Supplementary materials for "Multivariate functional response regression, with application to fluorescence spectroscopy in a cervical pre-cancer study"

Hongxiao Zhu[a,*], Jeffrey S. Morris[b], Fengrong Wei[c], Dennis D. Cox[d]

[a] *Department of Statistics, Virginia Tech, Blacksburg, VA 24061*
[b] *The University of Texas M.D. Anderson Cancer Center, Houston, TX 77230*
[c] *Department of Mathematics, University of West Georgia, Carrollton, GA 30118*
[d] *Department of Statistics, Rice University, Houston, TX 77005*

1. MCMC algorithm

Following model (4) and the prior setups in Section 3.3 in the main text, we design the following Markov Chain Monte Carlo algorithm for posterior sampling:

Step 0. Initialize $\widetilde{\boldsymbol{B}}, \{q_k\}, \{s_k\}, \{\zeta_{ak}\}$ and set prior parameters. In particular, we use Henderson's mixed model equations (Searle et al., 1992, page 275-286) to obtain initial estimates for $\widetilde{\boldsymbol{B}}$ and the variance components $\{q_k\}, \{s_k\}$.

Step 1. Update $(\gamma_{ak}|\widetilde{\boldsymbol{D}}_k, \widetilde{\boldsymbol{B}}_{(-a)k}, \{q_k\}, \{s_k\})$ and $(\widetilde{B}_{ak}|\gamma_{ak}, \widetilde{\boldsymbol{D}}_k, \widetilde{\boldsymbol{B}}_{(-a)k}, \{q_k\}, \{s_k\})$ for $a = 1, \ldots, A$, where $\widetilde{\boldsymbol{B}}_{(-a)k}$ is the vector of $\widetilde{\boldsymbol{B}}_k$ with the ath component removed.

From model (5) in the main text, we see that $\widetilde{\boldsymbol{D}}_k|\widetilde{\boldsymbol{B}}_k, \{q_k\}, \{s_k\} \sim N(\boldsymbol{X}\widetilde{\boldsymbol{B}}_k, \boldsymbol{\Sigma}_k)$, where $\boldsymbol{\Sigma}_k = q_k\boldsymbol{Z}\boldsymbol{Z}^T + s_k\mathbf{I}_N$. We first update γ_{ak} by calculating the conditional odds:

$$\begin{aligned}\text{Conditional Odds} &= \frac{f(\gamma_{ak}=1|\widetilde{\boldsymbol{D}}_k, \widetilde{\boldsymbol{B}}_{(-a)k}, \boldsymbol{\Sigma}_k)}{f(\gamma_{ak}=0|\widetilde{\boldsymbol{D}}_k, \widetilde{\boldsymbol{B}}_{(-a)k}, \boldsymbol{\Sigma}_k)} \\ &= \frac{f(\widetilde{\boldsymbol{D}}_k|\gamma_{ak}=1, \widetilde{\boldsymbol{B}}_{(-a)k}, \boldsymbol{\Sigma}_k)}{f(\widetilde{\boldsymbol{D}}_k|\gamma_{ak}=0, \widetilde{\boldsymbol{B}}_{(-a)k}, \boldsymbol{\Sigma}_k)} \cdot \frac{f(\gamma_{ak}=1)}{f(\gamma_{ak}=0)} \qquad (1) \\ &= \text{Conditional Bayes Factor} \cdot \text{Prior Odds}.\end{aligned}$$

*Corresponding author
Email addresses: hongxiao@vt.edu (Hongxiao Zhu), jefmorris@mdanderson.org (Jeffrey S. Morris), fwei@westga.edu (Fengrong Wei), dcox@rice.edu (Dennis D. Cox)

