Comparison of Ensemble Recalibration Methods in Flu Forecasting

Nutcha Wattanachit

03/17/2020

We compare 1) the equally-weighted ensemble, 2) the traditional linear pool (TLP), 3) the beta-transform linear pool (BLP), 4) the equally-weighted beta-transform linear pool, 5) the finite beta mixture 6) the finite beta mixture with equally-weighted component forecasts in the simulation studies and in the application of influenza forecasting. For both beta mixture approaches, the number of mixing beta components are $K = 2, 3$, and 4.

Methods

Let $f_1, ..., f_M$ be predictive density forecasts from M component forecasting models, the ensemble methods combine the component forecasting models as follows

Equally-weighted ensemble (EW)

The equally-weighted ensemble combines the component forecasting models with the aggregation predictive distribution function

$$f_{\text{EW}}(y) = \sum_{m=1}^{M} \frac{1}{M} f_m(y). \tag{1}$$

Traditional linear pool (TLP)

The TLP finds a set of optimal nonnegative weights w_i that maximize the likelihood of the aggregation predictive distribution function

$$f_{\text{TLP}}(y) = \sum_{m=1}^{M} w_m f_m(y), \tag{2}$$

where $\sum_{m=1}^{M} w_m = 1$. The TLP is underdispersed when the component models are probabilistically calibrated.

Beta-transform linear pool (BLP)

The BLP applies a beta transform on the combined predictive cumulative distribution function

$$F_{\text{BLP}}(y) = B_{\alpha,\beta}\Big(\sum_{m=1}^{M} w_m F_m(y)\Big), \tag{3}$$

Specifically, the BLP finds the transformation parameters $\alpha, \beta > 0$, and a set of nonnegative weights w_m that maximize the likelihood of the aggregated predictive distribution function

$$f_{\text{BLP}}(y) = \Big(\sum_{m=1}^{M} w_m f_m(y)\Big) b_{\alpha,\beta}\Big(\sum_{m=1}^{M} w_m F_m(y)\Big), \tag{4}$$

where $b_{\alpha,\beta}$ denotes the beta density and $\sum_{m=1}^{M} w_m = 1$.

Equally-weighted beta-transform linear pool (EW-BLP)

The EW-BLP applies a beta transform on the equally-weighted ensemble and has the predictive cumulative distribution function

$$F_{\text{EW-BLP}}(y) = B_{\alpha,\beta}\Big(\sum_{m=1}^{M} \frac{1}{M} F_m(y)\Big), \tag{5}$$

The EW-BLP finds the transformation parameters $\alpha, \beta > 0$ that maximize the likelihood of the aggregated predictive distribution function

$$f_{\text{EW-BLP}}(y) = \Big(\sum_{m=1}^{M} w_m f_m(y)\Big) b_{\alpha,\beta}\Big(\sum_{m=1}^{M} \frac{1}{M} F_m(y)\Big). \tag{6}$$

Finite beta mixture (BM$_k$)

The BM$_k$ extends the BLP method by using a finite beta mixture combination formula

$$F_{\text{BM}_k}(y) = \sum_{k=1}^{K} w_k B_{\alpha,\beta}\Big(\sum_{m=1}^{M} \omega_{km} F_m(y)\Big), \tag{7}$$

where the vector $w_1, ..., w_K$ comprises the beta mixture weights, $\alpha_1, ..., \alpha_K$ and $\beta_1, ..., \beta_K$ are beta calibration parameters, and for each beta component $\omega_k = (\omega_{k1}, ..., \omega_{kM})$ comprises the beta component-specific set of component model weights. The pdf representation of the method is

$$f_{\text{BM}_k}(y) = \sum_{k=1}^{K} w_k (\sum_{m=1}^{M} \omega_{km} f_m(y)\Big) b_{\alpha,\beta}\Big(\sum_{m=1}^{M} \omega_{km} F_m(y)\Big). \tag{8}$$

Finite beta mixture with equally weighted ensemble (EW-BM$_k$)

The EW-BM$_k$ uses a finite beta mixture combination formula to combine an equally-weighted ensemble as follows

$$F_{\text{EW-BM}_k}(y) = \sum_{k=1}^{K} w_k B_{\alpha,\beta}\Big(\sum_{m=1}^{M} \frac{1}{M} F_m(y)\Big), \tag{9}$$

where the vector $w_1, ..., w_K$ comprises the beta mixture weights and $\alpha_1, ..., \alpha_K$ and $\beta_1, ..., \beta_K$ are beta calibration parameters.

$$f_{\text{EW-BM}_k}(y) = \sum_{k=1}^{K} w_k (\sum_{m=1}^{M} \frac{1}{M} f_m(y)\Big) b_{\alpha,\beta}\Big(\sum_{m=1}^{M} \frac{1}{M} F_m(y)\Big). \tag{10}$$

Simulation studies

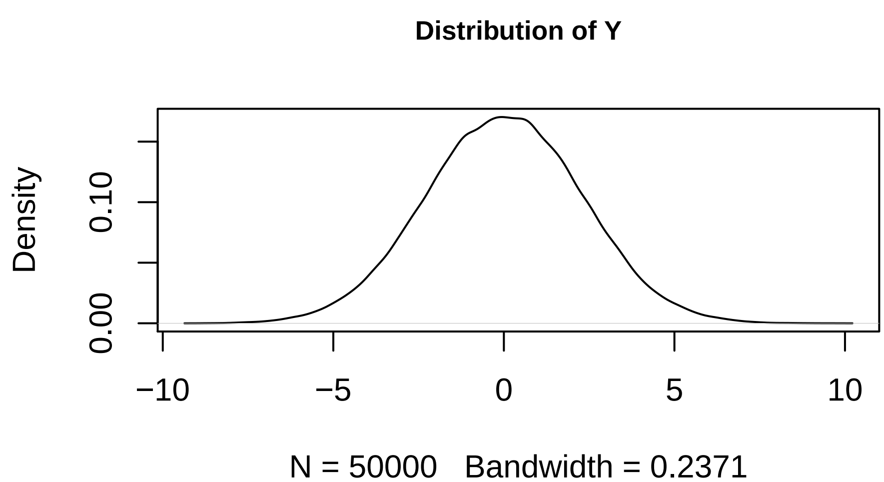
We investigate the out-of-sample performance of the aforementioned combination formulae in three simulation scenarios. For the mixture methods, we use 5-fold cross-validation to select the number of beta components and then implement the mixture methods with their corresponding selected number of beta components.

Scenario 1: Unbiased and calibrated components

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

$$Y = X_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \epsilon, \epsilon \sim N(0, 1)$$

where $a_1 = 1, a_2 = 1$, and $a_3 = 1.1$, and X_0, X_1, X_2, X_3, and ϵ are independent, standard normal random variables. The TLP's PITs are approximately beta distributed (underdispersed inverted U-shape) in this scenario, so BLP should be able to find optimal α and β to adjust the PITs. Specifically, this scenario serves to demonstrate the shortcoming of TLP and to motivate BLP. We expect BMC to do as well as BLP as it is more flexible (and thus has higher complexity), but BMC is not necessary.



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 training data set $(f_{1i}, f_{2i}, f_{3i}, Y_i) : i = 1, ..., n$ and evaluate on an independent test set. 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:

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

Table 1: Cross validation log scores for beta mixture methods

method	mean_train_ls	mean_valid_ls
BMC2	-1.870918	-1.871193
EW_BMC2	-1.872509	-1.872641
BMC3	-1.870656	-1.871127
EW_BMC3	-1.872510	-1.872640
BMC4	-1.870658	-1.871202
EW_BMC4	-1.872499	-1.872656
BMC5	-1.870057	-1.870685
EW_BMC5	-1.872488	-1.872656

Table 2: Weight Parameters

Method	w_1	w_2	w_3	w_4	w_5
TLP	NA	NA	NA	NA	NA
BLP	NA	NA	NA	NA	NA
EW	NA	NA	NA	NA	NA
EW-BLP	NA	NA	NA	NA	NA
BMC5	0.046	0.040	0.818	0.051	0.046
EW-BMC2	0.411	0.589	NA	NA	NA

