The Selection of Bayesian Polynomial Regression with INLA by using DIC, WAIC and CPO

H T Sutanto¹,², H Pramoedyo², W S Wardhani² and S Astutik²
¹Student of Mathematics Department, Brawijaya University, Jl. Veteran, Malang, Indonesia
²Statistics Department, Brawijaya University, Jl. Veteran, Malang, Indonesia
³Mathematics Departement, Universitas Negeri Surabaya
herysutanto@student.ub.ac.id

Abstract. The polynomial regression model is extended the multiple linear regression. The selection of Bayesian polynomial regression model with INLA required Criterion. Criterion is using the measure fit model with the available data. There are three criteria, namely DIC, WAIC and CPO. The smaller criterion value from DIC, WAIC and CPO on a model show the best Bayesian polynomial regression model with INLA.

1. Introduction
The regression analysis is the statistics model which use for study relationship between a dependent variable and an independent variable or more. The construction of regression model fit for statistics data requires “plot diagram” for the interest variable. The Using regression model have two objectives:

1. Predict relationship between one independent variable or more with one dependent variable.
2. Predict the expected value for the dependent variable when the independent variable is given

The one from the regression model using for study relationship between the dependent variable with independent variable in the polynomial function form is the polynomial regression. Polynomial regression is the multiple linear regression which is constructed with add the impact of independent variable from the little degree to the highest degree [1].

The difficult on the polynomial regression model is find the degree polynomial regression [2]. The polynomial degree with the classic statistics (frequentist), namely (1) Thompson on 1978 is using the variable selection (the forward selection or backward elimination) which use the t statistics for the variable coefficient test for the highest degree and (2) Akaike on 1973 the model selection the polynomial degree with Akaike Information Criterion.

In 2009, Rue development Integrated Nested Laplace Approximation (INLA) as alternative the Bayesian method. INLA is an approximation which have fast and exact compare MCMC. INLA is constructed use the Latent Gaussian Model with product output fast and exact compare MCMC.

The computation of Bayesian INLA is faster than MCMC, because:
1. The computation parameters joint posterior distribution model can be instead the each parameter posterior marginal distribution model [3].
2. The computation of GMRF (Gaussian Markov Random Fields) will decrease time for running.
3. Rue compute posterior marginal distribution for the parameter model with Laplace Approximation [3].

The selection best Bayesian polynomial regression model from Bayesian polynomial regression model candidate use the DIC, WAIC and CPO criterion. This criterion is used for measuring the fit
model with the given data. The first Bayes Information Criterion proposed by Schwarz on 1978, then the revised Kass and Raftery [5] developed WBIC. There are other criterion popular DIC (Deviance Information Criterion) by Spiegelhalter et al [6]. The first Akaike Information Criterion (AIC) proposed by Akaike on 1973, become WAIC1 (Watanabe, 2010) and WAIC2 (Watanabe, 2010). Then CPO (Conditional Predictive Ordinate) criterion proposed by Geisser and Eddy [7]

Regard to the background above, here the following problem in this study: 1) how determine the best bayesian polynomial regression model with INLA using DIC method, 2) how determine the best bayesian polynomial regression model with INLA using WAIC method and 3) how determine the best bayesian polynomial regression model with INLA using CPO method

2. Polynomial Regression

If Y is the response variable and x is the independent variable, the degree j polynomial regression model \((M_j)\):

\[ Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_j x^j + \varepsilon, \quad 0 < j \leq d \] (2.1)

If take sample with n number, then model for each observation:

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n \] (2.2)

With the assumption of \(\varepsilon_i \sim IID N(0, \sigma^2)\), then model on (2.2) became

\[ \tilde{Y} = X_j \tilde{\beta}_j + \tilde{\varepsilon} \] (2.3)

with \(\tilde{Y} = [Y_1 \ Y_2 \ \cdots \ Y_n]^T, \tilde{\beta}_j = [\beta_0 \beta_1 \ \cdots \ \beta_j]^T\),

\[ X_j = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^j \\ 1 & x_2 & x_2^2 & \cdots & x_2^j \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^j \end{bmatrix} \]

\[ \tilde{\varepsilon} = [\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n]^T \]