Further derivations show that

$$\text{Conditional Bayes Factor} = \frac{f(\widetilde{\boldsymbol{D}}_k|\gamma_{ak}=1,\widetilde{\boldsymbol{B}}_{(-a)k},\boldsymbol{\Sigma}_k)}{f(\widetilde{\boldsymbol{D}}_k|\gamma_{ak}=0,\widetilde{\boldsymbol{B}}_{(-a)k},\boldsymbol{\Sigma}_k)}$$
$$= (1+\frac{\zeta_{ak}}{V_{ak}})^{-1/2}\exp\left\{\frac{1}{2}\xi_{ak}^2(1+V_{ak}/\zeta_{ak})^{-1}\right\}, \quad (2)$$

where $V_{ak} = [\boldsymbol{X}_{(a)}^T(\boldsymbol{\Sigma}_k)^{-1}\boldsymbol{X}_{(a)}]^{-1}$, $\boldsymbol{X}_{(a)}$ is the ath column of matrix $\boldsymbol{X}$, $\xi_{ak} = \widehat{B}_{ak}/\sqrt{V_{ak}}$, and $\widehat{B}_{ak} = V_{ak}\boldsymbol{X}_{(a)}^T\boldsymbol{\Sigma}_k^{-1}\widetilde{\boldsymbol{D}}_k^+$ for $\widetilde{\boldsymbol{D}}_k^+ = \widetilde{\boldsymbol{D}}_k - \sum_{l=1,l\neq a}^{A}\boldsymbol{X}_{(l)}\widetilde{B}_{lk}$.

Given $\{\gamma_{ak}\}$, we then update $\widetilde{\boldsymbol{B}}_k$ by using the results $(\widetilde{B}_{ak}|\gamma_{ak}=0,\widetilde{\boldsymbol{D}}_k,\widetilde{\boldsymbol{B}}_{(-a)k},\boldsymbol{\Sigma}_k) = \delta_0$, and $(\widetilde{B}_{ak}|\gamma_{ak}=1,\widetilde{\boldsymbol{D}}_k,\widetilde{\boldsymbol{B}}_{(-a)k},\boldsymbol{\Sigma}_k) \sim N(\mu_{ak}^0, W_{ak}^0)$, where $\mu_{ak}^0 = \widehat{B}_{ak}/(1+V_{ak}/\zeta_{ak})$ and $W_{ak}^0 = V_{ak}/(1+V_{ak}/\zeta_{ak})$.

Step 2. Update $\{q_k\}$ and $\{s_k\}$ using Metropolis-Hastings. First, propose new values of q_k, s_k using log transform, e.g., $\log(\widetilde{q}_k) = \log(q_k) + e\epsilon$, for $\epsilon \sim N(0,1)$ and e is the pre-specified step size. Second, calculate the proposal ratio of the random walk proposal:

$$\frac{f(q_k|\widetilde{q}_k)}{f(\widetilde{q}_k|q_k)} = \frac{\widetilde{q}_k}{q_k}.$$

The new parameters $\{\widetilde{q}_k, \widetilde{s}_k\}$ are accepted with probability $\min\{o_{jk}, 1\}$, where

$$o_{jk} = \frac{f(\widetilde{\boldsymbol{D}}_k|\widetilde{q}_k,\widetilde{s}_k,\cdot)\pi(\widetilde{q}_k)\pi(\widetilde{s}_k)f(q_k|\widetilde{q}_k)f(s_k|\widetilde{s}_k)}{f(\widetilde{\boldsymbol{D}}_k|q_k,s_k,\cdot)\pi(q_k)\pi(s_k)f(\widetilde{q}_k|q_k)f(\widetilde{s}_k|s_k)}.$$

Here we have assumed $q_k \sim \text{IG}(a_k^q, b_k^q)$, therefore

$$\log\frac{\pi(\widetilde{q}_k)}{\pi(q_k)} = (a_k^q+1)(\log(q_k)-\log(\widetilde{q}_k)) + b_k^q(1/q_k - 1/\widetilde{q}_k).$$

