# BLP Simulation

*Nutcha Wattanachit*

*12/2/2019*

## Ensemble/pooling Methods

### TLP

Traditional linear pool finds optimal weights that maximmizes the likelihood of $f(y) = \sum_{i=1}^{k} w_i f_i(y)$.

### BLP

BLP finds $\alpha$, $\beta$, and weights that maximize the likelihood of

$$g_{\alpha,\beta} = (\sum_{i=1}^{k} w_i f_i(y)) b_{\alpha,\beta}(\sum_{i=1}^{k} w_i F_i(y)).$$

BLP Example: To obtain $\alpha$, $\beta$, and the weights for all component models, train the BLP model on half of the data. Then, use $\alpha$, $\beta$, and the weights from training to apply to the data held out for testing.

### Bias-corrected TLP (bcTLP)

This method corrects for bias of the component models, and then use the traditional method to generate the ensemble. The difference between this and the traditional TLP is that the component models inputted in bcTLP are bias-corrected. The bias correction method used is simple linear regression (Raftery, 2005). By regressing $y$ against the forecast output, we obtain the intercept and coefficient of each component,

$$g_i(y) = \alpha_i + \beta_i f_i(y)$$

.

and the final ensemble is

$$g(y) = \sum_{i=1}^{k} w_i g_i(y)$$

.

### Bias-corrected BLP (bcBLP)

This method also corrects for bias of the component models using linear regression, and then use the BLP method to generate the ensemble. The final ensemble is

$$g_{\alpha,\beta} = (\sum_{i=1}^{k} w_i g_i) b_{\alpha,\beta}(\sum_{i=1}^{k} w_i G_i(y)).$$

## BLP with Non-central Parameter(nBLP)

nBLP finds $\alpha$, $\beta$, non-central parameter $\lambda$, and weights that maximize the likelihood of $g_{\alpha,\beta,\lambda} = (\sum_{i=1}^{k} w_i f_i(y)) b_{\alpha,\beta}(\sum_{i=1}^{k} w_i F_i(y))$.

nBLP process: To obtain $\alpha$, $\beta$, $\lambda$, and the weights for all component models, train the nBLP model on half of the data. Then, use $\alpha$, $\beta$, $\lambda$, and the weights from training to apply to the data held out for testing.

## Component-wise BLP (cBLP)

This is the extension of the traditional BLP. We beta-transform each of the cumulative distribution functions of the component models. This is done by finding $\alpha$ and $\beta$ that maximize the likelihood of

$$\begin{aligned} G_{i,\alpha_i,\beta_i} &= B_{\alpha_i,\beta_i}[F_i(y)] \\ g_{i,\alpha_i,\beta_i} &= f_i(y) \times b_{\alpha_i,\beta_i}[F_i(y)] \end{aligned}$$

Then, to obtain $\alpha$, $\beta$, and the weights for 21 models, we apply BLP on the beta-transformed components:

$$\begin{aligned} G_{\alpha,\beta} &= B_{\alpha,\beta}\Big[\sum_{i=1}^{k} w_i B_{\alpha_i,\beta_i}(F_i(y))\Big] \\ g_{\alpha,\beta} &= \Big(\sum_{i=1}^{k} w_i b_{i,\alpha_i,\beta_i}(F_i(y)) f_i(y)\Big) b_{\alpha,\beta}\Big(\sum_{i=1}^{k} w_i B_{i,\alpha_i,\beta_i}(F_i(y))\Big) \end{aligned}$$

cBLP - Part 1: For each component model, train over all observations to get $\alpha_i$ and $\beta_i$. Then, apply $\alpha_i$ and $\beta_i$ to beta-transform the CDF. This ends the component-wise part.

cBLP Part 2: Apply the usual BLP process on the beta-transformed component models to get the BLP ensemble.

## Componentwise Bias-Corrected & Componentwise Recalibrated BLP (cbcBLP)

This method corrects for bias of the component models and also recalibrate them using beta transform. Then the BLP method is used to generate the ensemble.