3. Approximation of Bayesian Inference with INLA

If defining a Latent Gaussian Model for Bayesian inference in INLA on all latent parameters and hyperparameter with compute the marginal posterior distribution for each element of the parameter vector.

\[ p(\theta_i | y) = \int p(\theta_i, \Psi | y) \ d\Psi = \int p(\theta_i | \Psi, y) \ p(\Psi | y) \ d\Psi \quad i = 1, 2, \ldots, n_\theta \] (2.4)

And for each element of the hyperparameter vector

\[ p(\Psi_k | y) = \int p(\Psi | y) \ d\Psi_{-k} \quad k = 1, 2, \ldots, s \] (2.5)

With \(n_\theta = \) latent parameter number

\(s = \) hyperparameter number

This Integral-integral are not the analytic solution than the marginal posterior distribution are then approximated by

\[ \tilde{p}(\theta_i | y) = \int \tilde{p}(\theta_i | \Psi, y) \ \tilde{p}(\Psi | y) \ d\Psi \quad i = 1, 2, \ldots, n_\theta \] (2.6)

and

\[ \tilde{p}(\Psi_k | y) = \int \tilde{p}(\Psi | y) \ d\Psi_{-k} \quad k = 1, 2, \ldots, s \] (2.7)

The marginal posterior distribution approximation (2.5) and (2.6) then the hyperparameter joint posterior distribution with the Laplace approximation using equation (2.7) and the numerical integration on hyperparameter \(\Psi_{-k}\) for \(\tilde{p}(\Psi_k | y)\).

Thus compute

i) \(p(\Psi | y)\) from which also all the relevant marginal \(p(\Psi_k | y)\) can be obtained

ii) \(p(\theta_i | y)\) which is required to compute the parameter marginal posteriors \(p(\theta_i | y)\)

Step-step the Laplace approximation for the marginal posterior distribution (Blangiardo dan Cameletti ,2015)

i) Computation from the hyperparameter joint posterior distribution approximation
\[
p(\Psi|y) = \frac{p(\theta, \Psi|y)}{p(\theta|\Psi, y)}
\]

(2.8)

\[
= \frac{p(y|\theta, \Psi)p(\theta, \Psi)}{p(y)} \frac{1}{p(\theta|\Psi, y)}
\]

\[
= \frac{p(y|\theta, \Psi)p(\theta|\Psi)p(\Psi)}{p(y)} \frac{1}{p(\theta|\Psi, y)}
\]

\[
\propto \frac{p(y|\theta, \Psi)p(\theta|\Psi)p(\Psi)}{p(\theta|\Psi, y)}
\]

\[
\approx \frac{p(y|\theta, \Psi)p(\theta|\Psi)p(\Psi)}{p(\theta|\Psi, y)}
\]

(2.9)

ii) The vector parameter latent as \( \theta = (\theta_1, \theta_{-i}) \) and Laplace approximation

\[
p(\theta_1|\Psi, y) = \frac{p((\theta_1, \theta_{-i})|\Psi, y)}{p(\theta_{-i}|\theta_1, \Psi, y)}
\]

(2.10)

\[
\approx \frac{p(\theta_1, \Psi|y)}{p(\theta_{-i}|\theta_1, \Psi, y)}
\]

\( \theta_{-i} = \theta^*_i(\theta_1, \Psi) =: \tilde{\theta}_i(\Psi|y) \)

After obtained \( \tilde{\theta}_i(\Psi|y) \) equation (2.10) and \( \tilde{\Psi}(y) \) equation (2.9), then marginal posterior distribution from \( p(\theta|y) \) are then approximated by

\[
\tilde{p}(\theta|y) = \int \tilde{p}(\theta|\Psi, y) \tilde{p}(\Psi|y) d\Psi
\]

(2.11)