Similar formula holds for the ratio $\pi(\widetilde{s}_k)/\pi(s_k)$. The log likelihood ratio takes the form

$$\log\frac{f(\widetilde{\boldsymbol{D}}_k|\widetilde{q}_k,\widetilde{s}_k,\cdot)}{f(\widetilde{\boldsymbol{D}}_k|q_k,s_k,\cdot)} = \frac{|\widetilde{\boldsymbol{\Sigma}}_k|^{-1/2}\exp\{-1/2(\widetilde{\boldsymbol{D}}_k - \boldsymbol{X}\widetilde{\boldsymbol{B}}_k)^T\widetilde{\boldsymbol{\Sigma}}_k^{-1}(\widetilde{\boldsymbol{D}}_k - \boldsymbol{X}\widetilde{\boldsymbol{B}}_k)\}}{|\boldsymbol{\Sigma}_k|^{-1/2}\exp\{-1/2(\widetilde{\boldsymbol{D}}_k - \boldsymbol{X}\widetilde{\boldsymbol{B}}_k)^T\boldsymbol{\Sigma}_k^{-1}(\widetilde{\boldsymbol{D}}_k - \boldsymbol{X}\widetilde{\boldsymbol{B}}_k)\}}$$

where $\boldsymbol{\Sigma}_k = q_k\mathbf{Z}\mathbf{Z}^\mathbf{T} + s_k\mathbf{I}_N$, and $\widetilde{\boldsymbol{\Sigma}}_k = \widetilde{q}_k\mathbf{Z}\mathbf{Z}^\mathbf{T} + \widetilde{s}_k\mathbf{I}_N$.

Step 3. Update $\widetilde{\boldsymbol{U}}_k$ from $(\widetilde{\boldsymbol{U}}_k|\cdot) \sim N(\boldsymbol{\mu}_{\mathbf{U}_k}, \mathbf{W}_{\mathbf{U}_k})$, where $\mathbf{W}_{\mathbf{U}_k} = (s_k^{-1}\mathbf{Z}^T\mathbf{Z} + q_k^{-1}\mathbf{I}_M)^{-1}, \boldsymbol{\mu}_{\mathbf{U}_k} = \mathbf{W}_{\mathbf{U}_k}\mathbf{Z}^T s_k^{-1}(\widetilde{\boldsymbol{D}}_k - \boldsymbol{X}\widetilde{\boldsymbol{B}}_k)$. This step is optional.

Repeat Steps 1-3 until reaching a pre-specified maximum number of iterations.

2. Setting initial values and hyper-prior parameters

The proposed Bayesian MFMM requires multiple sets of hyper-prior parameters, and the MCMC algorithm requires initial values for $\widetilde{\boldsymbol{B}}, \{q_k\}, \{s_k\}$ and $\{\zeta_{ak}\}$. We discuss the setting of these parameters as follows:

1. *Initial values for* $\widetilde{\boldsymbol{B}}, \{q_k\}, \{s_k\}$ *and* $\{\zeta_{ak}\}$. Initial values of $\boldsymbol{B}, \{q_k\}, \{s_k\}$ were calculated by applying Henderson's Mixed Model Equations (MMEs) (Searle et al., 1992, pages 275-286) independently across the index k. A MME assumes that:

$$\mathbf{d}_k = \mathbf{X}\mathbf{b}_k + \mathbf{Z}\mathbf{u}_k + \mathbf{e}_k,$$

with $E\{\mathbf{u}_k\} = \mathbf{0}, \mathrm{Var}(\mathbf{u}_k) = q_k I_M$, $E\{\mathbf{e}_k\} = \mathbf{0}, \mathrm{Var}(\mathbf{e}_k) = s_k I_N$. Both $\mathbf{u}_k$ and $\mathbf{e}_k$ are assumed to be Gaussian distributed. The algorithm follows that of Viktor Witkovsky (available at Matlab Central's website: `http://www.mathworks.com/matlabcentral/fileexchange/200-mixed`). The solution of MME includes $\{\widehat{\mathbf{b}}_k\}$, $\{\widehat{q}_k, \widehat{s}_k\}$. These estimates were treated as the initial values for $\{\widetilde{\boldsymbol{B}}_k\}$, $\{q_k\}, \{s_k\}$ in our MFMM model (4). The initial values for $\{\zeta_{ak}\}$ are estimated by $\mathrm{var}(\widehat{\mathbf{b}}_k)$ which is available as a byproduct of the MME approach.

