Supplementary Material to

Bayesian Conditional Transformation Models

by

Manuel Carlan, Thomas Kneib and Nadja Klein

This supplement consists of

Part A: Proof of Theorem 2.1

Part B: Further Results for the Applications

Part C: Further Simulation Results

Part A Proofs

Part A.1 Proof of Theorem 2.1

Proof of Theorem 2.1. First note that it is sufficient to show that $h_j(\cdot|\boldsymbol{x}) : \mathbb{R} \to \mathbb{R}$ is monotonically increasing for $j = 1, \dots J$, since for j_1, j_2 and y_1, y_2 with $y_1 < y_2$ and $h_{j_1}(y_1|\boldsymbol{x}) \leq h_{j_1}(y_2|\boldsymbol{x})$, $h_{j_2}(y_1|\boldsymbol{x}) \leq h_{j_2}(y_2|\boldsymbol{x}) \Rightarrow h_{j_1}(y_1|\boldsymbol{x}) + h_{j_2}(y_1|\boldsymbol{x}) \leq h_{j_1}(y_2|\boldsymbol{x}) + h_{j_2}(y_2|\boldsymbol{x})$. In what follows we thus suppress the subscript j for notational simplicity. Given a knot sequence $\xi_{11} < \xi_{12} < \dots$ of equally spaced knots in y direction, according to (4) and Höllig and Hörner (2013) we can write

$$h(y|\boldsymbol{x}) = \sum_{d_1=1}^{D_1} \sum_{d_2=1}^{D_2} \gamma_{d_1 d_2} B_{1d_1}^{n_1}(y) B_{2d_2}^{n_2}(\boldsymbol{x}),$$

where $B_{1d_1}^{n_1}$ denotes a B-spline basis of order n_1 and $B_{2d_2}^{n_2}$ is left unspecified but assumed to be an appropriate basis function representation along the $\boldsymbol{x}$ dimension. Now, according to Höllig and Hörner (2013) the partial derivative of h along the y dimension is given by

$$\partial_y h(y|\boldsymbol{x}) = \frac{\partial h(y|\boldsymbol{x})}{\partial y} \equiv h'(y|\boldsymbol{x}) = \sum_{d_1=2}^{D_1} \sum_{d_2=1}^{D_2} \alpha_{d_1,\xi_1}^{n_1} (\gamma_{d_1 d_2} - \gamma_{d_1-1,d_2}) B_{1d_1}^{n_1-1}(y) B_{2d_2}^{n_2}(\boldsymbol{x}),$$

where

$$\alpha_{d_1,\xi_1}^{n_1} = \frac{n_1}{\xi_{1,d_1+n_1} - \xi_{1d_1}}.$$

Since by definition the $\alpha_{d_1,\xi_1}^{n_1}$ and all B-spline basis functions are nonnegative, a sufficient condition for $h'(y|\boldsymbol{x}) \geq 0$ (which is equivalent to a monotonically increasing transformation function $h(y|\boldsymbol{x})$) is

$$\gamma_{d_1 d_2} - \gamma_{d_1-1,d_2} \geq 0. \tag{1}$$

Therefore, an increasing sequence of all parameters $\gamma_{d_1 d_2}$ will produce a monotonically increasing function. To achieve (1), the constrained model coefficients, $\gamma_{d_1 d_2}$, are defined as

$$\gamma_{1d_2} = \beta_{1d_2}, \quad \gamma_{jd_2} = \beta_{1d_2} + \sum_{i=2}^{j} \exp(\beta_{id_2}), \quad j = 1, \dots, D_1, \quad d_2 = 1, \dots, D_2, \tag{2}$$

where the β_{id_2} are the unknown unconstrained parameters. In matrix notations (2) may be written as $\boldsymbol{\gamma} = \boldsymbol{\Sigma}\tilde{\boldsymbol{\beta}}$, with

$$\tilde{\boldsymbol{\beta}} = (\beta_{11}, \dots, \beta_{1D_2}, \exp(\beta_{21}), \dots \exp(\beta_{2D_2}), \dots, \exp(\beta_{D_1 1}), \dots, \exp(\beta_{D_1 D_2}))^\top$$

and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{D_1} \otimes \boldsymbol{I}_{D_2}$, where $\boldsymbol{I}_{D_2}$ is an identity matrix of size D_2, $\boldsymbol{\Sigma}_{D_1}$ is a lower triangular matrix of size D_1 with $\Sigma_{D_1,kl} = 0$ if $k < l$ and $\Sigma_{D_1,kl} = 1$ if $k \geq l$. □

