We now describe the update of the m^{th} regression tree $(T_m^{(j)}, \boldsymbol{\mu}_m^{(j)})$ in the ensemble $\mathcal{E}_j$ used to approximate β_j. Let $\boldsymbol{\mathcal{E}}^-$ denote the collection of all remaining $M(p+1)-1$ regression trees, where, for notational compactness, we have suppressed the dependence of this collection on the indices j and m. Additionally, let $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{in_i})^\top$ be the vector of subject i's *partial residuals* where

$$r_{it} = R_{it} + x_{itj} g(\boldsymbol{z}_{it}; T_m^{(j)}, \boldsymbol{\mu}_m^{(j)}),$$

where we have again suppressed the dependence of r_{it} on j and m for brevity.

Before describing the update of the tree, we require some additional notation. For an arbitrary decision tree T with $L(T)$ leaves, for each $i = 1, \ldots, n$, let $I_i(\ell; T)$ be the set of indices t such that $\boldsymbol{z}_{it}$ is contained in leaf ℓ of tree t. In other words, $I_i(\ell; T)$ records which observations from subject i are mapped to leaf ℓ in tree T. Further, let $\boldsymbol{X}_i(T)$ be the $n_i \times L(T)$ matrix whose (t, ℓ) entry is equal to x_{itj} if $t \in I_i(\ell, T)$ and is zero otherwise. With this additional notation, we have $\boldsymbol{r}_i = \boldsymbol{R}_i + \boldsymbol{X}_i(T_m^{(j)})\boldsymbol{\mu}_m^{(j)}$ for each $i = 1, \ldots, n$.

Marginal likelihood of $T_m^{(-j)}$, conditional on $\boldsymbol{\mathcal{E}}^-$

The conditional conjugacy allows us to marginalize out $\boldsymbol{\mu}$ and compute $p(\boldsymbol{Y} | T, \boldsymbol{\mathcal{E}}^-, \sigma^2)$. To this end, observe that

$$\begin{aligned} p(\boldsymbol{Y} | T, \boldsymbol{\mu}, \boldsymbol{\mathcal{E}}^-, \sigma^2) &= \prod_{i=1}^{n} p(\boldsymbol{y}_i | T, \boldsymbol{\mu}, \boldsymbol{\mathcal{E}}^-, \sigma^2) \\ &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{n_i}{2}} |\Omega_i|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} (\boldsymbol{r}_i - \boldsymbol{X}_i(T)\boldsymbol{\mu})^\top \Omega_i (\boldsymbol{r}_i - \boldsymbol{X}_i(T)\boldsymbol{\mu}) \right\} \end{aligned}$$

We additionally have

$$p(\boldsymbol{\mu} | T) = (2\pi\tau^2)^{-\frac{L(T)}{2}} \exp\left\{ -\frac{(\boldsymbol{\mu} - \mu_0 \mathbf{1}_{L(T)})^\top (\boldsymbol{\mu} - \mu_0 \mathbf{1}_{L(T)})}{2\tau^2} \right\}$$

We therefore compute

$$p(\boldsymbol{Y}, \boldsymbol{\mu}|T, \boldsymbol{\mathcal{E}}^{-}, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \times \left(\prod_{i=1}^{N} |\Omega_i|\right)^{\frac{1}{2}} \times \left(2\pi\tau^2\right)^{-\frac{L(T)}{2}} \times \exp\left\{-\frac{\tau^{-2}(\boldsymbol{\mu} - \mu_0 \mathbf{1}_{L(T)})^\top (\boldsymbol{\mu} - \mu_0 \mathbf{1}_{L(T)})}{2\tau^2}\right\}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\boldsymbol{r}_i - \boldsymbol{X}_i(T)\boldsymbol{\mu})^\top \Omega_i (\boldsymbol{r}_i - \boldsymbol{X}_i(T))\right\}$$

$$\begin{aligned} &= \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \times \left(\prod_{i=1}^{n} |\Omega_i|\right)^{\frac{1}{2}} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \boldsymbol{r}_i^\top \Omega_i \boldsymbol{r}_i\right\} \\ &\times \left(2\pi\tau^2\right)^{-\frac{L(T)}{2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}^\top P(T) \boldsymbol{\mu} - 2\boldsymbol{\mu}\Theta(T) + \tau^{-2}\mu_0^2 L(T)\right]\right\} \end{aligned} \tag{1}$$

where

$$P(T) = \tau^{-2} I_{L(T)} + \sigma^{-2} \sum_{i=1}^{n} \boldsymbol{X}_i(T)^\top \Omega_i \boldsymbol{X}_i(T)$$
$$\Theta(T) = \tau^{-2} \mu_0 \mathbf{1}_{L(T)} + \sigma^{-2} \sum_{i=1}^{n} \boldsymbol{X}_i(T)^\top \Omega_i \boldsymbol{r}_i.$$

