Web-based Supplementary Materials for the Quantile Function on Scalar Regression Analysis for Distributional Data

Hojin Yang[†], Veerabhadran Baladandayuthapani[†], Arvind U.K. Rao[‡],

and Jeffrey S. Morris[†]

Department of Biostatistics[†]

Department of Bioinformatics and Computational Biology[‡]

The University of Texas MD Anderson Cancer Center, Houston, TX 77030

hyang6@mdanderson.org, Veera@mdanderson.org, ARUppore@mdanderson.org,

and jefmorris@mdanderson.org

Abstract In this supplementary file, we conduct additional simulations and data analysis. We also provide additional explanations and details on estimation.

1.   DETAILS OF ESTIMATION

Proof of Theorem 2.1.   We consider the relation between beta and Binomial distributions

$$\sum_{k=0}^{n} c_{k,n} \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} x^k (1-x)^{n-k} = \sum_{k=0}^{n} \alpha_k \binom{n}{k} x^k (1-x)^{n-k} \tag{1}$$

Notice that the formula in right hand side in (1) is the random Bernstein polynomial of order n. Since $q(x)$ is continuous on the closed and bounded interval $[0,1]$, it is uniformly continuous and thus for any positive ϵ, there exists a $\delta(\epsilon)$ such that $|x - y| < \delta(\epsilon)$ implies $|q(x) - q(y)| < \epsilon$. Fix $x \in [0,1]$. Let $X_1, \ldots, X_n$ be a sample from Bernoulli(x) distribution. Let $\hat{x}_n$ be the sample average, $\hat{x}_n = \sum_{i=1}^{n} X_i / n$. Then, it is easy to see $Eq(\hat{x}_n) = q_n(x)$. Hence,

$$|q(x) - q_n(x)| \leq E|q(\hat{x}_n) - q(x)| + |Eq(\hat{x}_n) - q_n(x)|$$

$$= E\left\{ |q(\hat{x}_n) - q(x)| I(|\hat{x}_n - x| < \delta(\epsilon)) \right\} + E\left\{ |q(\hat{x}_n) - q(x)| I(|\hat{x}_n - x| \geq \delta(\epsilon)) \right\}$$

$$\leq \epsilon + 2\|q\|_\infty P(|\hat{x}_n - x| \geq \delta(\epsilon)),$$

where $\|q\|_\infty = \sup_{x \in [0,1]} |q(x)|$. It follows from Chebychev's inequality that

$$P(|\hat{x}_n - x| \geq \delta(\epsilon)) \leq \frac{x(1-x)}{n\delta(\epsilon)^2} \leq \frac{1}{4n\delta(\epsilon)^2} \quad \text{for all} \ \ x \in [0,1].$$

Hence, we have $\sup_{x \in [0,1]} |q(x) - q_n(x)| \leq \epsilon + 2\|q\|_\infty \frac{1}{4n\delta(\epsilon)^2}$. Letting $n \to \infty$ and then $\epsilon \to 0$ yields

$$\lim_{n \to \infty} \sup_{x \in [0,1]} |q(x) - q_n(x)| = 0, \tag{2}$$

where this implies that $q_n(x)$ converges $q(x)$ uniformly (over $[0,1]$). It follows from the integrability of $q(x)$ by continuity and well known triangle inequality that

$$\left| \int_0^p q(x) - q_n(x) dx \right| \leq \int_0^p |q(x) - q_n(x)| dx.$$

From the result (2), we already know that given ϵ, there exists N such that $|q(x) - q_n(x)| < \epsilon$ for $n > N$ (not depend on x). Therefore, when $n > N$,

$$\left| \int_0^p q(x) - q_n(x) dx \right| \leq \int_0^p |q(x) - q_n(x)| dx \leq p\epsilon.$$

1

Since $\max_{p \in [0,1]} p = 1$, this implies that for any $p \in [0, 1]$,

$$\lim_{n \to \infty} Q_n(p) = \lim_{n \to \infty} \int_0^p q_n(x) = \int_0^p q(x)dx = Q(p),$$

which is what we want to prove. $\square$

1.1   Details of Denoising

In practice, we have observed that the first number of orthogonal basis functions are relatively smooth, but the later basis functions can be quite noisy, sometimes with high-frequency oscillations. As we do not believe these oscillations capture meaningful features of the empirical quantile functions, we regularize the orthogonal basis functions using wavelet denoising to adaptively remove these oscillations.

Given a choice of mother wavelet function $\varphi(p)$, wavelets are formulated by the operations of dilation and translation given by

$$\varphi_{j,l}(p) = 2^{j/2}\varphi(2^j p - l)$$

with integers j, l indicating scale and location, respectively. We can decompose any arbitrary function $\psi_k^\perp(p) \in L^2(\Pi(\mathcal{P}))$ into the generalized Fourier series as

$$\psi_k^\perp(p) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d_{k,j,l}\varphi_{j,l}(p), \tag{3}$$

where $d_{k,j,l} = \int \psi_k^\perp(p)\varphi_{j,l}(p)dp = \langle \psi_k^\perp, \varphi_{j,l} \rangle$ are the wavelet coefficients corresponding to $\psi_k^\perp$. Wavelet coefficient $d_{k,j,l}$ describes features of the function $\psi_k^\perp$ at the spatial locations indexed by l and scales indexed by j. A fast algorithm, the discrete wavelet transform (DWT), can be used to compute these wavelet coefficients in linear time for data sampled on an equally spaced grid whose size L is a power of two, yielding a set of L wavelet coefficients, with L_j wavelet coefficients at each of J wavelet scales and L_0 scaling coefficients at the lowest scale. We apply this wavelet transform to the the basis functions $\psi_k^\perp(p)$ sampled on an equally-spaced fine grid on p, for example using a grid of size $L = 2^{10} = 1024$ for our GBM data.

Functions can be adaptively denoised by shrinking these wavelet coefficients nonlinearly towards zero (Donoho, Johnstone, Kerkyacharian and Picard 1995). Various shrinkage/thresholding rules can be used to accomplish this, such as hard thresholding with a threshold of $\sigma\sqrt{2\log L}$ introduced by Donoho et al. (1995), which yields a risk within a log factor of the ideal risk. In that case, the wavelet shrunken and denoised basis function $\psi_k^\dagger(p)$ can be constructed as

$$\psi_k^\dagger(p) = \sum_{j=0}^{J} \sum_{l=1}^{L_j} d_{k,j,l}^\dagger \varphi_{j,l}(p), \tag{4}$$

such that $d_{k,j,l}^\dagger = d_{k,j,l}$ if $|d_{k,j,l}| > \sigma\sqrt{2\log L}$ and $d_{k,j,l}^\dagger = 0$ If $|d_{k,j,l}| \leq \sigma\sqrt{2\log L}$. When σ is unknown, it is often replaced by an empirical estimator that is the median absolute deviation of the wavelet coefficients at the highest frequency level J.

1.2 Details of MCMC

Motivated by a belief that the covariate effects should be more regular than the empirical quantile functions themselves, we assume sparsity-inducing priors on the B_{ak}^* coefficients. We use a spike-Gaussian slab (Lempers 1971; Mitchell and Beauchamp 1988) distribution. The spike at 0 induces sparsity while the Gaussian prior applies a roughness penalty. Motivated by the belief that certain quantlets are *a priori* more likely to be important for representing covariate effects, we partition the set of K quantlet dimensions into H clusters of basis functions, each with their own set of prior hyperparameters. This allows us, for example, to allow a higher prior probability for certain quantlet dimensions to be important such as the the Gaussian basis levels $\{\psi_1, \psi_2\}$ and the quantlets explaining a high proportion of the relative variability in the empirical quantile functions. Recalling that quantlets are indexed in descending order of their proportion of relative variability explained, we can group together the Gaussian coefficients as one cluster, and then split the rest sequentially into H clusters each containing sets of basis functions whose relative variability explained are of similar order of magnitude.

Let Π be a $K \times J$ matrix with element $\Pi(k, j) = \xi_k(p_j)$ and Φ be a $K \times J$ matrix with element $\Phi(k, j) = \psi_k^\perp(p_j)$ for $k = 1, \ldots, K$ and $J = 1, \ldots, J$, where $\xi_k(p_j)$ and $\psi_k^\perp(p_j)$ are

elements in $\mathcal{D}^\mathcal{C}$ and $\mathcal{D}^\perp$. Note that we use Gram-Schmidt to orthogonalize the basis set $\mathcal{D}^\mathcal{C}$ to generate an orthogonalized basis set $\mathcal{D}^\perp$. Then, we measure the variability for each element as $\mathrm{diag}(\Lambda)$ based on the spectral decomposition structure, $\Lambda = \Phi\Pi^T\Pi\Phi^T$. We can split the elements in $\psi_k^\perp(\cdot) \in \mathcal{D}^\perp$ into H clusters each containing sets of basis functions sharing a similar variability explained, where the hierarchical cluster algorithm on the variability is used to determine it. Although we use the quantlets basis function $\psi_k(\cdot)$ which is obtained by denoising and re-standardizing for the corresponding element $\psi_k^\perp(\cdot)$, the identical clustering information is used in the estimation procedure.

Specifically, let $\mathcal{K}$ be a set of indices $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{H}$ be a set of indices $\mathcal{H} = \{1, \ldots, H\}$ such that $\mathcal{K} = \cup_{h=1}^{H}\mathcal{K}_h$ with $\mathcal{K}_h \cap \mathcal{K}_{h'} = \emptyset$ for $h \neq h'$, where $\mathcal{K}_h = \{1 \leq k \leq K | f(k) = h\}$ for the clustering map, $f(k) = h$. Let $\mathcal{H}_h = \{\pi(k) | k \in \mathcal{K}_h\} \equiv \{1, \ldots, |\mathcal{K}_h|\}$ be the ordered set consisting of the integers such that for all k and $k' \in \mathcal{K}_h$, if $k < k'$ implies $\pi(k) < \pi(k')$. By defining the index $h_{k,l}$ to indicate the quantlets ψ_k as the lth component of $\mathcal{H}_h$ within the h cluster, the prior on B_{ak}^* is given by

$$B_{ak}^* \equiv B_{ah_k,l}^* \sim \gamma_{ah_k,l}N(0, \tau_{ah_kl}^2) + (1 - \gamma_{ah_k,l})I_0 \tag{5}$$

$$\gamma_{ah_k,l} \sim \text{Bernoulli}(\pi_{ah}),$$

where I_0 is a point mass distribution at zero, and $\gamma_{ah_k,l}$ is an indicator of whether the kth quantlet basis coefficient is important for representing the effect for the ath covariate within the h cluster as the lth component. The hyperparameter π_{ah} indicates the prior probability that a quantlet coefficient in set $\mathcal{K}_h$ is important, and $\tau_{ah_k,l}^2$ the prior variance, and regularization factor, for coefficient B_{ak}^* conditional on it being chosen as important.

In order to fit model in the quantlet space model using a Bayesian approach, we also need to specify priors on the variance components $\{\sigma_k^2, k = 1, \ldots, K\}$. We place a vague proper inverse gamma prior on each diagonal element σ_k^2 given by $\sigma_k^2 \sim \text{inverse-gamma}(\nu_0/2, \nu_0/2)$, where ν_0 is some relatively small positive constants. Other relatively vague priors could also be used. If one wanted to allow Σ^* to be unconstrained, an Inverse Wishart prior could be assumed for the $K \times K$ matrix. The likelihood funciton is given $\boldsymbol{Q}_{.k}^* \sim N(\boldsymbol{X}\boldsymbol{B}_{.k}^*, \sigma_k^2 I)$ in the

4

projected space for each $k = 1, \ldots, K$.

The parameters π_{ah} and $\tau^2_{ah_k,l}$ can be estimated using an empirical Bayes method, assuming $\tau^2_{ak} \equiv \tau^2_{ah_k,l} = V_{ah_k,l} \Gamma_{ah}$ for some parameters Γ_{ah}, which allows for full flexibility in these regularization parameters within the $\mathcal{K}_h$ group for the ath covariate. This structure also enables us to integrate out the *quantlets* coefficients and compute the marginalized likelihood for Γ_{ah} and π_{ah} as:

$$l(\pi_{ah}, \Gamma_{ah}) \propto (1 + \Gamma_{ah})^{-\sum_{l \in \mathcal{H}_h}^{|\mathcal{K}_h|} \gamma_{ah_k,l}/2} \exp\left\{ -1/2 \sum_{l \in \mathcal{H}_h}^{|\mathcal{K}_h|} \zeta^2_{ah_k,l} \gamma_{ah_k,l}/(1 + \Gamma_{ah}) \right\}$$
$$\times (\pi_{ah})^{\sum_{l \in \mathcal{H}_h}^{|\mathcal{K}_h|} \gamma_{ah_k,l}} (1 - \pi_{ah})^{|\mathcal{K}_h| - \sum_{l \in \mathcal{H}_h}^{|\mathcal{K}_h|} \gamma_{ah_k,l}}.$$

On the marginalized likelihood, the MLEs of Γ_{ah} and π_{ah_k} can be obtained by

$$\hat{\pi}_{ah} = \sum_{l \in \mathcal{H}_h}^{|\mathcal{K}_h|} \gamma_{ah_k,l}/|\mathcal{K}_h|, \quad \widehat{\Gamma}_{ah} = \max\left(0, \sum_{l \in \mathcal{H}_h}^{|\mathcal{K}_h|} \zeta^2_{ah_k,l} \gamma_{ah_k,l}/\sum_{l \in \mathcal{H}_h}^{|\mathcal{K}_h|} \gamma_{ah_k,l} - 1\right)$$

$$\widehat{O}_{ah_k,j} = \frac{\hat{\pi}_{ah}}{1 - \hat{\pi}_{ah}}(1 + \widehat{\Gamma}_{ah})^{-1/2} \exp\left\{ \frac{1}{2} \zeta^2_{ah_k,l} \frac{\widehat{\Gamma}_{ah}}{1 + \widehat{\Gamma}_{ah}} \right\}, \quad \hat{\gamma}_{ah_k,l} = \frac{\widehat{O}_{ah_k,l}}{1 + \widehat{O}_{ah_k,l}}.$$

These empirical Bayes estimates can be computed for each iteration of the MCMC procedure. The Bayes estimates of π_{ah} and $\tau_{ah_k,l}$ are given by $\hat{\pi}_{ah}$ and $\widehat{V}_{ah_k,l} \widehat{\Gamma}_{ah}$.

1.3 Details of Predicted PDF

If desired, one can construct the estimate of the conditional probability density function given covariates $\boldsymbol{X}$ from

$$\hat{f}(x|\boldsymbol{X}) = M^{-1} \sum_{m=1}^{M} \delta / \left(\boldsymbol{X}^T \hat{\beta}^{(m)}(p) - \boldsymbol{X}^T \hat{\beta}^{(m)}(p - \delta) \right),$$

where δ is a fixed positive constant and $x = \inf(y : y \geq \boldsymbol{X}^T \hat{\beta}^{(m)}(p))$. Note that the above formula is derived from $f(Q(p)))dQ(p)/dp = 1$ by changing the variable. Remark that the conditional quantile function $X_i^T \hat{\beta}^{(m)}(p)$ for some samples may not enforce strict monotonicity, which leads to the negativity density value for its computation. However, when we take the coarse grid points for x with the sufficient gap between any two adjacent points, equivalently

δ to be the large value, which would allow $X_i^T \hat{\beta}^{(m)}(p) - X_i^T \hat{\beta}^{(m)}(p - \delta) > 0$, the valid probability density function is available. In practice, we use $\max\left(0, X_i^T \hat{\beta}^{(m)}(p) - X_i^T \hat{\beta}^{(m)}(p - \delta)\right)$ as the differential for $Q(p)$ for an arbitrary δ.