Part B Further Results for the Applications

Part B.1 Framingham heart study

Model	Specification
bctm_vcm	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{20}(y)^\top \otimes (1, age), (year, sex))^\top, \pi_{\vartheta}(\boldsymbol{\vartheta})\big)$
bctm_vcm_sd	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{20}(y)^\top \otimes (1, age), (year, sex))^\top, \pi_{\vartheta_{sd}}(\boldsymbol{\vartheta})\big)$
bctm_tensor	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{10}(y)^\top \otimes \boldsymbol{b}_{10}(\text{age})^\top, (year, sex))^\top, \pi_{\vartheta}(\boldsymbol{\vartheta})\big)$
bctm_tensor_sd	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{10}(y)^\top \otimes \boldsymbol{b}_{10}(\text{age})^\top, (year, sex))^\top, \pi_{\vartheta_{sd}}(\boldsymbol{\vartheta})\big)$
gamlss	$\eta_k = \beta_{k,0} + x_{\text{sex}}\beta_{k,\text{sex}} + x_{\text{age}}\beta_{k,\text{age}} + x_{\text{year}}\beta_{k,\text{year}}, \quad k = 1, \ldots, K, \quad K = 4$
bctm_vcm_re	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{20}(y)^\top \otimes (1, age), (1 \otimes \boldsymbol{b}(patient)^T), (year, sex))^\top, \pi_{\vartheta}(\boldsymbol{\vartheta})\big)$
bctm_vcm_re_sd	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{20}(y)^\top \otimes (1, age), (1 \otimes \boldsymbol{b}(patient)^T), (year, sex))^\top, \pi_{\vartheta_{sd}}(\boldsymbol{\vartheta})\big)$
bctm_tensor_re	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{10}(y)^\top \otimes \boldsymbol{b}_{10}(\text{age})^\top, (1 \otimes \boldsymbol{b}(patient)^T), (year, sex))^\top, \pi_{\vartheta}(\boldsymbol{\vartheta})\big)$
bctm_tensor_re_sd	$\big(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{10}(y)^\top \otimes \boldsymbol{b}_{10}(\text{age})^\top, (1 \otimes \boldsymbol{b}(patient)^T), (year, sex))^\top, \pi_{\vartheta_{sd}}(\boldsymbol{\vartheta})\big)$
gamlss_re	$\eta_k = \beta_{k,0} + x_{\text{sex}}\beta_{k,\text{sex}} + x_{\text{age}}\beta_{k,\text{age}} + x_{\text{year}}\beta_{k,\text{year}} + f_{skew\_re}(patient), \quad k = 1, \ldots, K, \quad K = 4$

Table B.1: Framing heart study. Model specifications with corresponding basis dimensions in the subscript. For the MLT specifications, the basis functions are Bernstein polynomials and for the other models B-splines. The subscripts denote the number of basis functions each.

Part B.1.1 Model selection

Tab. B.2 shows the model selection criteria DIC, WAIC and log-scores for all competing models. Here, bctm_vcm, bctm_vcm_sd and bctm_vcm_re denote the response varying coefficient BTCM with IG prior, the same model but with SD prior and the bctm_vcm augmented by patient-specific random effects. Further bctm_tensor, bctm_tensor_sd, bctm_tensor_re and bctm_tensor_re_sd denote the corresponding full BCTMs based on a tensor product represenation for the variable *age*. Log-scores are based on 10-fold cross validation.

Model	DIC	WAIC	Log-scores
bctm_vcm	2760	2760	-1122.539
bctm_vcm_sd	2791	2792	-1120.025
bctm_tensor	2716	2715	**-1068.967**
bctm_tensor_sd	2709	2704	-1077.803
gamlss		2713	-
bctm_vcm_re	1569	1524	-
bctm_vcm_re_sd	1571	1526	-
bctm_tensor_re	**1563**	**1517**	-
bctm_tensor_re_sd	1569	1518	-
gamlss_re		1616	-

Table B.2: Framing heart study. Model selection criteria for the competing models. Log-scores are based on 10-fold cross validation.

Part B.1.2 Effectiveness and stability

Tab. B.3 show the mean effective sample sizes (of 2000 samples after burn-in) for the coefficients of the tensor product spline in the full BCTMs with different priors and with and without random effects. Gelman and Rubin's convergence diagnostic ($\hat{R}$; Gelman and Rubin (1992)) based on four MCMC chains of the same models are given in Tab. B.4. Results have been obtained using the R package `coda` (Plummer et al., 2006).

Model	bctm_tensor	bctm_tensor_sd	bctm_tensor_re	bctm_tensor_re_sd
Eff. samp. size	937	1320	600	1430

Table B.3: Framing heart study. Mean effective sample sizes (of 2000 samples after burn-in) for the coefficients of the tensor product spline in the full BCTMs with different priors and with and without random effects.

Model	bctm_tensor	bctm_tensor_sd	bctm_tensor_re	bctm_tensor_re_sd
all $\hat{R} < 1.1$	✓	✓	✓	✓
mean $\hat{R}$	1.01	1.00	1.01	1.00

Table B.4: Framing heart study. Gelman and Rubin's convergence diagnostic ($\hat{R}$; Gelman and Rubin (1992)) based on four MCMC chains of the full BCTMs with different priors and with and without random effects.

Part B.2 Leukemia survival

Tab. B.5 shows the estimated posterior means of the log-negative hazard ratios (collected in $\boldsymbol{\gamma}_2$), medians and credible intervals of the same model. Similar to the results in Zhou et al. (2020), we find that *tpi*, *age* and *wbc* are significant risk factors for surviving leukemia.

Model mlt is specified as the transformation model $(F_{MEV}, ((\boldsymbol{a}(t)^T \otimes 1)^T, (1 \otimes \boldsymbol{x}^T))^T)$, model mlt_re is specified as $(F_{MEV}, ((\boldsymbol{a}(t)^T \otimes 1)^T, (1 \otimes (\boldsymbol{b}(s)^T, \boldsymbol{x}^T))^T)$ and model mlt_nph is specified as $(F_{MEV}, ((\boldsymbol{a}(t)^T \otimes \boldsymbol{b}(age))^T, (1 \otimes (\boldsymbol{b}(s)^T, \boldsymbol{x}^T))^T)$. Specifically, we specified:

```
library(tram)
mlt <- Coxph(Surv(time,cens) ~ age + sex+ wbc + tpi , data = data, order=20)
mlt_re <- CoxphME(Surv(time,cens) ~ age + sex + wbc +tpi + (1|district),
          data = data, order=20)
```

for the mlt, and mlt_re, while for the mlt_nph, we used

```
datay <- numeric_var("timec", support = c(1, max(data$time)), bounds = c(0, Inf))
age_var <- numeric_var("age")
data$timec <- with(data, Surv(time , cens))
B_datay <- Bernstein_basis(var = datay, order = 9, ui = "increasing")
B_age <- Bernstein_basis(var = age_var, order=9, ui="none")
fm_data <- Surv(timec, cens) ~   sex+ wbc +tpi
B_shift <- as.basis(fm_data[-2L], data, remove_intercept=F)
ctm_nph <- ctm(B_datay, shifting = B_shift, interacting = B_age, data = data,
            todistr = "MinExtrVal")
mlt_nph <- mlt(ctm_nph, data = data, scale=F)
```

Thus, the mlt/mlt_re have 20 coefficients for $h(y)$ and the mlt_nph has $10\times$ 10 coefficents for $h(y|age)$ similar to the BCTM specifications.

