Comparison of Ensemble Recalibration Methods in Flu Forecasting

Nutcha Wattanachit

12/14/2020

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 Bayesian finite beta mixture 6) the Bayesian 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

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

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

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

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

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

Bayesian 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

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

Bayesian finite equally weighted beta mixture (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.

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

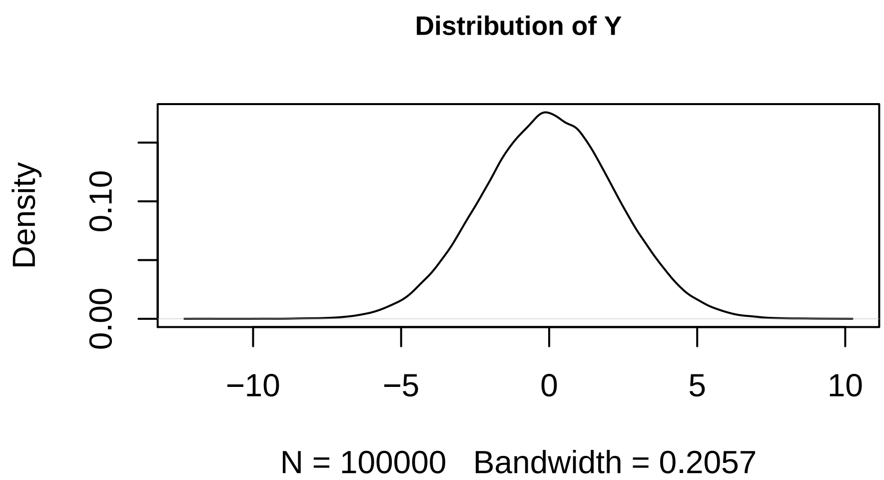
Simulation studies

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 random sample $(f_{1i}, f_{2i}, f_{3i}, Y_i) : i = 1, ..., n$ of size $n = 80,000$ and evaluate on an independent test sample of $n = 20,000$. 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: Model and Beta Mixing Weight Parameters

	ω_1	ω_2	ω_3	α	β
TLP	0.252	0.276	0.473	NA	NA
BLP	0.287	0.304	0.410	1.454	1.451
EW	0.333	0.333	0.333	NA	NA
EW-BLP	0.333	0.333	0.333	1.456	1.454

	w_1	w_2	w_3	w_4	w_5
BMC2	0.307	0.693	NA	NA	NA
EW-BMC2	0.306	0.694	NA	NA	NA
BMC3	0.043	0.261	0.696	NA	NA
EW-BMC3	0.099	0.301	0.600	NA	NA
BMC4	0.042	0.335	0.423	0.201	NA
EW-BMC4	0.101	0.300	0.397	0.202	NA
BMC5	0.041	0.221	0.322	0.369	0.047
EW-BMC5	0.101	0.202	0.300	0.347	0.050

Table 2: Mixture Parameters

	α_1	β_1	α_2	β_2	α_3	β_3	α_4	β_4	α_5	β_5
BMC2	1.190	1.172	1.617	1.625	NA	NA	NA	NA	NA	NA
EW-BMC2	1.208	1.186	1.607	1.618	NA	NA	NA	NA	NA	NA
BMC3	0.798	0.788	1.385	1.832	1.610	1.450	NA	NA	NA	NA
EW-BMC3	1.216	1.154	1.257	1.259	1.647	1.658	NA	NA	NA	NA
BMC4	0.800	0.788	1.427	1.818	1.737	1.424	1.435	1.454	NA	NA
EW-BMC4	1.272	1.176	1.336	1.399	1.767	1.765	1.296	1.256	NA	NA
BMC5	0.802	0.788	1.444	1.450	1.827	1.423	1.445	1.808	1.417	1.315
EW-BMC5	1.289	1.190	1.339	1.299	1.354	1.456	1.806	1.782	1.217	1.125

