# Latent class modeling using matrix-valued covariates with application to identifying early placebo responders based on EEG signals

Bei Jiang
Department of Mathematical and Statistical Sciences
University of Alberta

Joint work with

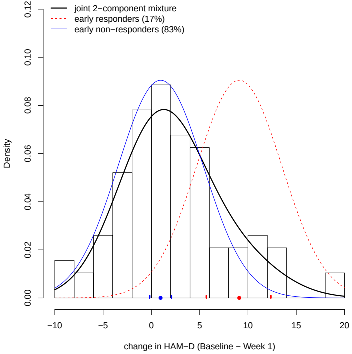
Eva Petkova (NYU), Thad Tarpey (Wright State), Todd Ogden (Columbia)

BIRS 2016, Banff

February 1, 2016

# Motivating study: placebo response

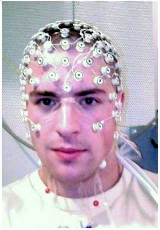
- Placebo response, i.e., *a positive medical response due to placebo effect, as if there were an active medication*, to antidepressant treatment is highly prevalent.
- For example, 96 placebo or drug treated depression patients

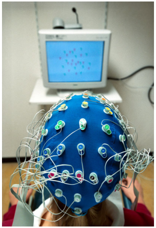



- Hamilton Depression (HAM-D) scale is a clinical measure to rate severity of depression
- Higher HAM-D scores indicate more severe depression
- An antidepressant takes more than 2 weeks to show real drug effect
- Subjects may cluster into two clinically relevant subgroups: early responder vs. non-responder.

# Motivating study: scientific goals

- Interest has focused on studying patient's characteristics that could contribute to placebo response
- However, typically measured clinical phenotypes, e.g. symptom severity and treatment history, have shown low predictive power.
- Now explore predictive ability of neuroimaging phenotypes, e.g, the electrical brain activity under certain tasks measured through Electroencephalography (EEG)





https://en.wikipedia.org/wiki/Electroencephalography
http://www.lsa.umich.edu/psych/danielweissmanlab/whatiseeg.htm

# Motivating study: statistical goals

- For clinical outcome HAM-D scores, formulate a latent class model
  - take into account uncertainty of latent class membership
- EEG covariate to predict latent class membership
  - $14 \times 45$ matrix (order-2 tensor)
  - captures brain activity measured at 14 electrodes, crossed with 45 frequencies within the theta (4 - 7 hz) and alpha (7 - 15 hz) bands.

# Motivating study: statistical goals

- Our approach: extend latent class models to incorporate matrix-valued EEG covariate
  - utilizes low rank CANDECOMP/PARAFAC (CP) decomposition (Kolda and Bader 2009) to represent the target coefficient matrix
    - reduce model dimensionality
    - explicitly capture bilinear structure
  - CP decomposition was previously considered by Hung and Wang (2013); Zhou et al. (2014) (penalized maximum likelihood approach)
  - In contrast, we adopt a hierarchical approach in formulating the CP decomposition and a Bayesian method for estimation (next)
    - provides a flexible way to incorporate prior knowledge on the patterns of covariate effect heterogeneity
    - provides a data-driven method of regularization
    - associated measures of uncertainty can be quantified through credible intervals

# The proposed methodology

- The manifest model for clinical outcome $y_i$

$$y_i = \eta_0 + \eta_1 \gamma_i + \epsilon_i, \ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

  - $\gamma_i = 1$ indicates a placebo responder; $\gamma_i = 0$ a non-responder;
  - Constrain $\eta_0 + \eta_1 > 0$: placebo responders are expected to experience improved mood and hence positive clinical outcome values.