	Mean	s.d.	Median	2.5%	97.5%
tpi	0.114	0.035	0.113	0.046	0.180
age	0.590	0.042	0591	0.511	0.673
sex	0.035	0.034	0.037	-0.031	0.102
wbc	0.208	0.033	0.207	0.142	0.275

Table B.5: Leukemia survial. Estimated posterior means, standard deviations and credible intervals of the log negative hazard ratios obtained from the spatial PH model $\big(\boldsymbol{\vartheta}, F_{\text{MEV}}, (\boldsymbol{a}(t)^\top \otimes \boldsymbol{b}(age))^\top, (1 \otimes (\boldsymbol{b}(s)^\top, \boldsymbol{x}^\top))^\top, \pi_\vartheta(\boldsymbol{\vartheta})\big)$.

	Parameter	BCTM	MLT	MPT
1	score	-0.055(0.011)	-0.057(0.011)	-0.055(0.010)
2	adeno vs. large	1.366(0.550)	1.356(0.561)	1.303(0.559)
3	small vs. large	1.474(0.510)	1.456(0.533)	1.362(0.527)
4	squamous vs. large	-0.147(0.605)	-0.188(0.598)	-0.173(0.580)

Table B.6: Veteran Lung Cancer Trial - Estimated posterior medians and posterior standard deviations (BCTM and MPT) or standard errors (MLT) in brackets obtained from the PO model.

Part B.3 Veteran's Administration lung cancer trial

The following analysis is based on the well-known Veteran's Administration lung cancer trial that was introduced by Prentice (1973) in conjunction with the findings in Hanson and Yang (2007). We considered a subsample of $n = 97$ individual patients who did not receive prior therapy. The survival times t of six patients are right-censored, in which case the likelihood is adapted accordingly. A semiparametric proportional odds model is a simple transformation model which takes the form of the mean shift BCTM $\{\boldsymbol{\vartheta}|F_{\text{SL}}, ((\boldsymbol{a}(t)^\top \otimes 1), (1 \otimes \boldsymbol{x}^\top))^\top, \pi_\vartheta(\boldsymbol{\vartheta})\}$ with transformation function

$$h(t|\boldsymbol{x}) = \boldsymbol{a}(t)^\top \boldsymbol{\gamma}_1 + \boldsymbol{x}^\top \boldsymbol{\gamma}_2,$$

where $F_{\text{SL}}(z) = (1 + \exp(-z))^{-1}$ denotes the standard logistic distribution function and $\boldsymbol{x}$ contains the Karnofsky performance score (*score*), measuring functional impairment and three binary variables indicating different cancer types (*adeno*, *small*, *squamous*; *large* is used as base category), as well as a treatment indicator (*treatment*). Table B.6 shows estimated posterior medians and standard deviations obtained from the BCTM, the MLT and the Mixture of Polya

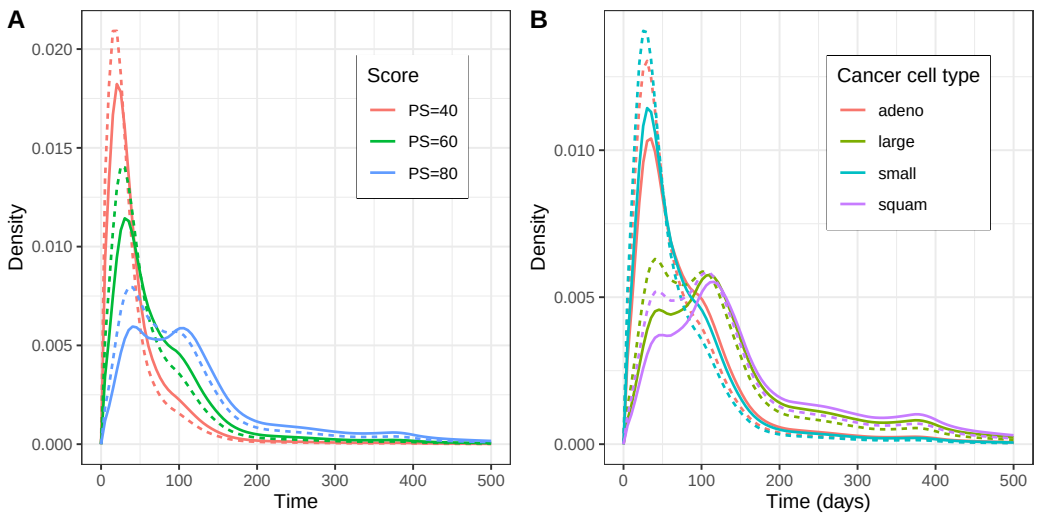

Figure B.1: Veteran Lung Cancer Trial - Predictive densities of survival time with (dotted) and without (regular) treatment effect. The left panel (A) shows results for *small* cell type. In the right panel (B), the Karnofsky performance score was fixed at 60.