2. OTHER RESULTS FROM SIMULATION

2.1 Simulation for Basis Representation

We let N, T, G, D, M_1, M_2, M_3, and M_4 be random variables from the standard normal, student-t, gamma, dirichlet distribution, and mixture distributions, respectively. We first generated four profiles of quantile functions, $Q_N(p)$, $Q_T(p)$, $Q_G(p)$, $Q_D(p)$, $Q_{M_1}(p)$, $Q_{M_2}(p)$, $Q_{M_3}(p)$, and $Q_{M_4}(p)$ defined on $\mathcal{P} = [\delta, 1 - \delta]$ and J fixed grid points $\{p_1, \dots, p_J\} \in \mathcal{P}$, where δ and J were set to be 1/1000 and 999, respectively. Specifically, we independently simulated each $Q_{(\cdot)}(p)$ according to the generating process

$$Q_{(\cdot)}(p) = \inf\left(y : F_{(\cdot)}(y) \geq p\right), \tag{6}$$

following the normal distribution $N(0, 1)$, student t distribution with one degree of freedom, shifted gamma distribution with shape 3 and scale 1 to -3, dirichlet distribution with base measure to be the kernel density estimator of the first observation in GBM data, two mixture skewed-normal distributions for which the mixture components were $SN(-3.06, 3.67, 0)$ and $SN(9.11, 7.89, -4)$ with 0.5 and 0.5 probabilities for one while $SN(-7.1, 2.4, 0)$ and $SN(-3.11, 7.89, 4)$ with 0.3 and 0.7 probabilities for the other, and two mixture normal distributions for which the mixture components were $N(-2.5, 2.5)$, $N(4, 3)$ and $N(9.5, 2.1)$ for one while $N(-2.5, 1.5)$, $N(4, 3.56)$ and $N(9.5, 1.1)$ for the other with 0.3, 0.5 and 0.2 probabilities, respectively. To illustrate, two panels of Figure 2 show eight probability density trajectories for $F_N(y)$, $F_T(y)$, $F_G(y)$, $F_D(y)$, $F_{M_1}(y)$, $F_{M_2}(y)$, $F_{M_3}(y)$, and $F_{M_4}(y)$. We see that each distribution has different characteristic in that $F_T(y)$ has heavy tails, $F_G(y)$ is skewed to the right, $F_D(y)$ has high frequency, and $F_{M_k}(y)$ has multiple peaks for $k = 1, \dots 4$, compared to $F_N(y)$, which is standard in this scenario.

We constructed a *quantlets* representation for each distribution as follows. We generated

the parameter space as a set of a sequence pairs, $\Theta = \{\theta_k = (a_k, b_k)\}_{k=1}^{11,881}$, uniformly sampled

on $[0.1, 1000]^2$ and an *overcomplete dictionary*, $\mathcal{D}^O = \{\xi_k : \theta_k \in \Theta\}$, where ξ_2 (Gaussian) is

not included in $\mathcal{D}^O$ to allow for a fair comparison in this scenario so that $P_{N\perp}$ is the identity

operator as the orthogonal complement to the empty space. We restrict the maximum value

of the parameter space as the total number of the probability gird points J and the minimum

value of the probability grid points p_1. We prefer to have a large number of K^O but restrict

its minimum as $K^O > J$. This setting is motivated by the structure of the random Bernstein

polynomials (Petrone 1999; Bornkamp and Ickstadt 2009). We used the Lasso method to

find the individual *dictionary*, $\mathcal{D}_i$ and used it as *quantlets* from $\mathcal{D}^O$ for each distribution case.

Note that we did not need to find $\mathcal{D}^U$ or $\mathcal{D}^C$ because there was only a single observation for

each distribution case.

We compared our *quantlets* with other competing representations such as B-spline (Schumaker

2007), integrated spline (Ramsay 1988) denoted by I-spline and convex spline (Meyer 2008)

denoted by C-spline basis representations. In Figure 4, we see the shape of I-spline and

C-spline basis functions. To compare the different methods, we examined one type of perfor-

mance measure for prediction accuracy and computed the empirical mean integrated squared

error of the test data set as $n^{-1} \sum_{i=1}^n (Q_{(\cdot)}(p_i)^{test} - \widehat{Q}_{(\cdot)}(p_i)^{test})^2$, where $\widehat{Q}_{(\cdot)}(p_i)$ is the pre-

dicted basis representation built from the aforementioned training set of 300 grid points

and $(Q_{(\cdot)}(p_1)^{test}, \ldots, Q_{(\cdot)}(p_n)^{test})$s' are observations in the test set of 999 grid points for

each distribution case. We used a B-spline basis of order 4, degree 3 on 10 knots uni-

formly spaced in $\mathcal{P}$ to generate I-spline and C-spline in this simulation. Also, to investi-

gate the degree of monotonicity, we compute the degree of ϵ-monotonicity, defined to be

$P_\epsilon^{\mathcal{M}} = \int_0^1 I[\widehat{Q}_{(\cdot)(p)} - \max_{p'<p}\{\widehat{Q}_{(\cdot)}(p)\} > \epsilon]dp$ for some ϵ considered negligibly small in the

context of the scale of Y in the current data set. When $P_\epsilon^{\mathcal{M}} \approx 0$, it says a strong monotonicity

and we use the monotonicity measure $1 - P_{0.01}^{\mathcal{M}}$ for each distribution.

Table 3 presents the empirical mean integrated squared error for each method calculated

and ϵ-monotonicity from the test data sets of 999 grid points on $\mathcal{P}$. We see that our method

significantly outperformed the competing methods for all distribution cases. Although B-

spline method showed lower IMSE compared to other competing methods, it does not show the monotone property in that it shows wiggly fitted regions near the curvature points. Hence, we conclude that our method is better than the existing spline approaches. Figure 3 plots the true quantile functions (gray dot line) along with the fits for quantlets (red), B-spline (dashed), I-spline (dot) and C-spline (dashed-dot) and shows that the quantlets can be the best representation for the shapes of these quantile functions compared to all others.

2.2   Other Additional Results from Simulations

There are additional results for the simulation conducted in the main paper. We ran the MCMC algorithm for $2,000$ iterations, keeping every one after a burn-in of 200 and then, transformed all the estimates in the *quantlets* space into the *data* space. The results are shown in Figure 9 based on each method: (A) naive quantile regression method (separate classical quantile regressions for each p by using *rq* function in *quantreg* R package (Koenker 2005)), (B) naive quantile functional regression approach (separate functional regressions for each subject-specific quantile p), (C) principal components method (quantile functional regression using PCs as basis functions), (D) *quantlet* without sparse regularization, (E) *quantlet* with sparse regularization, and (F) Gaussian model (quantlet approach but keeping only the first two coefficients). Compared to the other methods, our method (E) provided smoother estimators and tight 95% joint confidence intervals for all the parameters. Figure 11 depicts the simultaneous band scores $P_{SimBaS}(p)$ for the two contrast functions associated with the scale effect $\beta_3(p)$ and skewness effect $\beta_4(p)$, with regions of p. Since the two true contrast functions $\beta_3(p)$ and $\beta_4(p)$ have one and two zero points while the null hypothesis is $\beta_a(p) = 0$, respectively, as shown in Figure 9, those points need to be detected with the higher $P_{SimBaS}(p)$ at their zero points (not reject the null hypothesis). Compared to the other methods, our method (E) showed lower type II error at the level of significance, $\alpha = 0.05$ (solid black line).

Table 4 includes true conditional moment parameters such as the mean, standard de-

8

viation and skewness, and corresponding point and interval estimators for the four groups derived from the closed form of the formulas in (see Section 2.7) for each method. Although the performance of the point estimators seems to be similar for all cases, the performance of the interval estimators is clearly better when using the *quantlets* basis approaches compared to the naive approach because those intervals contain true parameters for the four groups. Summarizing Gaussianity score for the four groups, assessed by the relative energy, is reported in Table 5. We see that the first three groups not involved with the skew parameter α can be explained as Gaussianity with the higher score whereas the fourth group, which is involved with the skew parameter, is hard to explain by Gaussianity, where values for the 95% confidence intervals are reported in parentheses. Figure 5 reveals the quantlets basis functions in the simulation.

Because our method involves a lot of computational burden, we investigated the computational aspect of our method. Figure 23 depicts the run time for computing the basis set as the function of sample size (N) and probability grid size (m_i) from the simulated data in the multi modality scenario. We see that for the grid size, $m_i \leq 2^8$, the number of the subjects does not yield the heavy computation. However, for the grid size, $m_i \geq 2^{10}$, the amount of the computations is dramatically increased as the number of the subject increase. Hence, we recommend to use the smaller number of the probability grids for the data set with the large number of the observations.

2.3 Multi Modality Scenario

We also conducted the additional simulation based on multi modality scenario, in order to see the performance of our method as a balanced assessment. Specifically, we generated random samples for four groups of subjects whose mean quantile function was assumed to be from four mixture distributions, where two mixture skewed normal distributions consist of $SN(-3.06, 3.67, 0)$ and $SN(9.11, 7.89, -4)$ with 0.5 and 0.5 probabilities for one while $SN(-7.1, 2.4, 0)$ and $SN(-3.11, 7.89, 4)$ with 0.3 and 0.7 probabilities for the other, and two mixture normal distributions consist of $N(-2.5, 2.5)$, $N(4, 3)$ and $N(9.5, 2.1)$ with 0.3,

0.5 and 0.2 probabilities for one while $N(-2.5, 1.5)$, $N(4, 3.56)$ and $N(9.5, 1.1)$ for the other with the same probabilities.

Panels A, B and C of Figure 1 show the densities, mean quantile functions by group, and quantile functional regression coefficient estimates, respectively, corresponding to these distributions, where all observed quantile functions are depicted with the gray-lines in each panel. All other settings in this simulation are the same to those of the first simulation for the model equation, covariates, noise process, and sample size. We chose a *common* set, $\mathcal{D}^{\mathcal{C}}$, that retained 17 basis functions, which resulted in a near-lossless basis set with $\rho^0 = 0.997$ (see $K_{\mathcal{C}} = 17$ in panel D). After orthogonalization, denoising, and re-standardization, and the fitted *quantlet* projection almost perfectly coincided with the observed data for all of the empirical quantile functions (panel E). After running the MCMC algorithm, posterior estimate for each $\beta_a(p)$ is contained as the dashed-line in panel F of Figure 1.

As expected, Table 1 shows that our method (E) outperforms all other competing methods in that it leads to tighter bands with good coverage. Also, From Table 2, we see that test results found using our method (E), were near identical to true results and our method does not lose power relative to feature extraction approaches (G) when the distributional differences are indeed contained in the moments, where (μ, σ, ξ) for each group is set to be $(-0.06, 5.30, 0.02)$, $(-0.07, 6.40, 0.39)$, $(3.05, 5.05, -0.05)$, and $(3.05, 5.03, 0.07)$, respectively.

3. OTHER RESULTS FROM APPLICATION

There are additional results for the GBM study conducted in the main paper. Compared to principal components basis function (Figure 6), the quantlets (Figures 5 and 8) have some level of interpretability in that the first two basis functions define the space of all Gaussian quantile functions (ψ_1, and ψ_2). We see that in Figure 7 orthogonal basis $\psi_k^{\perp}$ (black line) is wiggly up and down, compared to quantlets ψ_k (blue line). Also, note that the next two quantlets for the GBM data seem to pick up on fundamental distributional characteristics like the kurtosis and skewness (ψ_3, and ψ_4). For Gaussian data, only the first two basis functions will be needed, while comparing with dimensions $k = 3, \ldots, K$ provides a measure
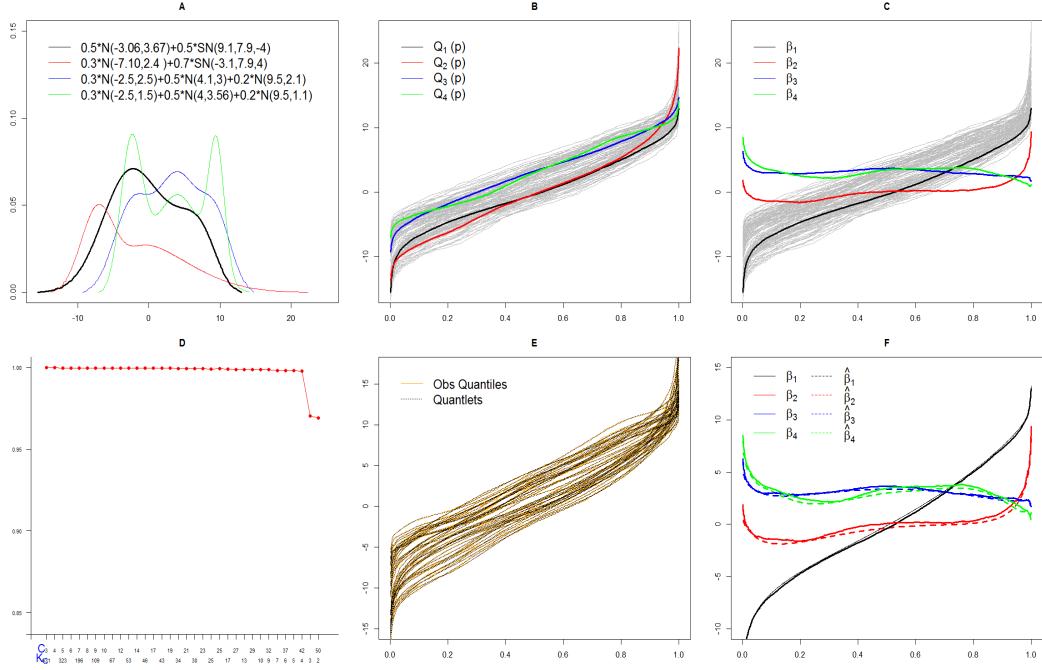
Table 1: Results for Simulation 2: Area and coverage for the joint 95% confidence intervals: (A) naive quantile regression approach, (B) naive quantile functional regression approach, (C) principal component method, (D) *quantlet* space without sparse regularization, (E) *quantlet* space with sparse regularization, and (F) Gaussian *quantlet* space approach.

Type	A	B	C	D	E	F
$\beta_1(p)$	3.528 (0.160)	2.041 (0.094)	0.296 (0.001)	2.821 (0.062)	1.012 (1.000)	2.813 (0.047)
$\beta_2(p)$	3.505 (0.363)	1.981 (0.229)	2.360 (0.062)	2.648 (0.059)	1.492 (1.000)	2.918 (0.127)
$\beta_3(p)$	3.367 (0.593)	1.954 (0.529)	2.855 (0.003)	4.282 (0.101)	1.488 (1.000)	0.536 (0.052)
$\beta_4(p)$	3.384 (0.485)	1.988 (0.332)	2.292 (0.023)	5.039 (0.630)	1.556 (1.000)	0.601 (0.001)

Table 2: Simulation 2: Testing for conditional moment statistics in simulation: (A) naive quantile regression approach, (B) naive quantile functional regression approach, (C) principal component method, (D) *quantlet* space without sparse regularization, (E) *quantlet* space with sparse regularization, (F) Gaussian *quantlet* space approach, and (G) feature extraction approach, where the values in this table are the posterior probability scores derived by its corresponding method for each test (the first column).

H_0	True	A	B	C	D	E	F	G
$\mu_1 = \mu_2$	$\mu_1 = \mu_2$	0.000	0.000	0.000	0.000	0.102	0.000	0.096
$\mu_3 = \mu_4$	$\mu_3 = \mu_4$	0.000	0.000	0.352	0.000	0.274	0.000	0.260
$\sigma_1 = \sigma_2$	$\sigma_1 \neq \sigma_2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\sigma_3 = \sigma_4$	$\sigma_3 = \sigma_4$	0.000	0.000	0.347	0.000	0.381	0.352	0.074
$\xi_1 = \xi_2$	$\xi_1 \neq \xi_2$	0.000	0.000	0.328	0.120	0.004	0.438	0.000

Figure 1: Simulated data in the multi modality scenario and their *quantlet* representations: (A) density functions of the population, (B) mean quantile functions by group and (C) quantile functional regression coefficients, (D) the near-lossless criterion varying with the different number of basis functions, (E) the relation between empirical quantile functions and *quantlet* fits, and (F) posterior estimates for each $\beta(p)$.

of the degree of *non-Gaussianity* in the distribution.