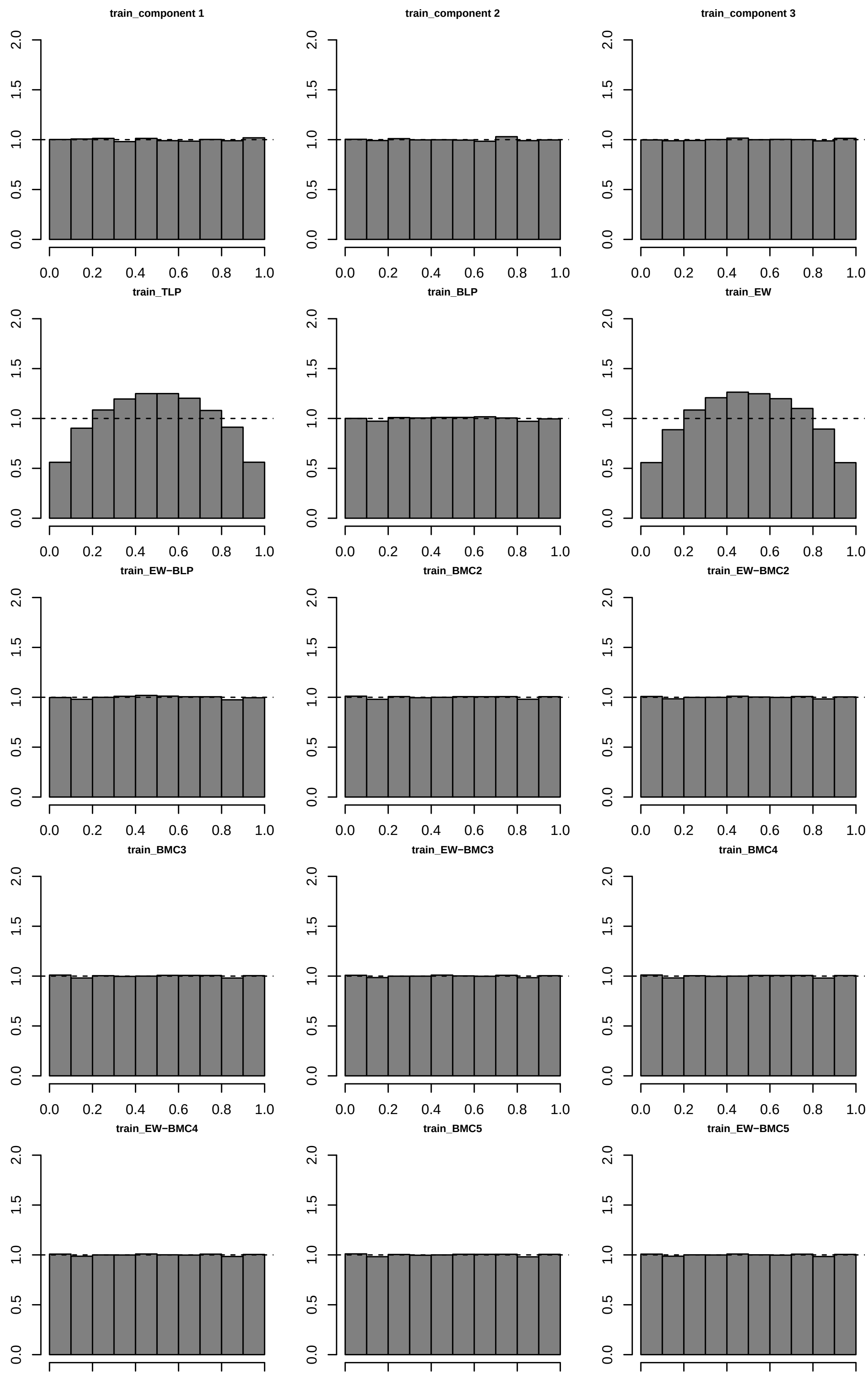
Table 3: Mixture Parameters

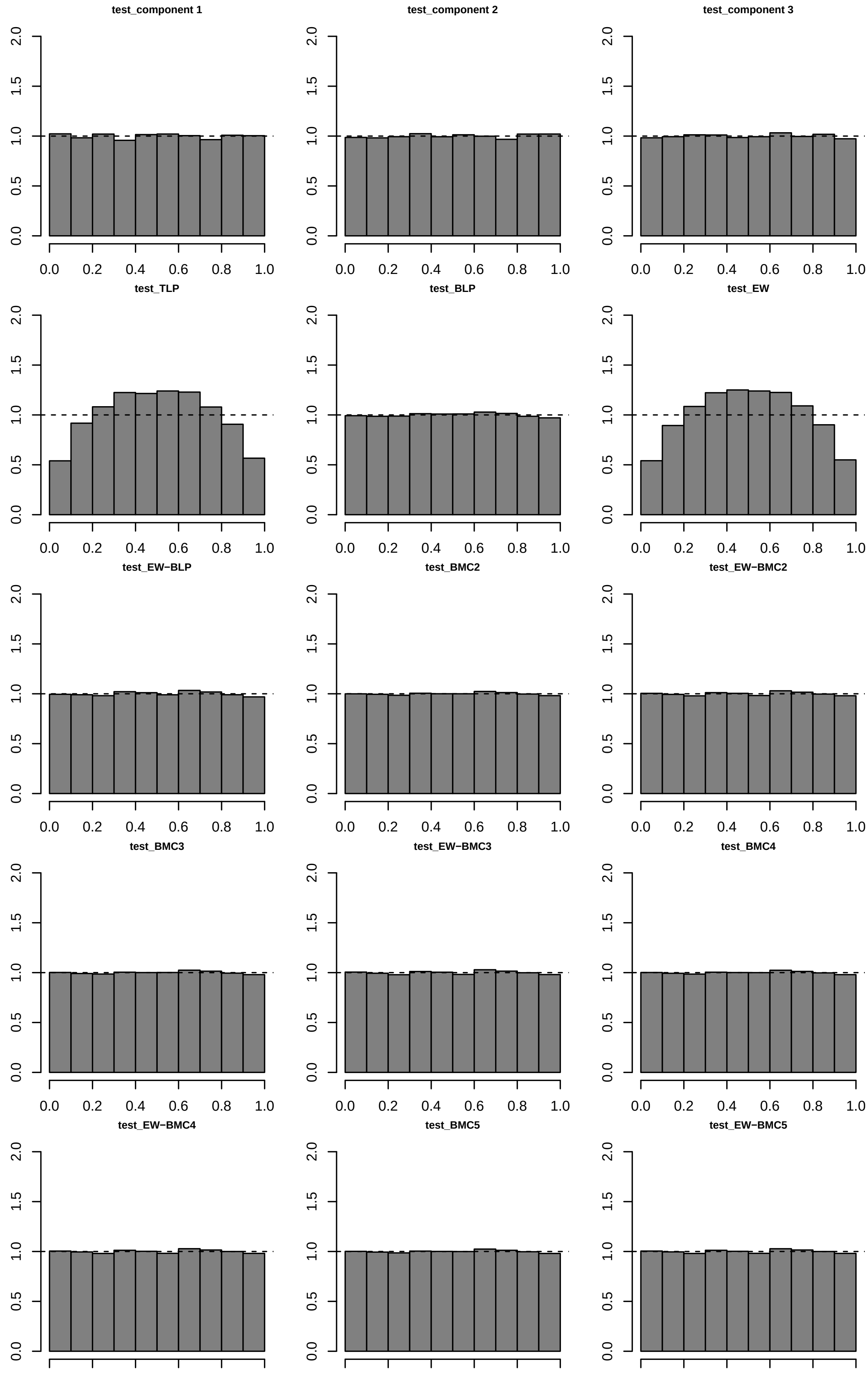
	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}	ω_{31}	ω_{32}	ω_{33}	ω_{41}	ω_{42}	ω_{43}	ω_{51}	ω_{52}	ω_{53}
BMC2	0.261	0.261	0.477	0.297	0.320	0.383	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BMC2	0.333	0.333	0.333	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA
BMC3	0.005	0.001	0.993	0.325	0.285	0.390	0.285	0.325	0.390	NA	NA	NA	NA	NA	NA
EW-BMC3	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA
BMC4	0.005	0.001	0.994	0.317	0.293	0.391	0.276	0.333	0.391	0.304	0.308	0.387	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	NA	NA	NA
BMC5	0.004	0.001	0.995	0.302	0.308	0.391	0.272	0.337	0.392	0.313	0.296	0.391	0.299	0.324	0.378
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 4: Log Score