Observe that the terms in the first line of Equation (1) do not depend on T or μ. We will collect these terms into a constant C, where we suppress the dependence on $\sigma^2, \boldsymbol{\Omega}$ and the $\boldsymbol{r}_i$'s from the notation.

Integrating out $\boldsymbol{\mu}$ from Equation (1), we obtain

$$p(\boldsymbol{Y}|T, \boldsymbol{\mathcal{E}}^{-1}, \sigma^2) = C \times \tau^{-L(T)} \times \exp\left\{\frac{\Theta(T)^\top P(T)^{-1} \Theta(T) + L\mu_0^2 \tau^{-2}}{2}\right\}$$

Computing $P(T)$ in the general case. Recall that

$$P(T) = \tau^{-2} I_{L(T)} + \sigma^{-2} \sum_{i=1}^{n} \boldsymbol{X}_i(T)^\top \Omega_i \boldsymbol{X}_i(T).$$

Observe that the (ℓ, ℓ') entry of $\boldsymbol{X}_i^\top \Omega_i \boldsymbol{X}_i(T)$ is given by

$$\begin{aligned}
(\boldsymbol{X}_i^\top \Omega_i \boldsymbol{X}_i(T))_{\ell,\ell'} &= \sum_{t=1}^{n_i} \sum_{t'=1}^{n_i} \omega_{t,t'} (\boldsymbol{X}_i(T))_{\ell,t} (\boldsymbol{X}_i(T))_{\ell',t'} \\
&= \sum_{t=1}^{n_i} \sum_{t'=1}^{n_i} \omega_{itt'} x_{itj} x_{it'j} \mathbb{1}(t \in I_i(\ell;T), t' \in I_i(\ell';T)) \\
&= \sum_{t \in I_i(\ell;T)} \sum_{t' \in I_i(\ell';T)} \omega_{itt'} x_{itj} x_{it'j}
\end{aligned}$$

The independent error case. Suppose first that $\Omega_i = I_{n_i}$ for each $i = 1, \ldots, n$. For $\ell \neq \ell'$, since $\omega_{itt'} = \mathbb{1}(t = t')$ we have

$$\begin{aligned}
(\boldsymbol{X}_i(T)^\top \Omega_i \boldsymbol{X}_i(T))_{\ell,\ell'} &= \sum_{t=1}^{n_i} \sum_{t'=1}^{n_i} \omega_{itt'} x_{itj} x_{it'j} \mathbb{1}(t \in I_i(\ell;T), t' \in I_i(\ell';T)) \\
&= \sum_{t=1}^{n_i} x_{itj}^2 \mathbb{1}(t \in I_i(\ell;T), t \in I_i(\ell';T))
\end{aligned}$$

Since the t^{th} observation for subject i can be associated to only one leaf in the tree T, we conclude that $P(T)$ is diagonal with diagonal entries

$$(P(T))_{\ell,\ell} = \tau^{-2} I_{L(T)} + \sigma^{-2} \sum_{i=1}^{n} \sum_{t \in I_i(\ell;T)} x_{itj}^2.$$

Recall further that

$$\Theta(T) = \tau^{-2} \mu_0 \mathbf{1}_{L(T)} + \sigma^{-2} \sum_{i=1}^{n} \boldsymbol{X}_i^\top \Omega_i \boldsymbol{r}_i$$

When each $\Omega_i = I_{n_i}$ we have for each $i = 1, \ldots, n$,

$$\begin{aligned}
(\boldsymbol{X}_i(T)^\top \Omega_i \boldsymbol{r}_i)_\ell &= \sum_{t=1}^{n_i} x_{itj} r_{it} \mathbb{1}(t \in I_i(\ell)) \\
&= \sum_{t \in I_i(\ell;T)} x_{itj} r_{it}
\end{aligned}$$