Table 3: Beta mixture parameters

Method	α_1	β_1	α_2	β_2	α_3	β_3	α_4	β_4	α_5	β_5
TLP	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BLP	1.450	1.448	NA	NA	NA	NA	NA	NA	NA	NA
EW	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	1.452	1.450	NA	NA	NA	NA	NA	NA	NA	NA
BMC5	1.281	7.350	8.013	1.324	1.964	1.957	6.973	1.448	1.64	8.477
EW-BMC2	1.267	1.257	1.623	1.628	NA	NA	NA	NA	NA	NA

Table 4: Component weight parameters -

Method	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}	ω_{31}	ω_{32}	ω_{33}	ω_{41}	ω_{42}	ω_{43}	ω_{51}	ω_{52}	ω_{53}
TLP	0.275	0.267	0.458	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BLP	0.302	0.296	0.401	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BMC5	0.004	0.000	0.995	0.002	0.002	0.996	0.304	0.298	0.398	0.506	0.492	0.002	0.474	0.524	0.002
EW-BMC2	0.333	0.333	0.333	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 5: Log score

	TLP	BLP	EW	EW-BLP	BMC5	EW-BMC2
Training	-1.912	-1.871	-1.914	-1.873	-1.869	-1.873
Test	-1.912	-1.871	-1.914	-1.873	-1.870	-1.873