## Simulation studies

The data generating process for the observation $Y$ in the regression model is

$$Y = X_0 + a_1X_1 + a_2X_2 + a_3X_3 + \epsilon, \epsilon \sim (0, 1)$$

where $a_1, a_2$, and $a_3$ are real constants that vary across different simulation studies, and $X_0, X_1, X_2, X_3$, and $\epsilon$ are independent, standard normal random variables. The individual predictive densities have partial access of the above set of covariates. $f_1$ has access to only $X_0$ and $X_1$, $f_2$ has access to only $X_0$ and $X_2$, and $f_3$ has access to only $X_0$ and $X_3$. We want to combine $f_1, f_2$, and $f_3$ to predict $Y$. In this setup, $X_0$ represent shared information, while other covariates represent information unique to each individual model.

We estimate the pooling/combination formulas on a random sample $(f_{1i}, f_{2i}, f_{3i}, Y_i) : i = 1, ..., n$ of size $n = 50,000$ and evaluate on an independent test sample of the same size.

### Scenario 1: calibrated components (Baseline scenario).

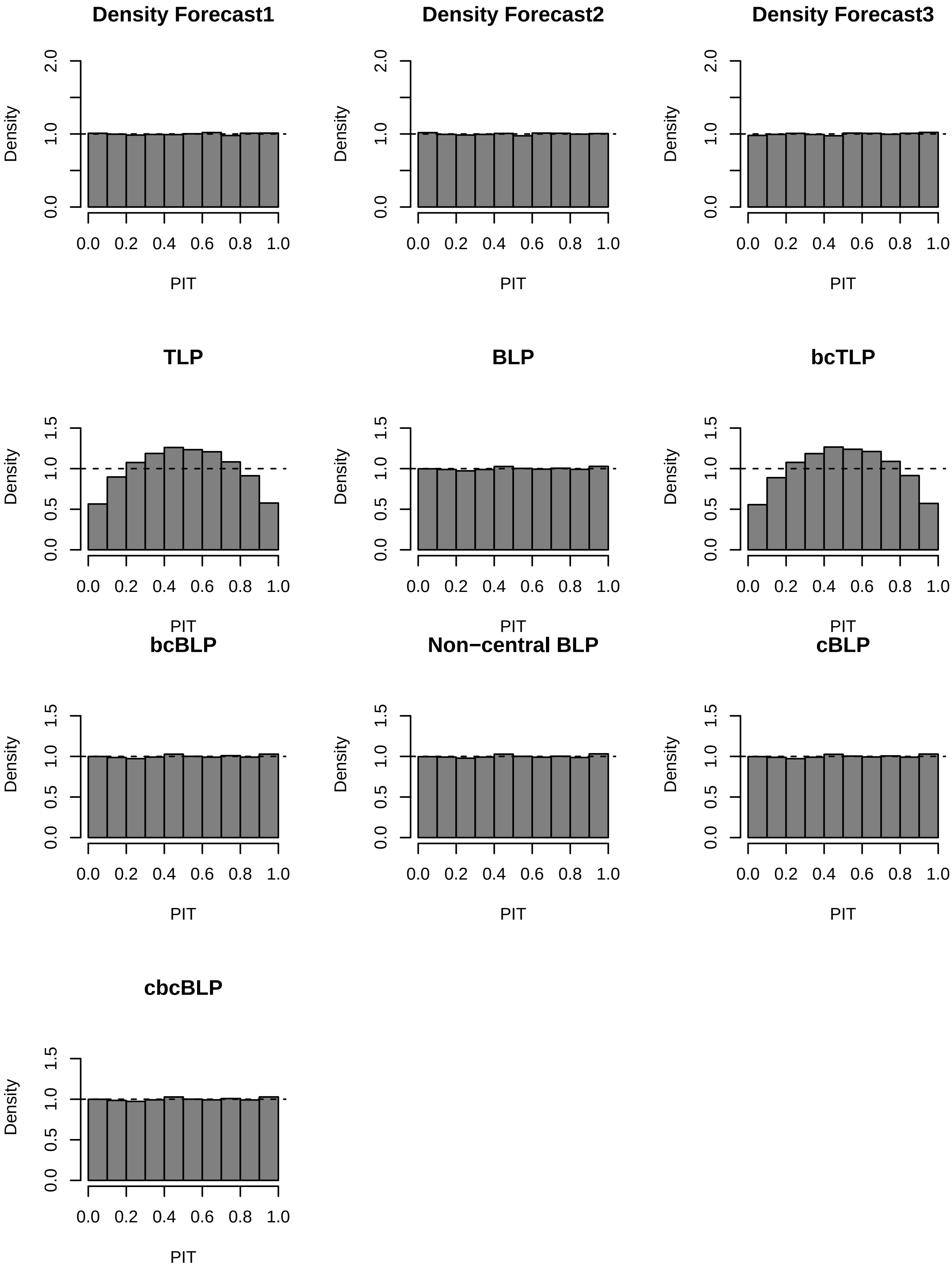
In this scenario, $a_1 = a_2 = 1$ and $a_3 = 1.1$, so that $f_3$ is a more concentrated, sharper density forecast than $f_1$ and $f_2$ (Gneiting and Ranjan (2013)) and they are defined as follows:

