Supplementary Material File for Semiparametric Bayesian Regression Via Potts Model

Alejandro Murua and Fernando A. Quintana*

*Département de mathématiques et de statistique,
Université de Montréal,
e-mail: murua@dms.umontreal.ca*

*Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile.
e-mail: quintana@mat.uc.cl*

This supplementary material presents further details on the Potts regression model. Section A includes the proofs of Lemmas 1, 2, and 3 of Section 3.1, and the proof of Proposition 1 of Section 3.2. Section B provides more information on the simulation scenarios described in Section 4. Section C gives details on the comparison between Potts regression and PPMx under the third scenario. Finally, in Section D, we complete the results of the analysis of Health Outcomes dataset of Section 5. In particular, Section D displays the list of country clusters found by our model.

A. Proofs of Technical Results

This section shows the proofs of the two lemmas and the proposition stated in Section 3 of the main document. In what follows, we simplify the notation by just denoting θ_j^* , Σ_j^* , $\underline{Y}_j$ and $\underline{X}_j$ as θ , Σ , Y and X , respectively.

*Corresponding author: quintana@mat.uc.cl

Proof of Lemma 1

Note that by properties of the Kronecker product

$$\begin{aligned}
(I_p \otimes \mathbf{x}'_i) \Sigma^{-1} (I_p \otimes \mathbf{x}'_i) &= ((I_p \otimes \mathbf{x}_i) \Sigma^{-1}) (I_p \otimes \mathbf{x}'_i) \\
&= (\Sigma^{-1} \otimes \mathbf{x}_i) (I_p \otimes \mathbf{x}'_i) \\
&= \Sigma^{-1} \otimes (\mathbf{x}_i \mathbf{x}'_i) \\
(I_p \otimes \mathbf{x}'_i) \Sigma^{-1} \mathbf{y}_i &= (\Sigma^{-1} \mathbf{y}_i) \otimes \mathbf{x}_i \\
&= \text{vec} ((\mathbf{x}_i \otimes \mathbf{y}'_i) \Sigma^{-1}) \\
&= \text{vec} (\mathbf{x}_i \mathbf{y}'_i \Sigma^{-1}).
\end{aligned}$$

Consequently

$$\begin{aligned}
\sum_{i \in S_j} (I_p \otimes \mathbf{x}'_i) \Sigma^{-1} (I_p \otimes \mathbf{x}'_i) &= \Sigma^{-1} \otimes \mathbf{X}' \mathbf{X}, \\
\sum_{i \in S_j} (I_p \otimes \mathbf{x}'_i) \Sigma^{-1} \mathbf{y}_i &= \text{vec} (\mathbf{X}' \mathbf{Y} \Sigma^{-1}).
\end{aligned}$$

Using these identities and the fact that $(\Sigma^{-1} \otimes \mathbf{X}' \mathbf{X})^{-1} = \Sigma \otimes (\mathbf{X}' \mathbf{X})^{-1}$, we obtain the least-squares estimator of $\boldsymbol{\theta}$,

$$\hat{\boldsymbol{\theta}} = \Sigma \otimes (\mathbf{X}' \mathbf{X})^{-1} \text{vec} (\mathbf{X}' \mathbf{Y} \Sigma^{-1}) = \text{vec} ((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}),$$

where we have also used the fact that $(\mathbf{C}' \otimes \mathbf{A}) \text{vec} (\mathbf{B}) = \text{vec} (\mathbf{A} \mathbf{B} \mathbf{C})$. This concludes the proof of the first part of the Lemma.

Next, let $\boldsymbol{\theta}_{mtx} = (\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 | \cdots | \boldsymbol{\theta}_p) \in \mathbb{M}_{q \times p}$, $\hat{\boldsymbol{\theta}}_{mtx} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$, and $\mathbf{W} = (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}}_{mtx})' (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}}_{mtx}) = \mathbf{Y}' (\mathbf{I}_{n_j} - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}') \mathbf{Y}$. It is straightforward to verify that

$$\sum_{i \in S_j} (\mathbf{y}_i - I_p \otimes \mathbf{x}'_i \boldsymbol{\theta}) (\mathbf{y}_i - I_p \otimes \mathbf{x}'_i \boldsymbol{\theta})' = (\mathbf{Y} - \mathbf{X} \boldsymbol{\theta}_{mtx})' (\mathbf{Y} - \mathbf{X} \boldsymbol{\theta}_{mtx}).$$

Noting that $\mathbf{X}' (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}}_{mtx}) = 0$, one can write

$$\begin{aligned}
(\mathbf{Y} - \mathbf{X} \boldsymbol{\theta}_{mtx})' (\mathbf{Y} - \mathbf{X} \boldsymbol{\theta}_{mtx}) &= \\
&= (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}}_{mtx} - \mathbf{X} (\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx}))' (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}}_{mtx} - \mathbf{X} (\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx})) \\
&= \mathbf{W} + (\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx})' \mathbf{X}' \mathbf{X} (\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx}).
\end{aligned}$$