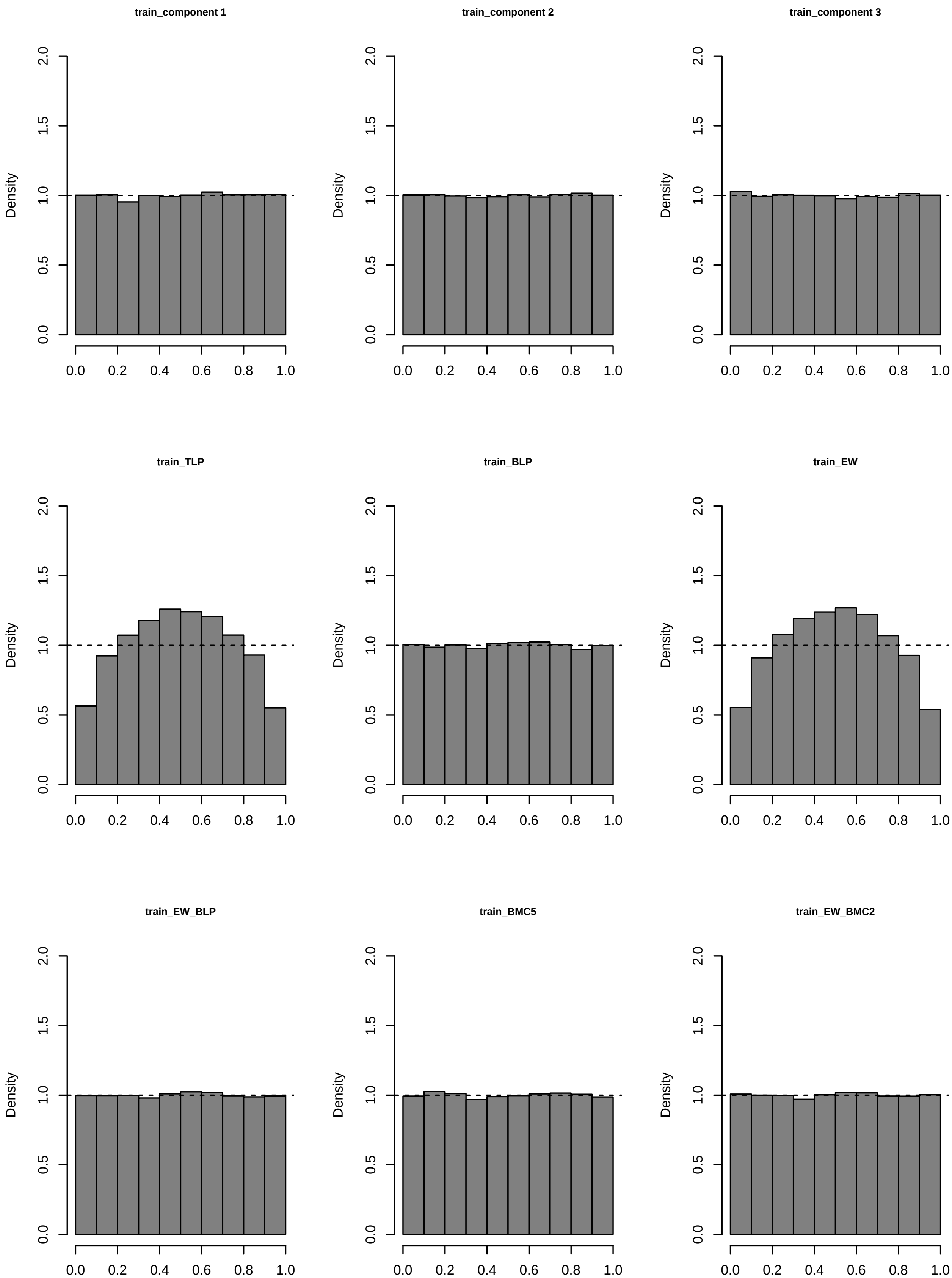


Figure 1: Train PITs

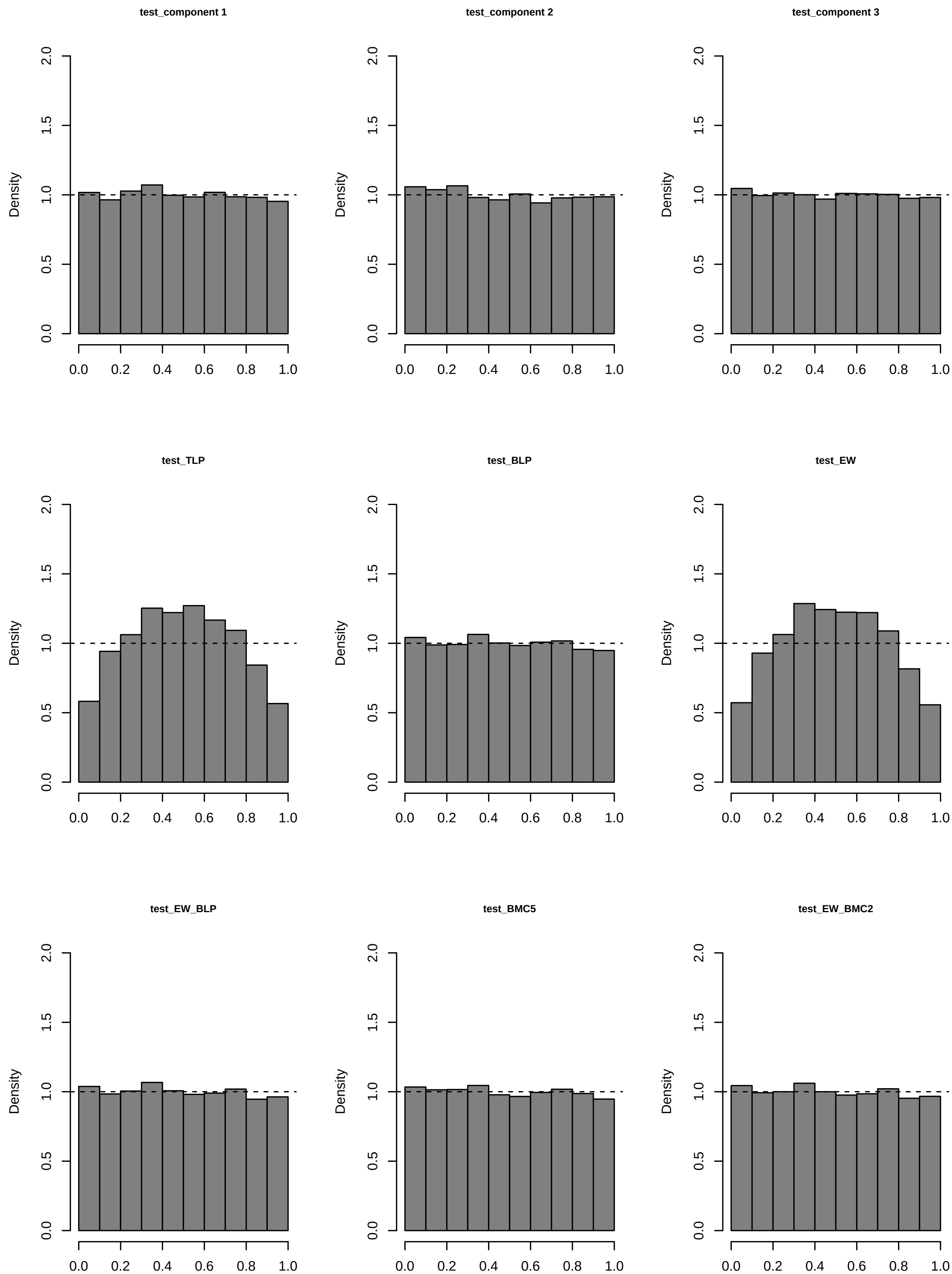


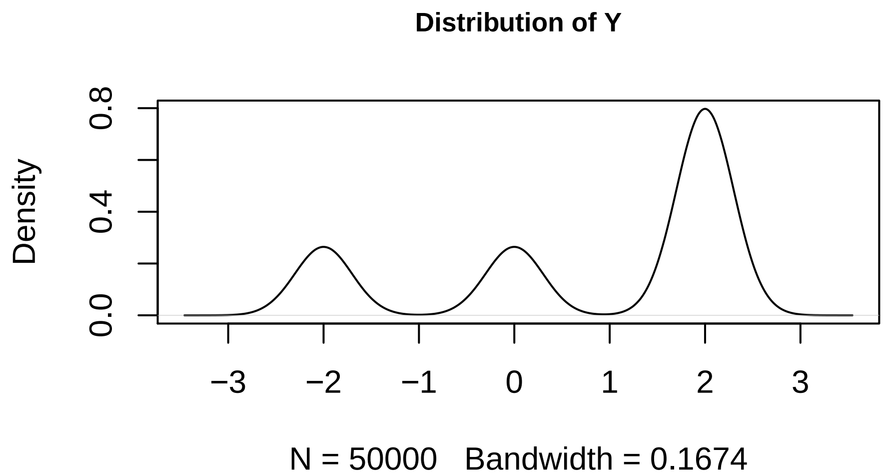
Figure 2: Test PITs

Scenario 2: Multimodal DGP (Normal mixture) and close-$\mathcal{M}$

The data generating process for the observation y_t is

$$y_t \overset{i.i.d.}{\sim} p_1\mathrm{N}(-2, 0.25) + p_2\mathrm{N}(0, 0.25) + p_3\mathrm{N}(2, 0.25), t = 1, ..., 100, 0000$$

where $p_1 = 0.2, p_2 = 0.2$, and $p_3 = 0.6$. In this scenario, the three component models are in the data generating process and the TLP's PITs are approximately beta distributed (uniformly distributed, specifically). This scenario serves to show the situation in which TLP is an optimal method of combining forecast distributions. We expect BLP and BMC to perform as equally well as TLP with higher complexity. In other words, this is when BLP and BMC are not needed.



The individual predictive densities are defined as follows:

$$f_1 \overset{i.i.d.}{\sim} \mathrm{N}(-2, 0.25)$$
$$f_2 \overset{i.i.d.}{\sim} \mathrm{N}(0, 0.25)$$
$$f_3 \overset{i.i.d.}{\sim} \mathrm{N}(2, 0.25)$$

Table 6: Cross validation log scores for beta mixture methods

method	mean_train_ls	mean_valid_ls
BMC2	-0.9812800	-0.9815715
EW_BMC2	-1.0016930	-1.0018338
BMC3	-0.9812788	-0.9815771
EW_BMC3	-0.9970163	-0.9969911
BMC4	-0.9812728	-0.9815954
EW_BMC4	-0.9966951	-0.9972928
BMC5	-0.9812703	-0.9815867
EW_BMC5	-0.9875902	-0.9875473

Table 7: Weight Parameters

Method	w_1	w_2	w_3	w_4	w_5
TLP	NA	NA	NA	NA	NA
BLP	NA	NA	NA	NA	NA
EW	NA	NA	NA	NA	NA
EW-BLP	NA	NA	NA	NA	NA
BMC2	0.821	0.179	NA	NA	NA
EW-BMC5	0.317	0.014	0.229	0.221	0.219

Table 8: Beta mixture parameters

Method	α_1	β_1	α_2	β_2	α_3	β_3	α_4	β_4	α_5	β_5
TLP	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BLP	1.000	1.003	NA	NA	NA	NA	NA	NA	NA	NA
EW	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	1.256	0.789	NA	NA	NA	NA	NA	NA	NA	NA
BMC2	1.010	1.040	0.948	0.851	NA	NA	NA	NA	NA	NA
EW-BMC5	11.113	2.811	0.912	0.744	0.942	0.745	0.944	0.745	0.943	0.745

Table 9: Component weight parameters -

Method	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}	ω_{31}	ω_{32}	ω_{33}	ω_{41}	ω_{42}	ω_{43}	ω_{51}	ω_{52}	ω_{53}
TLP	0.198	0.200	0.602	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BLP	0.197	0.200	0.603	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BMC2	0.212	0.166	0.622	0.121	0.369	0.510	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BMC5	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333

Table 10: Log score

	TLP	BLP	EW	EW-BLP	BMC2	EW-BMC5
Training	-0.981	-0.981	-1.132	-1.043	-0.981	-0.999
Test	-0.991	-0.991	-1.139	-1.053	-0.991	-1.011