	Training	Test
TLP	-1.910	-1.908
BLP	-1.869	-1.866
EW	-1.913	-1.911
EW-BLP	-1.871	-1.868

	Training	Test
BMC2	-1.869	-1.866
EW-BMC2	-1.871	-1.868
BMC3	-1.869	-1.866
EW-BMC3	-1.871	-1.868
BMC4	-1.869	-1.866
EW-BMC4	-1.871	-1.868
BMC5	-1.869	-1.866
EW-BMC5	-1.871	-1.868


train_component 1
train_component 2
train_component 3
train_TLP
train_BLP
train_EW
train_EW−BLP
train_BMC2
train_EW−BMC2
train_BMC3
train_EW−BMC3
train_BMC4
train_EW−BMC4
train_BMC5
train_EW−BMC5
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0

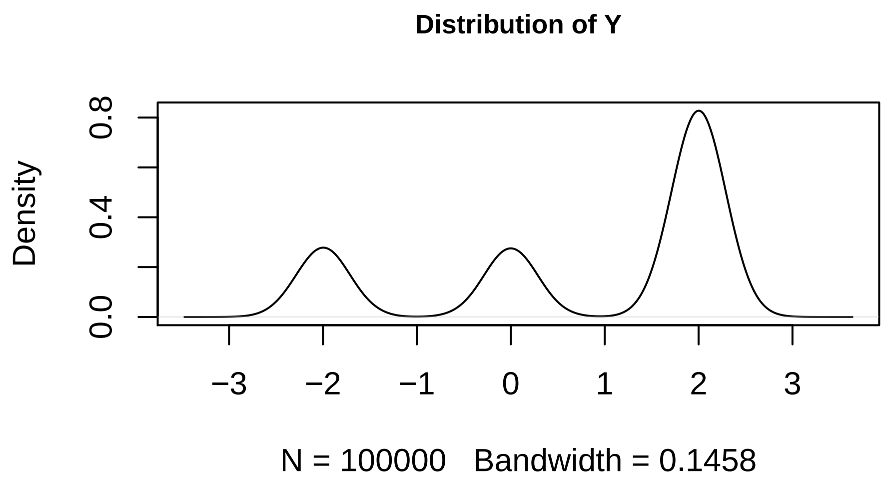

test_component 1
test_component 2
test_component 3
test_TLP
test_BLP
test_EW
test_EW−BLP
test_BMC2
test_EW−BMC2
test_BMC3
test_EW−BMC3
test_BMC4
test_EW−BMC4
test_BMC5
test_EW−BMC5
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0

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 5: Model and Beta Mixing Weight Parameters

	ω_1	ω_2	ω_3	α	β
TLP	0.201	0.198	0.601	NA	NA
BLP	0.205	0.199	0.597	1.011	1.000
EW	0.333	0.333	0.333	NA	NA
EW-BLP	0.333	0.333	0.333	1.249	0.786

	w_1	w_2	w_3	w_4	w_5
BMC2	0.300	0.700	NA	NA	NA
EW-BMC2	0.273	0.727	NA	NA	NA
BMC3	0.100	0.300	0.600	NA	NA
EW-BMC3	0.226	0.667	0.107	NA	NA
BMC4	0.100	0.300	0.400	0.200	NA
EW-BMC4	0.111	0.342	0.312	0.235	NA
BMC5	0.000	0.194	0.674	0.131	0.000
EW-BMC5	0.121	0.244	0.271	0.312	0.051

Table 6: Mixture Parameters

	α_1	β_1	α_2	β_2	α_3	β_3	α_4	β_4	α_5	β_5
BMC2	0.937	0.886	1.049	1.059	NA	NA	NA	NA	NA	NA
EW-BMC2	1.181	3.352	4.186	1.222	NA	NA	NA	NA	NA	NA
BMC3	1.003	0.996	1.007	1.001	1.014	0.999	NA	NA	NA	NA
EW-BMC3	12.732	2.214	0.946	0.785	73.427	27.310	NA	NA	NA	NA
BMC4	1.004	0.997	1.012	0.999	1.015	0.999	1.004	1.001	NA	NA
EW-BMC4	0.941	0.744	0.943	0.744	11.026	2.774	0.943	0.745	NA	NA
BMC5	0.000	0.000	1.006	0.994	1.018	0.994	1.092	1.037	0.000	0.000
EW-BMC5	0.942	0.744	0.943	0.745	0.944	0.744	11.025	2.774	0.936	0.744

