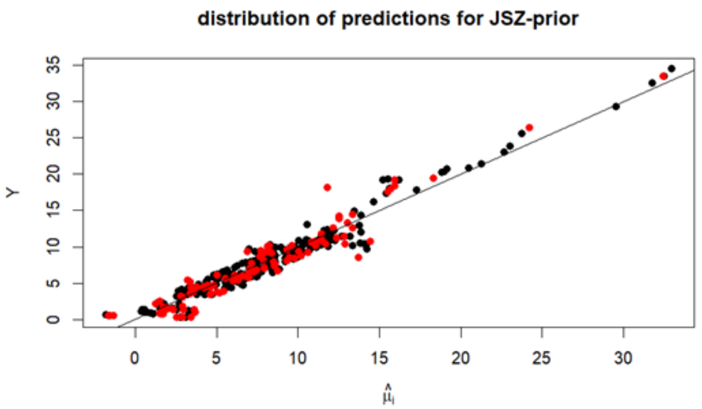

Figure 20 Distribution of predictions

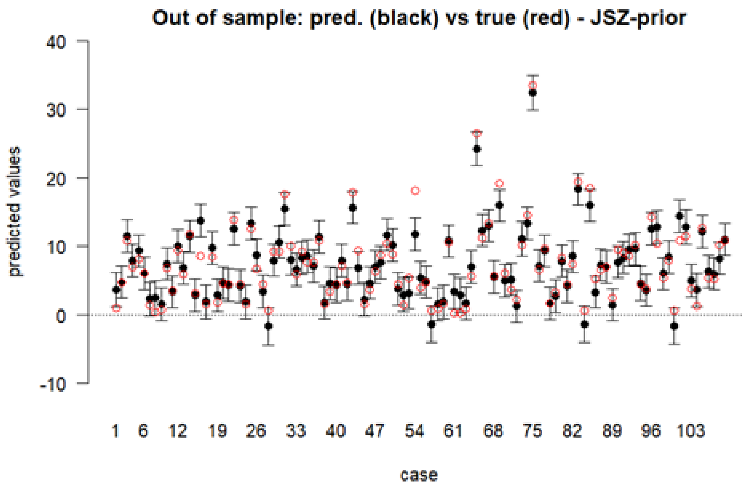

Figure 21 Out of sample predictions

The prediction errors in this case are:

- MSE= 2.559
- RMSE = 1.600
- MAE = 1.194
- MAPE = 30.257%

It can be noted that in this case the JZS prior shows a better performance with respect to the g-prior.

Predictions using Bayesian Information Criterion (BIC):

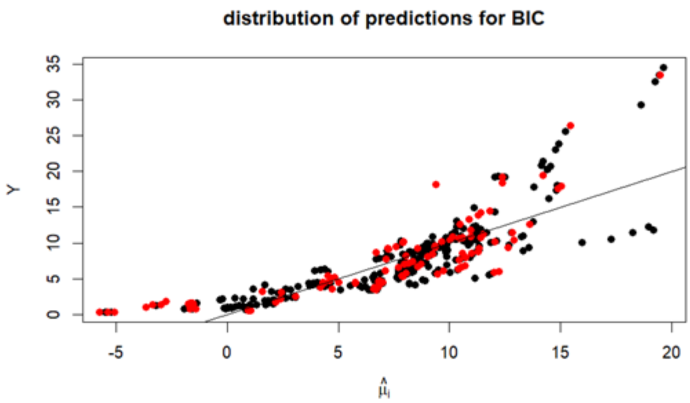

Figure 22 Distribution of predictions

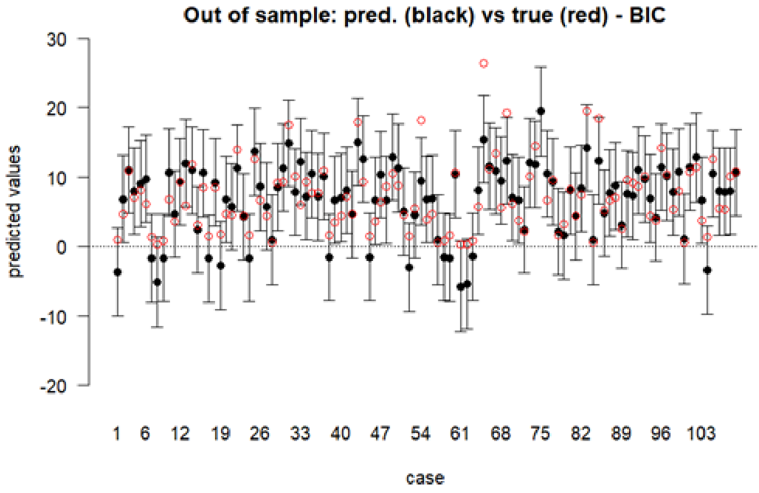

Figure 23 Out of sample predictions

The prediction errors are:

- MSE = 11.287
- MSE = 3.360
- MAE = 2.497
- MAPE = 51.457%

The previous prediction errors highlight that, with variable transformations, the prediction has been improved leading to a better fit. The prediction distribution plots show a more uniform distribution of the points with respect to the diagonal. Moreover, the out-of-sample prediction plots show less cases where the true values lie outside the 95% credible interval of the predictive distribution.
However, in the case of predictions using the BIC criterion, we are still far from achieving a good fit. This could be attributed to the use of a non-informative prior on the coefficients and the consideration of only the best model (through "HPM"), rather than utilizing all possible models in the predictions as in the g-prior and JZS prior approaches.