2. *Parameters for the Inverse-Gamma priors.* We determine the Inverse-Gamma parameters $\{(a_k^s, b_k^s)\}$, $\{(a_k^q, b_k^q)\}$, and $\{(a_k^\zeta, b_k^\zeta)\}$ by letting the mode of Inverse-Gamma distributions equal to the initial values and letting the variance of the Inverse-Gamma distributions equal to a pre-specified large value (e.g., 10^3). With these constraints, we can solve the two parameters explicitly for each k.

3. *Parameters for the Beta priors.* We determine the Beta parameters (a_k^π, b_k^π) of π_{ak} following the similar way to determining the Inverse-Gamma parameters. In particular, we let the mode of the Beta distribution equal to the initial value $\widehat{\pi}_{ak}$ and let the variance of the Beta distribution be a pre-specified value (e.g., 0.08). The initial values $\{\widehat{\pi}_{ak}\}$ are calculated from the (conditional) odds using equations (1),(2), and the initial values of $\widetilde{\boldsymbol{B}}, \{q_k\}, \{s_k\}$ and $\{\zeta_{ak}\}$.

3. Code sharing

We have shared matlab code which calls a C executable for implementing the proposed approach. In the shared code, we demonstrated an example that runs simulate 1 without replication. We have shared this code through the *RunMyCode* repository (`http://www.runmycode.org/`).

4. More results for sensitivity analysis

As discussed in Section 6 of the main text, we have performed a sensitivity analysis to study whether the results presented in Section 5 in the main text are sensitive to different model choices or hyperparameter settings. We have performed the analyses (i)-(iii) described in the main text. In Figure 1, we plot the posterior mean and the 95% SCB for the contrast effect $\underline{C}$, and flagged regions detected using the SimBaS and BFDR approaches on the posterior mean. The results from the original analysis is shown in (a.1)-(a.2) in the first row, and the results from analyses (i)-(iii) are demonstrated in rows 2-4 in Figure 1.

Figure 1 shows that, all analyses result in posterior means and the 95% SCBs similar to the original analysis. The flagged regions using BFDR ($\delta = 0.02$) are simlar across all analysis. The flagged locations using the SimBaS approach demonstrate similar patterns across all analysis, with slight variations shown at a few low-intensity regions on exictation curves 380-390 and 460-480 nms.

5. Convergence diagnostics for MCMC sampling

As discussed in Section 7 in the main text, the convergence of the MCMC samples needs to be monitored and tested. We monitored the behavior of the posterior samples by checking the trace plots, the autocorrelation plots, and calculating the effective sample size (ESS) of the retained samples. Stable trace plot and low autocorrelation indicate good mixing, and a higher effective sample size indicates a better mixing. We tested the convergence of the chains by calculating the Geweke's Z-statistics (Geweke, 1992) based on the retained samples