Therefore

$$\begin{aligned}
& \text{trace} \left[\Sigma^{-1} \sum_{i \in S_j} (\mathbf{y}_i - \mathbf{I}_p \otimes \mathbf{x}_i' \boldsymbol{\theta})(\mathbf{y}_i - \mathbf{I}_p \otimes \mathbf{x}_i' \boldsymbol{\theta})' \right] \\
&= \text{trace} [\mathbf{W} \Sigma^{-1}] + \text{trace} \left[(\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx})' \mathbf{X}' \mathbf{X} (\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx}) \Sigma^{-1} \right] \\
&= \text{trace} [\mathbf{W} \Sigma^{-1}] + \text{vec} (\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx})' \text{vec} (\mathbf{X}' \mathbf{X} (\boldsymbol{\theta}_{mtx} - \hat{\boldsymbol{\theta}}_{mtx}) \Sigma^{-1}) \\
&= \text{trace} [\mathbf{W} \Sigma^{-1}] + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' (\Sigma^{-1} \otimes \mathbf{X}' \mathbf{X}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \quad (\text{A.1})
\end{aligned}$$

where we have used again that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B})$. The likelihood form given by (11) follows as a result of (A.1). $\blacksquare$

Proof of Lemma 2

The joint posterior of $(\boldsymbol{\theta}, \Sigma)$ is proportional to

$$\begin{aligned}
& L(\boldsymbol{\theta}, \Sigma \mid \mathbf{y}_j^*, \mathbf{x}_j^*) \times |\Sigma|^{-(\nu+p+1)/2} \exp \left\{ -\frac{1}{2} \text{trace} [\Lambda \Sigma^{-1}] \right\} \\
& \times |\Sigma|^{-q/2} |\boldsymbol{\tau}|^{p/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' [\Sigma^{-1} \otimes \boldsymbol{\tau}] (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\}. \quad (\text{A.2})
\end{aligned}$$

(a) The conditional posterior of $\boldsymbol{\theta}$ given Σ . Using (A.2) and (11), this conditional posterior is proportional to

$$\exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' (\Sigma^{-1} \otimes \mathbf{X}' \mathbf{X}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right\} \times \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' [\Sigma^{-1} \otimes \boldsymbol{\tau}] (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\}.$$

Therefore it is easy to see that the conditional posterior is Normal with variance-covariance matrix

$$\boldsymbol{\tau}^{(j)} = (\Sigma^{-1} \otimes \mathbf{X}' \mathbf{X} + \Sigma^{-1} \otimes \boldsymbol{\tau})^{-1} = (\Sigma^{-1} \otimes (\mathbf{X}' \mathbf{X} + \boldsymbol{\tau}))^{-1} = \Sigma \otimes (\mathbf{X}' \mathbf{X} + \boldsymbol{\tau})^{-1},$$

and mean

$$\boldsymbol{\mu}^{(j)} = \boldsymbol{\tau}^{(j)} (\Sigma^{-1} \otimes \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\theta}} + \Sigma^{-1} \otimes \boldsymbol{\tau} \boldsymbol{\mu}).$$

Now use the Kronecker product properties

$$\begin{aligned}
(\Sigma \otimes (\mathbf{X}'\mathbf{X} + \boldsymbol{\tau})^{-1})(\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X}) &= \Sigma \Sigma^{-1} \otimes (\mathbf{X}'\mathbf{X} + \boldsymbol{\tau})^{-1} \mathbf{X}'\mathbf{X} \\
&= \mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X} + \boldsymbol{\tau})^{-1} \mathbf{X}'\mathbf{X} \\
(\Sigma \otimes (\mathbf{X}'\mathbf{X} + \boldsymbol{\tau})^{-1})(\Sigma^{-1} \otimes \boldsymbol{\tau}) &= \Sigma \Sigma^{-1} \otimes (\mathbf{X}'\mathbf{X} + \boldsymbol{\tau})^{-1} \boldsymbol{\tau} \\
&= \mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X} + \boldsymbol{\tau})^{-1} \boldsymbol{\tau},
\end{aligned}$$

to obtain the first identity in (12). The second identity follows by using the property $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ for matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$.