Trees Prior (MPT) taken from Hanson and Yang (2007) with very similar results. Based on the BCTM, an increase in performance score of 10 increases the odds of surviving past any particular time point by about 75% ($\exp((-10) \cdot (-0.055)) \approx 1.75$). Panel A in Fig B.1 shows estimated predictive densities for different combinations of performance scores of the PO model with (dotted) and without (regular) treatment. We can see that in general, there is a positive impact of the performance score and the treatment on survival. Panel B indicates very similar survival times for squamous and large cell cancer types that differ significantly frome the more right-skewed densities of *adeno* and *small*. The resulting WAIC is 1086. We also supplemented the model with a nonlinear effect for *score* which did not decrease the WAIC (1085.97).

For illustrative purposes, based on the whole dataset, we considered an additional proportional odds (PO) BCTM only including the treatment effect $\left(\boldsymbol{\vartheta}, F_{\text{SL}}, (\boldsymbol{a}(t)^T, \mathcal{I}(treatment)^T), \pi_\vartheta(\boldsymbol{\vartheta})\right)$ resulting in the transformation function

$$h(y|\boldsymbol{x}) = \boldsymbol{a}(t)^\top \boldsymbol{\gamma}_1 + \mathcal{I}(treatment)\gamma_2.$$

Here, β is the log-odds ratio for the treatment. We can drop the proportional odds assumption by including an interaction term in the model $\left(F_{\text{SL}}, (\boldsymbol{a}(t)^\top \otimes (1, \mathcal{I}(treatment)^\top), \pi_\vartheta(\boldsymbol{\vartheta})\right)$ (NPO) with transformation function

$$h(t|treatment, \boldsymbol{x}) = \boldsymbol{a}(t)^\top \boldsymbol{\gamma}_1 + \mathcal{I}(treatment)\boldsymbol{a}(t)^\top \boldsymbol{\gamma}_2.$$

The second term can be interpreted as the deviation from a constant log-odds ratio treatment effect. The PO model resulted in a WAIC (DIC) of 191.0 (191.0), while the NPO model re-

sulted in a WAIC (DIC) of 192.6 (192.2), indicating that the PO model is preferable in this case. It is straightforward to include nonlinear covariate effects at the cost of more challenging interpretability.

Part C Further Simulation Results

Model	Specification
Lin. BCTM	$\left(\boldsymbol{\vartheta}, \Phi, ((1, y) \otimes (1, \boldsymbol{x}_p^\top))^\top, \pi_\vartheta(\cdot)\right)$
Lin. MLT	$\left(\boldsymbol{\vartheta}, \Phi, ((1, y) \otimes (1, \boldsymbol{x}_p^\top))^\top\right)$
Full BCTM	$\left(\boldsymbol{\vartheta}, \Phi, (\boldsymbol{a}_{10}(y)^\top \otimes (\boldsymbol{b}_{10}(x_1)^\top, \ldots, \boldsymbol{b}_{10}(x_{p+2})^\top))^\top, \pi_\vartheta(\boldsymbol{\vartheta})\right)$
Full MLT	$(\boldsymbol{a}_{\text{Bs},10}(y)^\top \otimes (\boldsymbol{b}_{\text{Bs},10}(x_1)^\top, \ldots, \boldsymbol{b}_{\text{Bs},10}(x_{p+2})^\top))^\top$
Oracle BAMLSS	$\eta_\mu = \beta_0 + x_2 \cdot f_{20}(x_1) + \sum_{k=0}^{p} f_{20}(x_{2+k})$ and $\eta_{\sigma^2} = \beta_0 + f_{20}(x_1)$
BAMLSS QR	$\eta_\tau = \sum_{k=0}^{p} f_{2+k,20,\tau}(x_{2+k})$

Table C.7: Simulation study. Model specifications with corresponding basis dimensions in the subscript. Parameter p specifies the number of noise parameters. For the MLT specifications, the basis functions are Bernstein polynomials and for the other models B-splines. The subscripts denote the number of basis functions each.

Part C.1 Coverage rates

This section contains results for $n = \{100, 500\}$ and effect estimates from the second simulation setting Furthermore, empirical coverage rates of pointwise 95% credible intervals for $n = 500$ are shown in Figure C.2.

Figure C.2: Simulation 2. Coverage rates of pointwise 95% credible/confidence intervals of BCTM and MLT in each covariate value $x_1, \ldots, x_4$ for $n = 500$.

Part C.2 Effect estimates

This section shows effect estimates obtained from the BCTM, the MLT and tramME for $n = 100$ and $n = 500$. Estimation of $f_1(x_1)$ resulted in numerical problems in most replications for tramME.