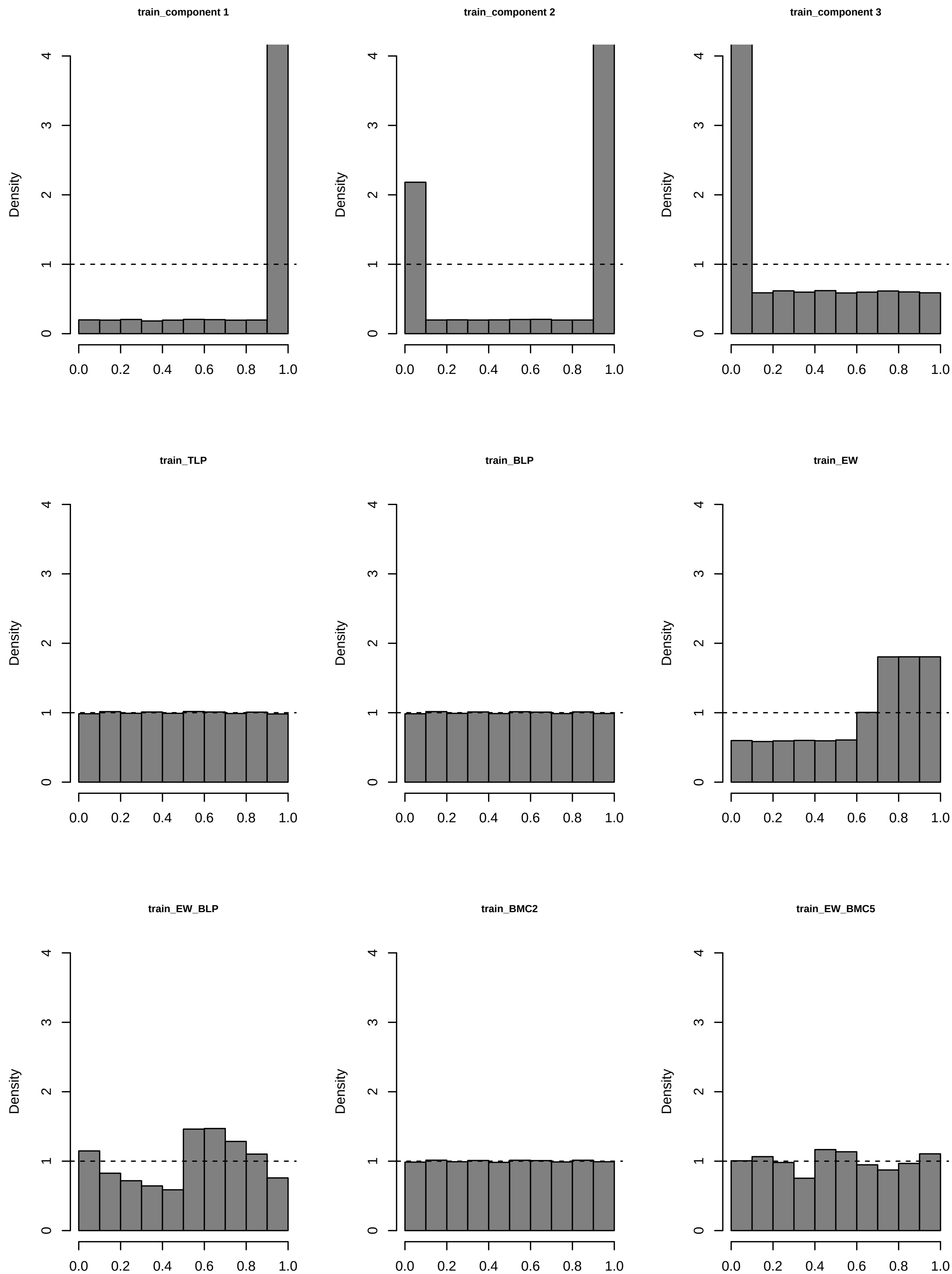


Figure 3: Train PITs

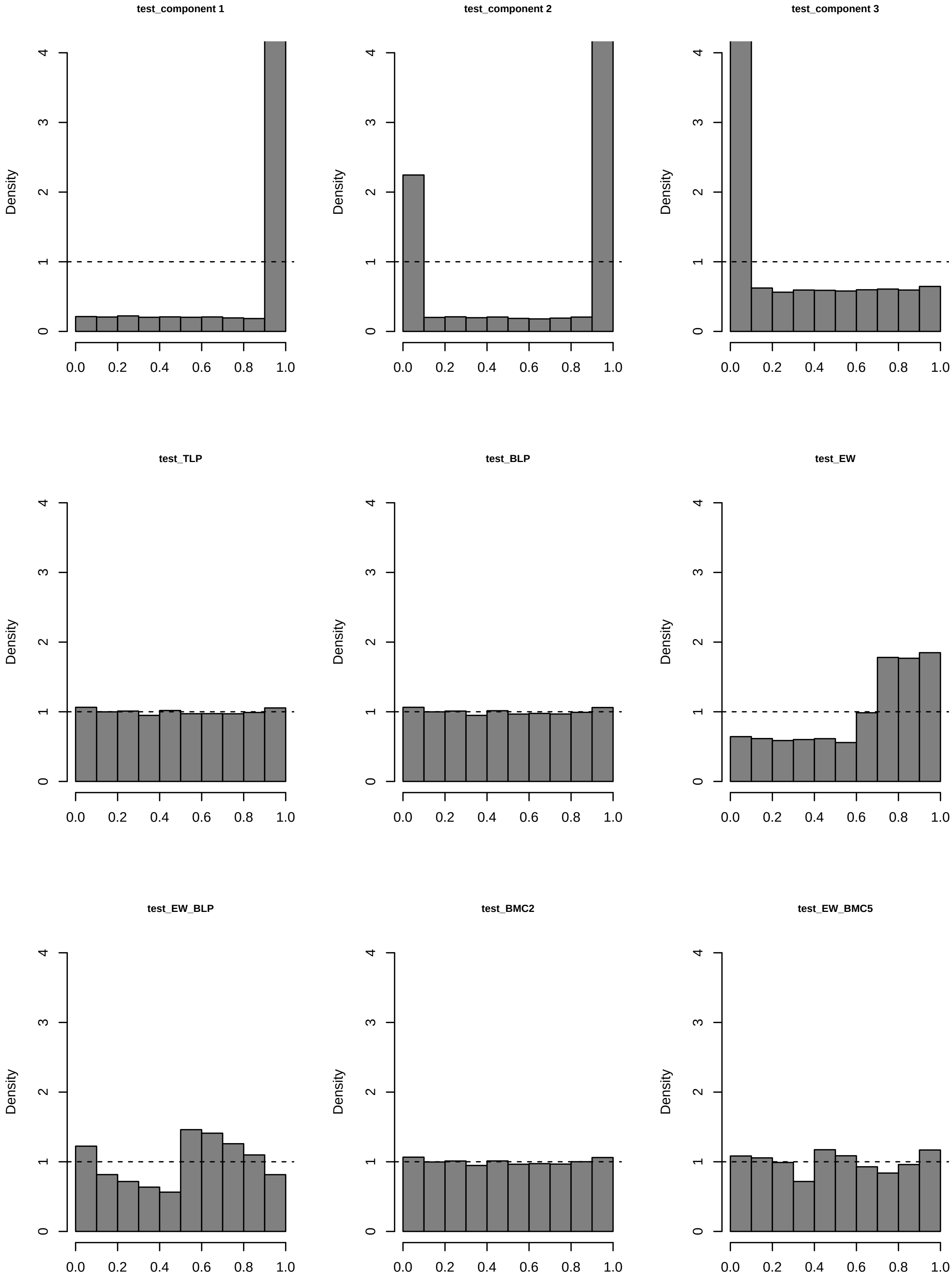
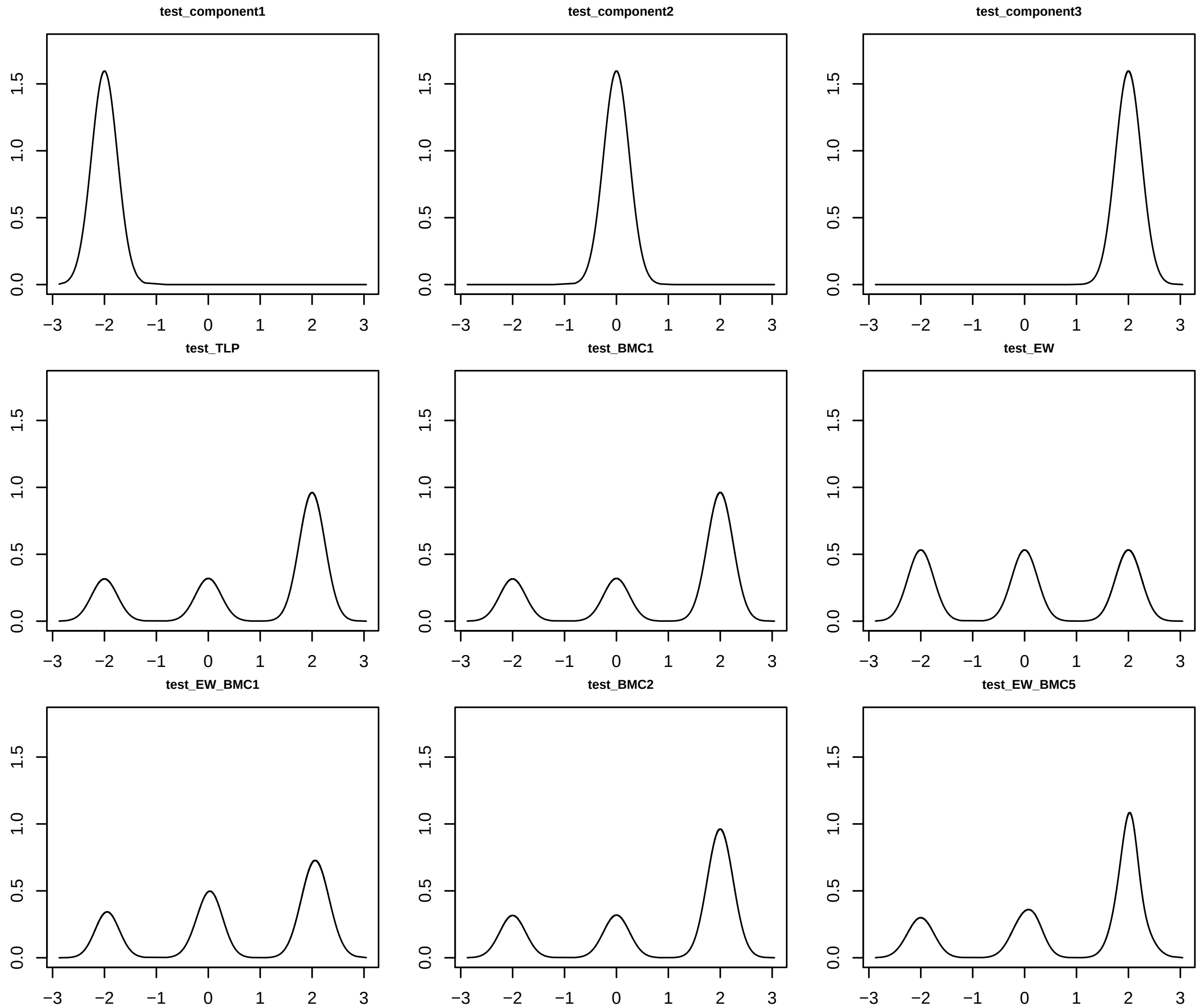