$$\begin{aligned} f_1 &= \mathrm{N}(X_0 + a_1X_1, 1 + a_2^2 + a_3^2) \\ f_2 &= \mathrm{N}(X_0 + a_2X_2, 1 + a_1^2 + a_3^2) \\ f_3 &= \mathrm{N}(X_0 + a_3X_3, 1 + a_1^2 + a_2^2) \end{aligned}$$

Table 1: Model Parameters and Log Score

| | w1 | w2 | w3 | alpha | beta | ncp |
|---|---|---|---|---|---|---|
| TLP | 0.271 | 0.264 | 0.465 | NA | NA | NA |
| BLP | 0.301 | 0.295 | 0.404 | 1.465 | 1.469 | NA |
| bcTLP | 0.272 | 0.264 | 0.464 | NA | NA | NA |
| bcBLP | 0.301 | 0.296 | 0.403 | 1.477 | 1.477 | NA |
| nBLP | 0.301 | 0.295 | 0.404 | 1.452 | 1.483 | 0.076 |
| cBLP | 0.300 | 0.296 | 0.405 | 1.459 | 1.460 | NA |
| cbcBLP | 0.301 | 0.296 | 0.403 | 1.469 | 1.469 | NA |

| | Training | Test |
|---|---|---|
| f1 | -1.998 | -2.004 |
| f2 | -1.999 | -2.006 |
| f3 | -1.965 | -1.969 |
| TLP | -1.907 | -1.912 |
| BLP | -1.864 | -1.872 |
| bcTLP | -1.906 | -1.911 |
| bcBLP | -1.861 | -1.870 |
| nBLP | -1.864 | -1.872 |
| cBLP | -1.864 | -1.872 |
| cbcBLP | -1.862 | -1.870 |


Density Forecast1
Density Forecast2
Density Forecast3
TLP
BLP
bcTLP
bcBLP
Non−central BLP
cBLP
cbcBLP
Density
PIT
0.0 0.2 0.4 0.6 0.8 1.0
0.0 1.0 2.0
0.0 0.5 1.0 1.5

## Scenario 2: Biased forecast scenario

In this scenario, $a_1 = a_2 = 1$ and $a_3 = 1.1$, and we add $N(2,1)$ to the mean of $f_1$ so that it is a biased forecast. The models are defined as follows:

$$f_1 = \mathrm{N}(X_0 + a_1X_1 + N(2,1), 1 + a_2^2 + a_3^2)$$
$$f_2 = \mathrm{N}(X_0 + a_2X_2, 1 + a_1^2 + a_3^2)$$
$$f_3 = \mathrm{N}(X_0 + a_3X_3, 1 + a_1^2 + a_2^2)$$