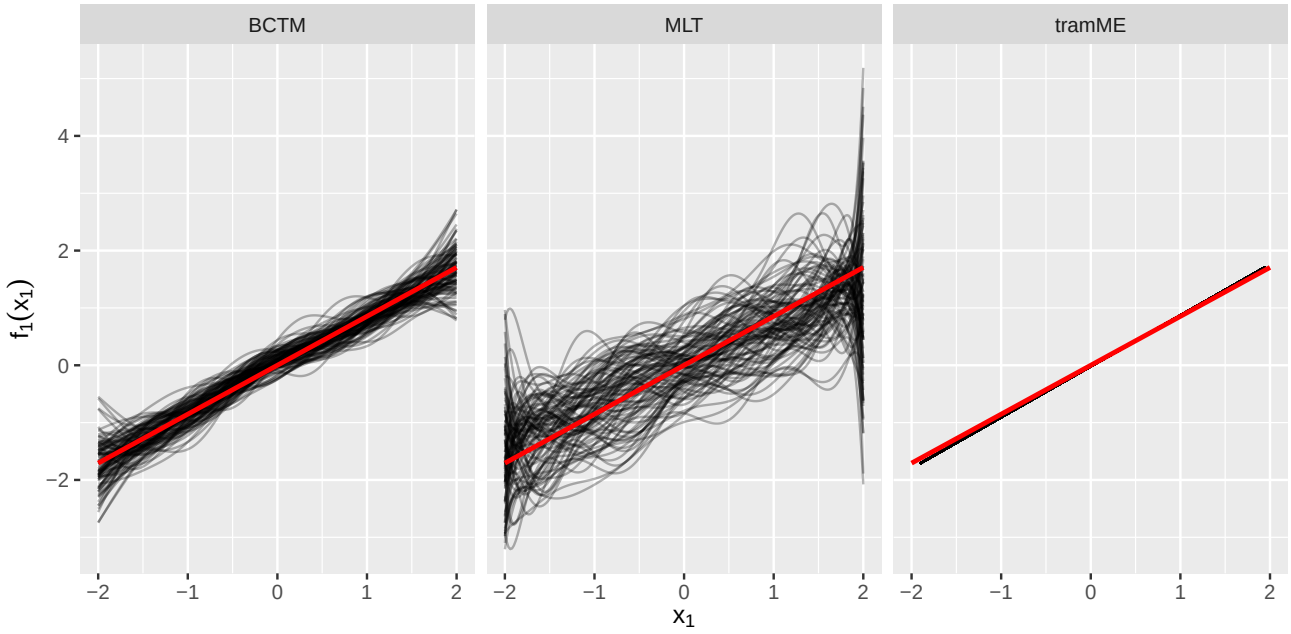

Figure C.3: Simulation 2. $n = 100$. Posterior mean estimates of $f_1(x_1)$ of 100 replications obtained from BCTM and MLT.

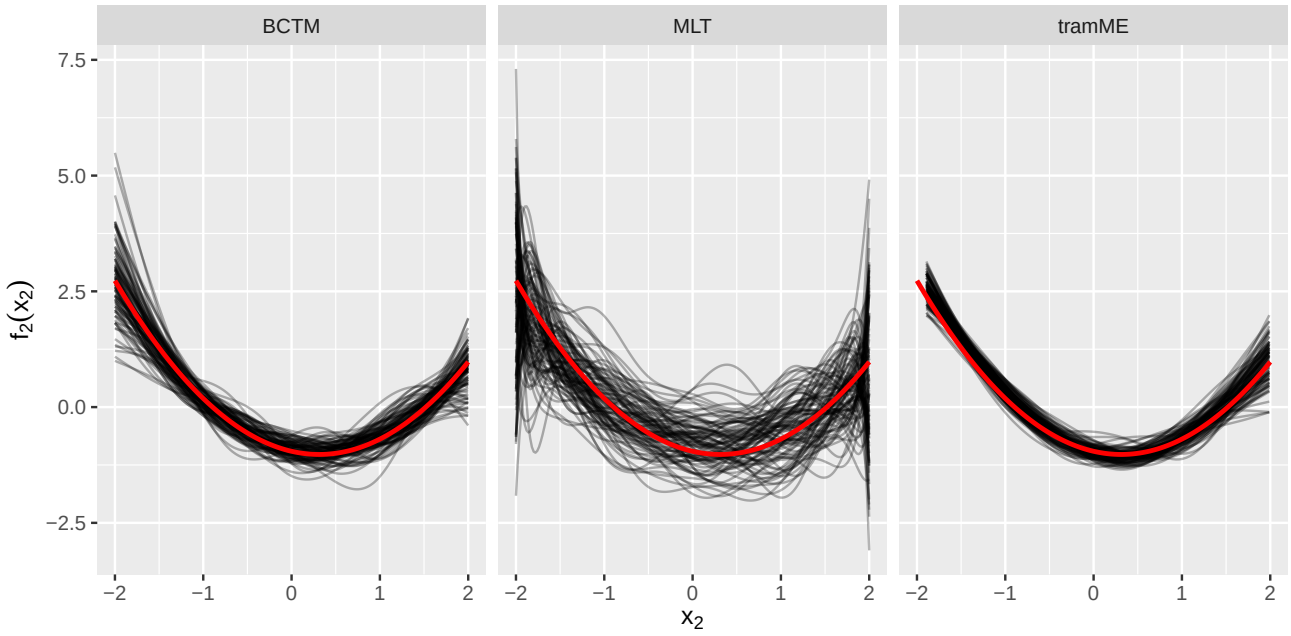

Figure C.4: Simulation 2. $n = 100$. Posterior mean estimates of $f_2(x_2)$ of 100 replications obtained from BCTM and MLT.