Equation (2.11) can be solved numerically through a finite weighted sum:

\[
\tilde{p}(\theta_1|y) \approx \sum \tilde{p}(\theta_1|\Psi(j), y) \tilde{p}(\Psi(j)|y) \Delta_j
\]

(2.12)

Solution the marginal posterior distribution process through INLA algorithm as follows:

Exploration for \( \tilde{\Psi}(y) \) : for the parameter joint posterior distribution, INLA search the parameter space and in selection to detect good points \( \{\Psi(j)\} \) for equation (2.12). Rue et al (2009) propose two different schemes, both requiring a reparameterization of the \( \Psi \) space in order to deal with more regular densities through the following steps:

1. Locate the mode \( \Psi^* \) of \( \tilde{p}(\Psi|y) \) from equation (2.9) for obtained initial value \( \Psi(f) \) on solution numerical integration on equation(2.12). Compute value \( \Psi(f) \) by optimizing \( \log \tilde{p}(\Psi|y) \) respect \( \Psi \) through Newton-Raphson method..

2. Search the mode \( \Psi^* \) with compute the negative Hessian \( H \) at the modal configuration. With \( \Sigma \) is covarains matrices for \( \Psi \), where \( \Sigma = H^{-1} \). Compute the eigen-decomposition \( \Sigma = V \Lambda^{1/2} V^T \)

3. With \( \Sigma = H^{-1} \). Compute the eigen-decomposition \( \Sigma = V \Lambda^{1/2} V^T \) where \( V \) is matrices eigen vector and \( \Lambda \) is matrices diagonal eigen value.

Then define the new variable \( z \) with standardized and mutually orthogonal components, such that:

\[
\Psi(z) = \Psi + V \Lambda^{1/2} z
\]

4. Posterior \( \tilde{p}(\theta_1|\Psi(j), y) \) evulation for each value \( \Psi(j) \), then \( \tilde{p}(\theta_1|y) \) compute using equation (2.12)
4. Model Selection

If the best model selection for several model possibility required a criterion is used for measure the model fit with the available data. Need criterion which is used for measure consistent with the model. There several criterions which is used for the best model selection.

a) The Bayes Information Criterion (BIC)

The Bayes Information Criterion (BIC) proposed by Schwarz (1978) with definition:

\[ S_{01} = \log(f(y|\hat{\theta}_{m1}, m1)) - \log(f(y|\hat{\theta}_{m0}, m0)) - \frac{1}{2} (d_{m1} - d_{m0}) \log(n) \]  \hspace{1cm} (2.13)

Where: \( n \) is sample size \( \hat{\theta}_m \) is the maximum likelihood estimator from parameter \( \theta_m \)

\( d_m \) is dimension for parameter \( \theta_m \)

On 1995, BIC criterion is revised by Kass and Wasserman and Kass and Raftery with definition

\[ \text{BIC}(m) = D(\hat{\theta}_m, m) + d_m \log n \]  \hspace{1cm} (4.14)

\( D(\hat{\theta}_m, m) \) is the deviant measure from model with definition

\[ D(\hat{\theta}_m, m) = -2 \log f(y|\hat{\theta}_m, m) \]

Then:

\[ S_{01} = \frac{-1}{2} \{ \text{BIC}(m_0) - \text{BIC}(m_1) \} \]  \hspace{1cm} (415)

where \( m_0 \) : the initial model which interest \( m_1 \): the other model

Marginal Likelihood with definition \( \int f(y|\theta) h(\theta) d\theta \) with \( h(\theta) \) is the prior density on parameter \( \theta \). Then BIC = \( -2 \log f(y|\hat{\theta}) + k \log n \) \hspace{1cm} (2.16)

WBIC is the free energy estimator \( F(D) \) for the singular model [8], with \( P_\beta(w|D) \) is a the posterior distribution under inverse temperature \( \beta \).