The summarizing Gaussianity score for the specific or reference group assessed by the relative sum is reported in Table 6. For instance, we see that the treatment group with the event time less than 12 months can be summarized as the higher Gaussian score compared to its reference group. It was hard to explain the quantile trajectories of the male group or the group without mesenchymal status as a Gaussian quantile process because their scores explained by the normal quantile process were not relatively high, which requires a nonparametric quantile process generated by mixed beta distributions to fully understand the entire quantile process.

The main results presented in the paper may depend on several modeling choices, containing the number of quantlets basis functions and determining the prior specification for

ν. Hence, we have conducted sensitivity analysis under different modeling choices. Figures 12 to 17 contain the posterior inference for functional coefficients for GBM data set: for each covariate (6), the left panel includes posterior mean estimate, point and joint credible bands, GBPV in heading along with SimBas less then .05 (orange line), and the right panel includes predicted densities for the two levels of the covariate along with the posterior probability scores for the moment different testings, where Figure 15 presents the naive quantile functional regression approach. We also see that it does not produce different the results for varying ν in Figures 16 and 17. Tables 7 and 8 show specific results to assess sensitivity for a wide range of possible values of K as well as different values of ν, where they include global Bayesian p-values, run times along with area of the joint 95% confidence intervals. Figure 20 depicts the functional coefficients if one naively applies regular (population) quantile regression methods across various quantiles p, and demonstrates that this approach gives nonsensical approaches for our application. As MCMC Diagnostic, Figure 18 contains Gewekes diagnostic histograms (Geweke et al. 1991) for four models. Under the null hypothesis of convergence, we would expect a uniform distribution of p-values. We do not see any enrichment of small p-values in these histograms, suggesting the chain converged. The diagnostics are given for (A) model 1 (K=194), (B) model 2 (K=27), (C) model 3 (K=7), and model 4 (K=2).

We lastly conducted a sensitivity analysis for lasso to see how selection of more or fewer dictionary elements via larger or smaller lasso parameters effects the ultimate number of quantlets. The three panels of Figure 24 show the common basis as the results from the choices including the large penalty (A), current penalty (B), and small penalty (C) of the lasso in GBM data. We see that the path of concordance value was different from each case and the more sparse selection resulted in the smaller possible basis choices, and vice versa for the reason that the possible basis choice is represented by the number of the points in each panel. However, by the current near-lossless criteria (horizontal line), we can reduce this variability to 15, 27, and 38 basis functions for each case. Also, from Figures 11, 21 and 22 in Supplementary material, we see that there are not dramatic changes on the final

results.

4.   **INVESTIGATION OF MONOTONICITY**

By definition, quantile functions are monotone non-decreasing, since any decreases in the quantile function would correspond to negative probability densities. There are a number of nonparametric smoothing methods in existing literature that impose monotonicity constraints on the functions, including integrated splines (Ramsay 1988) (I-splines) and convex splines (Meyer 2008) (C-splines), which are adaptions of B-spline basis functions that enforce monotonicity. A natural thought would be to consider utilizing basis sets like these for the quantile regression framework that could strictly enforce the monotonicity constraints. We considered this, but chose to use quantlets instead for several reasons.

As an illustration in Subsection (2.1), we generated empirical quantile functions from eight different parametric distributions and then fit C-spline, I-spline, and quantlet models to these data. The IMSE is orders of magnitude smaller for quantlets than I-splines or C-splines. We consider the flexibility of the basis set to capture the features of the data and the quantile functional regression coefficients $\beta_a(p)$ to be crucially important to this framework, so even a basis that constrains strict monotonicity may not be preferable if it lacks sufficient flexibility.

Second, from a modeling standpoint, in the quantile functional regression framework, monotonicity would have to be enforced for any possible combination of covariates $x_a, a = 1, \ldots, A$, a cumbersome and impractical constraint to impose.

Third, since the quantlets are constructed from empirical quantile functions that are by definition monotone non-decreasing, we have found that in practice, our quantile functional regression framework tends to lead to virtually monotone predicted quantile functions for the various combinations of covariates. While we would be concerned about a model producing gross non-monotonicities, we are not especially worried about very small magnitude non-monotonicities in the predicted values.

It may be possible to adapt our quantlet basis in some manner to enforce strict mono-

tonicity, but we leave that effort for future work.

To investigate the degree of monotonicity afforded by the model, we construct predicted quantile functions for a broad range of covariate values (exhaustively if possible), and compute the degree of ϵ-monotonicity, defined to be $P_\epsilon^{\mathcal{M}}(X) = \int_0^1 I[\widehat{Q}(p|X) - \max_{p'<p}\{\widehat{Q}(p|X)\} > \epsilon]dp$ for some ϵ considered negligibly small in the context of the scale of Y in the current data set. We have found in our simulations and real data analyses that $P_\epsilon^{\mathcal{M}} \approx 0 \; \forall X$, so it appears that for practical purposes, there is not a strong monotonicity problem in the models we have fit. If $P_\epsilon^{\mathcal{M}}(X)$ is large for a given model, then one should carefully assess the model fit before scientifically interpreting its results.

We reported the empirical rates of the ϵ-monotonicity as $1 - n^{-1}\sum_{i=1}^n P_\epsilon^{\mathcal{M}}(X_i)$ for our simulation and GBM data sets in Table 9, where n is the number of the possible levels of the predicted covariates, X_i. We first generated 30 additional predictors u_{ij} from the uniform distribution defined on $(0, 1)$ and replaced δ_{ij} by u_{ij} for $j = 2, 3, 4$, and $i = 1, \ldots, 10$ in the simulation and generated 82 predictors as possible combinations of the discrete variables at the age evaluated by minimum, Q1, Q2, Q3, mean, or maximum ages in the GBM data. Based on the ranges of the observed data sets which were given as $(-20, 20)$ and $(0, 100)$ for the simulation and GBM data, respectively, we set values for ϵ as shown in Table 9. We see that the fitted quantile functions based on our approach show the monotonicity with the small scale of ϵ compared to the range of the original data set. Figure 19 shows that the predicted quantile functions with bands for each level of the covariates in GBM dataset. We also see that they have ϵ-monotonicity in that their quantile functions have valid shape as the quantile function (nondecreasing shape).

5. SOFTWARE FOR IMPLEMENTATION

We provide description of the overall procedure to fit the quantile functional model and obtain inferential results for the simulation and real application. We upload **QFM.zip** file includes all the plots, estimates, and other inference results to reproduce works in this article. Among all files, we recommend to use the quantlets file, which produces the optimal

15

quantlets basis function as the output for the input data set at a given probability grid of values under the options ("irregular" or "regular"), one that computes empirical quantiles based on the length of each observation and another that does it based on the common length across all observations. The quantlets function requires the glmnmet function in *glmnet* R package (Friedman, Hastie and Tibshirani 2010) to figure out the union set of dictionary, gramSchmidt function in *pracma* R package (Borchers 2015) to obtain the orthonormal set for the common basis set, and wst function in *wavethresh* R package (Nason 2010) to utilize the non-decimated wavelet shrinkage method after the set of the *overcomplete* dictionary was generated in the way described in the Section 2.

Once we obtain the *quantlets* basis function, we can deal with it as the basis function and develop it to fit the functional regression model. There are several possible ways to estimate the unknown parameters and obtain the posterior samples to produce the further inferential results. One possible way is to fit the quantlet-space functional regression model as part of the Bayesian functional mixed model (BayesFMM) packages that have been developed in recent years (Morris and Carroll 2006; Zhu, Brown and Morris 2011; Zhu, Brown and Morris 2012; Meyer, Coull, Versace, Cinciripini and Morris 2015; Zhang, Baladandayutha-pani, Zhu, Baggerly, Majewski, Czerniak and Morris 2016; Zhu, Versace, Cinciripini and Morris 2018; Lee, Miranda, Baladandayuthapani, Rausch, Fazio, Downs and Morris 2018). We also mention WFMM executable as well as the BayesFMM packages, which is freely available at

https://biostatistics.mdanderson.org/SoftwareDownload,

where it does not need to formulate the random effect structure in the quantile functional regression model. To employ this, we need to create the input file WFMM-input.mat which includes the empirical *quantlets* coefficients and design matrix structure. Such a file will pipe into the WFMM software to fit our model. There is a key commend to run the WFMM software in DOS window as the following:

wfmm WFMM-input.mat WFMM-output.mat > WFMM-log.log

16

Remark that the input file should be placed in the same directory that the commands are executed. WFMM-output.mat will be produced by the above commend and contain the posterior samples of the quantile processes, which will be used for the further inference in R or Matlab environments.

All our codes in **QFM.zip** file are just for independent functional linear regression, while the FMM code can handle other structure including levels of random effects to model interfunctional correlation and nonparametric function.

Table 3: Results for the simulation 1: Empirical mean integrated squared error, $n^{-1}\sum_{i=1}^{n}(Q_{(\cdot)}(p_i) - \widehat{Q}_{(\cdot)}(p_i))^2$ and monotonicity, $1 - P_\epsilon^{\mathcal{M}}$ were computed based on each basis representation for distributions, where N, T, G, D, M_1, M_2, M_3, and M_4 indicate normal, $t_{(1)}$, gamma, dirichlet, mixtures of $SN(-3.06, 3.67, 0)$ and $SN(9.11, 7.89, -4)$ with 0.5 probability, $SN(-7.1, 2.4, 0)$ and $SN(-3.11, 7.89, 4)$ with 0.3 and 0.7 probabilities, $N(-2.5, 2.5)$, $N(4, 3)$ and $N(9.5, 2.1)$ and $N(-2.5, 1.5)$, $N(4, 3.56)$ and $N(9.5, 1.1)$ with 0.3, 0.5 and 0.2 probabilities distributions, respectively.

		N	T	G	D	M_1	M_2	M_3	M_4
	Quantlets	0.024	0.001	0.007	0.004	0.039	0.050	0.020	0.022
IMSE	**B-spline**	0.032	0.029	0.135	0.057	0.108	0.166	0.074	0.094
	I-spline	0.698	0.413	0.599	0.052	2.283	2.760	1.004	1.013
	C-spline	0.203	0.140	0.414	0.460	0.548	0.692	0.343	1.660
	Quantlets	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Monotonicity	**B-spline**	0.979	0.996	0.888	0.976	0.992	0.973	1.000	1.000
	I-spline	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	C-spline	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Conditional moment statistics and 95% confidence interval in simulation: (A) naive quantile regression approach, (B) naive quantile functional regression approach, (C) principal component method, (D) *quantlet* space without sparse regularization, and (E) *quantlet* space with sparse regularization.

True	A	B	C	D	E
$\mu_1 = 1$	0.93 (0.86,0.99)	1.20 (1.18,1.21)	1.2 (0.86,1.53)	1.19 (0.83,1.54)	1.18 (0.85,1.52)
$\mu_2 = 3$	2.71 (2.63,2.79)	2.92 (2.91,2.94)	2.93 (2.61,3.24)	2.93 (2.57,3.29)	2.94 (2.60,3.30)
$\mu_3 = 1$	0.66 (0.59,0.72)	0.99 (0.98,1.00)	0.99 (0.66,1.32)	1.00 (0.64,1.35)	0.99 (0.64,1.34)
$\mu_4 = 3$	2.59 (2.52,2.65)	2.90 (2.88,2.91)	2.89 (2.56,3.22)	2.90 (2.54,3.25)	2.89 (2.54,3.25)
$\sigma_1 = 5$	4.92 (4.84,4.99)	4.97 (4.96,4.98)	4.97 (4.84,5.11)	4.97 (4.89,5.05)	4.97 (4.9,5.05)
$\sigma_2 = 5$	4.96 (4.85,5.06)	4.96 (4.95,4.97)	4.96 (4.83,5.11)	4.96 (4.89,5.03)	4.96 (4.88,5.04)
$\sigma_3 = 6.5$	6.36 (6.27,6.46)	6.43 (6.42,6.44)	6.43 (6.29,6.56)	6.43 (6.35,6.51)	6.43 (6.36,6.51)
$\sigma_4 = 5$	5.05 (4.94,5.14)	4.93 (4.92,4.95)	4.94 (4.82,5.07)	4.94 (4.86,5.01)	4.93 (4.86,5.01)
$\xi_1 = 0.00$	-0.06 (-0.19,0.06)	0.01 (-0.01,0.02)	0.00 (-0.17,0.18)	0.01 (-0.21,0.22)	0.01 (-0.19,0.21)
$\xi_2 = 0.00$	-0.10 (-0.23,0.02)	0.00 (-0.01,0.02)	0.00 (-0.16,0.17)	0.00 (-0.21,0.23)	0.00 (-0.21,0.22)
$\xi_3 = 0.00$	-0.07 (-0.17,0.02)	0.00 (-0.01,0.02)	0.00 (-0.13,0.13)	0.00 (-0.16,0.17)	0.00 (-0.16,0.17)
$\xi_4 = -0.78$	-0.91 (-1.11,-0.73)	-0.74 (-0.76,-0.73)	-0.74 (-0.95,-0.56)	-0.74 (-0.97,-0.52)	-0.74 (-0.96,-0.52)

Table 5: Normality score of estimates for conditional subgroup in simulation.

Group	Estimate	Percentage (95% CI)
$(\xi, \omega, \alpha) = (1.0, 5.0, 0.0)$	$\hat{\beta}_1(p)$	68.8% (49.4, 88.4)
$(\xi, \omega, \alpha) = (3.0, 5.0, 0.0)$	$\hat{\beta}_1(p) + \hat{\beta}_2(p)$	89.5% (78.5, 97.5)
$(\xi, \omega, \alpha) = (1.0, 6.5, 0.0)$	$\hat{\beta}_1(p) + \hat{\beta}_3(p)$	81.9% (66.4, 94.9)
$(\xi, \omega, \alpha) = (9.1, 7.9, -4.0)$	$\hat{\beta}_1(p) + \hat{\beta}_4(p)$	34.7% (30.1, 39.0)

Table 6: Normality score of estimates for conditional subgroup in GBM application.

Group	Treatment	Quantlet Percent (95% CI)	Reference	Quantlet Percent (95% CI)
Sex	Male	63.3% $(40.0, 85.2)$	Female	77.9% $(59.2, 92.8)$
Age	84 years	65.3% $(44.7, 84.0)$	60 years	73.2% $(55.0, 89.1)$
DDIT3	Yes	70.9% $(44.9, 91.7)$	None	69.9% $(51.5, 86.6)$
EGFR	Yes	71.4% $(49.6, 90.0)$	None	71.3% $(50.2, 89.0)$
Mesenchymal	Yes	78.4% $(59.0, 94.0)$	None	61.3% $(39.7, 82.3)$
Survival status	$\leq$ 12 months	80.6% $(59.1, 96.2)$	> 12 months	63.8% $(44.6, 81.7)$

Figure 2: Density functions in simulation 1: panel (A) contains normal (skyblue), $t_{(1)}$ (black), shifted gamma (3,1) (purple), and dirichlet (gray dot) and panel (B) contains mixtures of $SN(-3.06, 3.67, 0)$ and $SN(9.11, 7.89, -4)$ with 0.5 and 0.5 probabilities (black), $SN(-7.1, 2.4, 0)$ and $SN(-3.11, 7.89, 4)$ with 0.3 and 0.7 probabilities (red), $N(-2.5, 2.5)$, $N(4, 3)$ and $N(9.5, 2.1)$ and $N(-2.5, 1.5)$, $N(4, 3.56)$ and $N(9.5, 1.1)$ with 0.3, 0.5 and 0.2 probabilities (blue and green), denoted by E, F, G, and H, respectively.
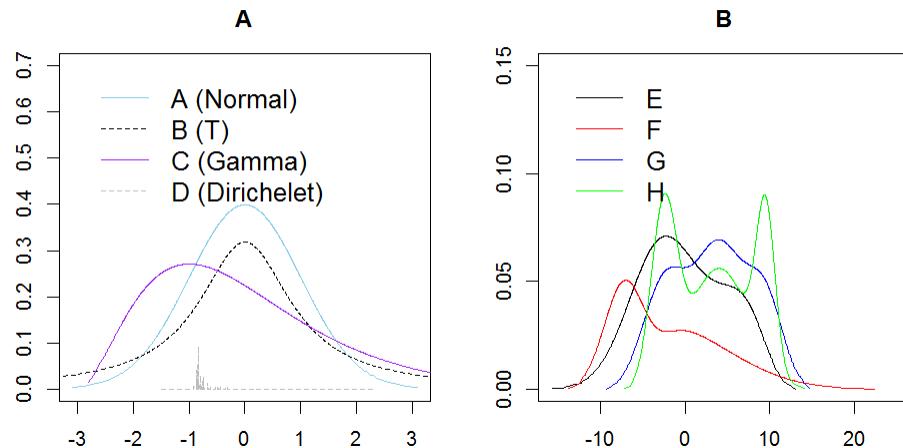
Table 7: GBM Results: Bayesian global p-values for quantlet models with various sizes of basis set K.