5 Relationship between CO2 and GDP

5.1 Features analysis

In this part, we compare the relationship between **GDP** and CO2 emissions per capita. The correlation between these 2 features, considering the original dataset, is 0.62.
The following graph displays the scatter plot between **GDP** and the target variable **co2percap**, along with a regression line fitted to the data.

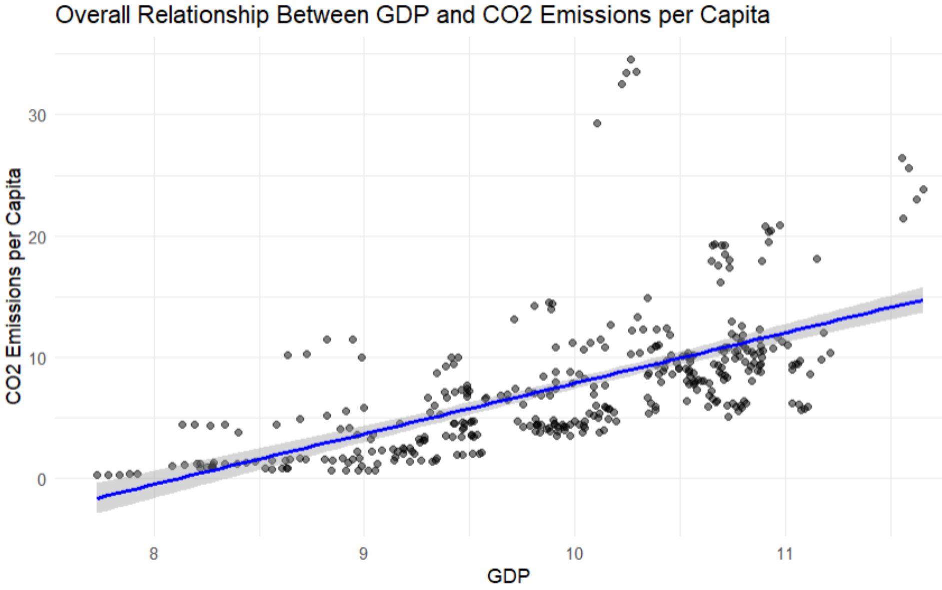

Figure 24 Scatter plot GDP vs co2percap

Now we split the data into 2 subsets based on the median **GDP** value: the first one is composed of the low income (**GDP** values less or equal than the median) and high income (**GDP** values greater than the median). In this way we ended up with 182 observations for both low income and high income. The outcome of this step is shown in Figure 25.

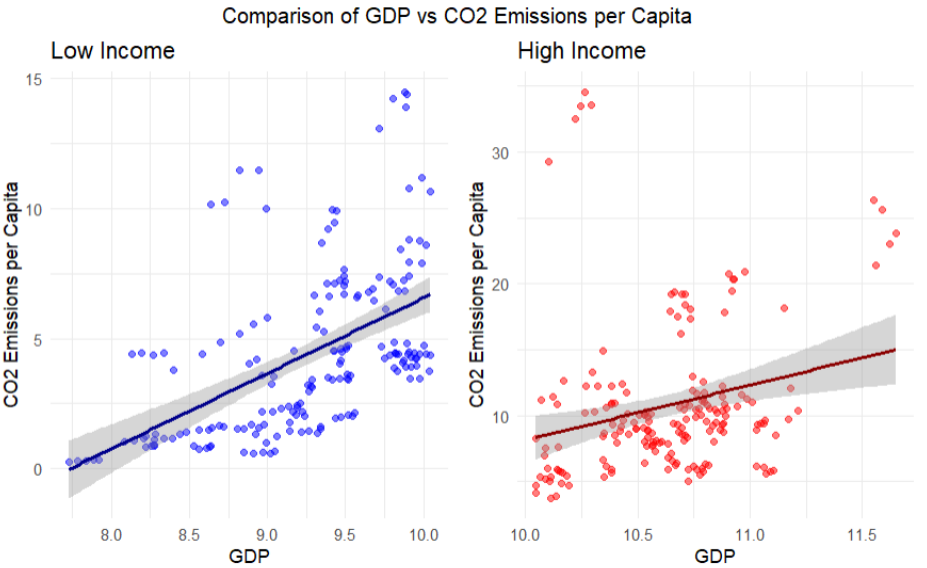

Figure 25 Scatter plot GDP vs co2percap with the splitted dataset

Additionally, the new correlation between **GDP** and **co2percap** is 0.53 for low income and 0.24 for high income, thus in both cases it decreases with respect to the original one (0.62).