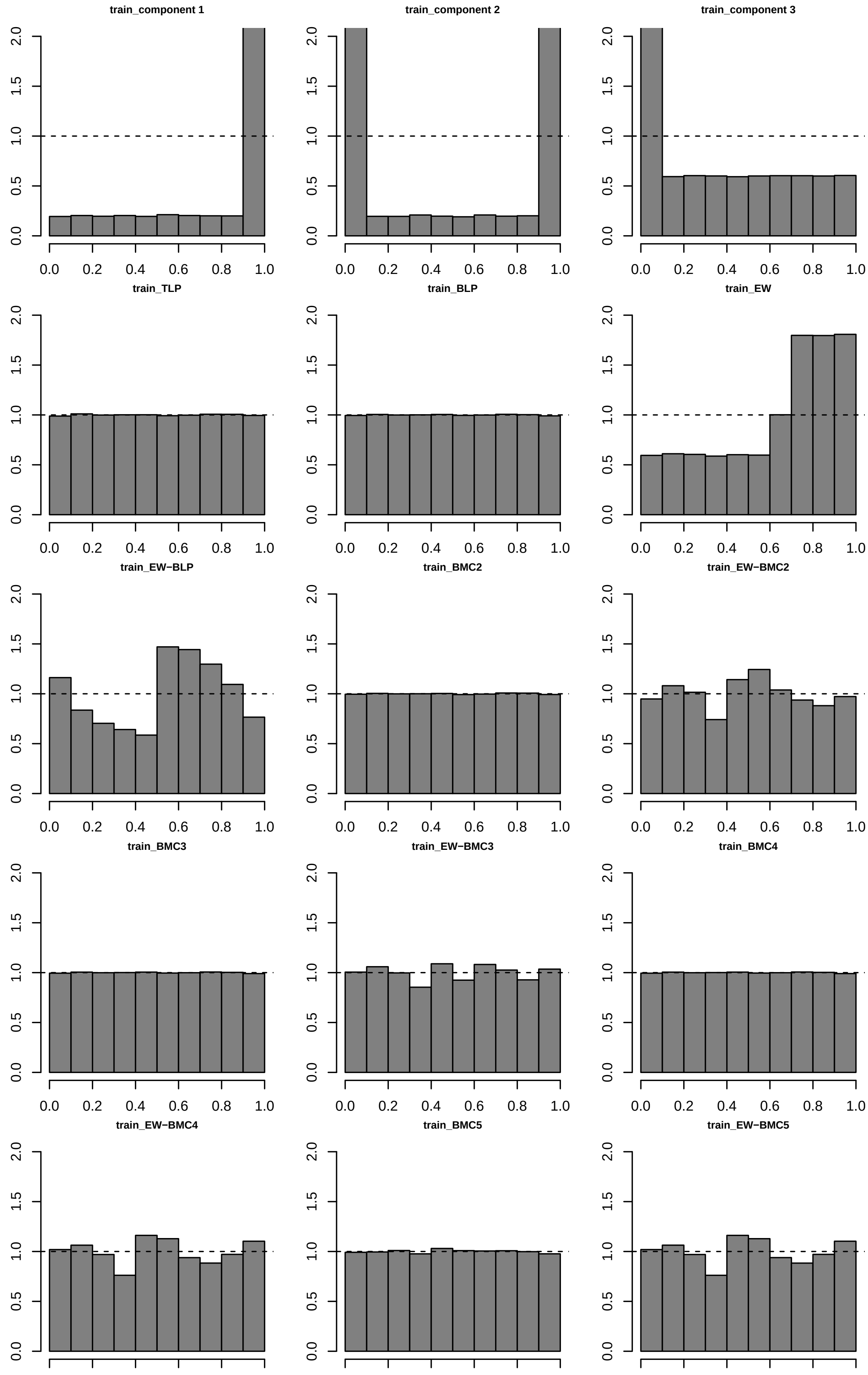
Table 7: Mixture Parameters

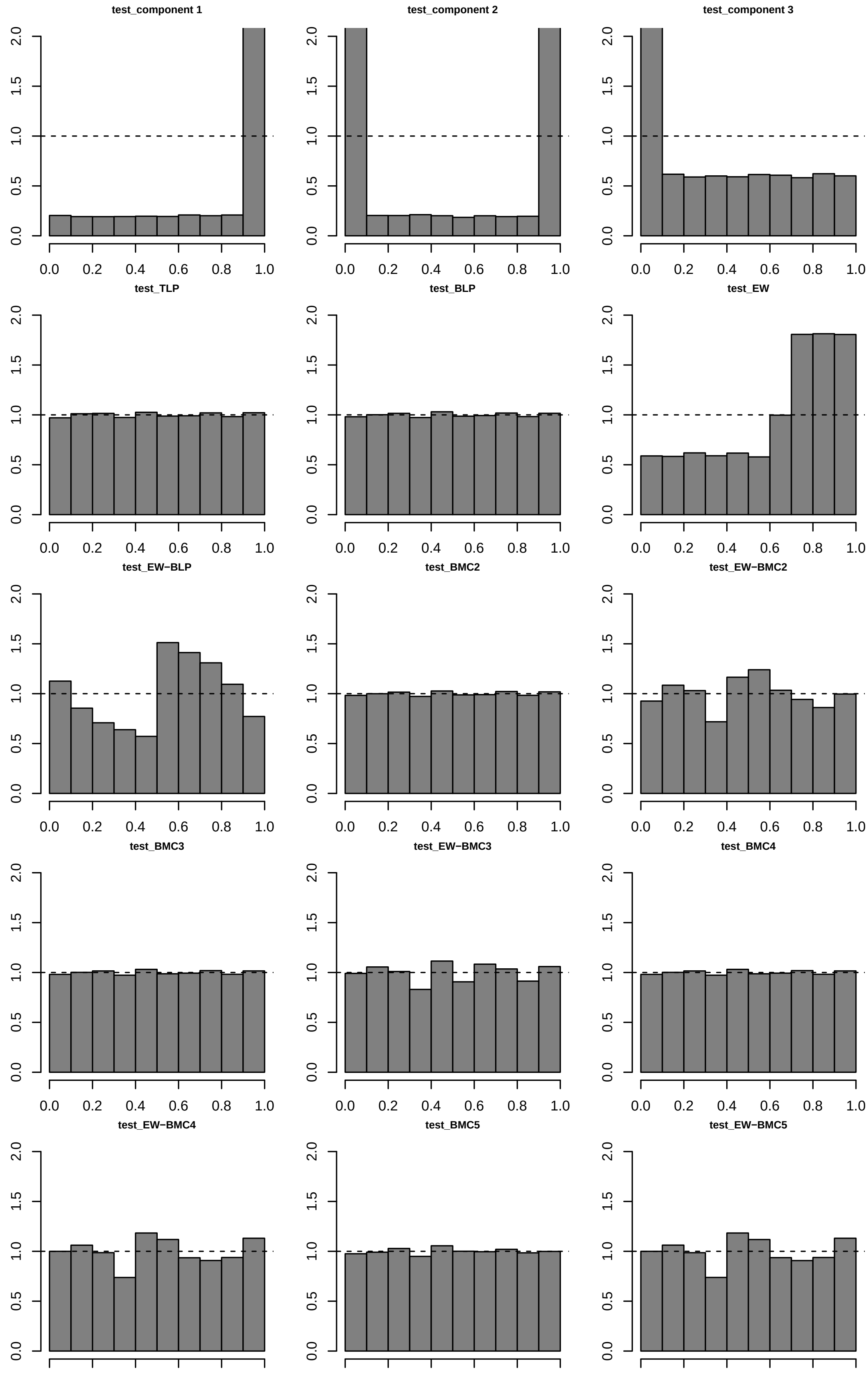
	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}	ω_{31}	ω_{32}	ω_{33}	ω_{41}	ω_{42}	ω_{43}	ω_{51}	ω_{52}	ω_{53}
BMC2	0.204	0.196	0.600	0.205	0.199	0.596	NA	NA	NA	NA	NA	NA	NA	NA	NA
EW-BMC2	0.333	0.333	0.333	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA	NA	NA	NA
BMC3	0.205	0.196	0.599	0.207	0.199	0.594	0.204	0.199	0.598	NA	NA	NA	NA	NA	NA
EW-BMC3	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	NA	NA	NA	NA	NA	NA
BMC4	0.205	0.196	0.599	0.207	0.199	0.594	0.204	0.199	0.597	0.201	0.198	0.600	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	NA	NA	NA
BMC5	0.995	0.005	0.000	1.000	0.000	0.000	0.000	0.118	0.882	0.075	0.925	0.000	0.270	0.006	0.724
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