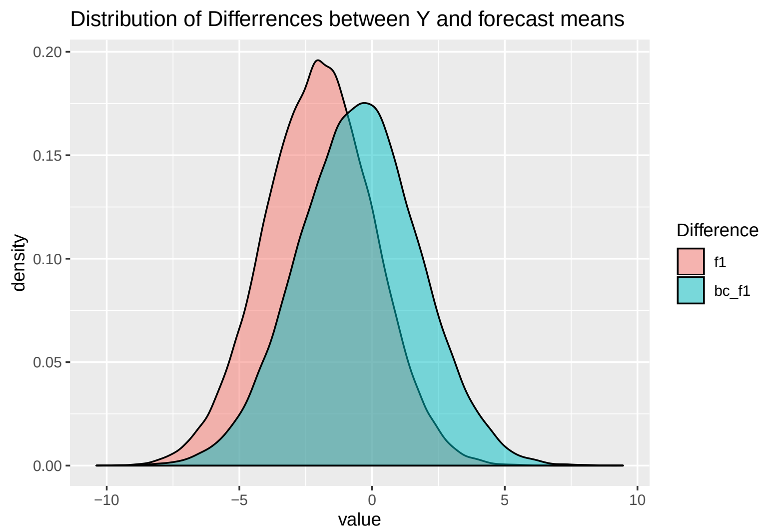
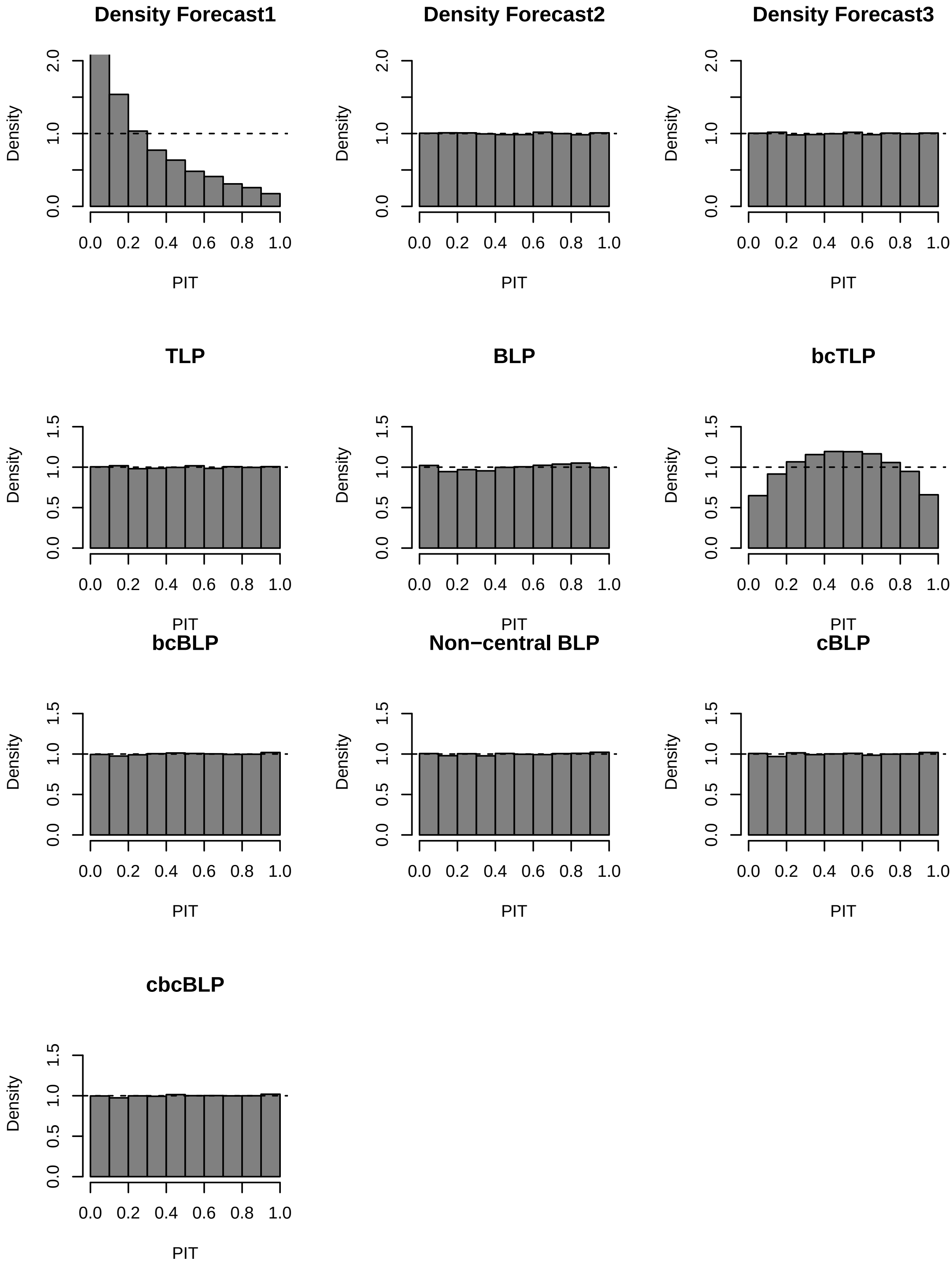