# The proposed methodology

- The low rank probit model:

$$\begin{aligned}\Phi^{-1}[p\{\gamma_i = 1\}] &= \boldsymbol{\theta}^T \boldsymbol{z}_i + \langle \boldsymbol{\Theta}, \boldsymbol{x}_i \rangle, \\ &= \boldsymbol{\theta}^T \boldsymbol{z}_i + \left\langle \sum_{r=1}^{R} \boldsymbol{\alpha}_r \boldsymbol{\beta}_r^T, \boldsymbol{x}_i \right\rangle, \\ &= \boldsymbol{\theta}^T \boldsymbol{z}_i + \left\langle \boldsymbol{A}\boldsymbol{B}^T, \boldsymbol{x}_i \right\rangle,\end{aligned}$$

  - $\boldsymbol{x}_i$ is $p \times q$ matrix covariate
  - $\boldsymbol{z}_i$ is a vector of scalar covariates
  - $\boldsymbol{\Theta} \in \mathbb{R}^{p \times q}$ denotes the target coefficient matrix
  - $\langle \boldsymbol{\Theta}, \boldsymbol{x}_i \rangle = \langle vec(\boldsymbol{\Theta}), vec(\boldsymbol{x}_i) \rangle$
  - CP decomposition: express $\boldsymbol{\Theta} = \sum_{r=1}^{R} \boldsymbol{\alpha}_r \boldsymbol{\beta}_r^T$, where $\boldsymbol{\alpha}_r \in \mathbb{R}^p$ and $\boldsymbol{\beta}_r \in \mathbb{R}^q$, $r = 1, \cdots, R < \min(p, q)$
  - let $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_R] \in \mathbb{R}^{p \times R}$ and $\boldsymbol{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_R] \in \mathbb{R}^{q \times R}$, we can re-write $\boldsymbol{\Theta} = \boldsymbol{A}\boldsymbol{B}^T$
  - Now reduced to estimating $\boldsymbol{A}$ and $\boldsymbol{B}$: a total of $R(p+q)$ parameters

# The proposed methodology

- Alternatively: re-express $\boldsymbol{A}$ and $\boldsymbol{B}$ w.r.t. their row vectors
  - $\boldsymbol{A} = [\tilde{\boldsymbol{\alpha}}_1, \cdots, \tilde{\boldsymbol{\alpha}}_p]^T$, $\boldsymbol{B} = [\tilde{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{\beta}}_q]^T$
  - $\tilde{\boldsymbol{\alpha}}_j \in \mathbb{R}^R$ and $\tilde{\boldsymbol{\beta}}_k \in \mathbb{R}^R$ can be interpreted as representing the effects due to the row and column components of the matrix covariate
  - $\boldsymbol{\Theta}_{j,k} = \langle \tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\beta}}_k \rangle$ is equivalent to modeling the two-way interaction effects
  - next we propose hierarchical priors on $\{\tilde{\alpha}_j\}_{j=1}^p$ and $\{\tilde{\beta}_k\}_{k=1}^q$

# The hierarchical formulation of CP decomposition

- For the row and column effect vectors in the CP decomposition, we consider the following hierarchical priors,

$$\tilde{\boldsymbol{\alpha}}_1, \cdots, \tilde{\boldsymbol{\alpha}}_p \overset{iid}{\sim} \text{MVN}(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha); \text{ and } \tilde{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{\beta}}_q \overset{iid}{\sim} \text{MVN}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$$

  - $\tilde{\boldsymbol{\alpha}}_j^T$ is the $j^{th}$ row of $\boldsymbol{A}$ and $\tilde{\boldsymbol{\beta}}_k^T$ is the $k^{th}$ row of $\boldsymbol{B}$
  - allows borrowing information and also provides a data-driven method of regularization.
- To complete the specification of these hierarchical priors, we define the following hyper-priors,

$$\boldsymbol{\mu}_\alpha, \boldsymbol{\mu}_\beta \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_0); \text{ and } \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_\beta \sim \text{inverse Wishart}(\boldsymbol{S}_0, s_0)$$

  - let $\boldsymbol{\Sigma}_0 = 9/4\boldsymbol{I}$ to bound $\Pr(\gamma_i = 1)$ to be away from 0 and 1
  - a diffuse prior for $\boldsymbol{\Sigma}_\alpha$ and $\boldsymbol{\Sigma}_\beta$ with $\boldsymbol{S}_0 = 10\boldsymbol{I}$, and $s_0 = R + 1$.