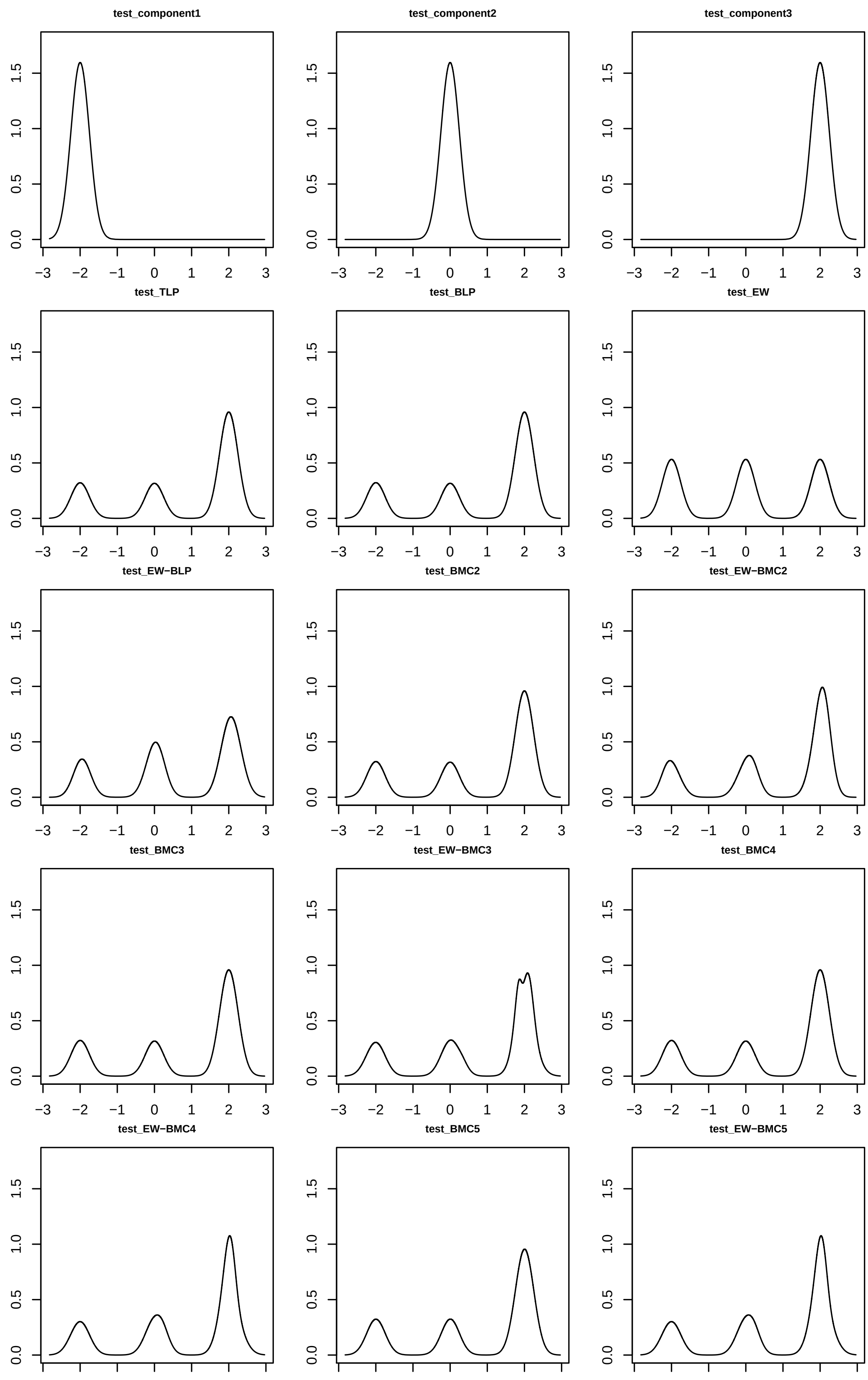
Table 8: Log Score

	Training	Test
TLP	-0.982	-0.980
BLP	-0.982	-0.980
EW	-1.131	-1.133
EW-BLP	-1.043	-1.043