Table 2: Model Parameters and Log Score

| | w1 | w2 | w3 | alpha | beta | ncp |
|---|---|---|---|---|---|---|
| TLP | 0.000 | 0.000 | 1.000 | NA | NA | NA |
| BLP | 0.140 | 0.391 | 0.469 | 1.420 | 1.676 | NA |
| bcTLP | 0.005 | 0.417 | 0.579 | NA | NA | NA |
| bcBLP | 0.088 | 0.411 | 0.500 | 1.372 | 1.373 | NA |
| nBLP | 0.138 | 0.390 | 0.472 | 1.309 | 1.822 | 0.653 |
| cBLP | 0.183 | 0.381 | 0.436 | 1.510 | 1.518 | NA |
| cbcBLP | 0.133 | 0.386 | 0.481 | 1.398 | 1.399 | NA |

| | Training | Test |
|---|---|---|
| f1 | -2.788 | -2.777 |
| f2 | -1.998 | -2.004 |
| f3 | -1.965 | -1.969 |
| TLP | -1.965 | -1.969 |
| BLP | -1.884 | -1.888 |
| bcTLP | -1.917 | -1.921 |
| bcBLP | -1.888 | -1.893 |
| nBLP | -1.883 | -1.888 |
| cBLP | -1.967 | -1.971 |
| cbcBLP | -1.895 | -1.900 |


Density Forecast1
Density Forecast2
Density Forecast3
TLP
BLP
bcTLP
bcBLP
Non−central BLP
cBLP
cbcBLP
Density
PIT
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.5 1.0 1.5
0.0 1.0 2.0