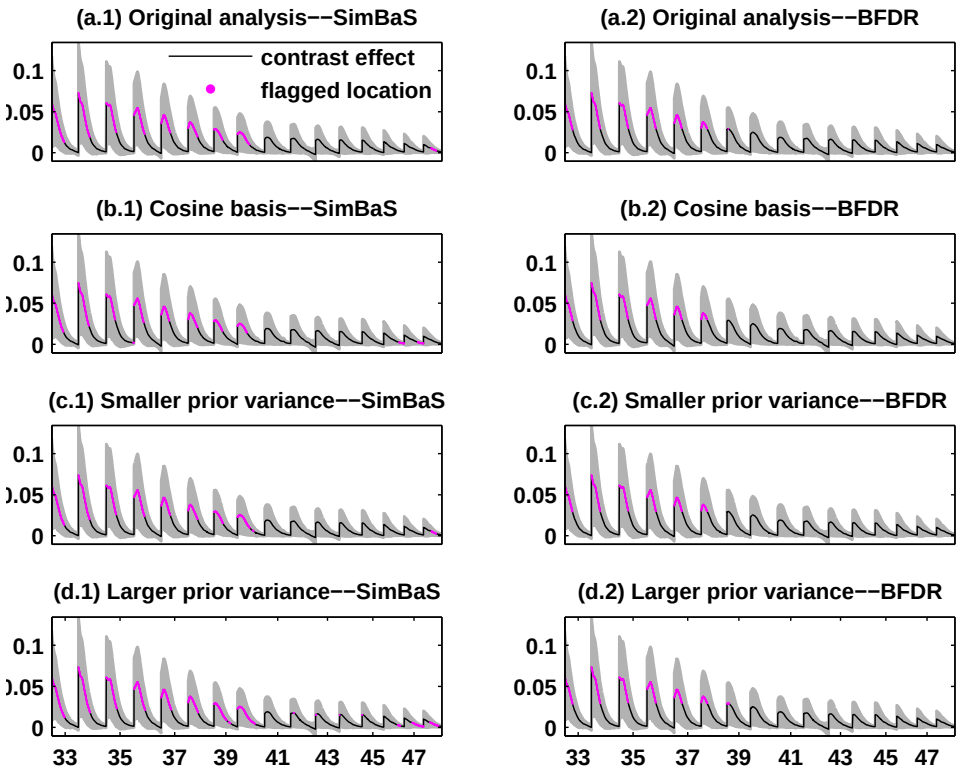

Figure 1: Regions detected using the SimBaS and BFDR ($\delta = 0.02$) approaches. Detected regions are flagged in magenta on posterior mean of the contrast effect $\underline{C}$. Row 1 contains the results from the original analysis, and rows 2, 3, and 4 are results for analyses (i), (ii) and (iii) respectively. The shaded gray regions are the 95% SCBs calcuated by adjusting EER across all 16 curves.

(i.e., throwing away samples in the burnin period). Since Geweke's Z-statistic has an asymptotically $N(0,1)$ distribution, Z-statistics with absolute values less than 2 indicate that the first 10% and the last 50% of the retained samples have no differences in terms of their means. When the test is performed on multiple components of a sequence/array, we expect to see that the type I error is less than the significant level α (e.g., $\alpha = 5\%$), i.e., $1-\alpha$ of the testings give Z-statistics < 2 in absolute value.

We performed these diagnostics on the simulation runs and the real data analysis. In Figure 2 - 3, we show summary plots for the diagnostic analysis for simulation 1 (based on one run). From these plots, we see that the parameters $\widetilde{\boldsymbol{B}}, \{\pi_{ak}\}$, and $\{\zeta_{ak}\}$ are mixing well, most having ESS greater than 1000 (based on the 2000 retained samples) and more than 90% of the Geweke's Z-statistics lie within the $[-2,2]$ interval. In contrast, some components of $\widetilde{\boldsymbol{U}}$ show higher autocorrelations (e.g., the component $\widetilde{\boldsymbol{U}}_{11}$) and the proportion that ESS$>$ 1000 for components in $\widetilde{\boldsymbol{U}}$ is lower (0.88). The good mixing of $\widetilde{\boldsymbol{B}}$ and its hyper-parameters may be explained by the fact that we have integrated out the random effect $\widetilde{\boldsymbol{U}}$ while updating $\widetilde{\boldsymbol{B}}$. Another observation is that, the ESS for $\{q_k\}, \{s_k\}$ are lower, with $\{q_k\}$ having around 200 and $\{s_k\}$ having around 400 ESS. This may be caused by the fact that Metropolis-Hastings samplers usually have lower than 50% acceptance rate, therefore the chains are not updating very frequently which results in lower ESS. Since our posterior inference has been focused on the fixed effects, we found 5000 iterations with a 3000 burnin period sufficient for our analysis. If the inference for the variance components is of interest, we suggest to run more iterations (e.g., keeping at least 10^4 samples after burnin).

Summary plots for diagnostics of the real data analysis are demonstrated in Figure 4. From these plots, we see that the mixing of the parameters are similar to that of the simulation studies. As in the simulation, we see that the ESS for the variance components $\{s_k\}$ is relatively low (around 200 based on 2000 samples).