	Training	Test
BMC2	-0.982	-0.980
EW-BMC2	-1.004	-1.004
BMC3	-0.982	-0.980
EW-BMC3	-0.990	-0.989
BMC4	-0.982	-0.980
EW-BMC4	-1.000	-1.000
BMC5	-0.982	-0.980
EW-BMC5	-1.000	-1.000


train_component 1
train_component 2
train_component 3
train_TLP
train_BLP
train_EW
train_EW−BLP
train_BMC2
train_EW−BMC2
train_BMC3
train_EW−BMC3
train_BMC4
train_EW−BMC4
train_BMC5
train_EW−BMC5


test_component 1
test_component 2
test_component 3
test_TLP
test_BLP
test_EW
test_EW−BLP
test_BMC2
test_EW−BMC2
test_BMC3
test_EW−BMC3
test_BMC4
test_EW−BMC4
test_BMC5
test_EW−BMC5
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0


test_component1
test_component2
test_component3
test_TLP
test_BLP
test_EW
test_EW−BLP
test_BMC2
test_EW−BMC2
test_BMC3
test_EW−BMC3
test_BMC4
test_EW−BMC4
test_BMC5
test_EW−BMC5

Scenario 3: Multimodal DGP (Normal mixture) and open-$\mathcal{M}$

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

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

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 9: Model and Beta Mixing Weight Parameters

	ω_1	ω_2	α	β
TLP	0.661	0.339	NA	NA
BLP	0.783	0.217	1.036	1.638
EW	0.500	0.500	NA	NA
EW-BLP	0.500	0.500	1.391	1.285

	w_1	w_2	w_3	w_4	w_5
BMC2	0.399	0.601	NA	NA	NA
EW-BMC2	0.201	0.799	NA	NA	NA
BMC3	0.000	0.601	0.399	NA	NA
EW-BMC3	0.123	0.315	0.561	NA	NA
BMC4	0.000	0.601	0.198	0.201	NA
EW-BMC4	0.052	0.146	0.601	0.201	NA
BMC5	NA	NA	NA	NA	NA
EW-BMC5	0.601	0.201	NA	NA	NA

Table 10: Mixture Parameters

	α_1	β_1	α_2	β_2	α_3	β_3	α_4	β_4	α_5	β_5
BMC2	0.953	31.689	12.881	12.821	NA	NA	NA	NA	NA	NA
EW-BMC2	6.872	75.943	7.277	3.644	NA	NA	NA	NA	NA	NA
BMC3	0.001	0.000	12.863	12.814	0.953	31.695	NA	NA	NA	NA
EW-BMC3	1.119	2.642	1.119	2.646	63.375	21.019	NA	NA	NA	NA
BMC4	0.000	0.002	12.807	12.760	3.092	114.826	6.805	78.737	NA	NA
EW-BMC4	58.462	85.885	108.808	138.833	55.939	18.732	6.754	74.345	NA	NA
BMC5	0.007	25.560	59.069	58.848	2.243	83.299	0.058	0.670	0.002	0.997
EW-BMC5	100.296	147.344	32.809	41.862	60.754	20.345	0.004	0.048	21.059	-20.913