Figure 4: Test PITs


test_component1
test_component2
test_component3
test_TLP
test_BMC1
test_EW
test_EW_BMC1
test_BMC2
test_EW_BMC5
0.0
0.5
1.0
1.5
−3
−2
−1
0
1
2
3

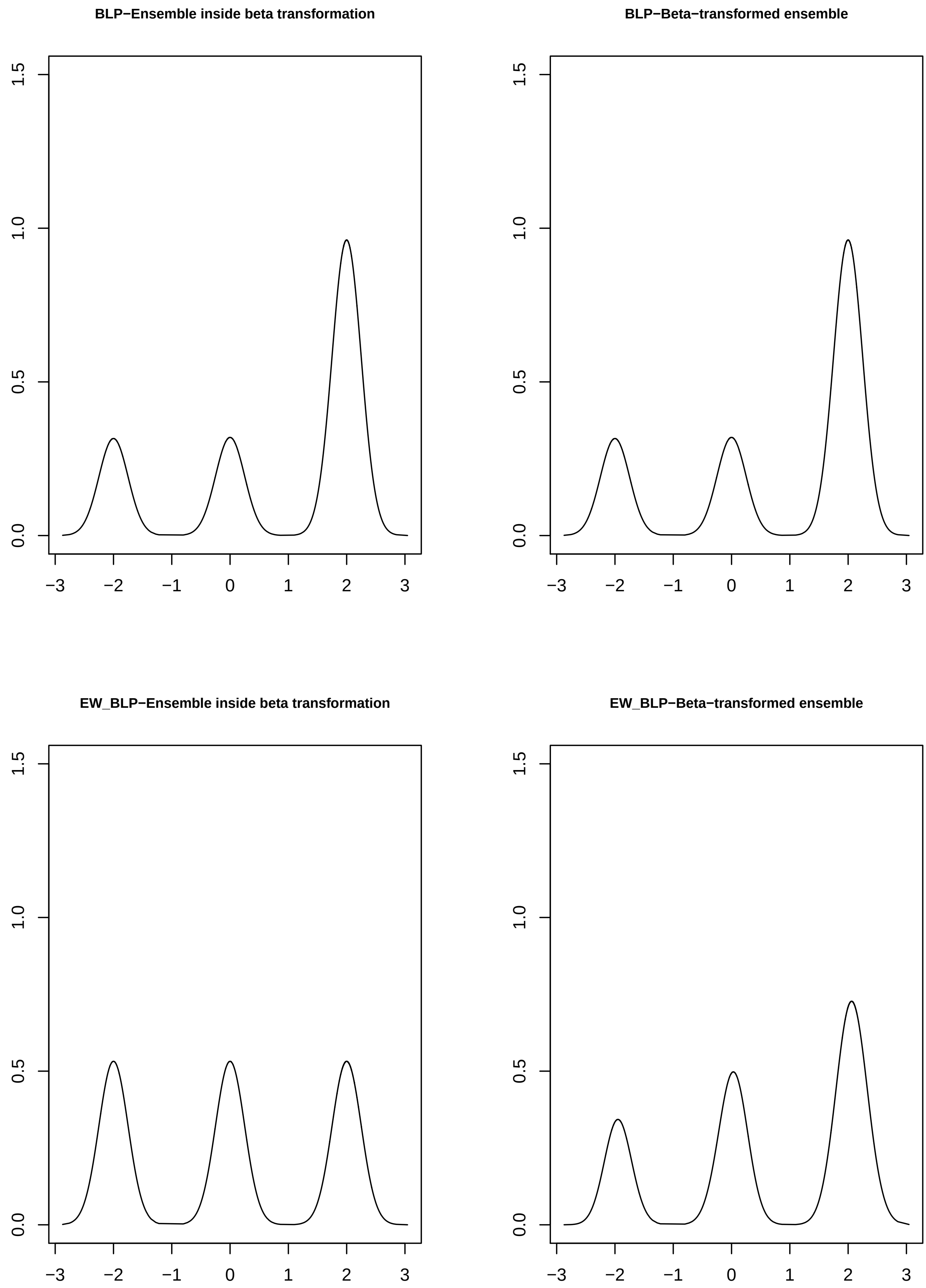


Figure 5: Ensemble details

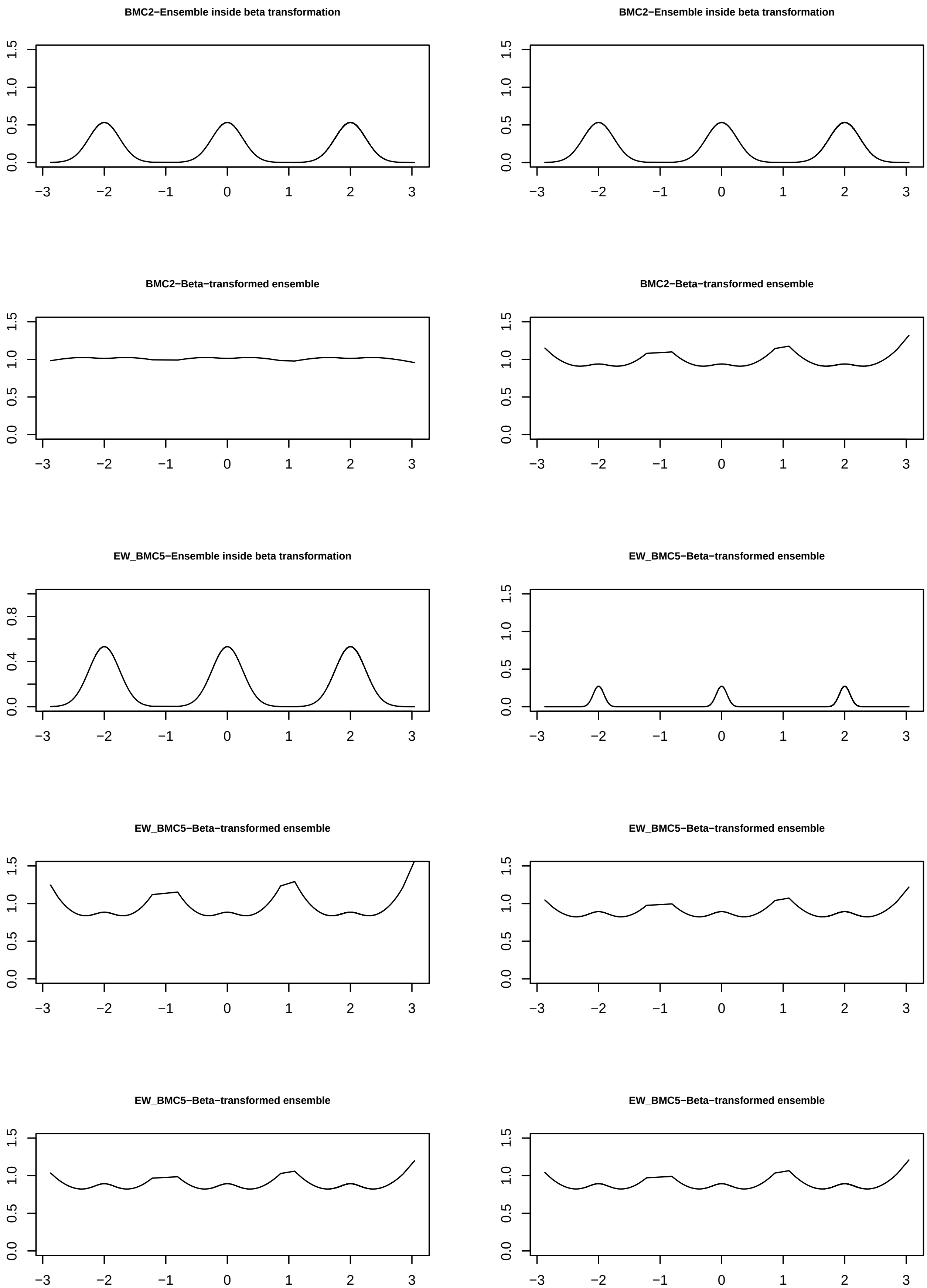


Figure 6: Ensemble details

Scenario 2.1: Multimodal DGP (Normal mixture) with empirical distributions

Using the same data generating process as the continuous distribution version, but with $t = 2000$ so it computes faster. Since we calculate the binned probability exactly from normal distributions, binned probabilities (empirical pdfs) are exactly the same for all t in each component model.