\[ \text{WBIC} \] is derived from BIC for the singular model with definition

\[ \text{WBIC} = -\int (\sum_{n=1}^{N} \log p(x^\mu|w)) P_\beta(w|D) \, dw P_\beta(w|D) \]  \hspace{1cm} (4.17)

with

\[ P_\beta(w|D) = \frac{1}{P_\beta(D)} p(w) \prod_{\mu=1}^{N} p(x^\mu|w)^\beta \]

b) Akaike Information Criterion (AIC)

Distance Kullback-Leibler is measure from the interest model to the true model, which have density \( g(y) \) with definition:

\[ K(f, g) = \int \left( \log g(y) - \log f(y|\hat{\theta}) \right) g(y) \, d(y) \]  \hspace{1cm} (4.18)

Akaike Informasi Criterion with definition

\[ \text{AIC} = -2 \log f(y|\hat{\theta}_{MLE}) + 2k \]  \hspace{1cm} (4.19)

With \( k \) is number parameter \( \hat{\theta}_{MLE} \) is the parameter estimator obtained from the maximum likelihood method

Value AIC is smaller showing the selection best model from model-model possibility.
WAIC is derived from the singular learn theory [9] as the new information Criterion for the singular model. Watanabe is introduced 4 quantity:

\[ BL_g (\text{Bayes generalization}), BL_t (\text{Bayes training loss}), GL_g (\text{Gibbs generalization loss}), GL_t (\text{Gibbs training loss}) \]

\[ BL_g = - \int p^*(x) \log p(x|D) \, dx \quad (4.20) \]
\[ BL_t = - \frac{1}{N} \sum_{\mu=1}^{N} \log p(x^\mu|D) \quad (4.21) \]
\[ GL_g = - \int \left( \int p^*(x) \log(x|w) \, dx \right) p(w|D) \, dw \quad (4.22) \]
\[ GL_t = - \int \left( \frac{1}{N} \sum_{\mu=1}^{N} \log p(x^\mu|D) \right) p(w|D) \, dw \quad (4.23) \]

where:
- \( p^*(x) \) = the true distribution
- \( p(w|D) \) = posterior distribution
- \( p(x|D) \) = predictive distribution
- \( p(w|D) = \frac{p(w)}{p(D)} \sum_{\mu=1}^{N} p(x^\mu|w) \)
- \( p(x|D) = \int p(x|w) p(w|D) \, dw \)

WAIC1 is estimator from \( BL_g \) derive AIC for the regular model (Watanabe, 2010) with \( WAIC1 = BL_t + 2(GL_t - BL_t) \)

WAIC2 is the estimator from \( GL_g \) [8] with

\[ WAIC2 = GL_g + 2(GL_t - BL_t) \quad (4.25) \]

**c) Deviance Information Criterion**

Kriteria deviance Information Criterion be proposed by Spiegelhalter et al [6] with definition:

\[ D(\theta) = -2 \log f(\theta|y) \]

The Bayesian model for random variable is used compute the expected deviance \( E[D(\theta)] \) under posterior distribution as the fit model measure. This measure is show the effective number parameter

\[ p_D = E[D(\theta)] - D[E(\theta)] = \bar{D} - D(\hat{\theta}) \]

then DIC= \( D(\hat{\theta}) + 2P_D \)

**d) CPO (Conditional predictive Ordinate)** is the cross-validation method for estimation the leave-one-out predictive distribution (Geisser and Eddy, 1979). The posterior predictive distribution for observation \( i \) with definition:

\[ p(y_i|y_{-i}) = \int p(y_i|\theta) \prod (\theta|y_{-i}) \, d\theta \]

where
- \( y_i \) is observation to \( i \) from \( y \)
- \( y_{-i} \) is observation to \( i \) out \( y \)

Then CPO is the posterior probability observation \( y_i \) when model without \( y_i \).

