1 The Model

1.1 Data Model

For years $t = 1, \dots, T$,

$$\left[\boldsymbol{y}_{ot} | \boldsymbol{\beta}_t, \sigma^2\right] = \frac{1}{\tau} \sum_{\boldsymbol{\gamma}_t \in \boldsymbol{\Gamma}} \left[\boldsymbol{y}_{ct} \middle| \boldsymbol{\beta}_{\boldsymbol{\gamma} t}, \sigma^2, \boldsymbol{\gamma}_t\right] \left[\boldsymbol{\gamma}_t | \boldsymbol{y}_{ct}, \sigma^2\right]$$

where τ is the number of models under consideration.

1.2 Parameter Model

$$\begin{aligned}
[\beta_{0t}] &\propto 1 \\
[\beta_{jt} | \sigma^2, \lambda_j, \gamma_j] &\overset{iid}{\sim} \begin{cases} 0 & \text{if } \gamma_j = 0 \\ \mathrm{N}\left(0, \frac{\sigma^2}{\lambda_j}\right) & \text{if } \gamma_j = 1 \end{cases} && \text{for } j = 1, \dots, p \\
\left[\sigma^2\right] &\propto \frac{1}{\sigma^2} \\
[\gamma_{jt}] &\propto \mathrm{Bern}\left(\pi_j\right) && \text{for } j = 1, \dots, p
\end{aligned}$$

where π_j and λ_j are fixed hyperpriors for $j = 1, \dots, p$.

1.3 Posterior

For a given model for time t indexed by $\boldsymbol{\gamma}_t$, the posterior distribution is

2 Posteriors

2.1 Posterior for $\boldsymbol{\beta}_{\boldsymbol{\gamma}_t}$

For $t = 1, \dots, T$,

$$\begin{aligned}
\left[\boldsymbol{\beta}_{\boldsymbol{\gamma}_t} \middle| \sigma^2, \boldsymbol{\gamma}_t, \boldsymbol{y}_{ot}\right] &\propto \left[\boldsymbol{y}_{ot} \middle| \boldsymbol{\beta}_{\boldsymbol{\gamma} t}, \sigma^2, \boldsymbol{\gamma}_t\right] \left[\boldsymbol{\beta}_{\boldsymbol{\gamma} t} \middle| \sigma^2, \boldsymbol{\gamma}_t\right] \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}\right)^T \left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}\right)\right\} \exp\left\{-\frac{1}{2\sigma^2} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}^T \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{\beta}_{\boldsymbol{\gamma}_t}^T \left(\boldsymbol{X}_{o\boldsymbol{\gamma}_t}^T \boldsymbol{H}_t^T \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t}\right) \boldsymbol{\beta}_{\boldsymbol{\gamma}_t} - 2\boldsymbol{\beta}_{\boldsymbol{\gamma}_t}^T \boldsymbol{X}_{o\boldsymbol{\gamma}_t}^T \boldsymbol{H}_t^T \boldsymbol{y}_{ot}\right]\right\}
\end{aligned}$$

which is $\mathrm{MVN}\left(\tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}, \boldsymbol{V}_{\boldsymbol{\beta}_t}\right)$ where $\tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t} = \left(\boldsymbol{X}_{o\boldsymbol{\gamma}_t}^T \boldsymbol{H}_t^T \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t}\right)^{-1} \boldsymbol{X}_{o\boldsymbol{\gamma}_t}^T \boldsymbol{H}_t^T \boldsymbol{y}_{ot}$ and $\boldsymbol{V}_{\boldsymbol{\beta}_t} = \left(\boldsymbol{X}_{o\boldsymbol{\gamma}_t}^T \boldsymbol{H}_t^T \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t}\right)$

2.2 Posterior for σ^2

$$\left[\sigma^2 | \boldsymbol{\gamma}_t, \boldsymbol{y}_{ot}\right] = \frac{\left[\sigma^2, \boldsymbol{\beta}_{\boldsymbol{\gamma}_t} \middle| \boldsymbol{\gamma}_t, \boldsymbol{y}_{ot}\right]}{\left[\boldsymbol{\beta}_{\boldsymbol{\gamma}_t} \middle| \sigma^2, \boldsymbol{\gamma}_t, \boldsymbol{y}_{ot}\right]}$$