Table 11: Cross validation log scores for beta mixture methods

x
BMC3
EW_BMC4

method	mean_train_ls	mean_valid_ls
BMC2	-3.792627	-3.747251
EW_BMC2	-3.818782	-3.772621
BMC3	-3.792136	-3.746503
EW_BMC3	-3.813448	-3.769789
BMC4	-3.789741	-3.746363
EW_BMC4	-3.809776	-3.766783
BMC5	-3.789478	-3.747252
EW_BMC5	-3.809634	-3.767397

Table 12: Weight Parameters

Method	w_1	w_2	w_3	w_4
TLP	NA	NA	NA	NA
BLP	NA	NA	NA	NA
EW	NA	NA	NA	NA
EW-BLP	NA	NA	NA	NA
BMC3	0.141	0.276	0.583	NA
EW-BMC4	0.075	0.080	0.615	0.23

Table 13: Beta mixture parameters

Method	α_1	β_1	α_2	β_2	α_3	β_3	α_4	β_4
TLP	NA	NA	NA	NA	NA	NA	NA	NA
BLP	0.506	2.008	NA	NA	NA	NA	NA	NA
EW	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	0.607	2.035	NA	NA	NA	NA	NA	NA
BMC3	0.083	13.274	0.725	3.993	0.487	5.330	NA	NA
EW-BMC4	0.968	64.938	0.082	14.852	0.773	9.565	0.315	11.957

Table 14: Component weight parameters -

Method	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}	ω_{31}	ω_{32}	ω_{33}	ω_{41}	ω_{42}	ω_{43}
TLP	0.209	0.199	0.592	NA	NA	NA	NA	NA	NA	NA	NA	NA
BLP	0.216	0.200	0.584	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA
BMC3	0.157	0.809	0.034	0.001	0.001	0.998	0.267	0.196	0.537	NA	NA	NA
EW-BMC4	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333

Table 15: Log score

	TLP	BLP	EW	EW-BLP	BMC3	EW-BMC4
Training	-3.794	-3.794	-3.934	-3.844	-3.791	-3.817
Test	-3.709	-3.710	-3.879	-3.796	-3.705	-3.725