Hence

$$(\Theta(t))_\ell = \tau^{-2} \mu_0 + \sigma^{-2} \sum_{i=1}^{n} \sum_{t \in I_i(\ell;T)} x_{itj} r_{it}$$

The compound symmetry case. Suppose that for each $i = 1, \ldots, n$

$$\Omega_i^{-1} = (1-\rho)I_{n_i} + \rho \mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top.$$

It is not difficult to verify that in fact

$$\Omega_i = \frac{1}{1-\rho} I_{n_i} - \frac{\rho}{(1-\rho)^2 + n_i\rho(1-\rho)} \mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top.$$

We therefore have

$$\boldsymbol{X}_i(T)^\top \Omega_i \boldsymbol{X}_i(T) = \frac{1}{1-\rho}\boldsymbol{X}_i(T)^\top \boldsymbol{X}_i(T) - \frac{\rho}{(1-\rho)^2 + n_i\rho(1-\rho)} \boldsymbol{X}_i(T)^\top \mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top \boldsymbol{X}_i(T)$$

From our computation above, we know that

$$\begin{aligned}(\boldsymbol{X}_i(T)^\top \boldsymbol{X}_i(T))_{\ell,\ell'} &= \sum_{t=1}^{n_i} x_{itj}^2 \mathbb{1}(t \in I_i(\ell;T), t \in I_i(\ell';T)) \\ &= \mathbb{1}(\ell = \ell') \times \sum_{t \in I_i(\ell)} x_{ijt}^2\end{aligned}$$

We further have

$$(\boldsymbol{X}_i(T)^\top \mathbf{1}_{L(T)})_\ell = \sum_{t \in I_i(\ell;T)} x_{itj}$$

Hence

$$(\boldsymbol{X}_i(T)^\top \mathbf{1}_{L(T)}\mathbf{1}_{L(T)}^\top \boldsymbol{X}_i(T))_{\ell,\ell'} = \left(\sum_{t \in I_i(\ell;T)} x_{itj}\right) \times \left(\sum_{t' \in I_i(\ell';T)} x_{it'j}\right)$$

We conclude that

$$\begin{aligned}(P(T))_{\ell,\ell'} = \mathbb{1}(\ell = \ell') \times &\left[\tau^{-2} + \sigma^{-2}\sum_{i=1}^n \sum_{t \in I_i(\ell)} x_{itj}^2\right] \\ &- \frac{\rho}{1-\rho} \times \sum_{i=1}^n \frac{1}{1-\rho+n_i\rho} \times \left(\sum_{t \in I_i(\ell;T)} x_{itj}\right)\left(\sum_{t' \in I_i(\ell';T)} x_{it'j}\right)\end{aligned}$$

Turning our attention to $\Theta(T)$, have

$$\boldsymbol{X}_i(T)^\top \Omega_i \boldsymbol{r}_i = \frac{1}{1-\rho}\boldsymbol{X}_i(T)^\top \boldsymbol{r}_i - \frac{\rho}{1-\rho} \times \frac{\rho}{1-\rho+n_i\rho} \times \boldsymbol{X}_i^\top \mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top \boldsymbol{r}_i$$

Notice that

$$(\boldsymbol{X}_i(T)^\top \boldsymbol{r}_i)_\ell = \sum_{t=1}^{n_i} x_{itj} r_{it} \times \mathbb{1}(t \in I_i(\ell; T)) = \sum_{t \in I_i(\ell;T)} x_{itj} r_{it}$$

and

$$(\boldsymbol{X}_i^\top \mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top \boldsymbol{r}_i)_\ell = \left(\mathbf{1}_{n_i}^\top \boldsymbol{r}_i\right) \times \sum_{t \in I_i(\ell;T)} x_{itj}$$

We conclude that

$$\Theta(T)_\ell = \tau^{-2}\mu_0 + \frac{\sigma^{-2}}{1-\rho} \times \sum_{i=1}^{n} \sum_{t \in I_i(\ell;T)} \left[x_{itj} r_{it} - \frac{\mathbf{1}_{n_i}^\top \boldsymbol{r}_i}{1-\rho+n_i\rho} x_{itj} \right]$$