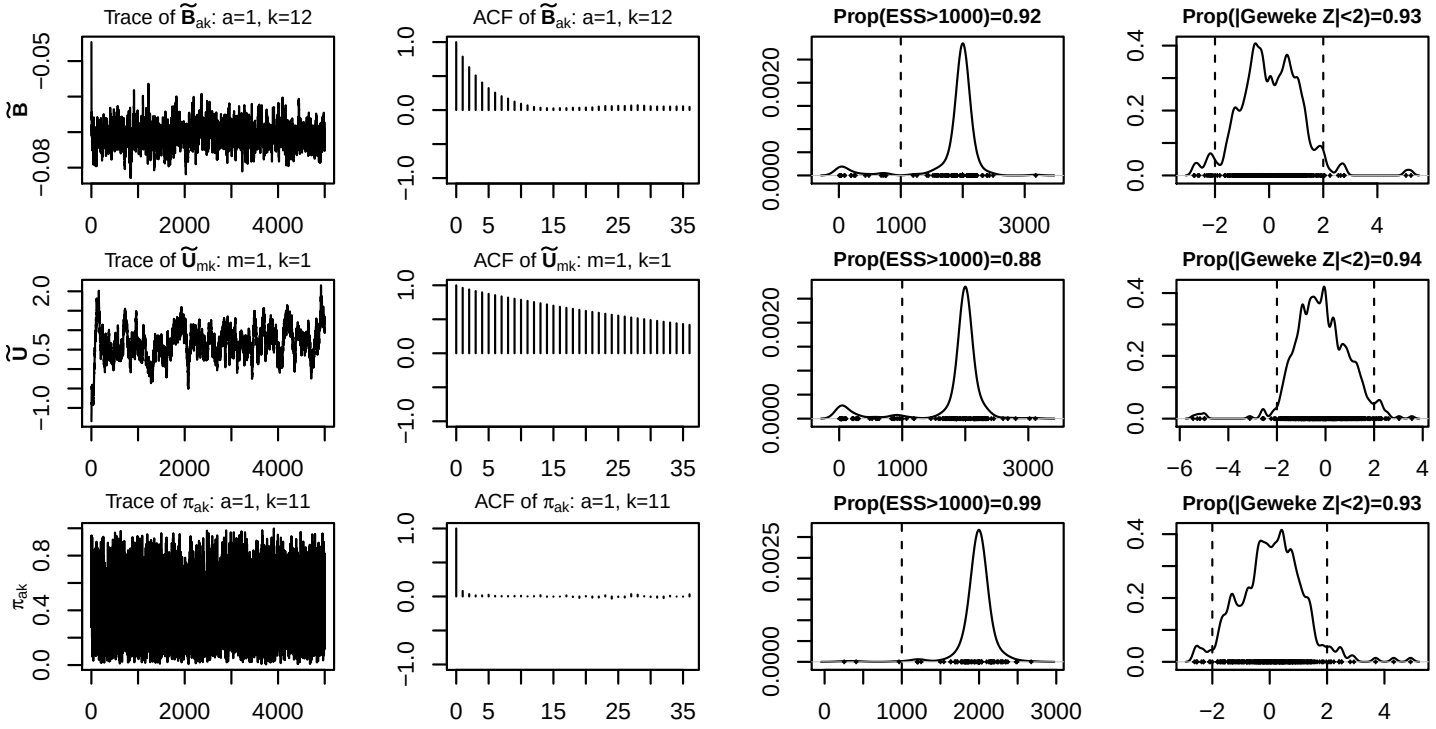

Figure 2: Summary plots of the convergence diagnostics for Simulation 1 (part I). Rows 1-3 are the summary plots for the parameters $\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{U}}$ and $\{\pi_{ak}\}$ respectively. In each row, the first two columns are the trace plot and autocorrelation plot for a selected component, the column 3 is the kernel density estimation of the effective sample sizes (ESS) calculated for all components, and the column 4 is the kernel density estimation of the Geweke's Z-statistics calculated for all components. In the titles of plots in columns 3-4, we have noted the proportion of all components with values greater than a threshold.