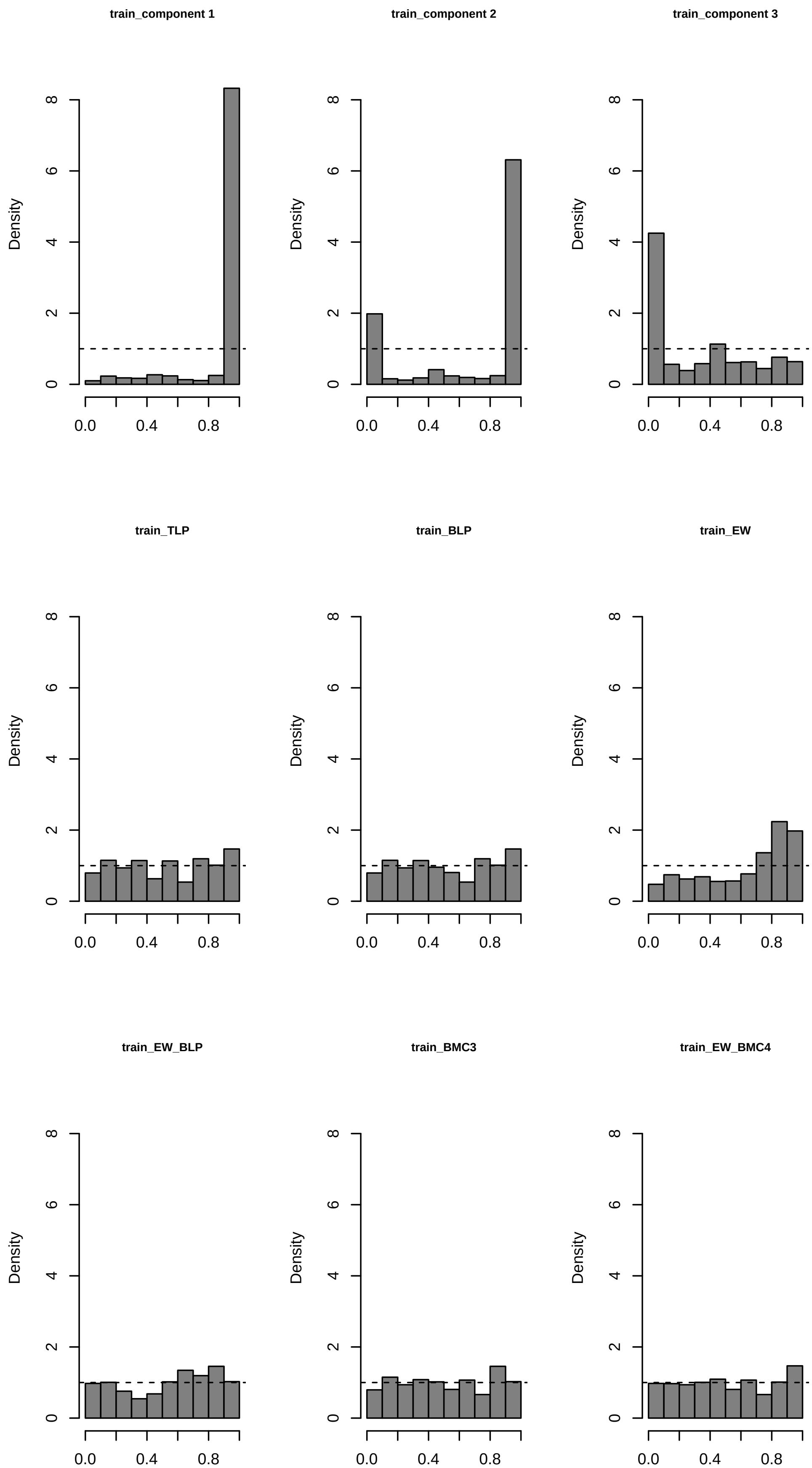


Figure 7: Train PITs

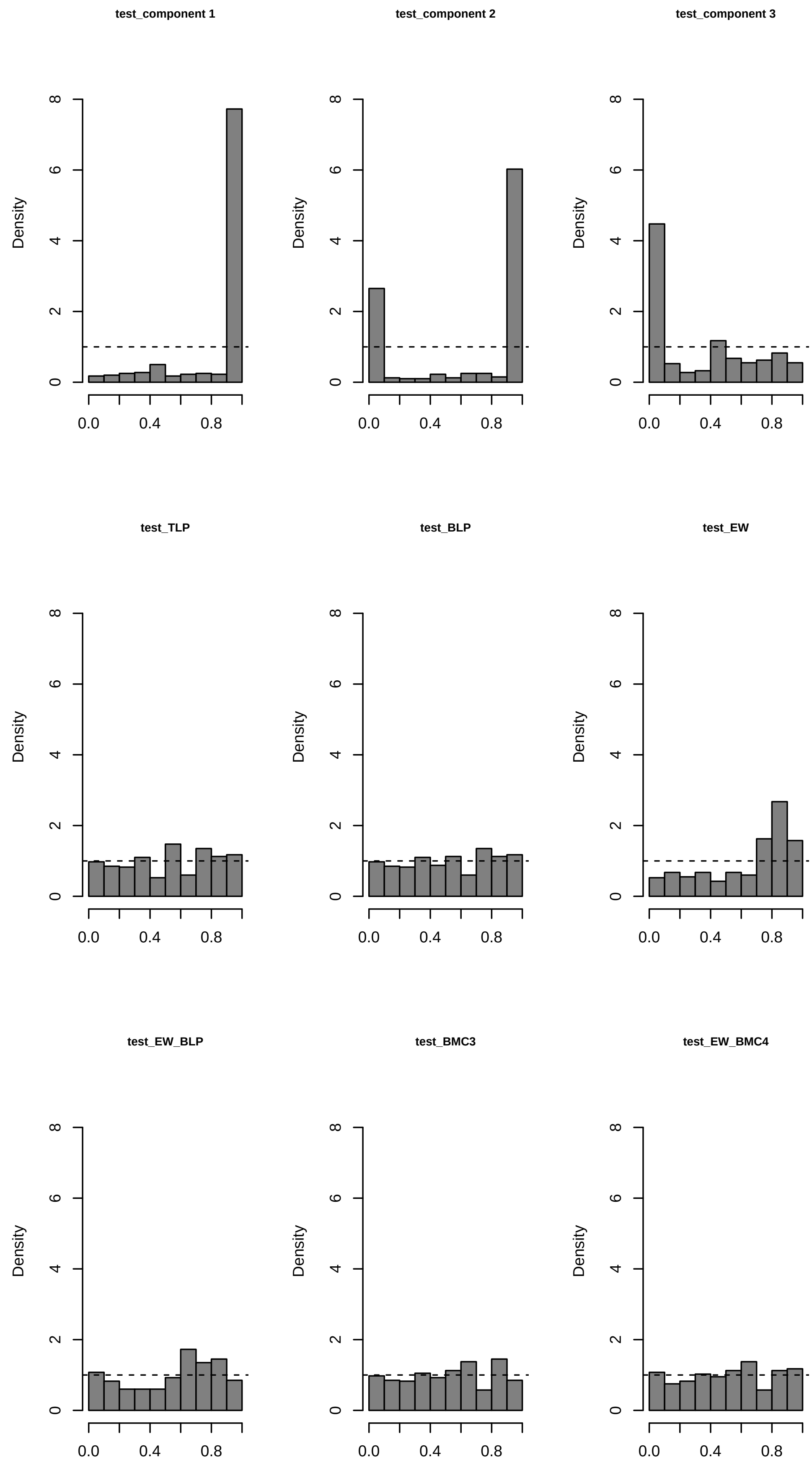


Figure 8: Test PITs

Scenario 3: Misspecified Normal mixture

The data generating process for the observations in this scenario is the same as in Scenario 2. There are two component models defined as follows

$$f_1 \overset{i.i.d.}{\sim} \mathrm{N}(1.5, 1)$$
$$f_2 \overset{i.i.d.}{\sim} \mathrm{N}(0.5, 1).$$
$$f_3 \overset{i.i.d.}{\sim} \mathrm{N}(-2, 1).$$

The component models are not part of the data generating process. In this scenario the TLP's PITs are not approximately beta distributed, so we expect BLP to not be able to find optimal α and β to calibrate the PITs. Specifically, this scenario serves to motivate BMC and show that BMC is highly flexible and can calibrate the PITs when BLP cannot. We also expect BMC with higher K to be more flexible than BMC with lower K.

Table 16: Cross validation log scores for beta mixture methods

method	mean_train_ls	mean_valid_ls
BMC2	-1.2073506	-1.2066975
EW_BMC2	-1.2065485	-1.2066941
BMC3	-0.9742500	-0.9746691
EW_BMC3	-1.1609765	-1.1590791
BMC4	-0.9745309	-0.9748045
EW_BMC4	-1.0680773	-1.0660947
BMC5	-0.9746870	-0.9751163
EW_BMC5	-0.9745933	-0.9749492

Table 17: Weight Parameters

Method	w_1	w_2	w_3	w_4
TLP	NA	NA	NA	NA
BLP	NA	NA	NA	NA
EW	NA	NA	NA	NA
EW-BLP	NA	NA	NA	NA
BMC3	0.197	0.202	0.601	NA
EW-BMC4	0.057	0.197	0.202	0.544

Table 18: Beta mixture parameters

Method	α_1	β_1	α_2	β_2	α_3	β_3	α_4	β_4
TLP	NA	NA	NA	NA	NA	NA	NA	NA
BLP	2.162	1.320	NA	NA	NA	NA	NA	NA
EW	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	1.557	0.910	NA	NA	NA	NA	NA	NA
BMC3	13.202	13.356	4.356	55.587	19.529	8.880	NA	NA
EW-BMC4	75.832	15.237	20.262	99.713	56.823	68.511	65.866	9.36

Table 19: Component weight parameters -

Method	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}	ω_{31}	ω_{32}	ω_{33}	ω_{41}	ω_{42}
TLP	0.778	0.000	0.222	NA	NA	NA	NA	NA	NA	NA	NA
BLP	0.562	0.000	0.438	NA	NA	NA	NA	NA	NA	NA	NA
EW	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA
EW-BLP	0.500	0.500	NA	NA	NA	NA	NA	NA	NA	NA	NA
BMC3	0.003	0.000	0.997	1.0	0.0	0	0.998	0.0	0.002	NA	NA
EW-BMC4	0.500	0.500	NA	0.5	0.5	NA	0.500	0.5	NA	0.5	0.5

Table 20: Log score

	TLP	BLP	EW	EW-BLP	BMC3	EW-BMC4
Training	-1.718	-1.657	-1.857	-1.742	-0.975	-0.977
Test	-1.722	-1.660	-1.858	-1.747	-0.993	-0.994