## Scenario 3 : Higher variance forecast scenario

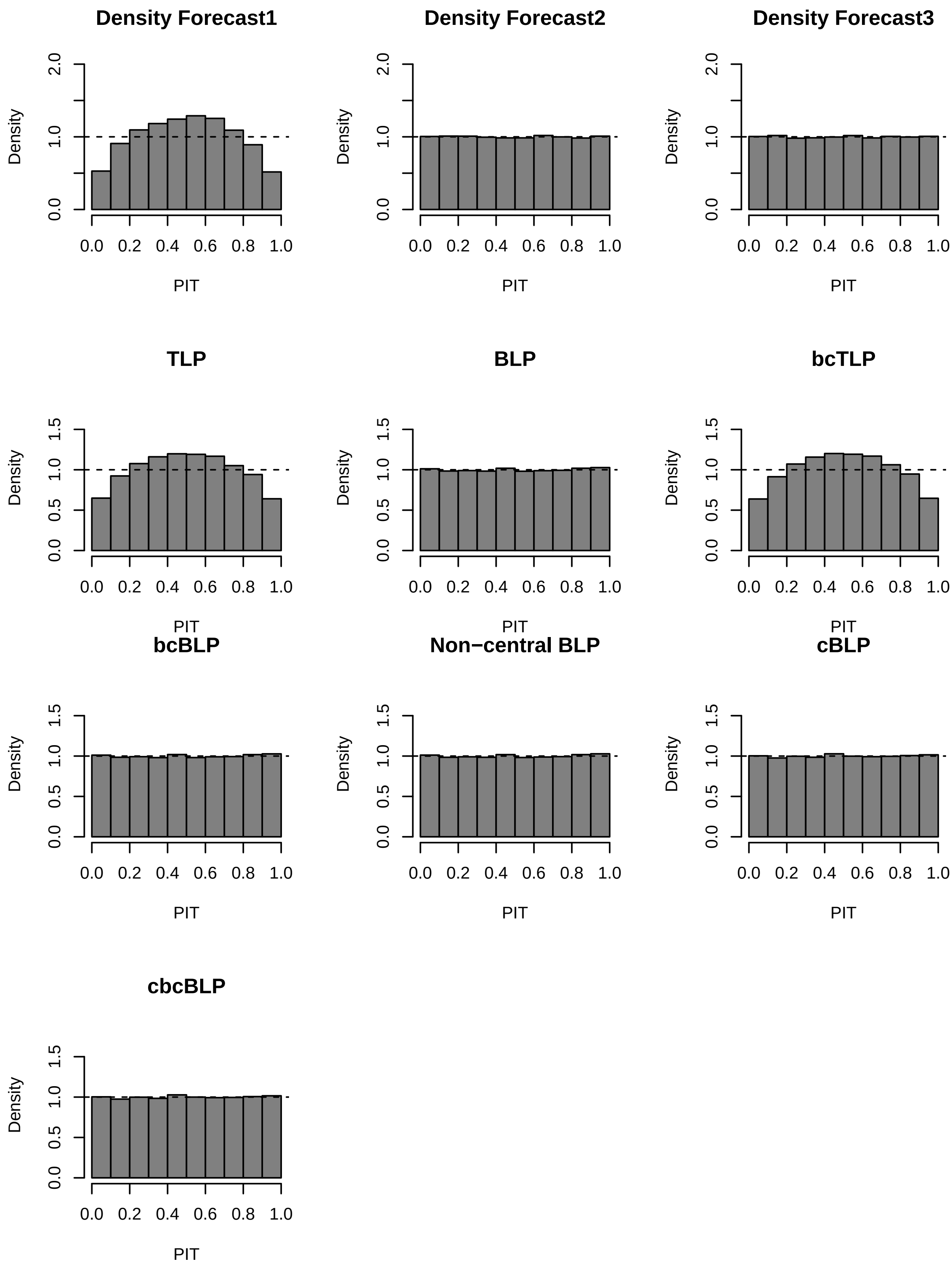
We modified the standard deviation of the first density forecast by adding a constant of 2 as follows:

$$f_1 = \mathrm{N}(X_0 + a_1 X_1, 1 + a_2^2 + a_3^2 + 2)$$
$$f_2 = \mathrm{N}(X_0 + a_2 X_2, 1 + a_1^2 + a_3^2)$$
$$f_3 = \mathrm{N}(X_0 + a_3 X_3, 1 + a_1^2 + a_2^2)$$

Table 3: Model Parameters and Log Score

| | w1 | w2 | w3 | alpha | beta | ncp |
|---|---|---|---|---|---|---|
| TLP | 0.013 | 0.411 | 0.576 | NA | NA | NA |
| BLP | 0.327 | 0.288 | 0.385 | 1.658 | 1.668 | NA |
| bcTLP | 0.013 | 0.412 | 0.575 | NA | NA | NA |
| bcBLP | 0.327 | 0.289 | 0.384 | 1.668 | 1.668 | NA |
| nBLP | 0.327 | 0.288 | 0.385 | 1.654 | 1.672 | 0.02 |
| cBLP | 0.204 | 0.357 | 0.439 | 1.443 | 1.444 | NA |
| cbcBLP | 0.204 | 0.357 | 0.439 | 1.448 | 1.448 | NA |

| | Training | Test |
|---|---|---|
| f1 | -2.051 | -2.053 |
| f2 | -1.998 | -2.004 |
| f3 | -1.965 | -1.969 |
| TLP | -1.918 | -1.921 |
| BLP | -1.864 | -1.869 |
| bcTLP | -1.917 | -1.921 |
| bcBLP | -1.863 | -1.868 |
| nBLP | -1.864 | -1.869 |
| cBLP | -1.866 | -1.871 |
| cbcBLP | -1.865 | -1.870 |