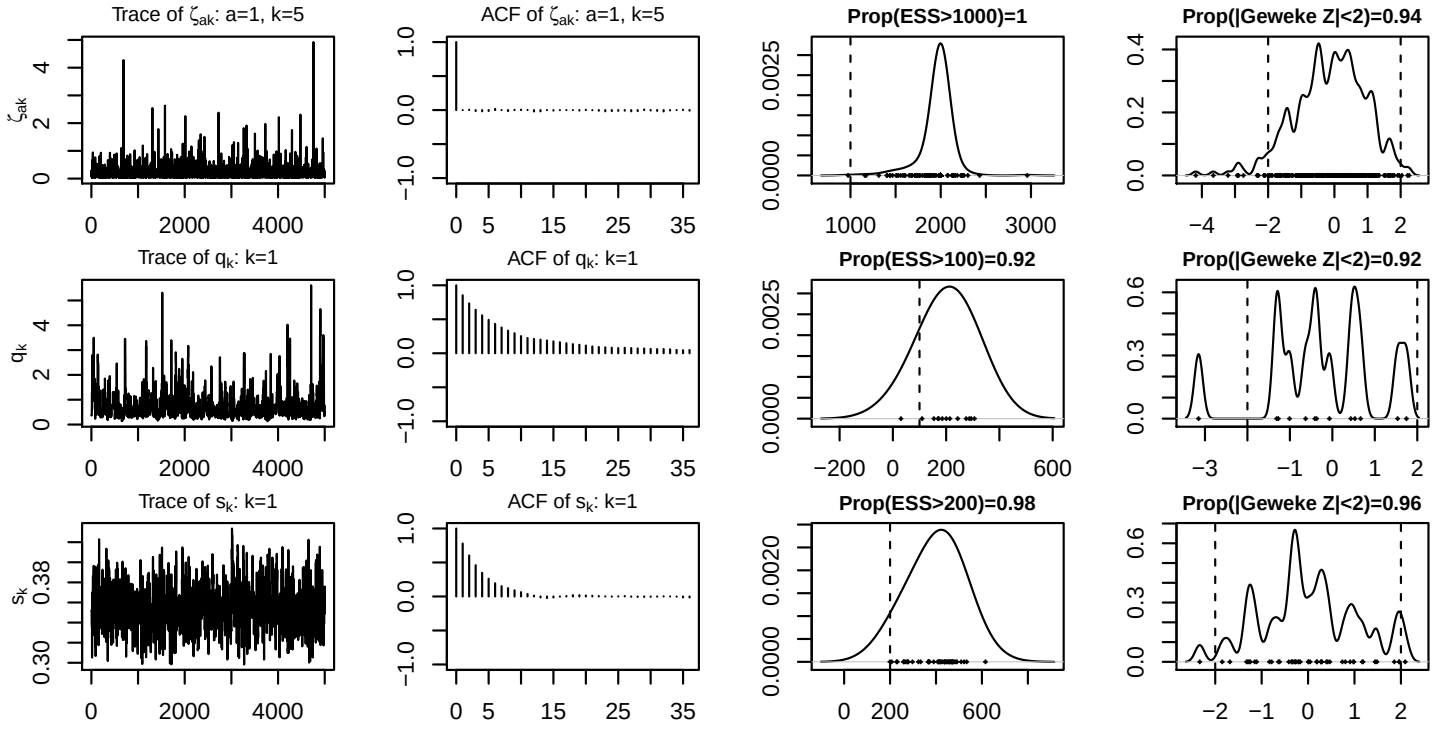

Figure 3: Summary plots of the convergence diagnostics for Simulation 1 (part II). Rows 1-3 are the summary plots for the parameters $\{\zeta_{ak}\}, \{q_k\}$ and $\{s_k\}$ respectively. In each row, the first two columns are the trace plot and autocorrelation plot for a selected component, the column 3 is the kernel density estimation of the effective sample sizes (ESS) calculated for all components, and the column 4 is the kernel density estimation of the Geweke's Z-statistics calculated for all components. In the titles of plots in columns 3-4, we have noted the proportion of all components with values greater than a threshold.