The biggest value from CPO is show a the best adjustment. For the complete model, each time CPO approximation is obtained with the sampling posterior distribution:

\[ \text{CPO}_i = \frac{1}{T^{-1} \sum_{t=1}^{T} p(y_i|\theta^t)^{-1}} \]

Where \( \theta^t \) is a sample from \( \Pi(\theta|y_i) \) and \( T \) is number sample posterior. Inverse from mean posterior for inverse likelihood. This method is efficient but required the numerical computation (Held, Schrodle and Rue, 2010)

If INLA is computed CPO the efficient less.

\[ \log \text{CPO} = -\frac{1}{n} \sum_{i=1}^{n} \log(p(y_i|y_{-i})) \]
The smallest value from log CPO is show the best model.

5. Discussion

Data and scatter diagram
If data on tabel 1 about machine setting and energy consumption number

| Number | Machine setting | Energy consumption |
|--------|-----------------|--------------------|
| 1      | 11.15           | 21.6               |
| 2      | 15.17           | 4                  |
| 3      | 18.9            | 1.8                |
| 4      | 19.4            | 1                  |
| 5      | 21.4            | 1                  |
| 6      | 21.7            | 0.8                |
| 7      | 25.3            | 3.8                |
| 8      | 26.4            | 7.8                |
| 9      | 26.7            | 4.3                |
| 10     | 29.1            | 36.2               |

Data on table 1 is plotted to obtain a scatter diagram like shown below.

The plot data from scatter diagram figure 1 is obtained from several polynomial regression possibilities: the linear regression model, the square regression model, the degree 3 polynomial regression model and the degree 4 polynomial regression model.

The Bayesian Linear Regression Model with INLA
The Bayesian Linear Regression Model with INLA with program.

```r
formula1 <- kons ~ mes
m.gaussianpol1 <- inla(formula1, data = datakonsmes, family = "gaussian",
control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))
summary(m.gaussianpol1)
```

Call:
```
c("inla(formula = formula1, family = "gaussian", data = datakonsmes, ", "
```
control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))

Time used:
  Pre = 1.28, Running = 0.33, Post = 0.249, Total = 1.86

Fixed effects:
  mean sd 0.025quant 0.5quant 0.975quant mode kld
(Intercept) 1.089 16.443 -31.817 1.088 33.945 1.088 0
mes 0.329 0.740 -1.153 0.329 1.808 0.329 0

The model has no random effects

Model hyperparameters:
  mean sd 0.025quant 0.5quant 0.975quant mode
Precision for the Gaussian observations

Expected number of effective parameters(stdev): 2.00(0.00)
Number of equivalent replicates: 5.00

Deviance Information Criterion (DIC) ...............: 82.41
Deviance Information Criterion (DIC, saturated) ....: 16.28
Effective number of parameters ......................: 3.14
Watanabe-Akaike information criterion (WAIC) ...: 85.71
Effective number of parameters ......................: 4.84
Marginal log-Likelihood: -54.35
CPO and PIT are computed
Posterior marginals for the linear predictor and
the fitted values are computed

The Bayesian Square Regression Model with INLA

The Bayesian Square Regression Model with INLA with program.
form2 <- kons ~ mes + meskw
m.gaussianpol2 <- inla(formula2,
  data = datakonsmes, family = "gaussian",
  control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))
summary(m.gaussianpol2)
Call:
c("inla(formula = formula2, family = "gaussian", data = datakonsmes, ", ", 
control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))")

Time used:
  Pre = 1.22, Running = 0.526, Post = 0.454, Total = 2.2

Fixed effects:
  mean sd 0.025quant 0.5quant 0.975quant mode kld
(Intercept) 130.056 26.574 76.449 130.163 182.947 130.323 0
mes -13.207 2.683 -18.550 -13.218 -7.802 -13.234 0
meskw 0.331 0.065 0.200 0.331 0.460 0.332 0

The model has no random effects

Model hyperparameters:
  mean sd 0.025quant 0.5quant 0.975quant mode
Precision for the Gaussian observations