Density Forecast1
Density Forecast2
Density Forecast3
TLP
BLP
bcTLP
bcBLP
Non−central BLP
cBLP
cbcBLP
Density
PIT
0.0 0.2 0.4 0.6 0.8 1.0
0.0 1.0 2.0
0.0 0.5 1.0 1.5

## Scenario 4 Biased + higher variance forecast scenario

In this scenario, $a_1 = a_2 = 1$ and $a_3 = 1.1$, but we modified the standard deviation of the first and second density forecast as follows:

$$f_1 = \mathrm{N}(X_0 + a_1 X_1 + N(2,1), 1 + a_2^2 + a_3^2)$$
$$f_2 = \mathrm{N}(X_0 + a_2 X_2, 1 + a_1^2 + a_3^2 + 2)$$
$$f_3 = \mathrm{N}(X_0 + a_3 X_3, 1 + a_1^2 + a_2^2)$$

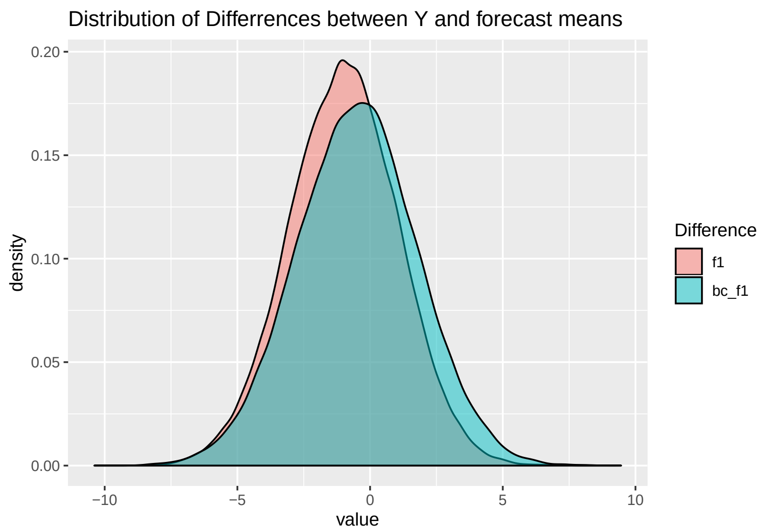
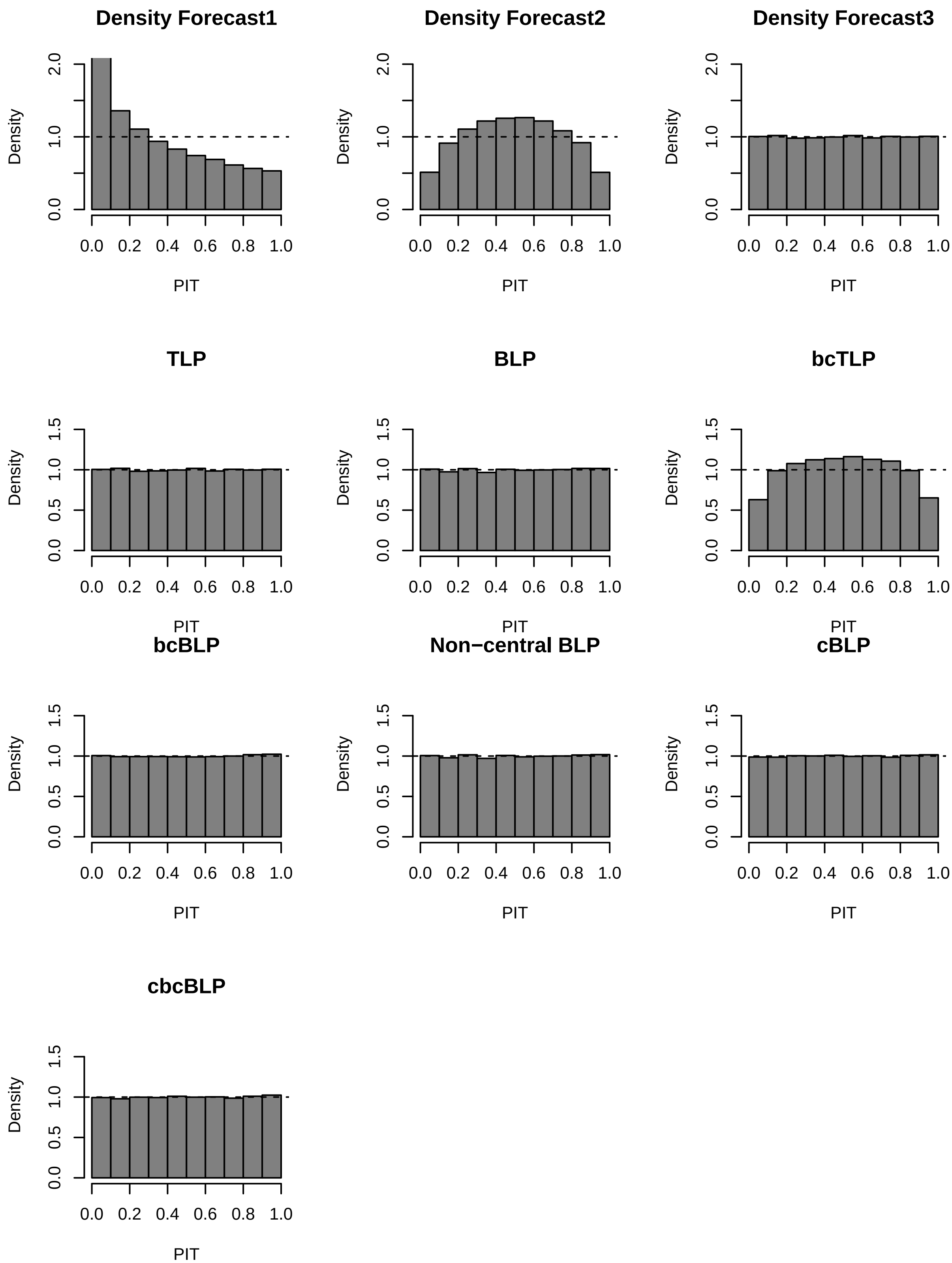