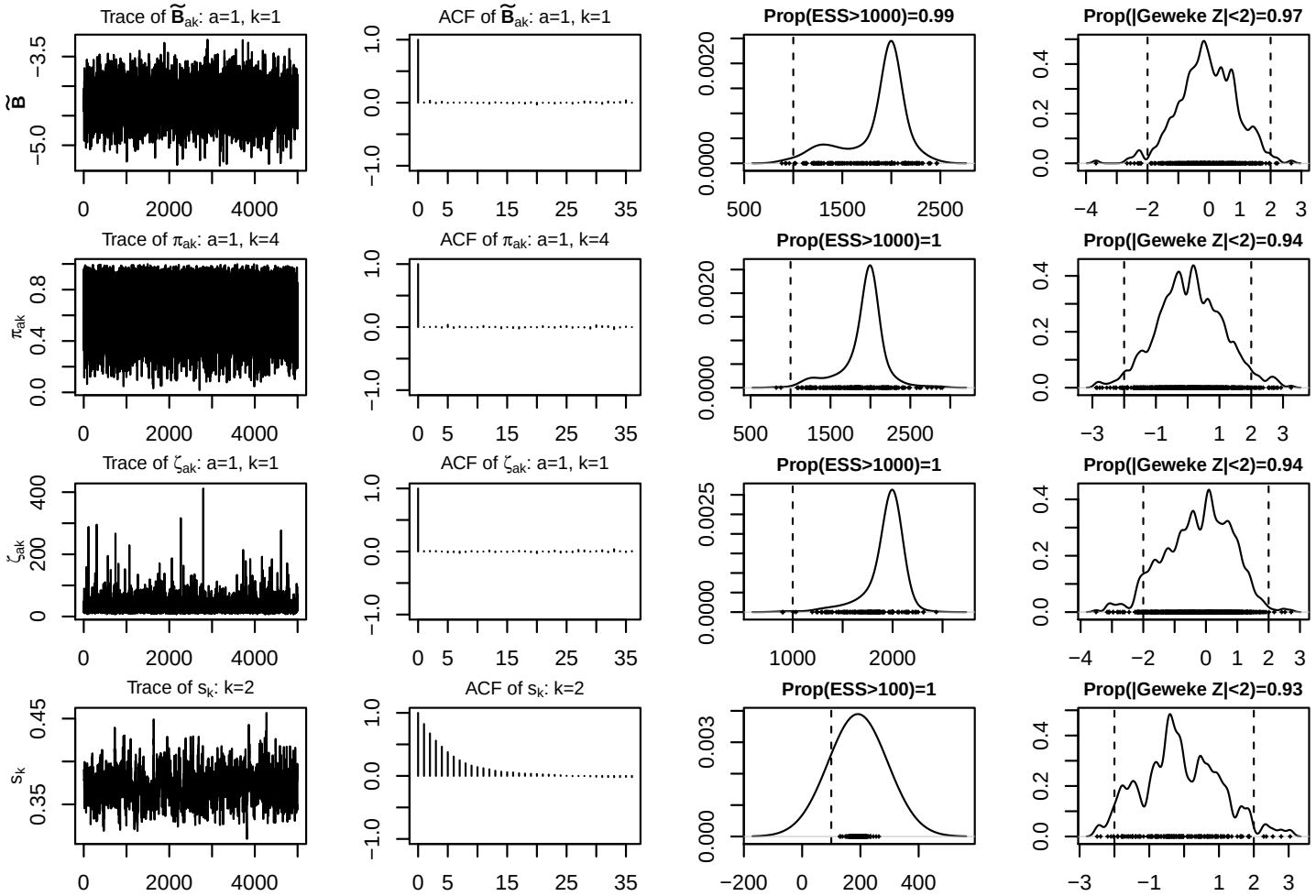

Figure 4: Summary plots of the convergence diagnostics for the real data analysis. Rows 1-4 are the summary plots for the parameters $\widetilde{\boldsymbol{B}}, \{\pi_{ak}\}, \{\zeta_{ak}\}$ and $\{s_k\}$ respectively. In each row, the first two columns are the trace plot and autocorrelation plot for a selected component, the column 3 is the kernel density estimation of the ESS calculated for all components, and the column 4 is the kernel density estimation of the Geweke's Z-statistics calculated for all components. In the titles of plots in columns 3-4, we have noted the proportion of all components with values greater than a threshold.

References

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Bayesian Statistics 4 .

Searle, S.R., Casella, G., McCulloch, C.E. (Eds.), 1992. Variance Components. John Wiley & Sons, New York.