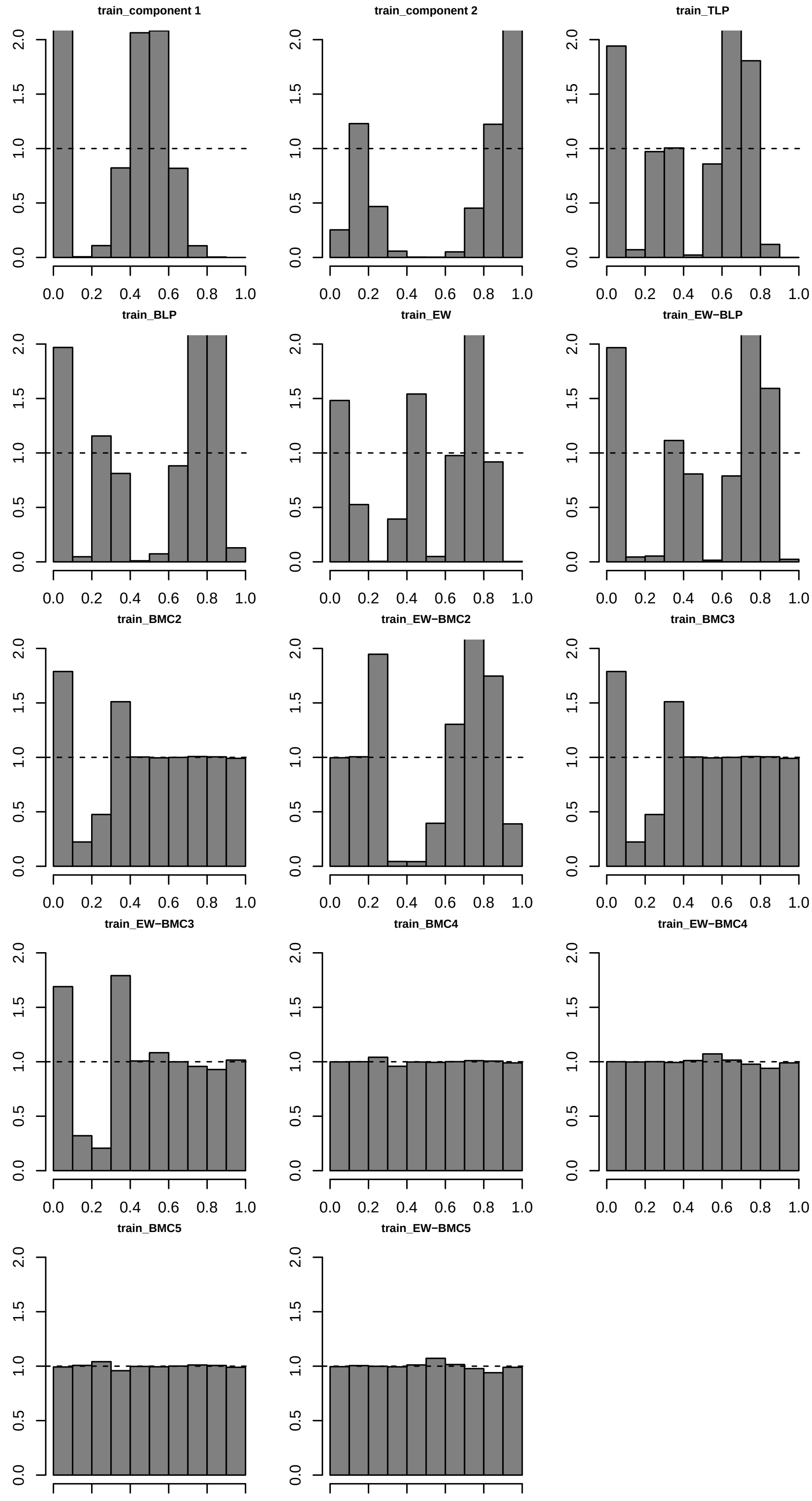
Table 11: Mixture Parameters

	ω_{11}	ω_{12}	ω_{21}	ω_{22}	ω_{31}	ω_{32}	ω_{41}	ω_{42}	ω_{51}	ω_{52}
BMC2	0.967	0.033	0.998	0.002	NA	NA	NA	NA	NA	NA
EW-BMC2	0.500	0.500	0.500	0.500	NA	NA	NA	NA	NA	NA
BMC3	0.961	0.039	0.998	0.002	0.967	0.033	NA	NA	NA	NA
EW-BMC3	0.500	0.500	0.500	0.500	0.500	0.500	NA	NA	NA	NA
BMC4	0.728	0.272	0.998	0.002	1.000	0.000	0.522	0.478	NA	NA
EW-BMC4	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	NA	NA
BMC5	1.000	0.478	0.522	0.000	0.272	0.601	0.002	0.198	0.0	0.201
EW-BMC5	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.5	0.500