# The hierarchical formulation of CP decomposition

- The CP decomposition of $\boldsymbol{\Theta}$ suffers from non-identifiability of $\boldsymbol{A}$ and $\boldsymbol{B}$ separately, since $\boldsymbol{A}\boldsymbol{B}^T = \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}\boldsymbol{B}^T$, for any $R \times R$ non-singular matrix $\boldsymbol{\Lambda}$.
- A consequence of this complication is that $\boldsymbol{\mu}_\alpha, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_\beta$ are not individually identifiable either.
- However, from a Bayesian perspective, good mixing and convergence can be achieved for all parameters in the identifiable $\boldsymbol{\Theta} = \boldsymbol{A}\boldsymbol{B}^T$.

# priors for other model parameters

- In the manifest model, we assume diffuse priors: $\eta_0 \sim N(0, \tau_0^2)$, $\eta_1 \sim N(0, \tau_0^2)\mathrm{I}(-\eta_0, \infty)$ with $\tau_0^2 = 100$ and $\sigma^2 \sim$ inverse gamma$(a_0, b_0)$ with $a_0 = b_0 = 0.01$.
- For covariate effect parameters in the probit model, we let $\boldsymbol{\theta} \sim N(\mathbf{0}, \boldsymbol{V}_0)$, where $\boldsymbol{V}_0 = (9/4)\boldsymbol{I}$ would bound the probability to be away from 0 and 1.

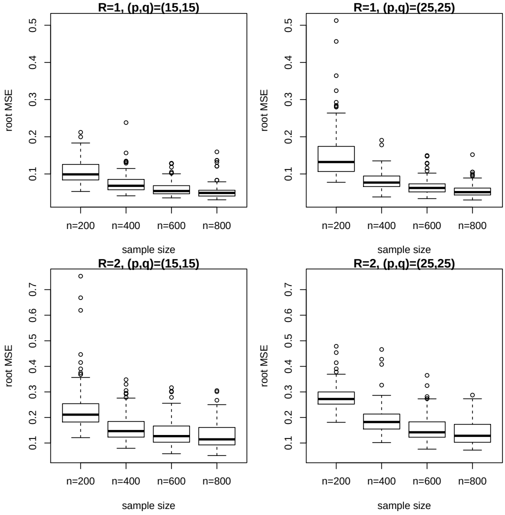
# Posterior computation

- note that $\left\langle \boldsymbol{A}\boldsymbol{B}^T, \boldsymbol{x}_i \right\rangle$ in the probit model can be rewritten as a linear function with respect to $\tilde{\alpha}_1, \cdots, \tilde{\alpha}_p$ or $\tilde{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{\beta}}_q$ as follows,

$$\left\langle \boldsymbol{A}\boldsymbol{B}^T, \boldsymbol{x}_i \right\rangle = \sum_{j=1}^{p} \tilde{\alpha}_j^T \boldsymbol{u}_{ij} = \sum_{k=1}^{q} \tilde{\boldsymbol{\beta}}_k^T \boldsymbol{v}_{ik}$$

  - $\boldsymbol{u}_{ij}$ denotes the $j^{th}$ row of $\boldsymbol{x}_i^T \boldsymbol{B} \in \mathbb{R}^{p \times R}$, $j = 1, \cdots, p$;
  - $\boldsymbol{v}_{ik}$ denotes the $k^{th}$ row of $\boldsymbol{x}_i^T \boldsymbol{A} \in \mathbb{R}^{q \times R}$, $k = 1, \cdots, q$.
- We introduce a latent variable $w_i$ such that $\gamma_i = \mathrm{I}(w_i > 0)$ and $w_i \sim \mathrm{N}(\boldsymbol{\theta}^T \boldsymbol{z}_i + \langle \boldsymbol{A}\boldsymbol{B}^T, \boldsymbol{x}_i \rangle, 1)$ (Albert and Chib,1993).
- $\{\tilde{\alpha}_j\}_{j=1}^{p}$ and $\{\tilde{\boldsymbol{\beta}}_k^T\}_{k=1}^{q}$ can be updated iteratively in a similar fashion as in a regular regression model