First consider the numerator of the above equation

$$
\begin{aligned}
\left[\sigma^2, \boldsymbol{\beta}_{\boldsymbol{\gamma}_t} \middle| \gamma_t, \boldsymbol{y}_{ot}\right] &\propto \left[\boldsymbol{y}_{ot} \middle| \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}, \sigma^2, \gamma_t\right] \left[\boldsymbol{\beta}_{\boldsymbol{\gamma}_t} \middle| \sigma^2, \gamma_t\right] \left[\sigma^2\right] \\
&\propto \left(\sigma^2\right)^{-\frac{n_o}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}\right)^T \left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}\right)\right\} \left(\sigma^2 \left|\boldsymbol{\Delta}^{+}_{\boldsymbol{\gamma}_t}\right|\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T_{\boldsymbol{\gamma}_t} \ldots\right. \\
&\propto \left(\sigma^2\right)^{-\frac{n_o-1}{2}-1} \exp\left\{-\frac{1}{\sigma^2} \frac{\left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}\right)^T \left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}\right) + \boldsymbol{\beta}^T_{\boldsymbol{\gamma}_t} \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t} \boldsymbol{\beta}_{\boldsymbol{\gamma}_t}}{2}\right\}
\end{aligned}
$$

Now we average over $\boldsymbol{\beta}_{\boldsymbol{\gamma}_t}$ by replacing $\boldsymbol{\beta}_{\boldsymbol{\gamma}_t}$ with its posterior mean $\tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t} = \left(\boldsymbol{X}^T_{o\boldsymbol{\gamma}_t} \boldsymbol{H}^T_t \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t}\right)^{-1} \boldsymbol{H}^T_t \boldsymbol{X}^T_{o\boldsymbol{\gamma}_t} \boldsymbol{y}_{ot}$ to get the posterior distribution

$$
\left[\sigma^2 \middle| \gamma_t, \boldsymbol{y}_{ot}\right] \propto \left(\sigma^2\right)^{-\frac{n_o-1}{2}-1} \exp\left\{-\frac{1}{\sigma^2} \frac{\left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}\right)^T \left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}\right) + \tilde{\boldsymbol{\beta}}^T_{\boldsymbol{\gamma}_t} \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}}{2}\right\}
$$

which is $\mathrm{IG}\left(\frac{n_o-1}{2}, \frac{\left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}\right)^T \left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}\right) + \tilde{\boldsymbol{\beta}}^T_{\boldsymbol{\gamma}_t} \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}}{2}\right)$. Now consider the quadratic term

$$
\left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}\right)^T \left(\boldsymbol{y}_{ot} - \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}\right) = \boldsymbol{y}^T_{ot} \boldsymbol{y}_{ot} - 2\tilde{\boldsymbol{\beta}}^T_{\boldsymbol{\gamma}_t} \boldsymbol{H}^T_t \boldsymbol{X}^T_{o\boldsymbol{\gamma}_t} \boldsymbol{y}_{ot} + \tilde{\boldsymbol{\beta}}^T_{\boldsymbol{\gamma}_t} \left(\boldsymbol{X}^T_{o\boldsymbol{\gamma}_t} \boldsymbol{H}^T_t \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t}\right) \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}
$$

2.3 Posterior for $\boldsymbol{y}_{at}$

This posterior is calculated in the same fashion as a posterior predictive distribution of new observations $\tilde{\boldsymbol{y}}$ given new covariates $\tilde{\boldsymbol{X}}$ for simple linear regression as shown $\left[\tilde{\boldsymbol{y}} \middle| \sigma^2, \boldsymbol{y}, \tilde{\boldsymbol{X}}\right] \sim \mathrm{N}\left(\tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}, \sigma^2\left(\boldsymbol{I} + \tilde{\boldsymbol{X}} \boldsymbol{V}_{\boldsymbol{\beta}} \tilde{\boldsymbol{X}}^T\right)\right)$ which is

$$
\left[\boldsymbol{y}_{at} \middle| \sigma^2, \boldsymbol{\gamma}_t, \boldsymbol{y}_{ot}\right] \sim \mathrm{N}\left(\boldsymbol{X}_{a\boldsymbol{\gamma}_t} \tilde{\boldsymbol{\beta}}_{\boldsymbol{\gamma}_t}, \sigma^2\left(\boldsymbol{I}_{n_a} + \boldsymbol{X}_{a\boldsymbol{\gamma}_t}\left(\boldsymbol{X}^T_{o\boldsymbol{\gamma}_t} \boldsymbol{H}^T_t \boldsymbol{H}_t \boldsymbol{X}_{o\boldsymbol{\gamma}_t} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}_t}\right)^{-1} \boldsymbol{X}^T_{a\boldsymbol{\gamma}_t}\right)\right)
$$