Expected number of effective parameters(stdev): 2.99(0.005)
Number of equivalent replicates: 3.34
Deviance Information Criterion (DIC) ..............: 69.15
Deviance Information Criterion (DIC, saturated) ....: 17.54
Effective number of parameters ......................: 4.27
Watanabe-Akaike information criterion (WAIC) ....: 71.37
Effective number of parameters ......................: 4.88
Marginal log-Likelihood: -52.94
CPO and PIT are computed

Posterior marginals for the linear predictor and the fitted values are computed

**Compare the Bayesian Linear Regression Model with INLA and The Bayesian Square Regression Model with INLA**

For Compare the Bayesian Linear Regression Model with INLA with The Bayesian Square Regression Model with INLA is used table 2.

| Model      | DIC     | WAIC    | CPO          | MLIK     |
|------------|---------|---------|--------------|----------|
| Gaussian   | 82.40976| 85.70997| 45.14273     | -54.33366|
| Gaussian + r. eff. | 69.15072| 71.37171| 39.96533     | -52.91957|

Because value DIC (69.15072) and WAIC(71.37171) on the bayesian square polynomial regression model with INLA less than value DIC(82.40976) and WAIC(85.70997) the Bayesian linear regression model with INLA then the best Bayesian square polynomial regression model with INLA.

**The Bayesian degree 3 polynomial regression Model with INLA**
The Bayesian degree 3 polynomial regression Model with INLA with program.

```r
formula2 <- kons ~ mes + meskw + mesp3
m.gaussianpol3 <- inla(formula2,
  data = datalconmes, family = "gaussian",
  control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))
summary(m.gaussianpol3)
```

Call:
`c("inla(formula = formula2, family = "gaussian", data = datalconmes, ",
  control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))")`

Time used:
Pre = 1.54, Running = 0.187, Post = 0.0939, Total = 1.82

Fixed effects:

| Effect     | mean   | sd     | 0.025quant | 0.5quant | 0.975quant | mode | kld     |
|------------|--------|--------|------------|----------|------------|------|---------|
| (Intercept)| -38.113| 73.092 | -179.873   | -39.528  | 112.280    | -41.815 | 0       |
| mes        | 15.075 | 11.977 | -9.537     | 15.309   | 38.332     | 15.685  | 0       |
| meskw      | -1.157 | 0.622  | -2.364     | -1.169   | 0.124      | -1.189  | 0       |
| mesp3      | 0.025  | 0.010  | 0.004      | 0.025    | 0.045      | 0.025   | 0       |

The model has no random effects

Model hyperparameters:

| mean   | sd     | 0.025quant | 0.5quant | 0.975quant | mode     |
|--------|--------|------------|----------|------------|----------|
| 0.062  | 0.031  | 0.018      | 0.057    | 0.136      | 0.047    |
the Gaussian observations

Expected number of effective parameters(stdev): 3.85(0.074)
Number of equivalent replicates : 2.59
Deviance Information Criterion (DIC) ..................: 64.78
Deviance Information Criterion (DIC, saturated) ....: 18.68
Effective number of parameters ....................: 5.26
Watanabe-Akaike information criterion (WAIC) ...: 65.92
Effective number of parameters ....................: 4.88

Marginal log-Likelihood: -58.14
CPO and PIT are computed
Posterior marginals for the linear predictor and
the fitted values are computed

Comparing the Bayesian Square Regression Model with INLA and The Bayesian Degree 3 Polynomial Regression Model with INLA

To compare the Bayesian square Regression Model with INLA with The Bayesian Degree 3 polynomial Regression Model with INLA is used table 3.

|Model            | DIC        | WAIC        | CPO         | MLIK          |
|-----------------|------------|-------------|-------------|---------------|
|Gaussian         | 69.15072   | 71.37171    | 39.96533    | -52.91957     |
|Gaussian + r. eff. | 64.78352  | 65.92286   | 41.15463    | -58.13238     |

Because value DIC (64.78352) and WAIC (65.92286) on Bayesian degree 3 polynomial regression model less than value DIC (69.15072) dan WAIC (71.37171) on Bayesian square regression model with INLA the best bayesian degree 3 polynomial regression model with INLA.