# Simulation study I: known rank



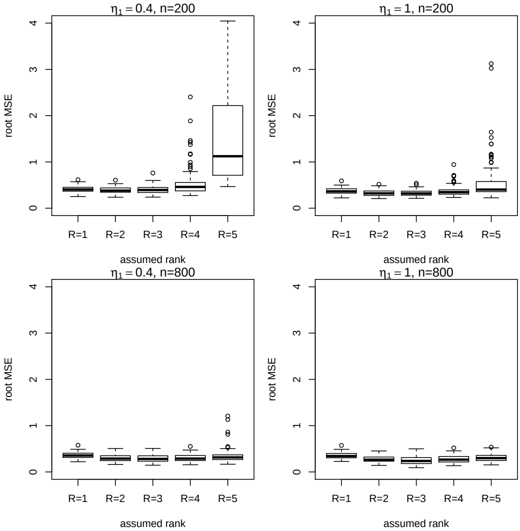

- the degree of overlapping in latent classes is fixed: $\eta_0 = 0$, $\eta_1 = 1.0$ & $\sigma^2 = 0.2^2$ (well separated).
- true values for $R$, $(p, q)$, and $n$ vary under different simulation scenarios.
- $\text{MSE} = \frac{1}{S}\sum_{s=1}^{S}\left\{\frac{1}{pq}\|\hat{\boldsymbol{\Theta}}^{(s)} - \boldsymbol{\Theta}^{(s)}_{\text{true}}\|_F^2\right\}$

# Simulation study I: known rank

Table: AUC for prediction of binary latent class indicator.

| | true rank $R = 1$ | | | | true rank $R = 2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $n = 200$ | $n = 400$ | $n = 600$ | n=800 | $n = 200$ | $n = 400$ | $n = 600$ | n=800 |
| **within sample AUC** | | | | | | | | |
| $(p, q) = (15, 15)$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $(p, q) = (25, 25)$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **out of sample AUC** | | | | | | | | |
| $(p, q) = (15, 15)$ | 0.86 | 0.90 | 0.90 | 0.91 | 0.80 | 0.88 | 0.91 | 0.92 |
| $(p, q) = (25, 25)$ | 0.87 | 0.93 | 0.94 | 0.95 | 0.74 | 0.88 | 0.92 | 0.94 |

# Simulation study II: misspecified rank




- true rank $R = 3$, $(p, q) = (15, 15)$; $n = 200$ or 800.
- $\eta_1 = 0.4$ and $\eta_1 = 1.0$ indicate high and low degrees of overlapping.
- the models are fit with varying assumed rank values

# Simulation study II: misspecified rank

Table: AUC for prediction of binary latent class indicator.

| | **high degree overlapping** | | | | | **low degree overlapping** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| assumed rank | R=1 | R=2 | R=3 | R=4 | R=5 | R=1 | R=2 | R=3 | R=4 | R=5 |
| | | | | | **within sample AUC** | | | | | |
| | 0.94 | 0.96 | 0.96 | 0.95 | 0.93 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | | | | **out of sample AUC** | | | | | |
| | 0.80 | 0.84 | 0.84 | 0.81 | 0.79 | 0.82 | 0.88 | 0.88 | 0.86 | 0.84 |

# Application to identify placebo responder subgroup using EEG data

- let $y_i$ denote the change in HAM-D (baseline - week 1)
- let $\boldsymbol{x}_i^* \in \mathbb{R}^{14 \times 45}$ denote Current Source Density amplitude spectrum values ($\mu V/m^2$) (Keyser and Tenke, 2006) at 14 electrodes in brain's posterior region, crossed with 45 frequency bands within the theta (4 - 7 hz) and alpha (7 - 15 hz) frequency waves
  - Multilinear Principal Component Analysis (MPCA) to reduce the original dimension of EEG data from $(p, q)$ to $(p_0, q_0)$.
  - The dimension $(p_0, q_0)$ and rank $R$ are determined by the widely applicable information criterion (WAIC) proposed by Watanabe (2010).
- let $\boldsymbol{z}_i$ denote gender and depression chronicity