	K	ρ^0	$\bar{\rho}$	ν_0	Sex	Age	DDIT3	EGFR	Mesenchymal	Survival$_{12}$	Run time
Model 1	546	.998	1.000	.006	0.031	0.284	0.020	0.646	0.173	0.102	9.32 min
Model 2	194	.997	1.000	.006	0.023	0.197	0.029	0.648	0.127	0.088	2.42 min
Model 3	101	.997	1.000	.006	0.018	0.196	0.051	0.629	0.159	0.080	1.30 min
Model 4	66	.996	.999	.006	0.017	0.271	0.018	0.646	0.188	0.106	46.20 sec
Model 5	50	.996	.999	.006	0.032	0.242	0.048	0.698	0.229	0.092	39.13 sec
Model 6	44	.996	.999	.006	0.027	0.239	0.023	0.657	0.175	0.097	33.11 sec
Model 7	42	.994	.999	.006	0.005	0.122	0.006	0.648	0.058	0.038	35.60 sec
Model 8	31	.993	.999	.006	0.008	0.177	0.009	0.627	0.116	0.061	26.19 sec
Model 9	27	.990	.998	.006	0.016	0.168	0.012	0.605	0.087	0.067	21.83 sec
Model 10	23	.989	.997	.006	0.036	0.242	0.014	0.684	0.128	0.076	18.78 sec
Model 11	19	.989	.997	.006	0.038	0.297	0.035	0.668	0.221	0.132	15.31 sec
Model 12	15	.988	.997	.006	0.019	0.226	0.020	0.683	0.103	0.111	12.89 sec
Model 13	13	.981	.996	.006	0.041	0.274	0.042	0.796	0.302	0.143	12.93 sec
Model 14	10	.981	.996	.006	0.036	0.294	0.027	0.694	0.218	0.113	11.97 sec
Model 15	9	.981	.996	.006	0.006	0.152	0.007	0.605	0.128	0.048	11.81 sec
Model 16	8	.964	.993	.006	0.027	0.239	0.025	0.640	0.129	0.104	10.65 sec
Model 17	7	.962	.993	.006	0.007	0.147	0.006	0.607	0.084	0.063	9.12 sec
Model 18	5	.860	.974	.006	0.014	0.160	0.009	0.561	0.096	0.063	0.022 sec
Model 19	2	.858	.966	.006	0.014	0.053	0.006	0.494	0.067	0.042	0.006 sec
Naive	1024				1.000	1.000	1.000	1.000	1.000	1.000	2.710 min
Model 9	27			.01	0.016	0.169	0.012	0.607	0.088	0.068	22.25 sec
Model 9	27			.0001	0.015	0.161	0.010	0.601	0.083	0.061	19.60 sec

Table 8: GBM Results: Area of the joint 95% confidence intervals for quantlet models with various sizes of basis set K.

	K	ρ_0	$\bar{\rho}$	ν_0	Intercept	Sex	Age	DDIT3	EGFR	Mesenchymal	Survival$_{12}$
Model 1	546	.998	1.000	.006	17.373	15.028	32.168	25.436	15.919	15.511	14.707
Model 2	194	.997	1.000	.006	17.127	14.549	29.665	25.073	14.940	14.448	14.056
Model 3	101	.997	1.000	.006	16.680	14.397	30.862	32.653	15.015	15.288	14.498
Model 4	66	.996	.999	.006	18.154	15.729	33.493	25.821	15.295	15.391	15.665
Model 5	50	.996	.999	.006	18.184	15.772	32.420	30.657	15.700	15.801	15.117
Model 6	44	.996	.999	.006	17.183	15.654	32.230	26.074	14.955	15.221	15.048
Model 7	42	.994	.999	.006	14.621	12.503	26.807	22.012	13.023	12.269	12.860
Model 8	31	.993	.999	.006	15.948	13.392	29.069	23.708	13.981	13.699	13.566
Model 9	27	.990	.998	.006	15.747	13.035	26.578	22.607	13.819	13.807	12.933
Model 10	23	.989	.997	.006	17.151	15.533	31.429	24.489	15.319	14.935	14.368
Model 11	19	.989	.997	.006	19.168	17.205	36.168	28.363	16.637	16.385	16.799
Model 12	15	.988	.997	.006	17.176	13.822	30.512	23.948	14.784	14.109	14.297
Model 13	13	.981	.996	.006	19.953	16.687	34.823	28.207	17.134	17.192	17.060
Model 14	10	.981	.996	.006	19.087	16.097	33.559	27.918	16.082	16.100	16.433
Model 15	9	.981	.996	.006	16.848	13.388	27.394	22.433	14.079	14.100	13.028
Model 16	8	.964	.993	.006	16.922	14.165	30.335	25.136	14.903	14.424	14.194
Model 17	7	.962	.993	.006	14.724	12.679	28.202	22.669	13.407	12.965	13.339
Model 18	5	.860	.974	.006	15.679	13.513	27.982	22.483	13.584	13.361	13.288
Model 19	2	.858	.966	.006	15.170	12.845	26.967	21.397	12.859	12.525	12.448
Naive	1024				26.013	22.210	47.030	37.527	22.680	22.546	21.837
Model 9			27	.01	15.815	13.093	26.683	22.705	13.881	13.865	12.989
Model 9			27	.0001	15.627	12.934	26.392	22.435	13.708	13.702	12.835

Table 9: ϵ-monotonicity of quantile functions for conditional subgroup.

	Simulation 1		Simulation 2		GBM data	
	$\epsilon = 0.001$	$\epsilon = 0.01$	$\epsilon = 0.03$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.5$
Naive	25.8%	96.8%	0.0%	0.0%	0.0%	43.9%
PCA	100.0%	100.0%	35.4%	83.8%	90.0%	93.9%
Quantlets	100.0%	100.0%	100.0%	100.0%	93.9%	96.3%