2.4 Posterior for $\gamma_j, j = 1, \ldots, p$

3 Full Conditionals

3.1 Full Conditional for σ^2

$$\begin{aligned}
[\sigma^2|\cdot] &\propto \int \left[\boldsymbol{y}_o, \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{X}_o\right] \left[\boldsymbol{y}_a \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{X}_a\right] \left[\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2 \middle|\boldsymbol{\gamma}\right] \, d\boldsymbol{y}_a \\
&\propto \int \left[\boldsymbol{y}_a \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{X}_a\right] \, d\boldsymbol{y}_a \left[\boldsymbol{y}_o, \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{X}_o\right] \left[\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2 \middle|\boldsymbol{\gamma}\right] \\
&\propto \left[\boldsymbol{y}_o, \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{X}_o\right] \left[\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2 \middle|\boldsymbol{\gamma}\right] \\
&\propto \left(\sigma^2\right)^{-\frac{n_o}{2}} \exp\left\{-\frac{1}{\sigma^2} \frac{\left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)^T \left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)}{2}\right\} \left(\sigma^2\right)^{-\frac{p_{\boldsymbol{\gamma}}}{2}} \exp\left\{-\frac{1}{\sigma^2}\frac{\boldsymbol{\beta}_{\boldsymbol{\gamma}}^T \boldsymbol{\Lambda}_{\boldsymbol{\gamma}} \boldsymbol{\beta}_{\boldsymbol{\gamma}}}{2}\right\} \left(\sigma^2\right)^{-1} \\
&\propto \left(\sigma^2\right)^{-\frac{n_o + p_{\boldsymbol{\gamma}}}{2} - 1} \exp\left\{-\frac{1}{\sigma^2} \frac{\left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)^T \left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right) + \boldsymbol{\beta}_{\boldsymbol{\gamma}}^T \boldsymbol{\Lambda}_{\boldsymbol{\gamma}} \boldsymbol{\beta}_{\boldsymbol{\gamma}}}{2}\right\}
\end{aligned}$$

which is IG $\left(\frac{n_o + p_{\boldsymbol{\gamma}}}{2}, \frac{\left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)^T \left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right) + \boldsymbol{\beta}_{\boldsymbol{\gamma}}^T \boldsymbol{\Lambda}_{\boldsymbol{\gamma}} \boldsymbol{\beta}_{\boldsymbol{\gamma}}}{2}\right)$

3.2 Full Conditional for $\boldsymbol{y}_a$

3.3 Full Conditional for $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$

$$\begin{aligned}
\left[\boldsymbol{\beta}_{\boldsymbol{\gamma}} \middle| \cdot\right] &\propto \int \left[\boldsymbol{y}_o \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}\right] \left[\boldsymbol{y}_a \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}\right] \left[\boldsymbol{\beta}_{\boldsymbol{\gamma}} \middle|\sigma^2, \boldsymbol{\gamma}\right] \, d\boldsymbol{y}_a \\
&\propto \left[\boldsymbol{y}_o \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}\right] \left[\boldsymbol{\beta}_{\boldsymbol{\gamma}} \middle|\sigma^2, \boldsymbol{\gamma}\right] \int \left[\boldsymbol{y}_a \middle|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma}\right] \, d\boldsymbol{y}_a \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)^T \left(\boldsymbol{y}_o - \boldsymbol{X}_{o\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)\right\} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_{\boldsymbol{\gamma}}^T \boldsymbol{\Delta}_{\boldsymbol{\gamma}} \boldsymbol{\beta}_{\boldsymbol{\gamma}}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{\beta}_{\boldsymbol{\gamma}}^T \left(\boldsymbol{X}_{o\boldsymbol{\gamma}}^T \boldsymbol{X}_{o\boldsymbol{\gamma}} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}}\right) \boldsymbol{\beta}_{\boldsymbol{\gamma}} - 2\boldsymbol{\beta}_{\boldsymbol{\gamma}}^T \left(\boldsymbol{X}_{o\boldsymbol{\gamma}}^T \boldsymbol{y}_o\right)\right]\right\}
\end{aligned}$$