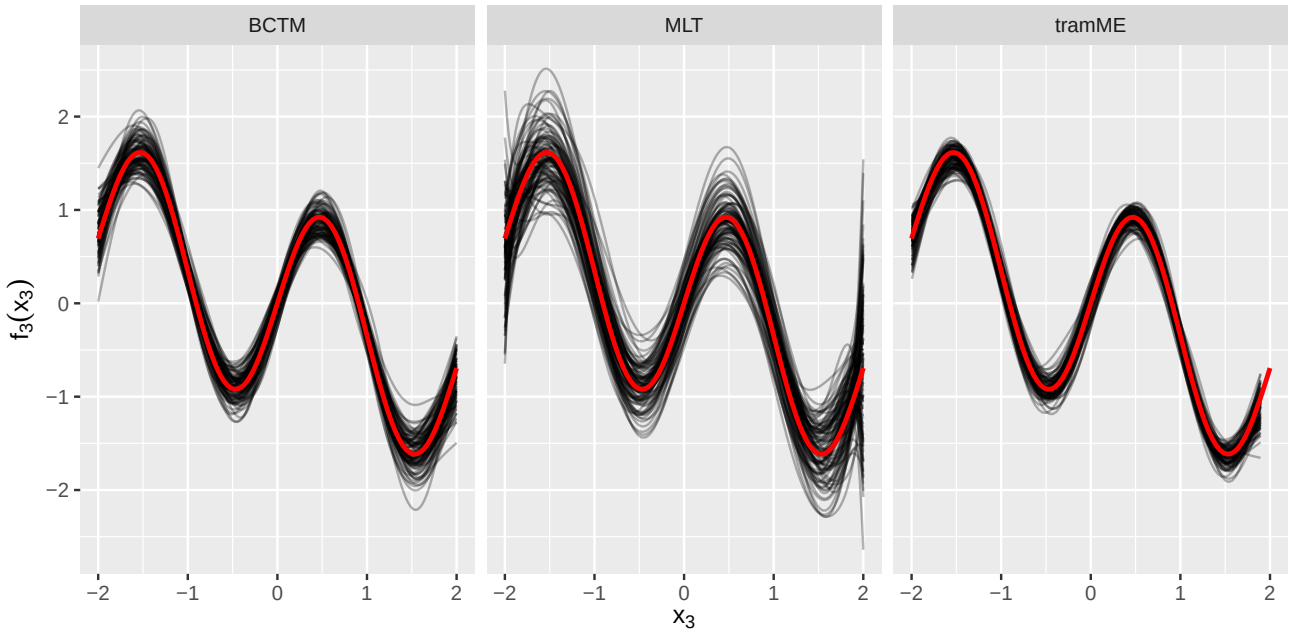

Figure C.5: Simulation 2. $n = 100$. Posterior mean estimates of $f_3(x_3)$ of 100 replications obtained from BCTM and MLT.

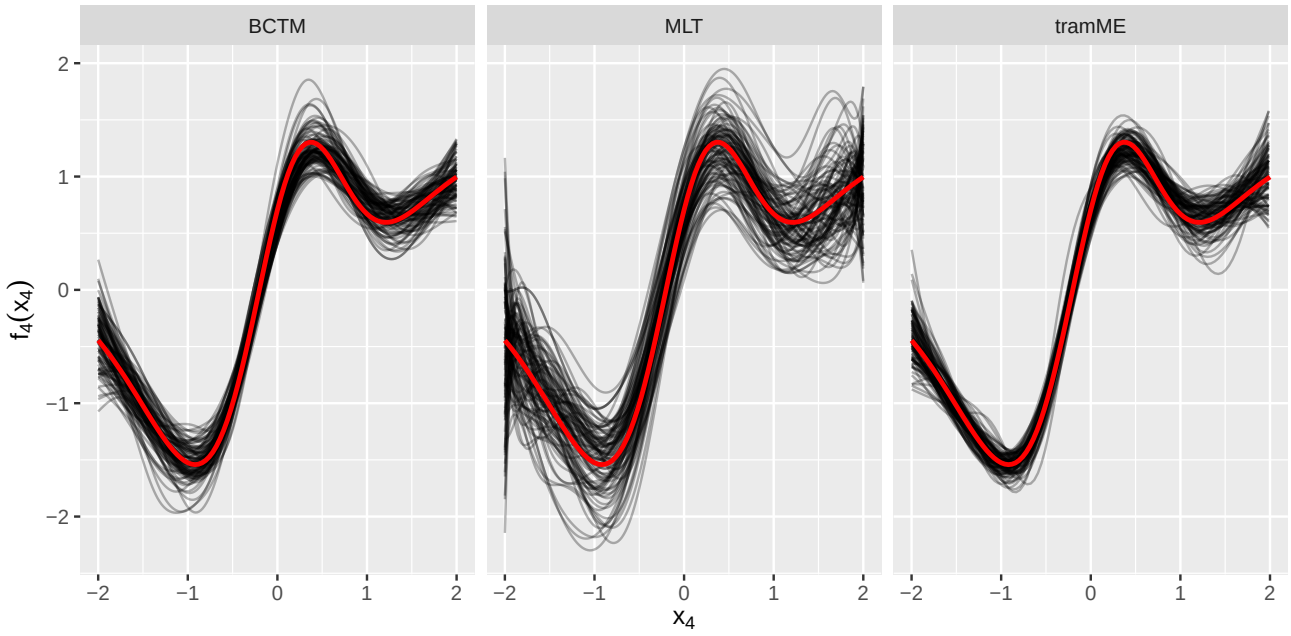

Figure C.6: Simulation 2. $n = 100$. Posterior mean estimates of $f_4(x_4)$ of 100 replications obtained from BCTM and MLT.