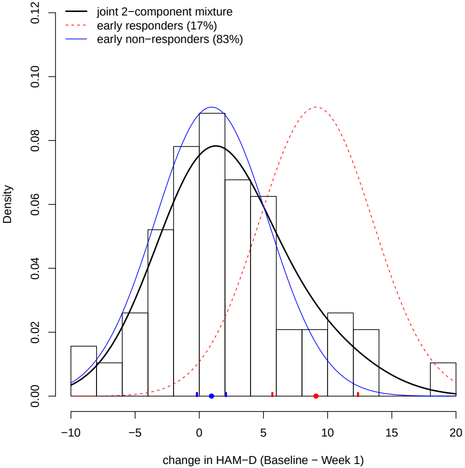
# Application to identify placebo responder subgroup using EEG data

Table: WAIC from fitting different models for the prediction of the placebo responder subgroup using EEG data.

| | rank R=1 | | | | | rank R=2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $p_0$ / $q_0$ | 2 | 3 | 4 | 5 | 6 | 2 | 3 | 4 | 5 | 6 |
| 2 | 603.4 | 605.5 | 601.3 | 600.2 | 601.9 | 604.4 | 604.6 | 601.0 | 601.3 | 599.1 |
| 3 | 602.8 | 601.2 | 594.2 | 593.9 | 594.8 | 604.0 | 599.2 | 596.2 | 596.5 | 603.8 |
| 4 | 602.7 | 603.9 | 583.7 | 590.2 | 595.1 | 604.1 | 601.7 | 595.1 | 596.0 | 602.0 |
| 5 | 601.1 | 602.4 | 591.6 | 598.0 | 598.1 | 603.5 | 602.3 | 598.5 | 597.8 | 598.0 |
| 6 | 598.1 | 600.1 | 577.3 | 577.9 | 582.1 | 600.6 | 601.7 | 589.9 | 588.6 | 588.7 |
| 7 | 582.9 | 591.1 | 571.9 | 578.7 | 578.7 | 599.2 | 595.4 | 588.1 | 593.9 | 596.2 |
| 8 | 584.7 | 589.4 | 573.3 | 574.7 | 575.0 | 599.3 | 597.3 | 590.1 | 589.8 | 594.0 |
| 9 | 587.1 | 594.4 | **568.8** | 573.1 | 573.5 | 602.3 | 600.6 | 591.8 | 594.4 | 592.8 |
| 10 | 588.4 | 594.2 | 571.1 | 574.1 | 583.8 | 600.5 | 600.0 | 590.6 | 588.8 | 590.3 |

Note: WAIC for no-mixture model is 845.6.

# Application to identify placebo responder subgroup using EEG data




- the proportion of placebo responders for the placebo arm and drug arm: 9/50=18% and 7/46=15% respectively.
- a chi-square test indicates no significant difference

# Application to identify placebo responder subgroup using EEG data

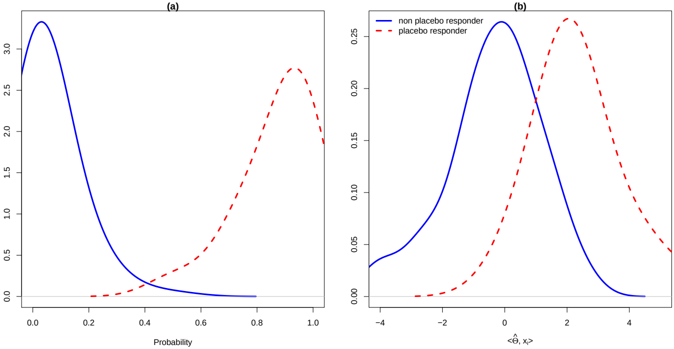
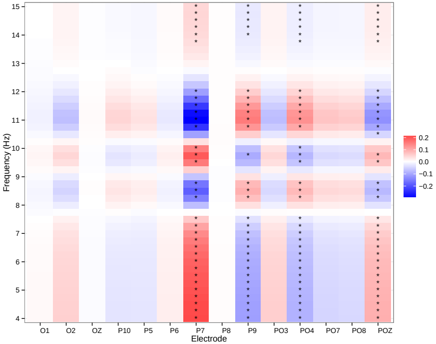



Figure: left: posterior density estimate of the probability of being in the placebo responder subgroup; right: posterior density estimate of $< \boldsymbol{\Theta}, \boldsymbol{x}_i >$.