which is MVN $\left(\boldsymbol{A}^{-1}\boldsymbol{b}, \boldsymbol{A}^{-1}\right)$ where $\boldsymbol{A}^{-1} = \left(\boldsymbol{X}_{o\boldsymbol{\gamma}}^T \boldsymbol{X}_{o\boldsymbol{\gamma}} + \boldsymbol{\Delta}_{\boldsymbol{\gamma}}\right)^{-1}$ and $\boldsymbol{b} = \boldsymbol{X}_{o\boldsymbol{\gamma}}^T \boldsymbol{y}_o$

3.4 Full Conditional for γ_j

For $j = 1, \ldots, p$ and using the fact that $\beta_j = \left(\boldsymbol{X}_{cj}^T \boldsymbol{X}_{cj}\right)^{-1} \boldsymbol{X}_{cj}^T \boldsymbol{y}_c$ and $\boldsymbol{X}_{cj}^T \boldsymbol{X}_{cj} = \delta_j$,

$$\begin{aligned}
[\gamma_j|\cdot] &\propto \left[\boldsymbol{y}_c, |\beta_{\gamma_j}, \sigma^2, \gamma_j, \boldsymbol{X}_o, \boldsymbol{X}_a, \boldsymbol{y}_a\right] \left[\beta_{\gamma_j}, \sigma^2 | \gamma_j\right] [\gamma_j] \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{y}_c - \boldsymbol{X}_{cj}\gamma_j\beta_j\right)^T \left(\boldsymbol{y}_c - \boldsymbol{X}_{cj}\gamma_j\beta_j\right)\right\} \left(\frac{\lambda_j}{\sigma^2}\right)^{\frac{\gamma_j}{2}} \exp\left\{-\frac{\gamma_j \lambda_j \beta_j^2}{2\sigma^2}\right\} \pi^{\gamma_j} \left(1 - \pi\right)^{1-\gamma_j} \\
&\propto \left[\left(\frac{\lambda_j}{\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\beta_j^2 \left(\boldsymbol{X}_{cj}^T \boldsymbol{X}_{cj} + \lambda_j\right) - 2\beta_j \boldsymbol{X}_{cj}^T \boldsymbol{y}_c\right]\right\} \frac{\pi}{1-\pi}\right]^{\gamma_j} \\
&\propto \left[\left(\frac{\lambda_j}{\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\beta_j^2 \left(\delta_j + \lambda_j\right) - 2\delta_j \beta_j^2\right]\right\} \frac{\pi}{1-\pi}\right]^{\gamma_j} \\
&\propto \Psi^{\gamma_j}
\end{aligned}$$

which is Bern$\left(\frac{\Psi}{1+\Psi}\right)$ where $\Psi = \left(\frac{\lambda_j}{\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\beta_j^2\left(\delta_j+\lambda_j\right)-2\delta_j\beta_j^2\right]\right\}\frac{\pi}{1-\pi}$

4 Data Augmentation

To perform the model selection and averaging, the "complete" design matrix

$$\boldsymbol{X}_c = \left[\begin{array}{c} \boldsymbol{X}_o \\ \boldsymbol{X}_a \end{array}\right]$$

which has orthogonal columns, hence $\boldsymbol{X}_c^T\boldsymbol{X}_c = \boldsymbol{I}$. The matrix $\boldsymbol{X}_a$ is chosen to be the Cholesky decomposition of $\boldsymbol{D} - \boldsymbol{X}_o^T\boldsymbol{X}_o$ where $\boldsymbol{D}$ is a diagonal matrix with $\delta+\varepsilon$ on the diagonal where δ is the largest eigenvalue of $\boldsymbol{X}_o^T\boldsymbol{X}_o$ and $\varepsilon = 0.001$ is added to avoid computationally unstable solutions.