Figure 3: Basis representations based on different methods: representations for *quantlets* (red), B-spine (dashed) I-spline (dot) and C-spline (dashed-dot) are given for the following quantiles, where the gray dot line is the true quantile function in each panel: (A) normal, (B) t, (C) gamma and (D) dirichlet, (E) mixture of $SN(-3.06, 3.67, 0)$ and $SN(9.11, 7.89, -4)$ with 0.5 probability, (F) mixture of $SN(-7.1, 2.4, 0)$ and $SN(-3.11, 7.89, 4)$ with 0.3 and 0.7 probabilities, (G) mixture of $N(-2.5, 2.5)$, $N(4, 3)$ and $N(9.5, 2.1)$ and (H) mixture of $N(-2.5, 1.5)$, $N(4, 3.56)$ and $N(9.5, 1.1)$ with 0.3, 0.5 and 0.2 probabilities, respectively.
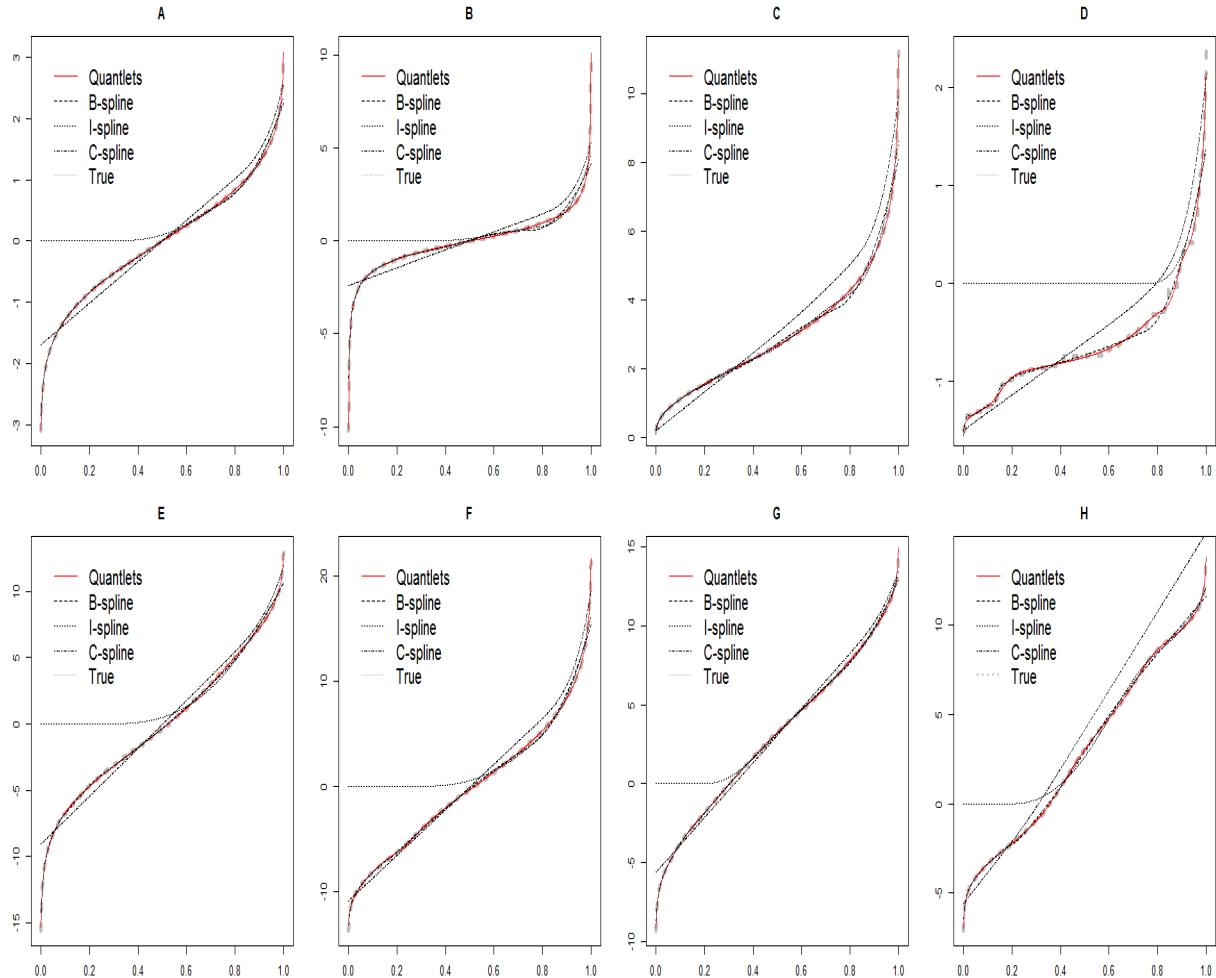
Figure 4: I-spline and C-spline basis functions: (A) I-spline and (B) C-spline.
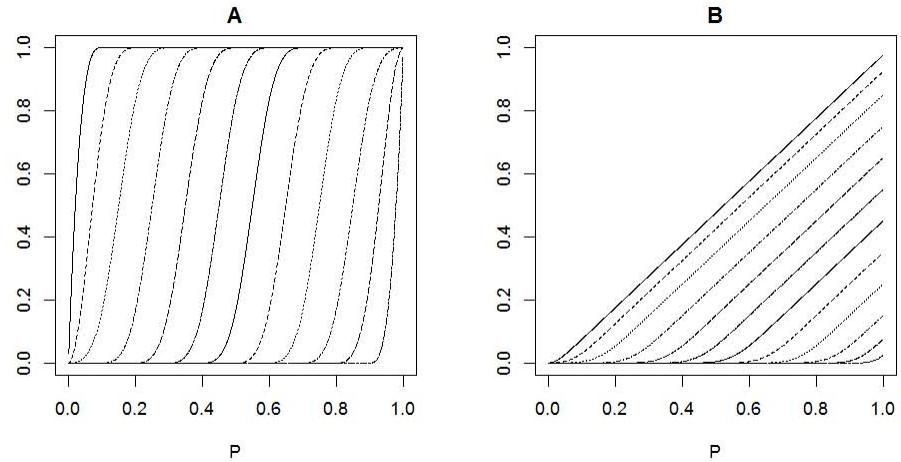
Figure 5: Quantlets basis functions in simulation.
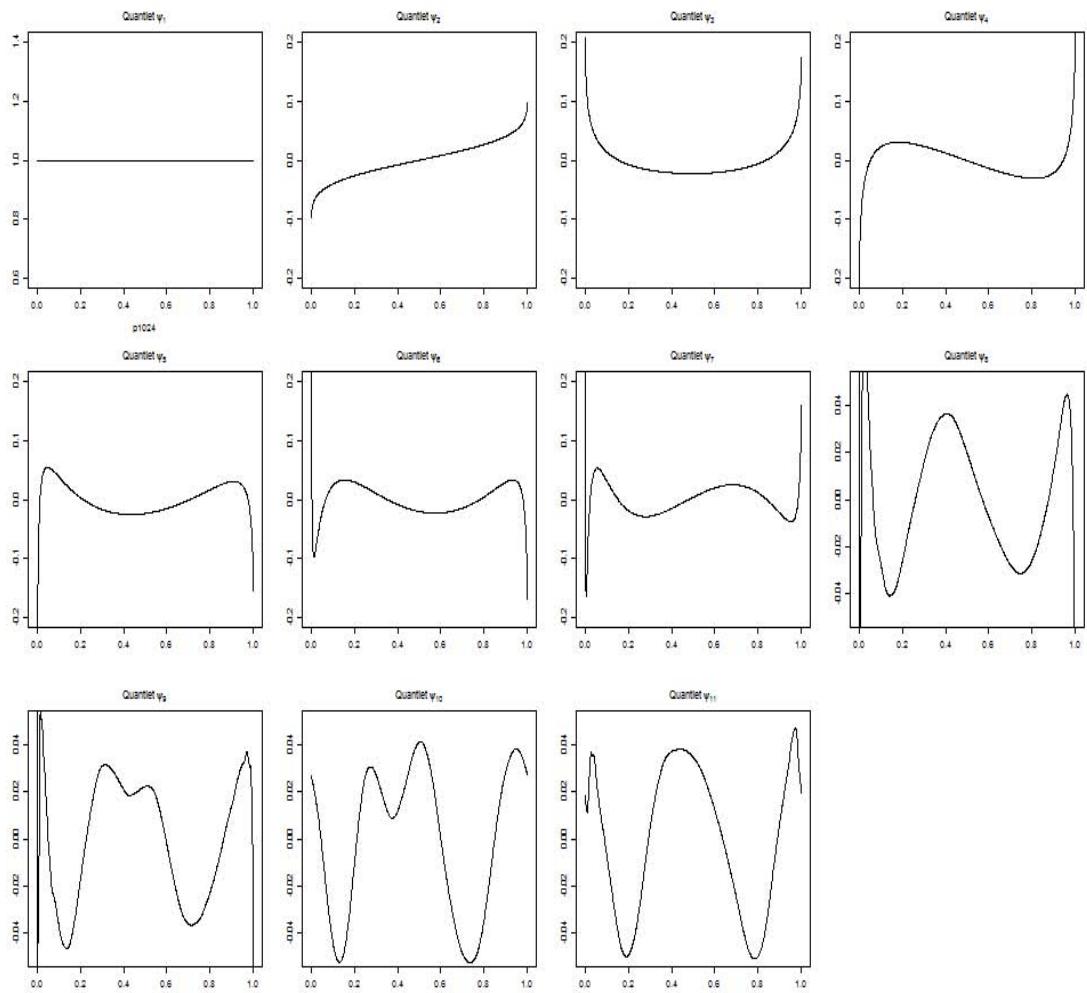
Figure 6: Principal Component basis functions in GBM application.
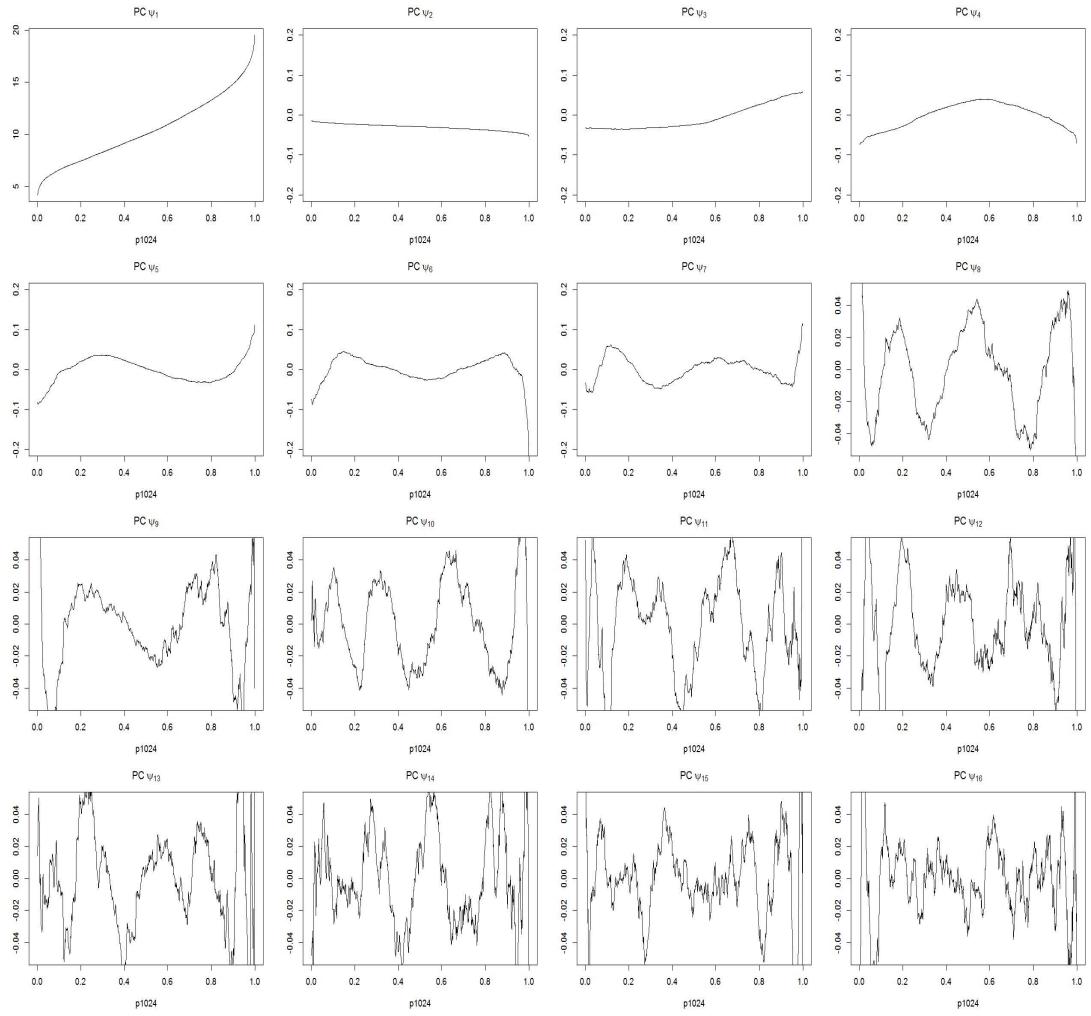
Figure 7: Wavelet denoising for 16 basis functions in GBM application: orthogonal basis (black) and wavelet denoising basis (blue).
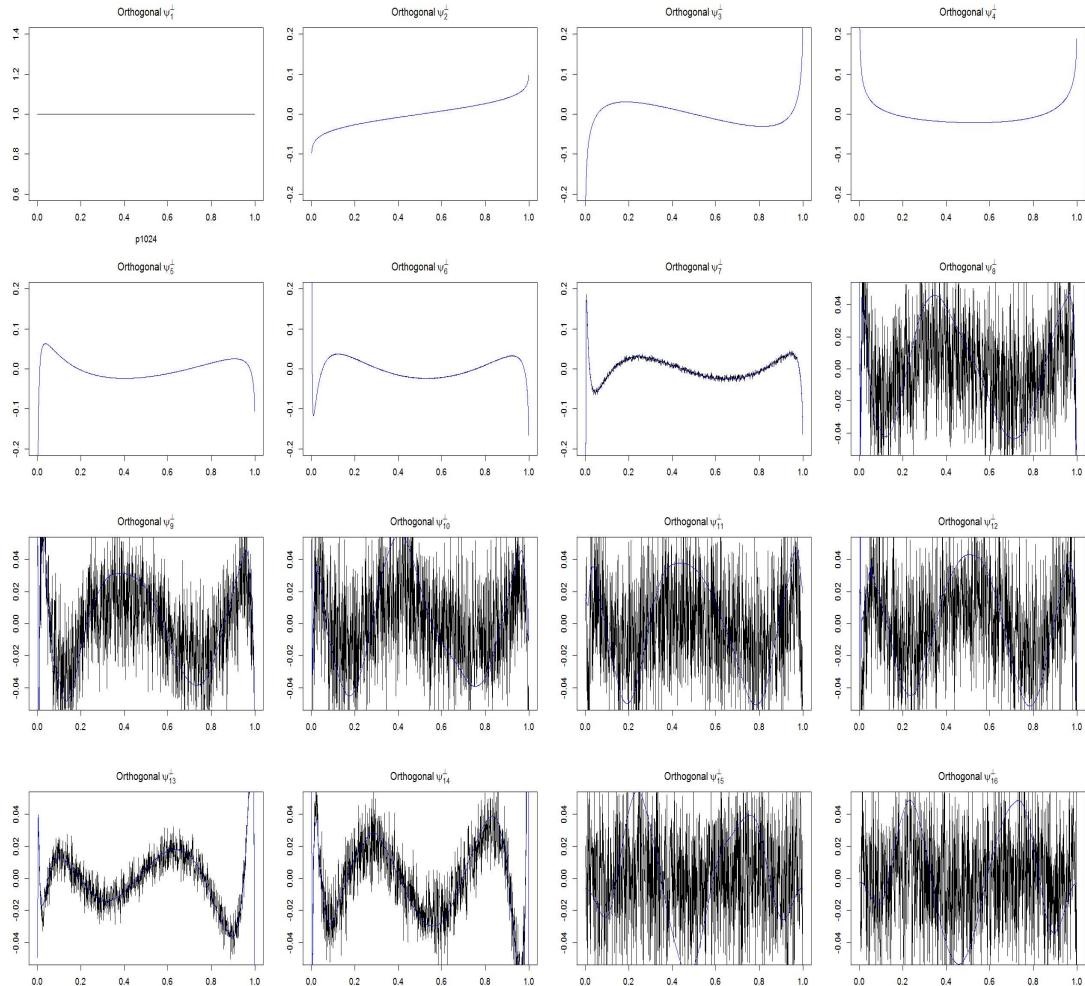
Figure 8: Quantlets basis functions in GBM application.
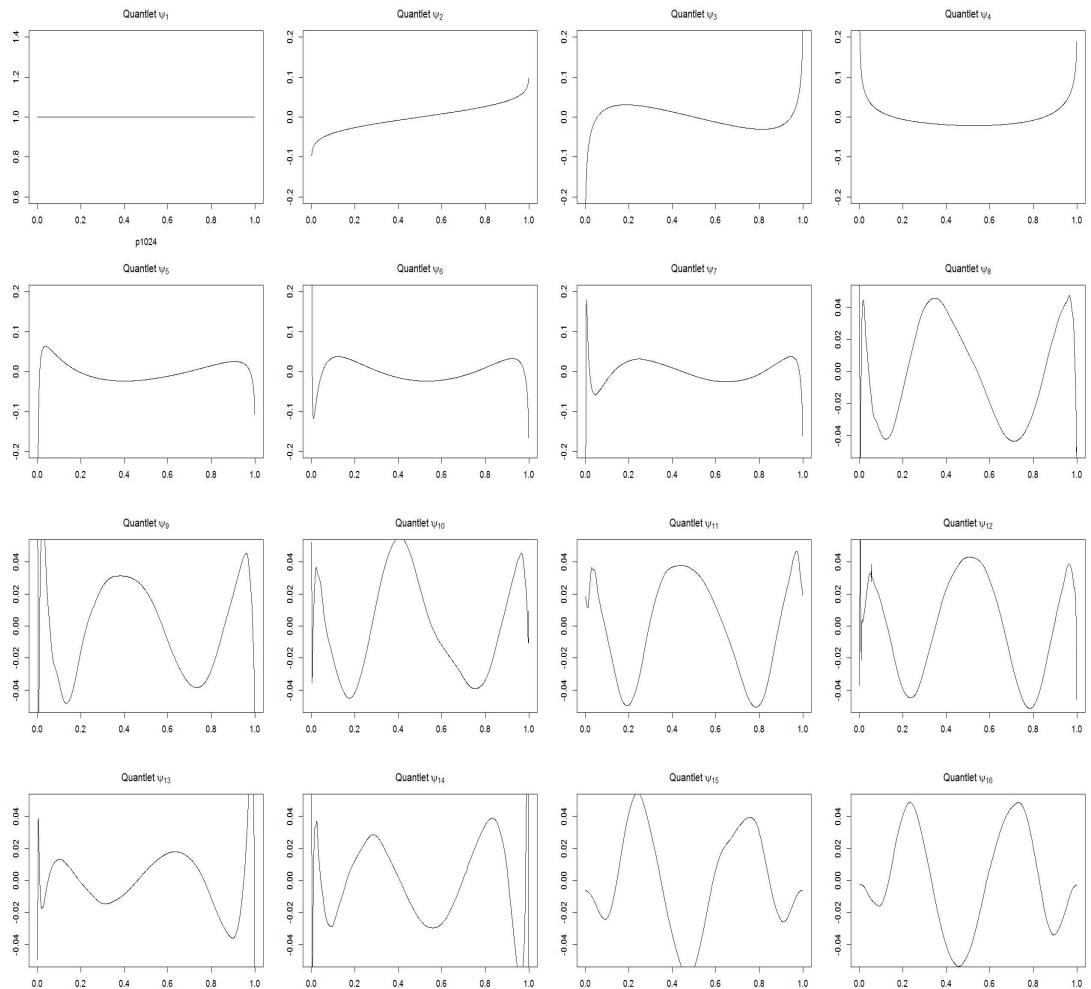
Figure 9: Simulation Results: Estimators and and 95% joint credible intervals for $\beta_1(p)$ (black) $\beta_2(p)$ (red), $\beta_3(p)$ (blue), $\beta_4(p)$ (green), their corresponding true coefficients (brown, orange, skyblue, and darkgreen, respectively), and fitted values by *quantlets* (gray) are derived from the (A) naive quantile regression approach, (B) naive quantile functional regression approach, (C) principal component method, (D) *quantlet* space without sparse regularization, (E) *quantlet* space with sparse regularization, and (F) Gaussian *quantlet* space approach.