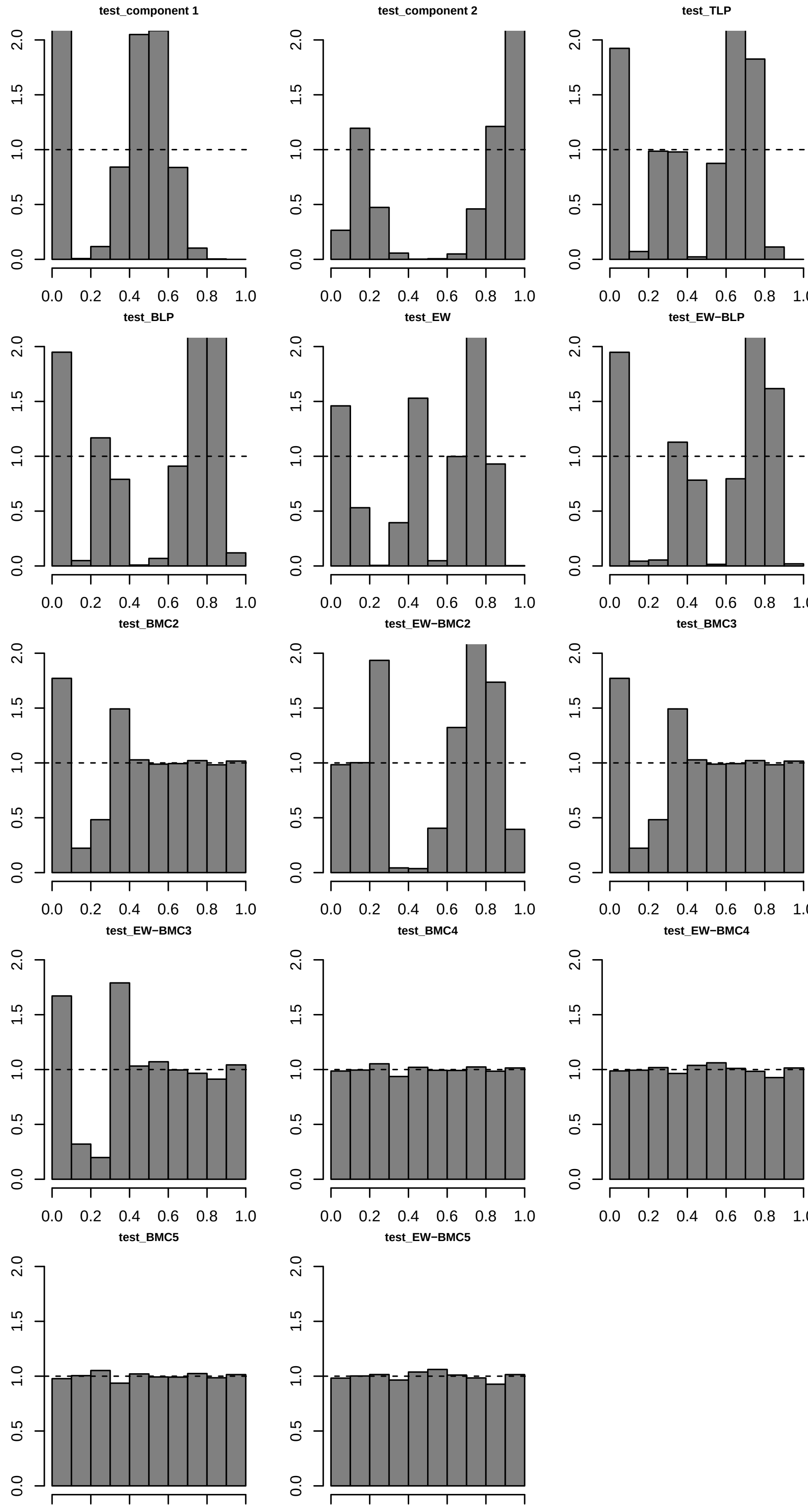
Table 12: Log Score

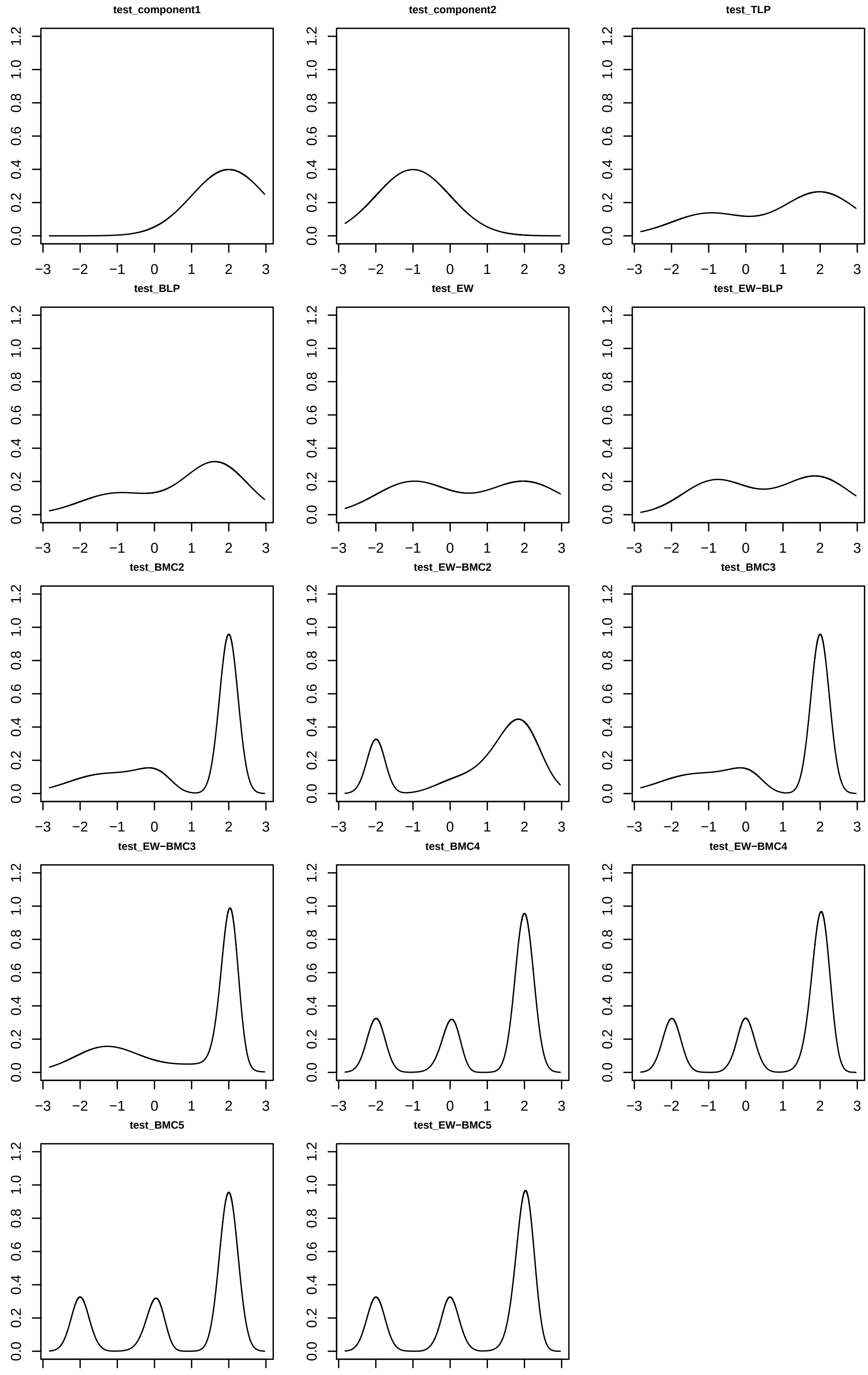
	Training	Test
TLP	-1.742	-1.739
BLP	-1.679	-1.676
EW	-1.786	-1.784
EW-BLP	-1.757	-1.755

	Training	Test
BMC2	-1.208	-1.202
EW-BMC2	-1.383	-1.383
BMC3	-1.208	-1.202
EW-BMC3	-1.291	-1.286
BMC4	-0.983	-0.982
EW-BMC4	-0.985	-0.983
BMC5	-0.983	-0.982
EW-BMC5	-0.985	-0.983



train_component 1
train_component 2
train_TLP
train_BLP
train_EW
train_EW−BLP
train_BMC2
train_EW−BMC2
train_BMC3
train_EW−BMC3
train_BMC4
train_EW−BMC4
train_BMC5
train_EW−BMC5


test_component 1
test_component 2
test_TLP
test_BLP
test_EW
test_EW−BLP
test_BMC2
test_EW−BMC2
test_BMC3
test_EW−BMC3
test_BMC4
test_EW−BMC4
test_BMC5
test_EW−BMC5


test_component1
test_component2
test_TLP
test_BLP
test_EW
test_EW−BLP
test_BMC2
test_EW−BMC2
test_BMC3
test_EW−BMC3
test_BMC4
test_EW−BMC4
test_BMC5
test_EW−BMC5
0.0 0.2 0.4 0.6 0.8 1.0 1.2
−3 −2 −1 0 1 2 3