5.2 GDP posterior analysis

We fit again the regression model and, in particular, we fitted it once for the data with the lower **GDP** and another time for the higher ones, both with JSZ prior. Considering first the dataset without variable transformation we got the following posterior distributions of the β corresponding to the **GDP**:

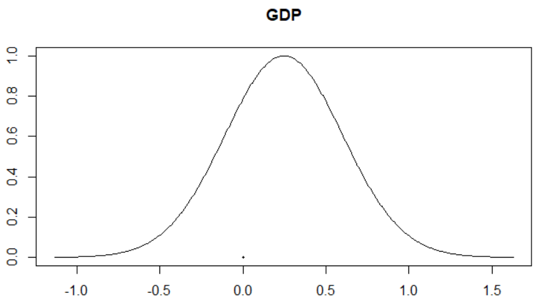

Figure 26 Low income

Figure 28 High income

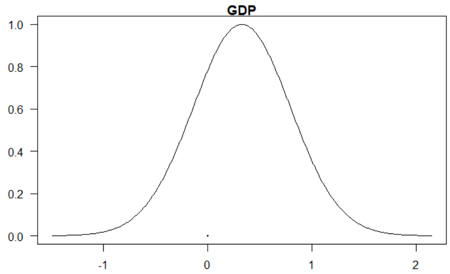

Figure 27 Full dataset

Analysing the low income data we ended up with a model which includes more covariates with respect to the original case, this is also due to the fact that we have less data. However as the first case the most participating covariates are: **EnergyUse**, **Lowcarbon_energy** and **internet**.
The **GDP** seems to do not participate in estimating the model since the zero is very near to the centre of its distribution.
On the other hand, the high income analysis shows a very similar result with respect to the original case. The **GDP** seems to do not participate in estimating the model also in this case, since its beta distribution is very similar to the low income one.

While the results concerning the dataset with the variable transformations are as follows:

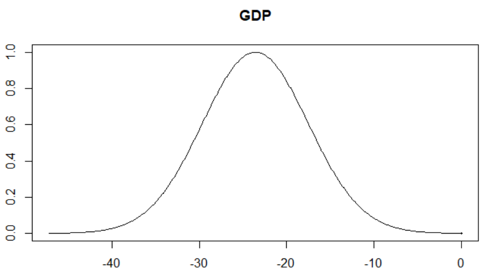

Figure 31 Low income

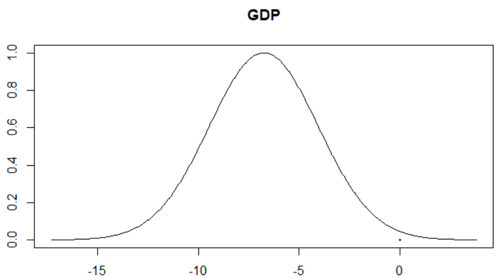

Figure 30 High income

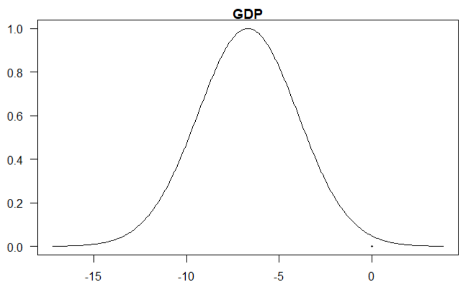

Figure 29 Full dataset

In this case we got that at low income the **GDP** has a great impact in fitting the model since the centre of the posterior distribution of its beta is very far from zero. Moreover we noticed that with the new features also **GDP_poly_2** has a strong impact, but less than **GDP**.
The high income analysis shows a very similar result with respect to the low income case. However it can be noticed that in this case the centre of the posterior distribution of **GDP**'s beta is closer to zero, additionally **GDP_poly_2** has gained a greater impact impact in fitting the model with respect to the low income case.

6 Conclusions

We observed that in terms of prediction accuracy, both the g-prior and JZS prior produced very similar results, confirming the robustness of the model despite variations in the value of g. Additionally, the model selected using the BIC criterion showed the poorest performance in predicting the target variable, highlighting that model selection via BMA is more effective.
Furthermore, applying variable transformations has demonstrated a significant improvement in the performance of all analyses.

Finally, the analysis of the relationship between GDP and CO2 per capita, with variable transformations applied, has shown that at lower income levels, CO2 emissions are strongly correlated with wealth. This is because increased income leads to higher energy consumption, often sourced from burning fossil fuels. Conversely, at higher income levels, GDP is less correlated with CO2 emissions. This is because wealthier countries can afford to adopt more expensive alternatives that result in lower environmental impact.