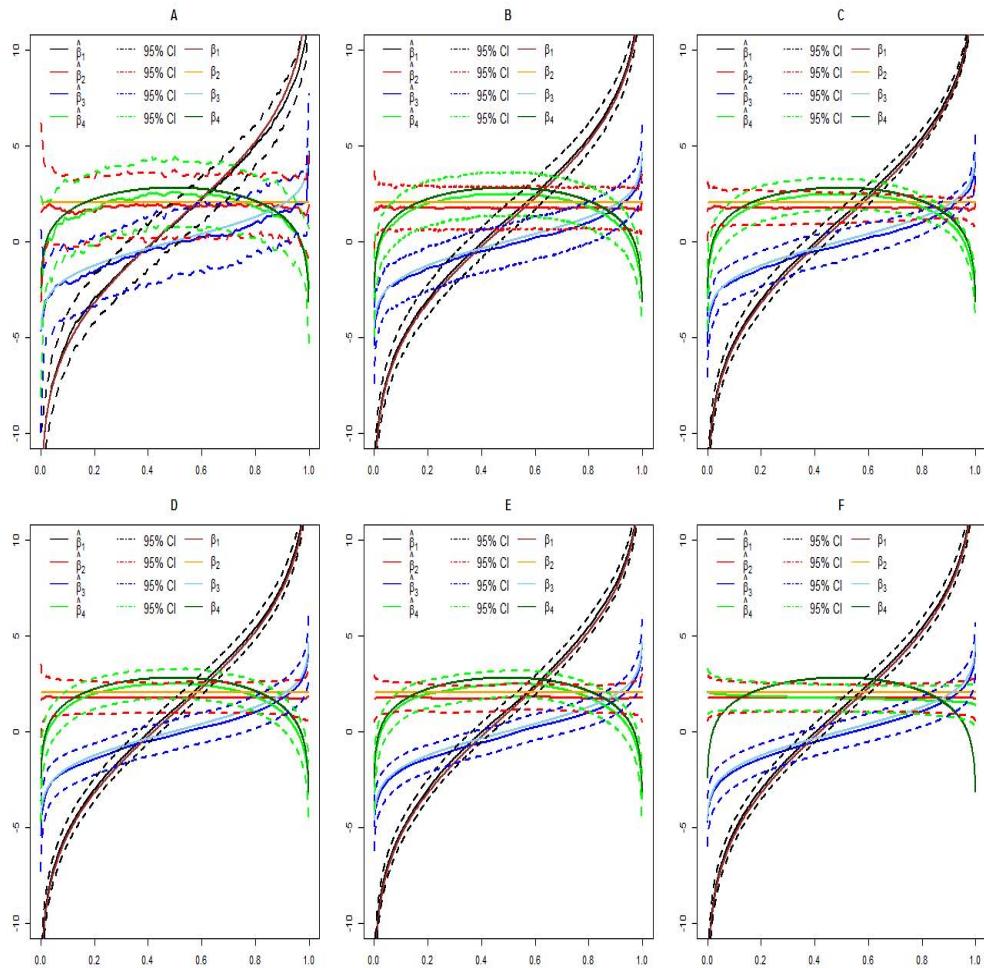
Figure 10: Intraquantile correlation estimated empirically (A) and assuming independence in the quantlet space (B) in GBM application.
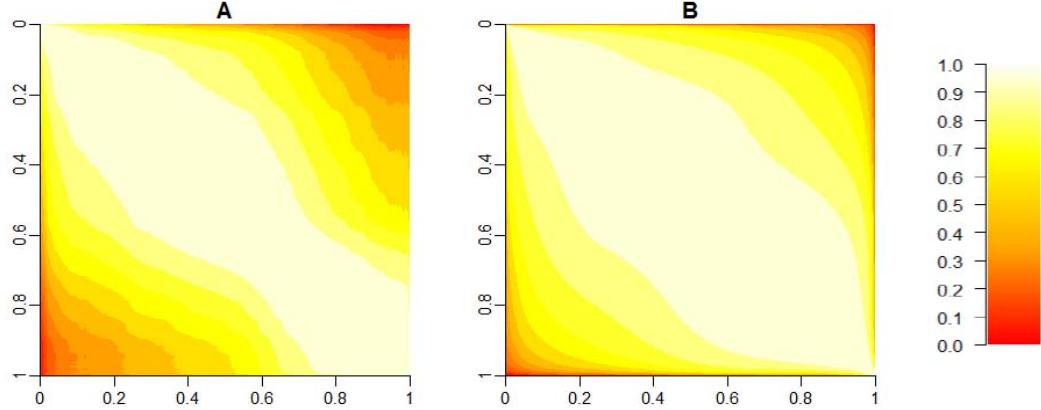
Figure 11: Simulation Results: SimBaS for $\beta_3(p)$ (blue) and $\beta_4(p)$ (green) at all $p \in \mathcal{P}$ are derived from the (B) quantile functional approach, (C) principal components method, (D) *quantlet* space without sparse regularization, and (E) *quantlet* space with sparse regularization, where vertical line (black) is significant level (0.05).
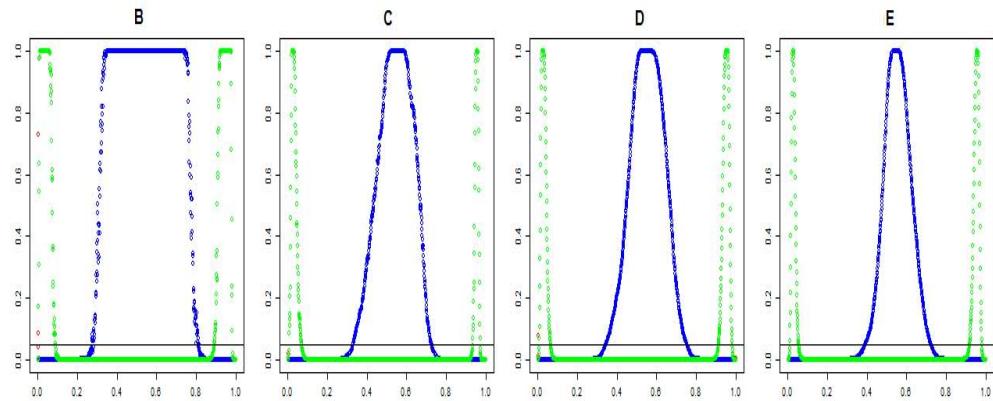
Figure 12: Posterior inference for the model ($K = 27$ and $\nu_0 = .006$) in GBM application: for each covariate (6), the left panel includes posterior mean estimate, point and joint credible bands, GBPV in heading along with SimBas less then .05 (orange line), and the right panel includes predicted densities for the two levels of the covariate along with the posterior probability scores for the moment different testings.
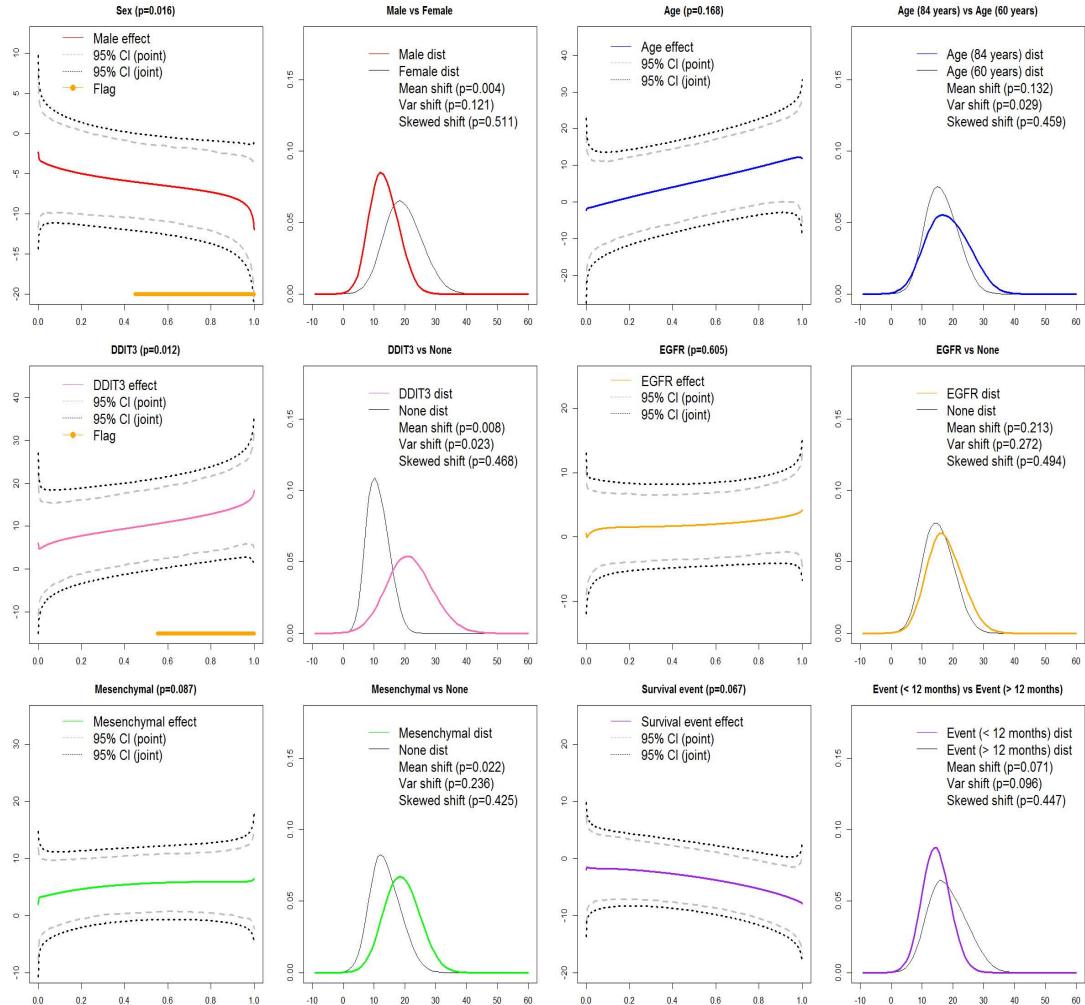
Figure 13: Posterior inference for the model ($K = 194$ and $\nu_0 = .006$) in GBM application.
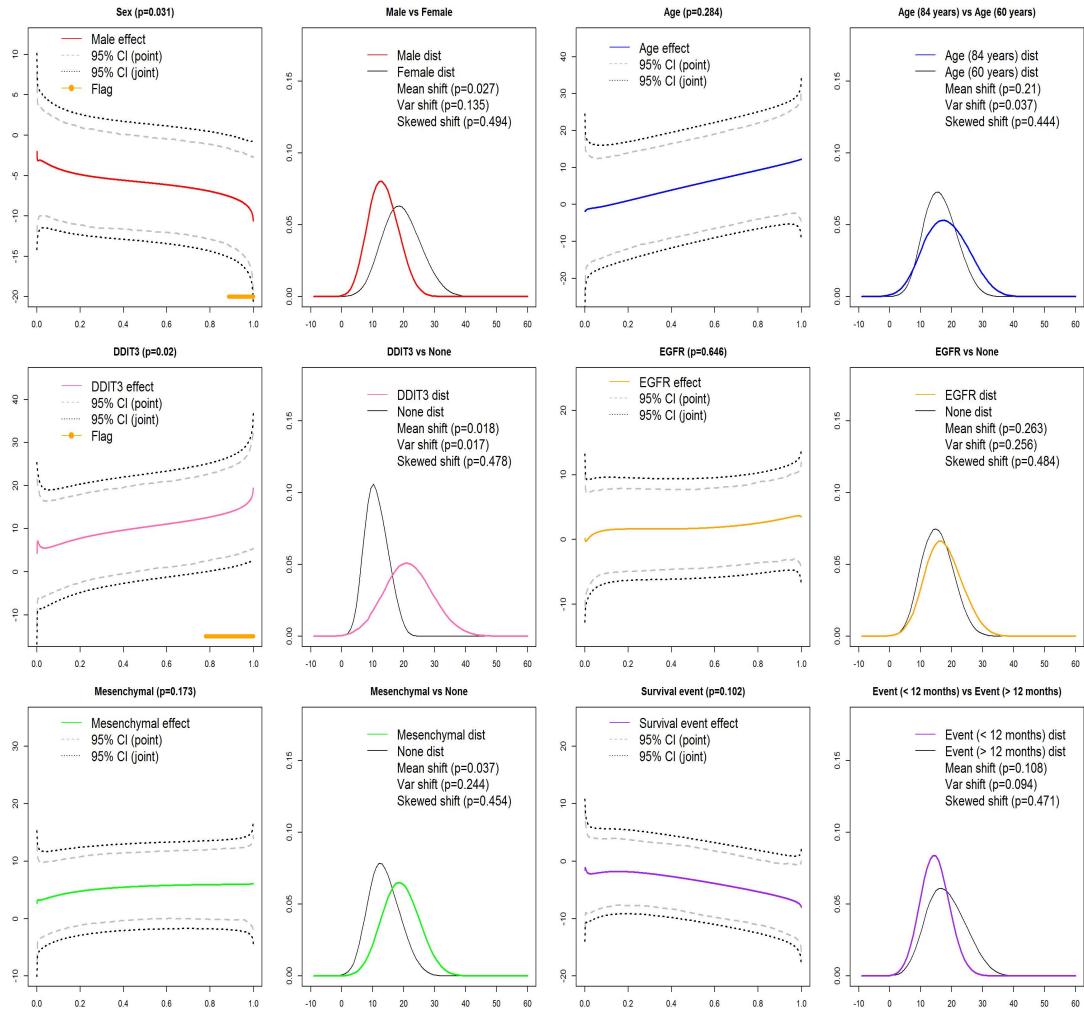
Figure 14: Posterior inference for the model ($K = 2$ and $\nu_0 = .006$) in GBM application.
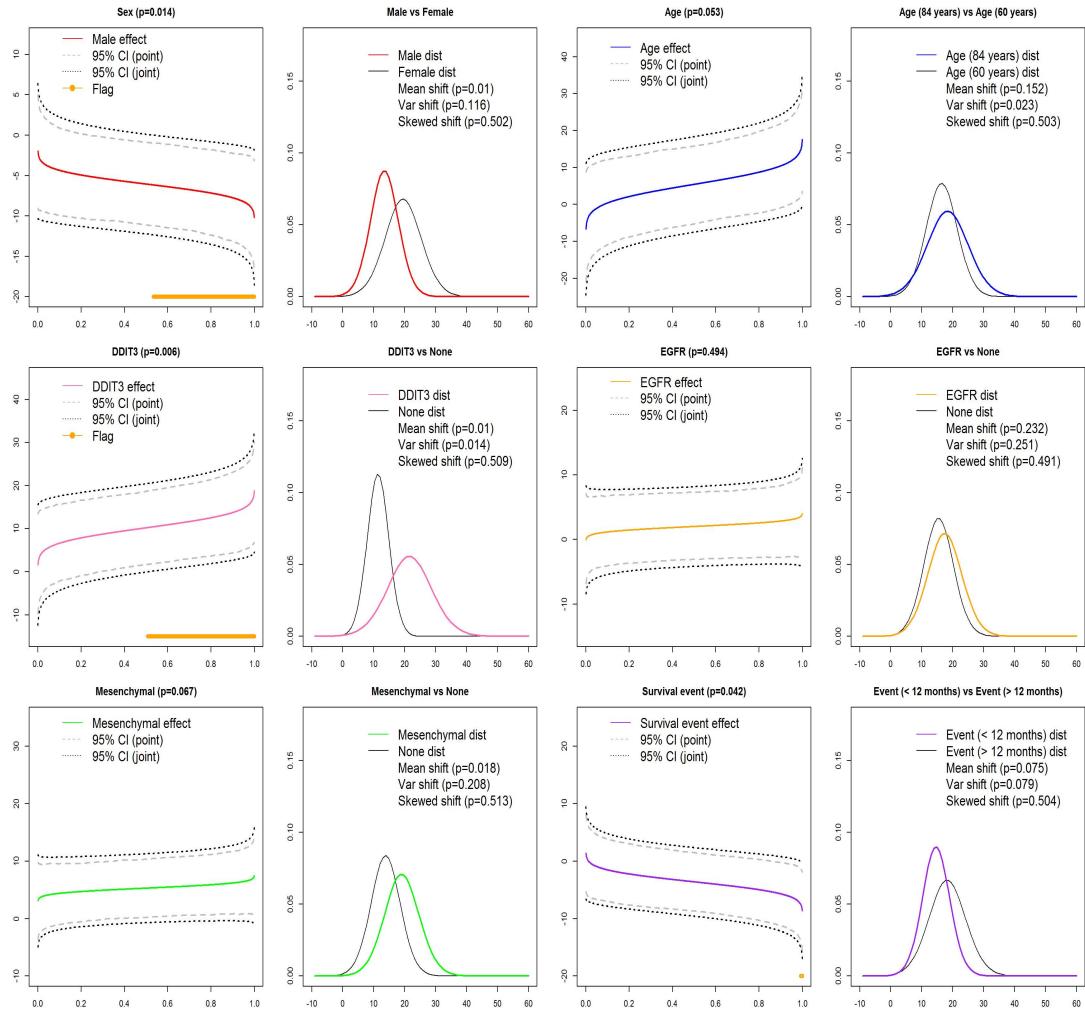