# Application to identify placebo responder subgroup using EEG data

- Heatmap of the estimated coefficient matrix (* indicates significance at the 0.05 level)




- being chronically depressed is less likely for placebo response ($\hat{\theta}_2 = -1.65$ (95% CI: -3.59, -0.12)), while gender is not a contributing factor ($\hat{\theta}_1 = -1.34$ (95% CI: -3.14, 0.02)).

# Summary

- We consider a low rank hierarchical latent class model to incorporate matrix-valued covariates.
- The proposed approach readily extends to incorporate the covariates that are multi-dimensional arrays in general regression settings.
- Our simulation studies have shown that our proposed hierarchical approach is robust against rank misspecification.
- The findings in the application raise hope for utilizing EEG measures to differentiate potential placebo responders from non-responders in clinical practice to further guide the selection of effective treatment for depression patients.

# Simulation study: setup

- For all simulation scenarios, $(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \gamma_i)$ are generated as follows,
  1. each element in $\boldsymbol{x}_i$, $\{\boldsymbol{x}_i\}_{j,k} \overset{iid}{\sim} \text{uniform}(-1, 1)$;
  2. let $\boldsymbol{z}_i = (1, z_{i1})^T$ with $z_{i1} \sim \text{uniform}(0, 1)$;
  3. let $\boldsymbol{\theta} = (0, 1)^T$, generate $\gamma_i$ given $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, with $\boldsymbol{\Theta}$ generated from: When rank $R > 1$,
     - 3.1 let $\boldsymbol{\mu}_\alpha = \boldsymbol{\mu}_\beta = (0, \cdots, 0)^T$ and $\boldsymbol{\Sigma}_\alpha = \boldsymbol{\Sigma}_\beta$ be diagonal with all diagonal elements equal to $0.5^2$; generate $\tilde{\boldsymbol{\alpha}}_j \overset{iid}{\sim} N(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha)$, $j = 1, \cdots, p$ and $\tilde{\boldsymbol{\beta}}_k \overset{iid}{\sim} N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, $k = 1, \cdots, q$;
     - 3.2 set $\boldsymbol{A} = [\tilde{\boldsymbol{\alpha}}_1, \cdots, \tilde{\boldsymbol{\alpha}}_p]^T$ and $\boldsymbol{B} = [\tilde{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{\beta}}_q]^T$, then $\boldsymbol{\Theta} = \boldsymbol{A}\boldsymbol{B}^T$.

     When rank $R = 1$,
     - 3.1 let $\mu_\alpha = \mu_\beta = 0$ and $\sigma_\alpha^2 = \sigma_\beta^2 = 0.5^2$; generate $\tilde{\alpha}_j \overset{iid}{\sim} N(\mu_\alpha, \sigma_\alpha^2)$, $j = 1, \cdots, p$ and $\tilde{\beta}_k \overset{iid}{\sim} N(\mu_\beta, \sigma_\beta^2)$, $k = 1, \cdots, q$;
     - 3.2 set $\boldsymbol{\alpha}_1 = (\tilde{\alpha}_1, \cdots, \tilde{\alpha}_p)^T$ and $\boldsymbol{\beta}_1 = (\tilde{\beta}_1, \cdots, \tilde{\beta}_q)^T$, then $\boldsymbol{\Theta} = \boldsymbol{\alpha}_1 \boldsymbol{\beta}_1^T$.
  4. We generate $y_i$ given $\gamma_i$, where we fix $\eta_0 = 0$ and $\sigma = 0.2$, but we vary the value of $\eta_1$ in different scenarios.