Bayesian Degree 4 Polynomial Regression Model with INLA

Bayesian Degree 4 Polynomial Regression Model with INLA with program.

```r
formula2 <- kons ~ mes + meskw + mesp3 + mesp4
Call: 
  c("inla(formula = formula3, family = "gaussian", data = dataokensmes, ", ", 
      control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE))"
) 
Time used: 
  Pre = 1.37, Running = 0.39, Post = 0.109, Total = 1.87
Fixed effects:
  mean sd 0.025quant 0.5quant 0.975quant  mode  kld
(Intercept) 234.942 152.679 -71.762 237.888 523.976 243.378
mes -46.476 32.923 -108.902 -47.056 19.390 -48.127
meskw  3.813 2.566 -1.341  3.863  8.671  3.954
mesp3 -0.147 0.086 -0.309 -0.149  0.028 -0.153
mesp4  0.002 0.001  0.000  0.002  0.004  0.002
```

The model has no random effects

Model hyperparameters:
  mean sd 0.025quant 0.5quant 0.975quant mode
Precision for 0.137 0.091 0.029 0.115 0.374 0.078

the Gaussian observations
Expected number of effective parameters (stdev): 4.30(0.128)
Number of equivalent replicates : 2.33

Deviance Information Criterion (DIC) ...............: 57.51
Deviance Information Criterion (DIC, saturated) ....: 17.81
Effective number of parameters ...................: 5.14
Watanabe-Akaike information criterion (WAIC) ...: 59.56
Effective number of parameters ...................: 5.55
Marginal log-Likelihood: -66.41
CPO and PIT are computed

Posterior marginals for the linear predictor and the fitted values are computed

**Compare Bayesian Degree 4 Polynomial Regression Model with INLA and Bayesian degree 3 polynomial regression Model with INLA**

Compare Bayesian Degree 4 Polynomial Regression Model with INLA and Bayesian degree 3 polynomial regression Model with INLA use table 4.

| Model | DIC       | WAIC     | CPO      | MLIK       |
|-------|-----------|----------|----------|------------|
| Gaussian | 64.78352 | 65.92286 | 41.15463 | -58.13238 |
| Gaussian + r. eff. | 57.51199 | 59.55811 | 38.38390 | -66.46425 |

Because DIC (57.51199) and WAIC (59.55811) on Bayesian degree 4 polynomial regression model with INLA less than value DIC (64.78352) and WAIC (65.92286) on Bayesian degree 3 polynomial regression with INLA then the best Bayesian degree 4 polynomial regression Model with INLA.

### 6. Conclusion

From the problem formulation and discussion can be taken conclusion. The best Bayesian degree 4 polynomial regression model with INLA fit for data relationship between machine setting variable with consumption energy variable. It is suggested to study more regarding the need of the four Bayesian polynomial regression model to compare joint then find the Bayesian polynomial regression model fit model with available data.

**References**

[1] Johnson B W and McCulloch R E 1987 Technometrics, 29 427
[2] Guttman I, Peña D, and Redondas, D 2005 Technometrics 47 23
[3] Lima E C, Gomes A A, and Tran H N 2020 J. of Molecular Liq. 113315.
[4] Rue H, Martino S, Lindgren F, Simpson D, Riebler A, and Krainski E T 2009 INLA: functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximation. *R* package version 0.0.
[5] Kass R E and Raftery A E 1995 *J. of the American Statistical Assoc.* 90 773
[6] Spiegelhalter D J, Best N G, Carlin B P, and Van Der Linde A 2002 *J. of the Royal Statistical Society: Series B (Statistical Methodology)* 64 583
[7] Gelfand A E and Dey D K 1994 *J. of the Royal Statistical Society: Series B (Methodological)* 56 501
[8] Watanabe S and Opper M 2010 *J. of Machine Learning Res.* 11
[9] Watanabe S 2009 *Algebraic geometry and statistical learning theory* (Vol. 25) (UK: Cambridge university press)