Figure 15: Inference from the naive quantile functional regression approach (separate functional regressions for each subject-specific quantile p) in GBM application.
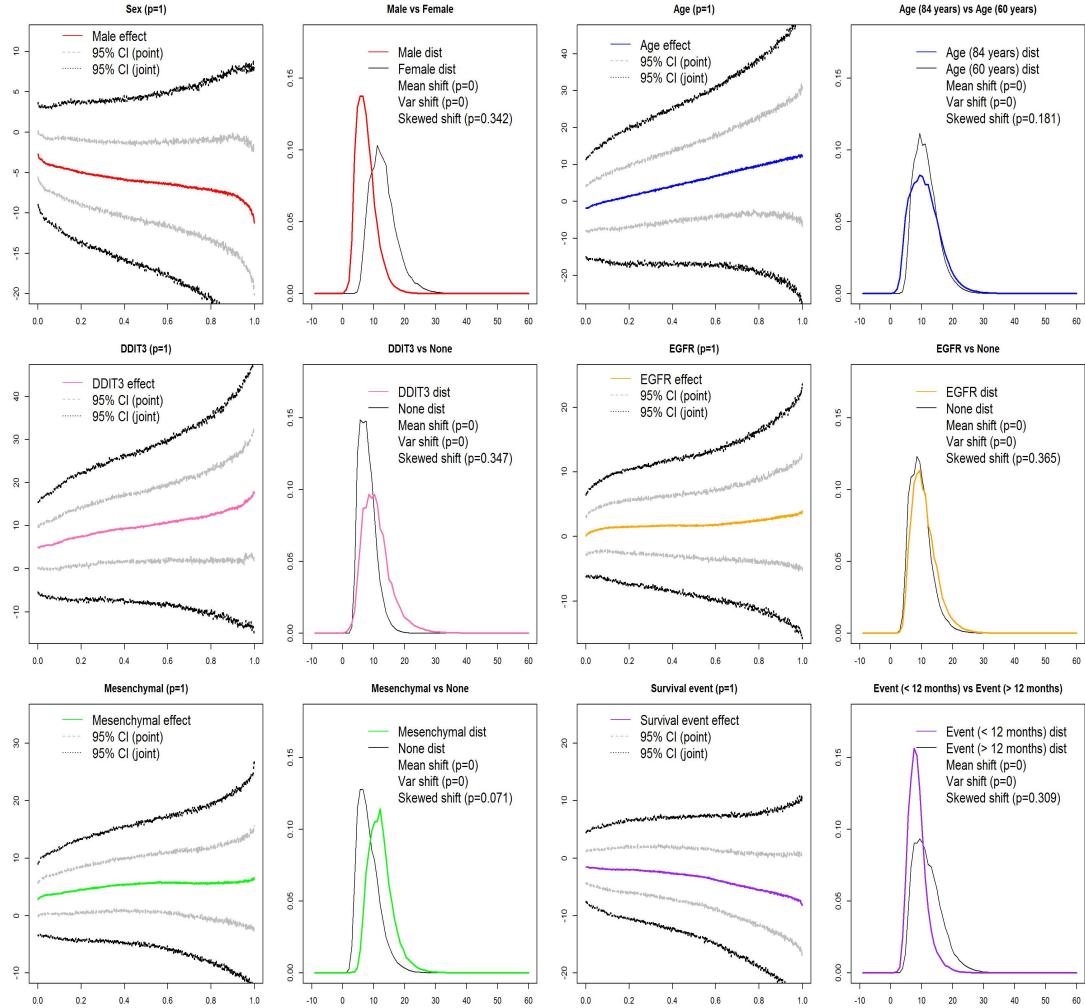
Figure 16: Posterior inference for the model ($K = 27$ and $\nu_0 = .01$) for in GBM application.
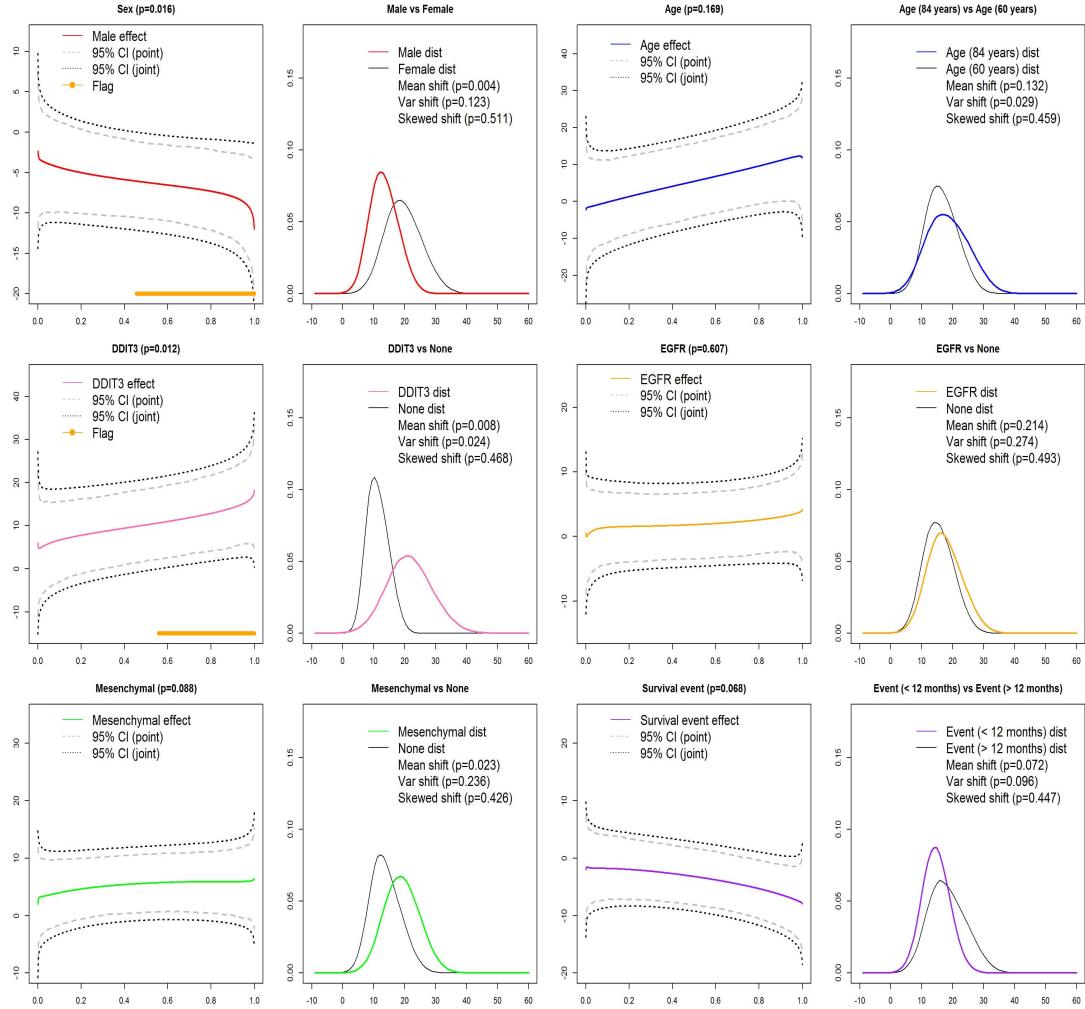
Figure 17: Posterior inference for the model ($K = 27$ and $\nu_0 = .0001$) in GBM application.
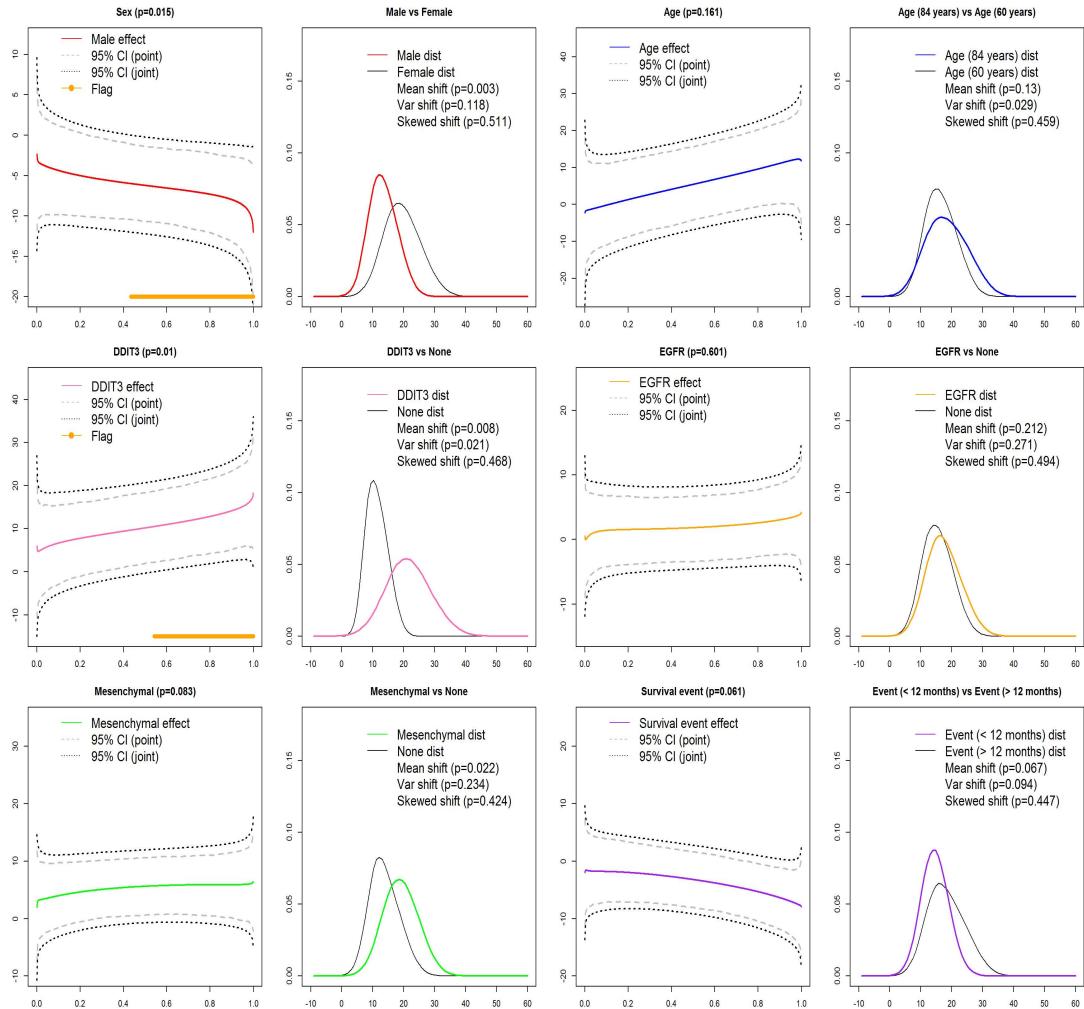
Figure 18: Gewekes diagnostic histograms for four models in GBM application. Under the null hypothesis of MCMC convergence, we would expect a uniform distribution in the p-values. We see no enrichment of small p-values in these histogram, suggesting chain convergence. Summaries are given for models (A) model 1 (K=194), (B) model 2 (K=27), (C) model 3 (K=7), and model 4 (K=2).
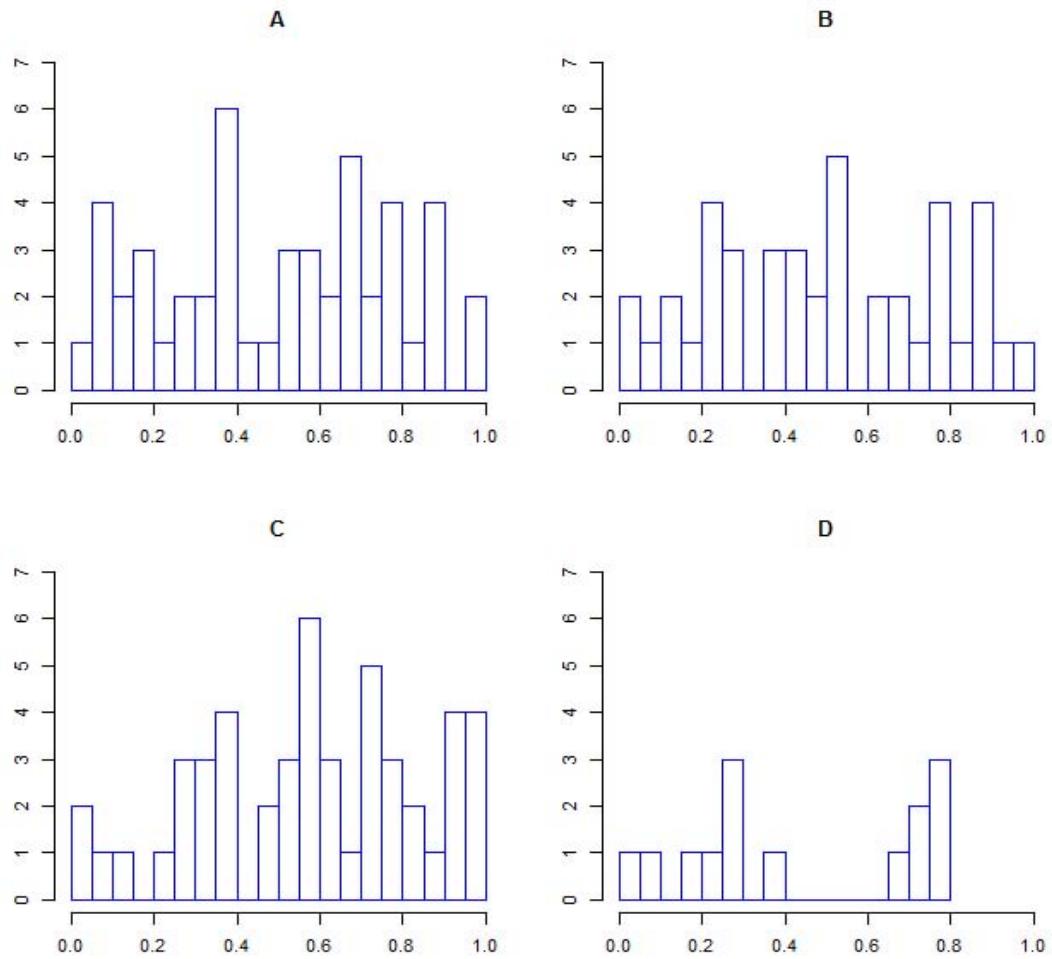
Figure 19: Predicted quantile functions with the point and joint 95% credible interval for the specific groups in GBM application: each row indicates the status of the sex and DDIT 3 whereas each column reports at the summary value of the age (min, Q1, Q2, Q3, and max).
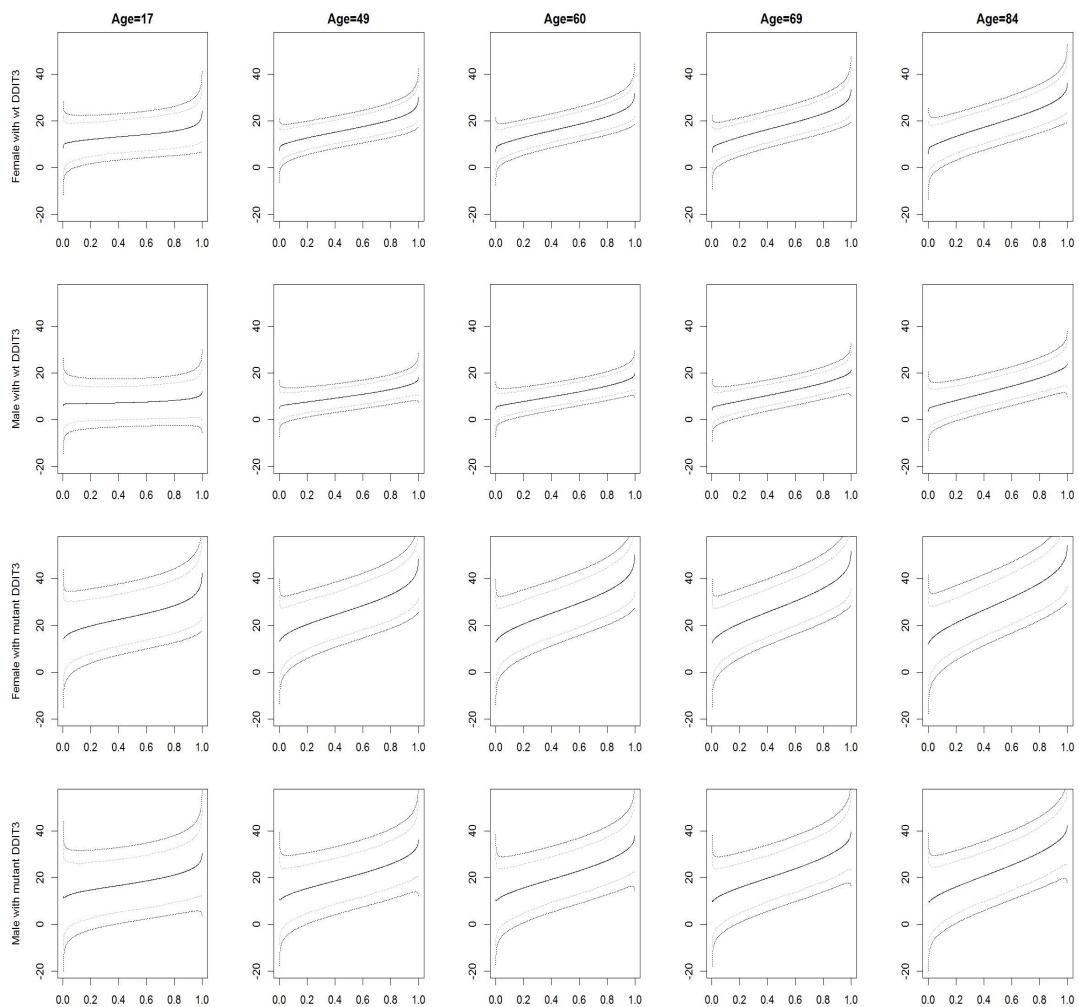
Figure 20: Inference from the naive quantile regression method (separate classical quantile regressions for each p in GBM application.
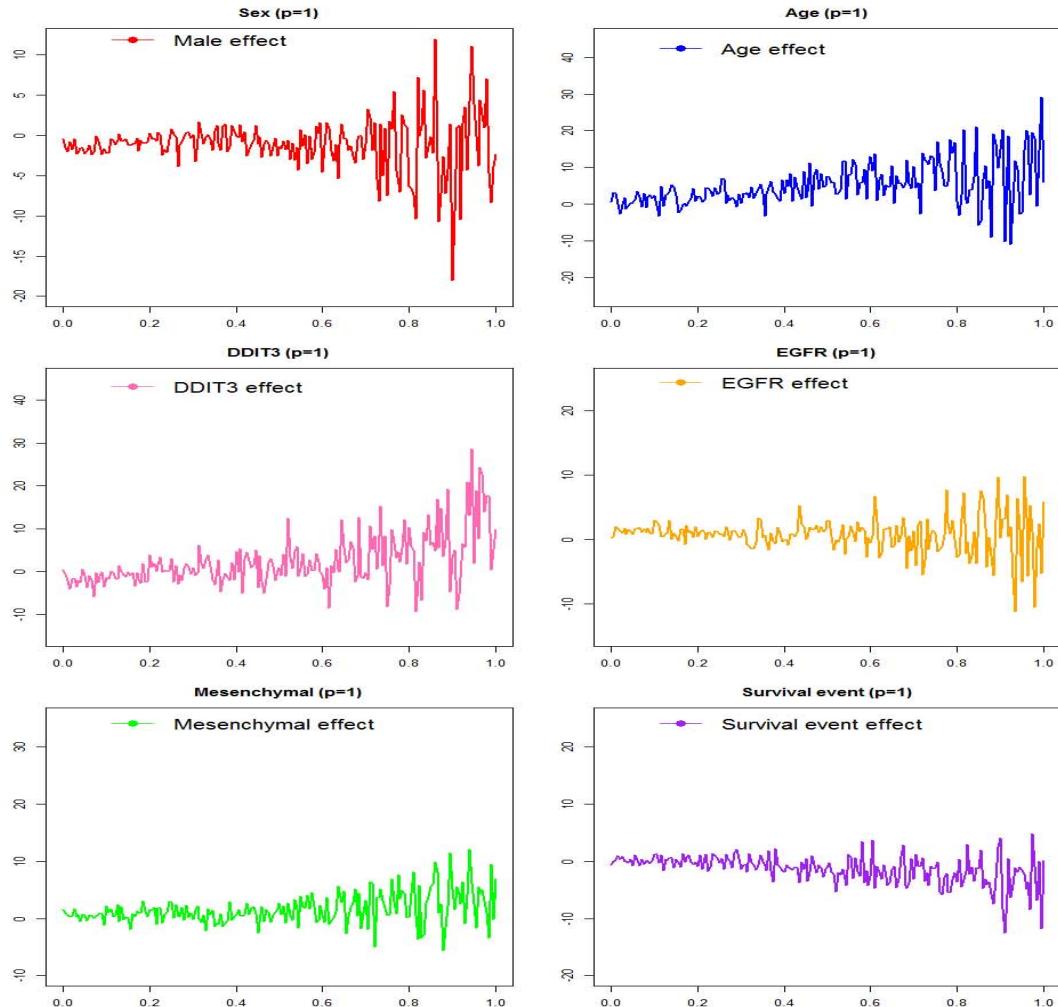
Figure 21: Posterior inference for the model ($K = 15$, $\lambda = 10\lambda^{(c)}$ and $\nu_0 = .006$) for in GBM application.
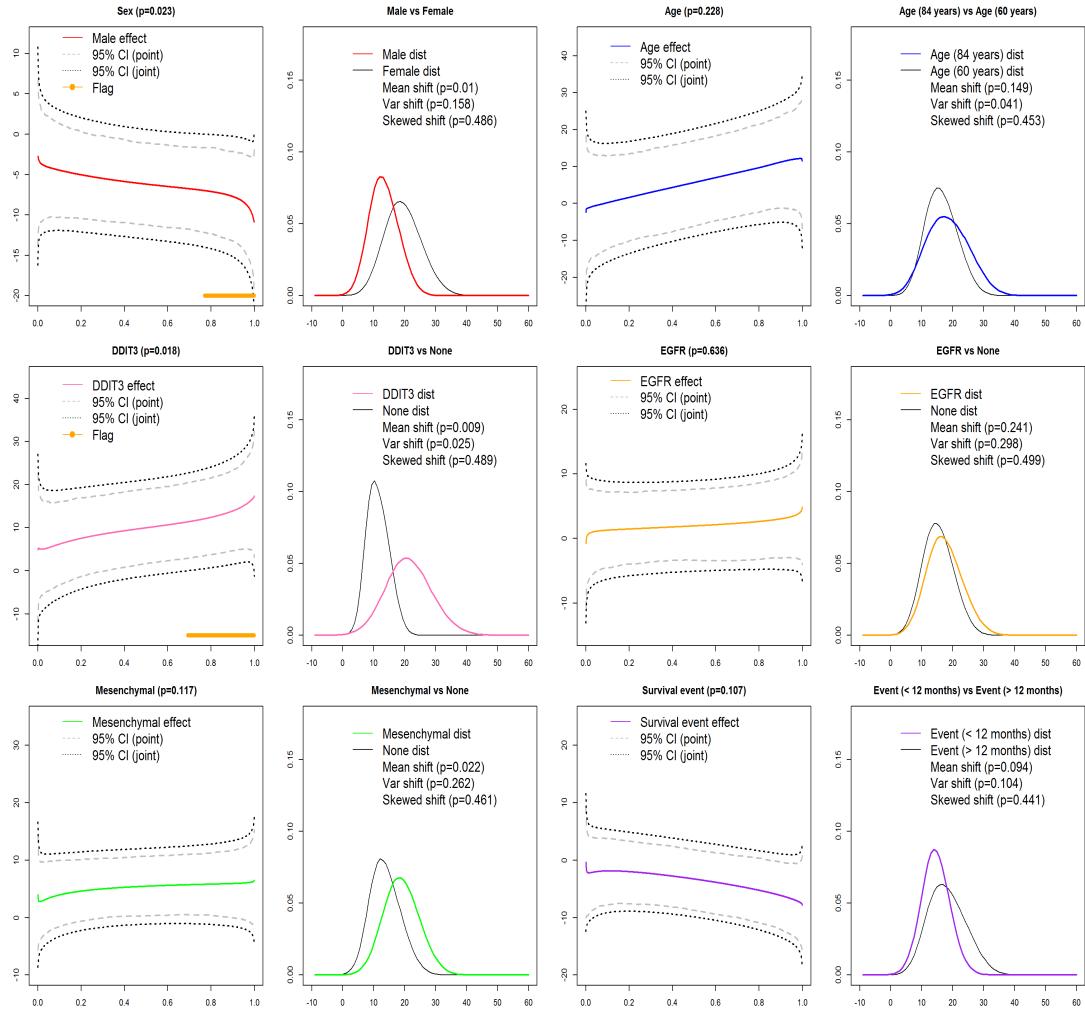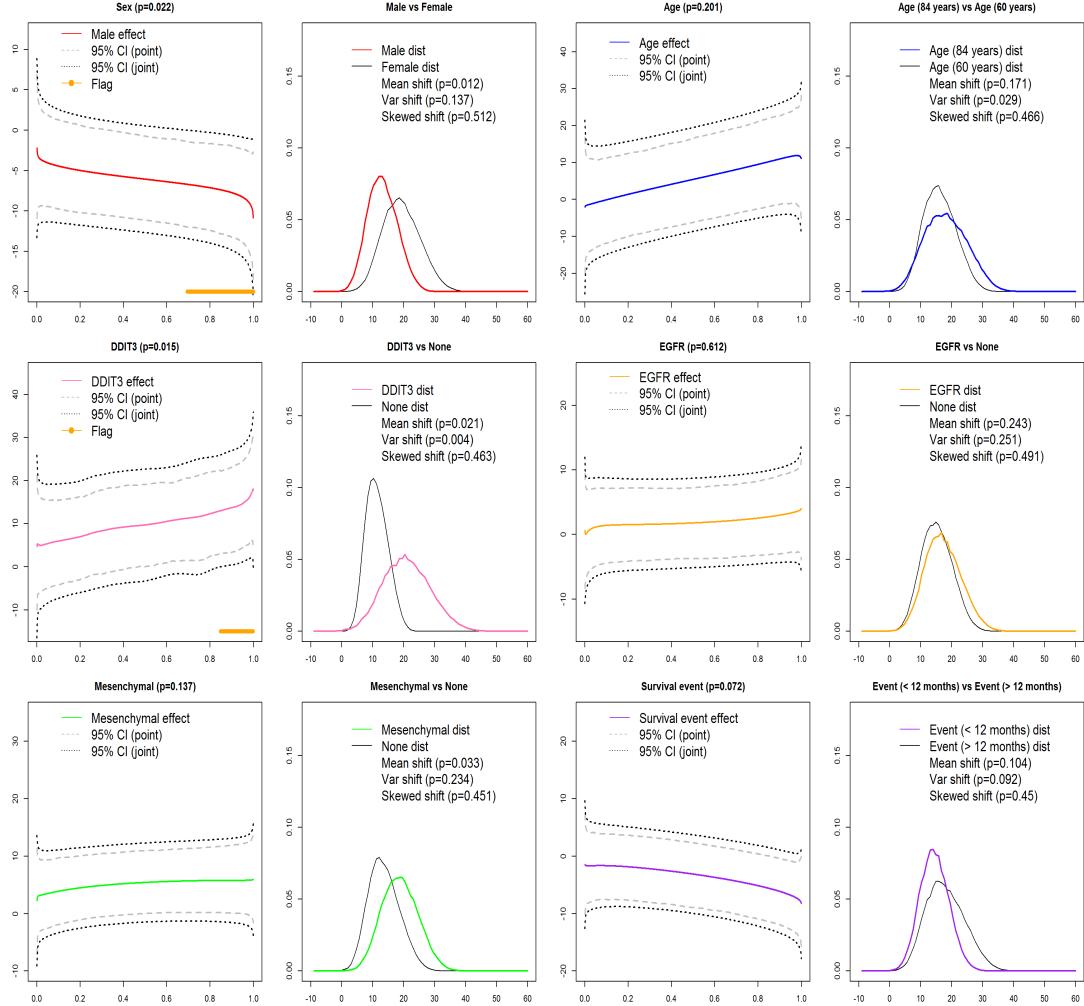
Figure 22: Posterior inference for the model ($K = 38$, $\lambda = .1\lambda^{(c)}$ and $\nu_0 = .006$) for in GBM application.