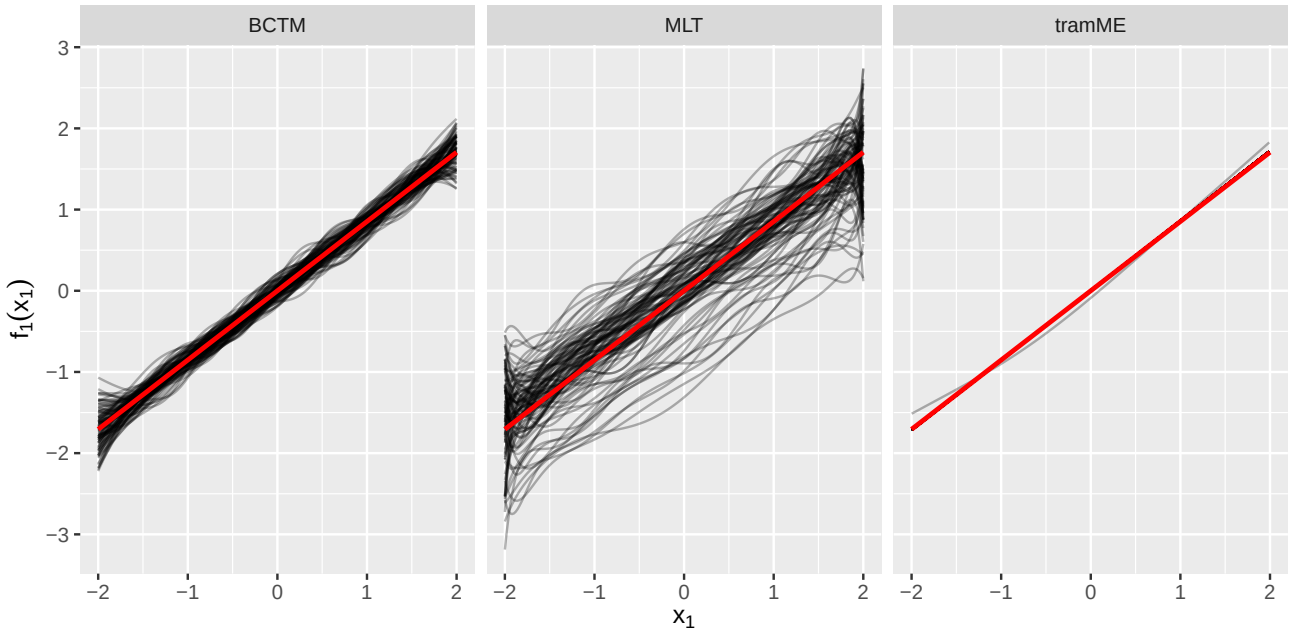

Figure C.7: Simulation 2. $n = 500$. Posterior mean estimates of $f_1(x_1)$ of 100 replications obtained from BCTM and MLT.

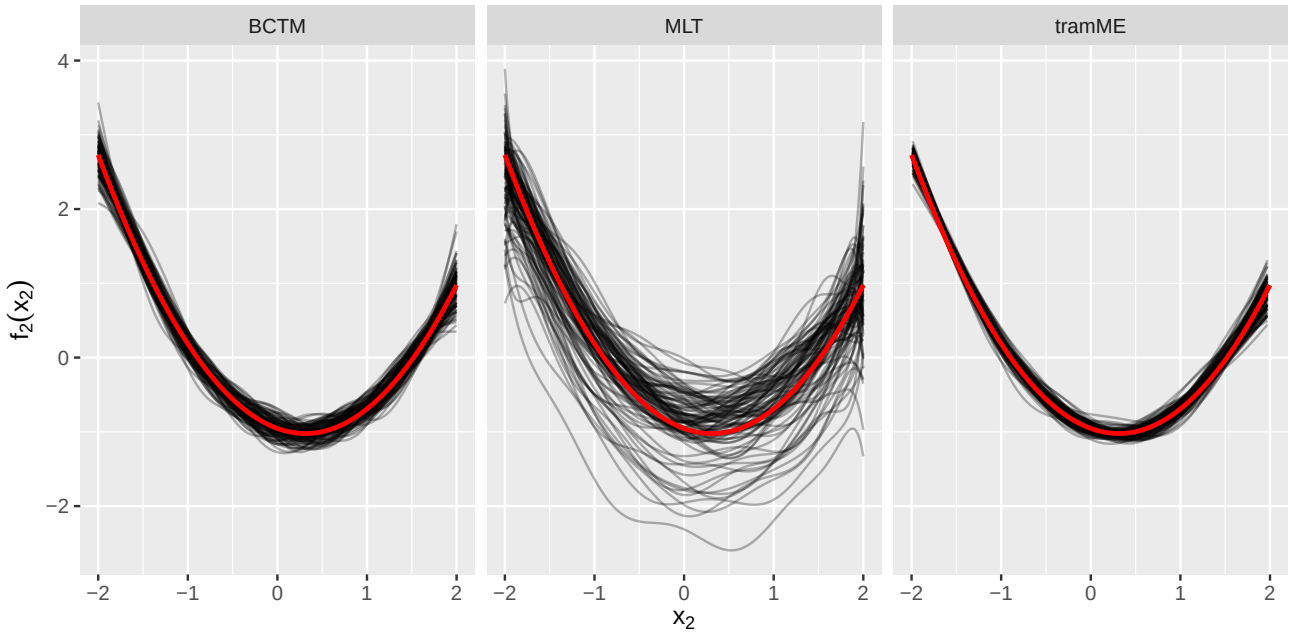

Figure C.8: Simulation 2. $n = 500$. Posterior mean estimates of $f_2(x_2)$ of 100 replications obtained from BCTM and MLT.