Table 4: Model Parameters and Log Score

| | w1 | w2 | w3 | alpha | beta | ncp |
|---|---|---|---|---|---|---|
| TLP | 0.000 | 0.000 | 1.000 | NA | NA | NA |
| BLP | 0.161 | 0.414 | 0.425 | 1.660 | 1.832 | NA |
| bcTLP | 0.130 | 0.150 | 0.720 | NA | NA | NA |
| bcBLP | 0.085 | 0.453 | 0.461 | 1.639 | 1.639 | NA |
| nBLP | 0.160 | 0.414 | 0.426 | 1.648 | 1.846 | 0.068 |
| cBLP | 0.249 | 0.264 | 0.487 | 1.524 | 1.528 | NA |
| cbcBLP | 0.205 | 0.281 | 0.514 | 1.401 | 1.401 | NA |

| | Training | Test |
|---|---|---|
| f1 | -2.317 | -2.310 |
| f2 | -2.050 | -2.054 |
| f3 | -1.965 | -1.969 |
| TLP | -1.965 | -1.969 |
| BLP | -1.878 | -1.882 |
| bcTLP | -1.942 | -1.945 |
| bcBLP | -1.888 | -1.893 |
| nBLP | -1.878 | -1.882 |
| cBLP | -1.915 | -1.920 |
| cbcBLP | -1.905 | -1.910 |


Density Forecast1
Density Forecast2
Density Forecast3
TLP
BLP
bcTLP
bcBLP
Non−central BLP
cBLP
cbcBLP
Density
PIT
0.0 0.5 1.0 1.5 2.0
0.0 0.2 0.4 0.6 0.8 1.0