REFERENCES

Borchers, H. (2015), "Pracma: practical numerical math functions," *R package version*, 1, 1–373.

Bornkamp, B., and Ickstadt, K. (2009), "Bayesian nonparametric estimation of continuous monotone functions with applications to dose–response analysis," *Biometrics*, 65, 198–205.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995), "Wavelet shrinkage: asymptopia?," *Journal of the Royal Statistical Society. Series B*, 57, 301–369.

Friedman, J., Hastie, T., and Tibshirani, R. (2010), "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, 33.

Figure 23: Run time for the basis computation: m_i indicates the number of the grid points.
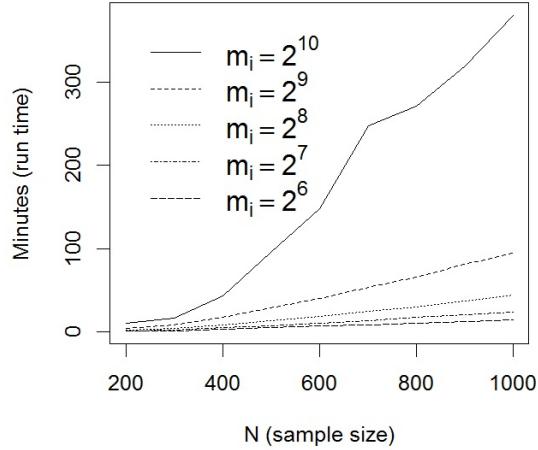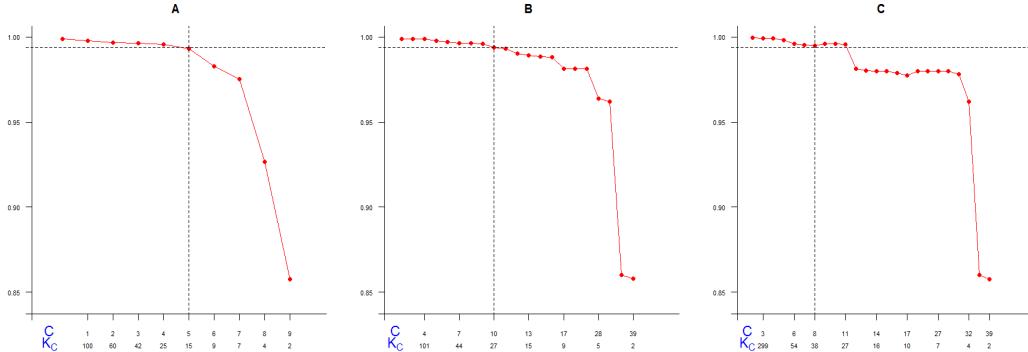
Figure 24: Near-lossless criterion for the different penalty (λ) of the lasso: (A) large value of λ , (B) current value of λ, and (C) small value of λ, where the large and small values are set to be $\lambda = 10\lambda^{(c)}$ and $\lambda = 0.1\lambda^{(c)}$ for the current penalty, $\lambda^{(c)}$.

Geweke, J. et al. (1991), *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments* Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.

Koenker, R. (2005), *Quantile regression* Cambridge university press.

Lee, W., Miranda, M., Baladandayuthapani, V., Rausch, P., Fazio, M., Downs, C., and Morris, J. S. (2018), "Bayesian semiparametric functional mixed models for longitudinal functional data with application to glaucoma data," *Journal of the American Statistical Association to appear*, doi:10.1080/01621459.2018.1476242.

Lempers, F. B. (1971), *Posterior probabilities of alternative linear models* Rotterdam University Press.

Meyer, M. C. (2008), "Inference using shape-restricted regression splines," *The Annals of Applied Statistics*, 2, 1013–1033.

Meyer, M. J., Coull, B. A., Versace, F., Cinciripini, P., and Morris, J. S. (2015), "Bayesian function-on-function regression for multilevel functional data," *Biometrics*, 71, 563–574.

Mitchell, T. J., and Beauchamp, J. J. (1988), "Bayesian variable selection in linear regression," *Journal of the American Statistical Association*, 83, 1023–1032.

Morris, J. S., and Carroll, R. J. (2006), "Wavelet-based functional mixed models," *Journal of the Royal Statistical Society. Series B*, 68, 179–199.

Nason, G. (2010), *Wavelet methods in statistics with R* New York: Springer.

Petrone, S. (1999), "Random bernstein polynomials," *Scandinavian Journal of Statistics*, 26, 373–393.

Ramsay, J. (1988), "Monotone regression splines in action," *Statistical science*, 3, 425–441.

Schumaker, L. (2007), *Spline functions: basic theory* Cambridge University Press.

Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A., and Morris, J. S. (2016), "Functional CAR models for large spatially correlated functional datasets," *Journal of the American Statistical Association*, 111, 772–786.

Zhu, H., Brown, P. J., and Morris, J. S. (2011), "Robust, adaptive functional regression in functional mixed model framework," *Journal of the American Statistical Association*, 106, 1167–1179.

Zhu, H., Brown, P. J., and Morris, J. S. (2012), "Robust classification of functional and quantitative image data using functional mixed models," *Biometrics*, 68, 1260–1268.

Zhu, H., Versace, F., Cinciripini, P., and Morris, J. S. (2018), "Robust functional mixed models for spatially correlated functional regression, with application to event-related potentials for nicotine-addicted individuals," *Neuroimage*, 181, 501–512.