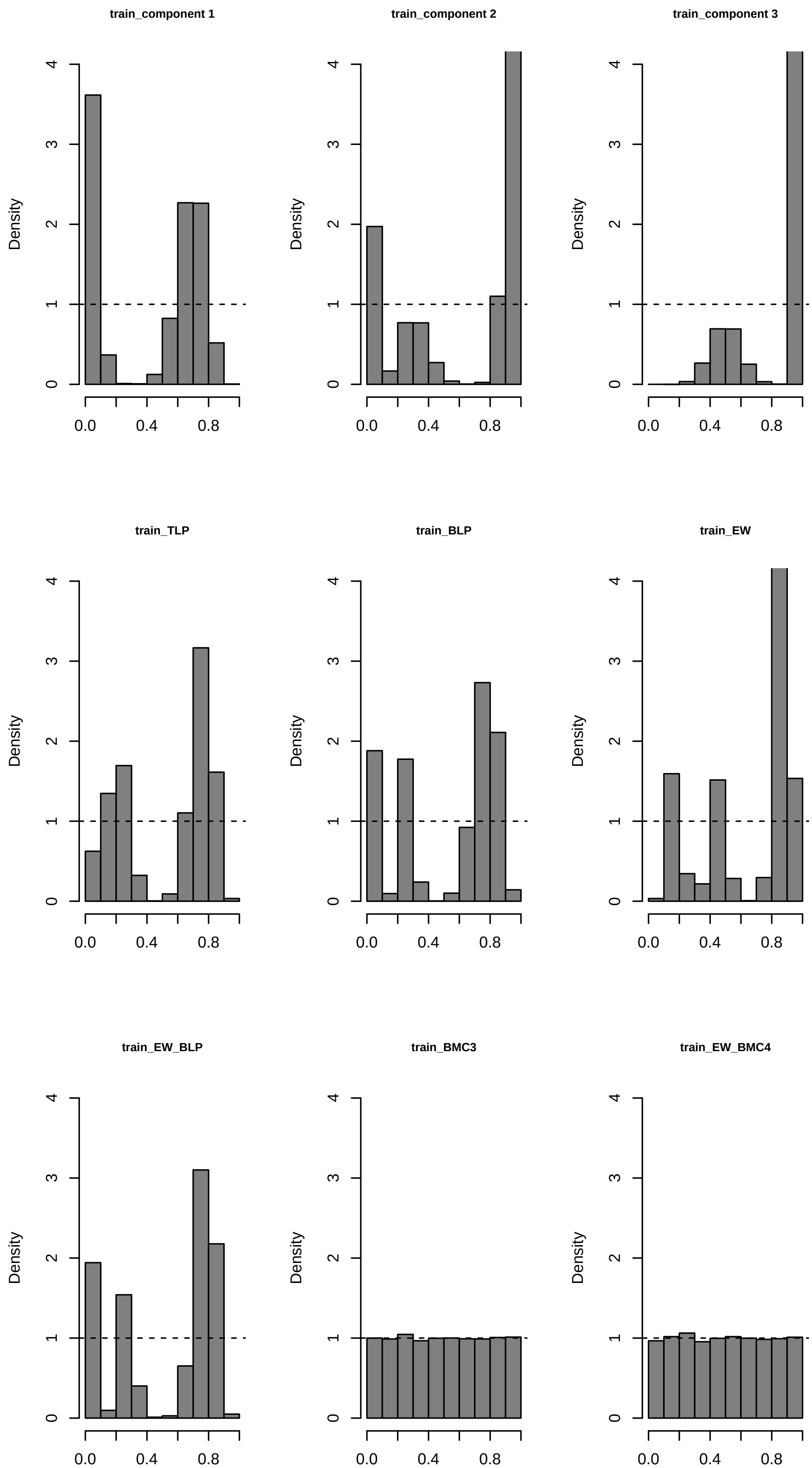


Figure 9: Train PITs

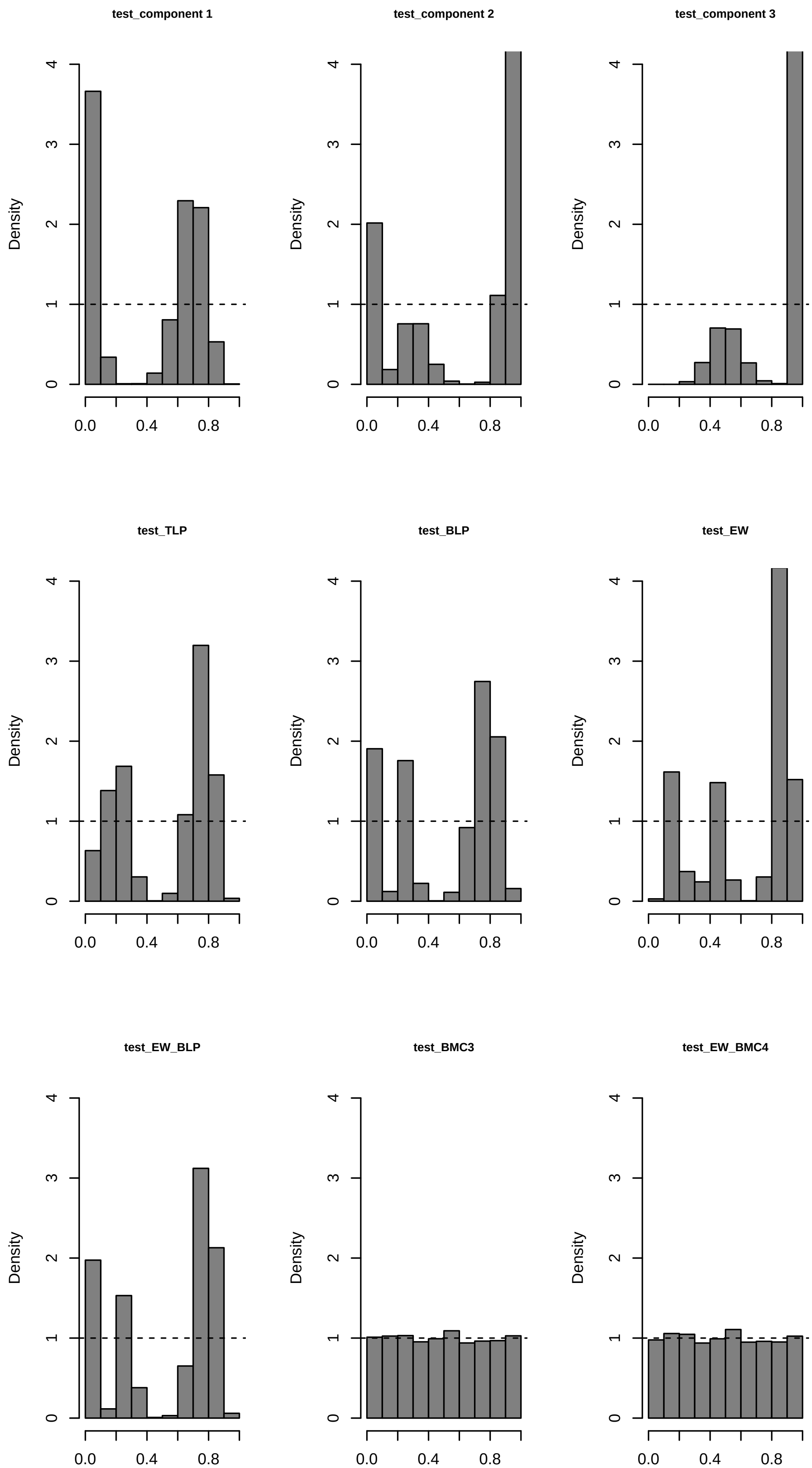
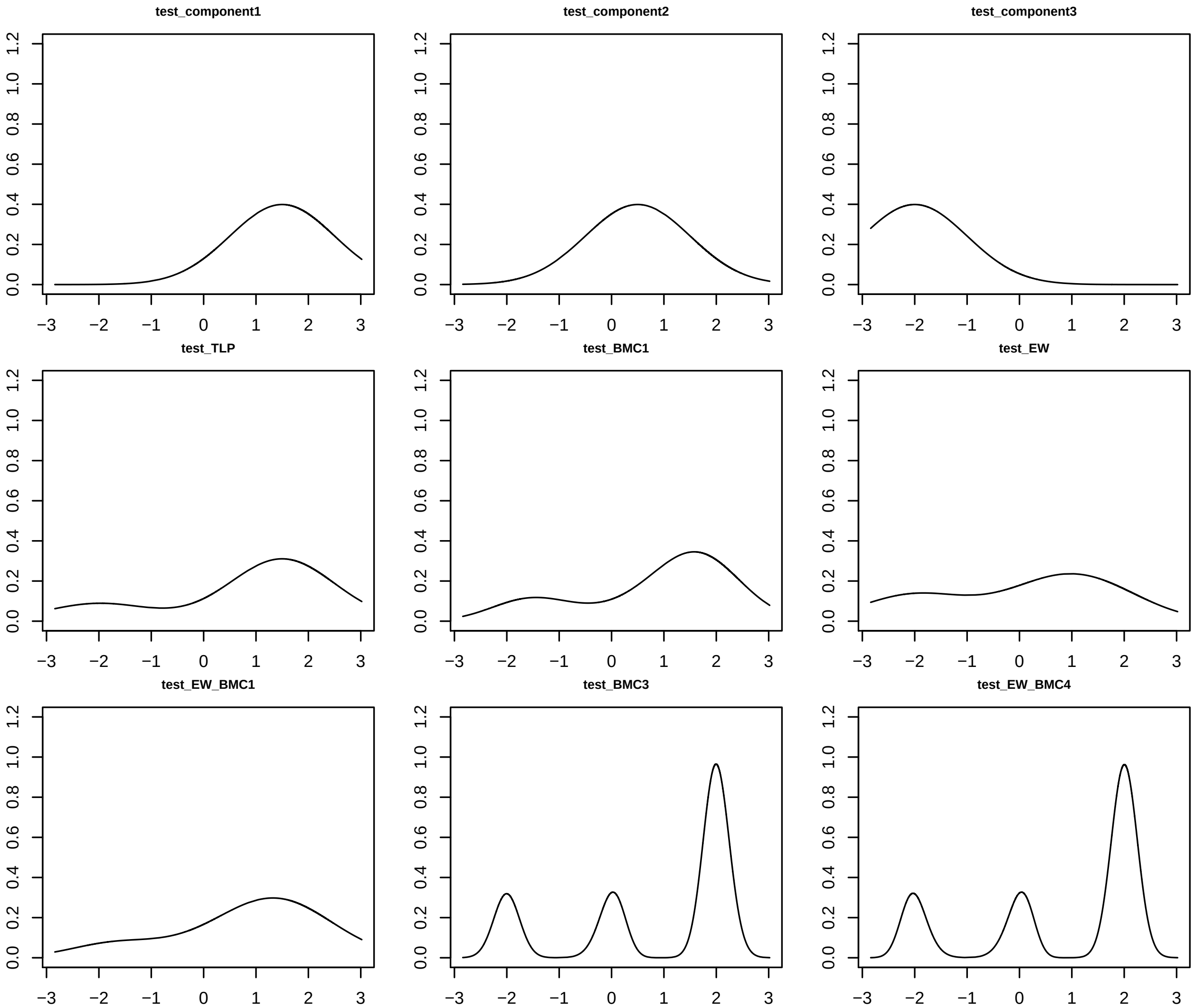


Figure 10: Test PITs



test_component1
test_component2
test_component3
test_TLP
test_BMC1
test_EW
test_EW_BMC1
test_BMC3
test_EW_BMC4
0.0 0.2 0.4 0.6 0.8 1.0 1.2
−3 −2 −1 0 1 2 3