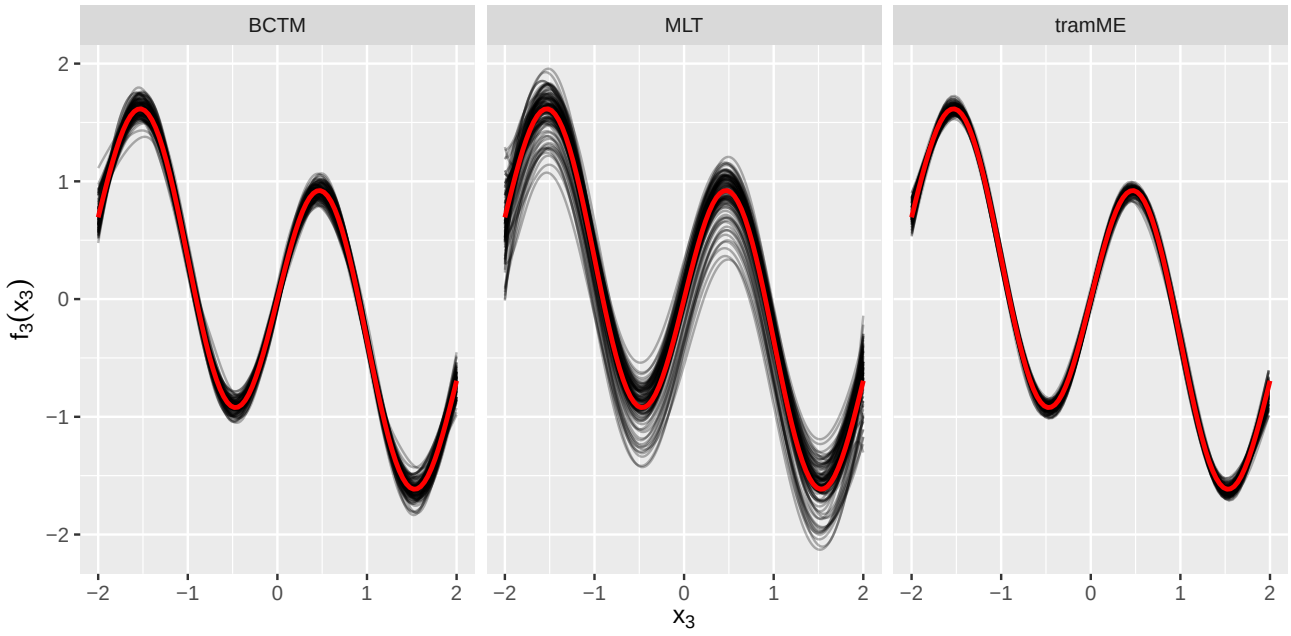

Figure C.9: Simulation 2. $n = 500$. Posterior mean estimates of $f_3(x_3)$ of 100 replications obtained from BCTM and MLT.

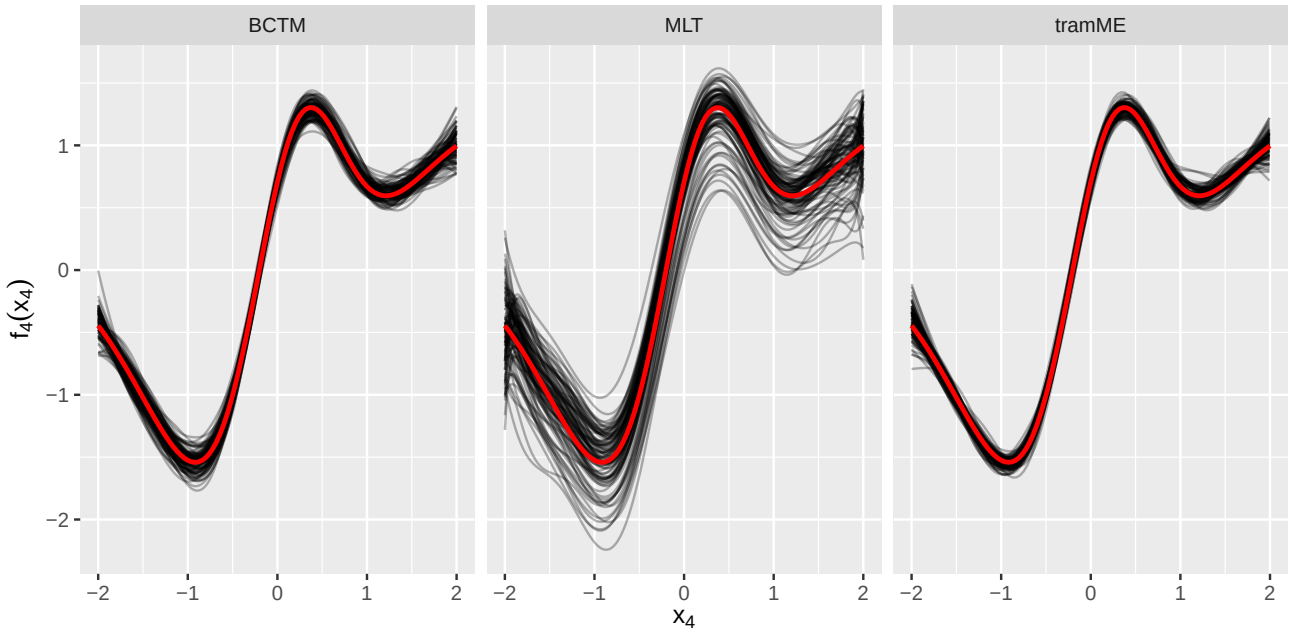

Figure C.10: Simulation 2. $n = 500$. Posterior mean estimates of $f_4(x_4)$ of 100 replications obtained from BCTM and MLT.

References

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical science* **7**(4): 457–472.

Hanson, T. and Yang, M. (2007). Bayesian semiparametric proportional odds models, *Biometrics* **63**(1): 88–95.

Höllig, K. and Hörner, J. (2013). *Approximation and Modeling with B-Splines*, Society for Industrial and Applied Mathematics, USA.

Plummer, M., Best, N., Cowles, K. and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC, *R News* **6**(1): 7–11.
URL: *https://journal.r-project.org/archive/*

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables, *Biometrika* **60**(2): 279–288.

Zhou, H., Hanson, T. and Zhang, J. (2020). spBayesSurv: Fitting Bayesian spatial survival models using R, *Journal of Statistical Software* **92**(9): 1–33.