(b) The posterior of Σ . Using (11), the posterior of Σ is proportional to

$$\begin{aligned}
&|\Sigma|^{-(n_j + \nu + p + 1 + q)/2} \exp \left\{ -\frac{1}{2} \text{trace} [(\mathbf{W} + \boldsymbol{\Lambda})\Sigma^{-1}] \right\} \\
&\times \int_{\boldsymbol{\Theta}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}^{(j)})' [\boldsymbol{\tau}^{(j)}]^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}^{(j)}) \right\} (2\pi)^{-pq/2} |\boldsymbol{\tau}^{(j)}|^{-1/2} d\boldsymbol{\theta} \\
&\times |\boldsymbol{\tau}^{(j)}|^{1/2} \exp \left\{ -\frac{1}{2} \hat{\boldsymbol{\theta}}' (\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X}) \hat{\boldsymbol{\theta}} - \frac{1}{2} \boldsymbol{\mu}' (\Sigma^{-1} \otimes \boldsymbol{\tau}) \boldsymbol{\mu} + \frac{1}{2} \boldsymbol{\mu}^{(j)'} [\boldsymbol{\tau}^{(j)}]^{-1} \boldsymbol{\mu}^{(j)} \right\}
\end{aligned}$$

Using the fact that for two matrices $\mathbf{A} \in \mathbb{M}_{p \times p}$ and $\mathbf{B} \in \mathbb{M}_{q \times q}$, $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^q |\mathbf{B}|^p$, one obtains $|\boldsymbol{\tau}^{(j)}|^{1/2} \propto |\Sigma|^{q/2}$. This yields

$$\begin{aligned}
P(\Sigma \mid \mathbf{y}_j^*, \mathbf{x}_j^*) &\propto |\Sigma|^{-(n_j + \nu + p + 1)/2} \exp \left\{ -\frac{1}{2} \text{trace} [(\mathbf{W} + \boldsymbol{\Lambda})\Sigma^{-1}] \right\} \\
&\times \exp \left\{ -\frac{1}{2} (\hat{\boldsymbol{\theta}}' (\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X}) \hat{\boldsymbol{\theta}} + \boldsymbol{\mu}' (\Sigma^{-1} \otimes \boldsymbol{\tau}) \boldsymbol{\mu} - \boldsymbol{\mu}^{(j)'} [\boldsymbol{\tau}^{(j)}]^{-1} \boldsymbol{\mu}^{(j)}) \right\}
\end{aligned}$$

Next, we write the last term as a trace. For this we use the identities

$$\begin{aligned}
\hat{\boldsymbol{\theta}}' (\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X}) \hat{\boldsymbol{\theta}} &= \text{vec} (\hat{\boldsymbol{\theta}}_{mtx})' \text{vec} (\mathbf{X}'\mathbf{X} \hat{\boldsymbol{\theta}}_{mtx} \Sigma^{-1}) \\
&= \text{trace} [\hat{\boldsymbol{\theta}}_{mtx}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\theta}}_{mtx} \Sigma^{-1}] \\
&= \text{trace} [\mathbf{Y}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \Sigma^{-1}] \\
\boldsymbol{\mu}' (\Sigma^{-1} \otimes \boldsymbol{\tau}) \boldsymbol{\mu} &= \text{vec} (\boldsymbol{\mu}_{mtx})' \text{vec} (\boldsymbol{\tau} \boldsymbol{\mu}_{mtx} \Sigma^{-1}) \\
&= \text{trace} [\boldsymbol{\mu}_{mtx}' \boldsymbol{\tau} \boldsymbol{\mu}_{mtx} \Sigma^{-1}] \\
\boldsymbol{\mu}^{(j)'} [\boldsymbol{\tau}^{(j)}]^{-1} \boldsymbol{\mu}^{(j)} &= \text{vec} (\boldsymbol{\mu}_{mtx}^{(j)})' \text{vec} ((\mathbf{X}'\mathbf{X} + \boldsymbol{\tau}) \boldsymbol{\mu}_{mtx}^{(j)} \Sigma^{-1}) \\
&= \text{trace} [\boldsymbol{\mu}_{mtx}^{(j)'} (\mathbf{X}'\mathbf{X} + \boldsymbol{\tau}) \boldsymbol{\mu}_{mtx}^{(j)} \Sigma^{-1}].
\end{aligned}$$

Therefore

$$\begin{aligned} \hat{\theta}'(\Sigma^{-1} \otimes X'X)\hat{\theta} + \mu'(\Sigma^{-1} \otimes \tau)\mu - \mu^{(j)'}[\tau^{(j)}]^{-1}\mu^{(j)} = \\ \text{trace} [(Y'X(X'X)^{-1}X'Y + \mu'_{mtx}\tau\mu_{mtx} - \mu_{mtx}^{(j)'}(X'X + \tau)\mu_{mtx}^{(j)})\Sigma^{-1}]. \end{aligned}$$

Note that

$$Y'X(X'X)^{-1}X'Y = Y'Y - Y'(I_{n_j} - X(X'X)^{-1}X')Y = Y'Y - W.$$

Claim.

$$\begin{aligned} Y'Y + \mu'_{mtx}\tau\mu_{mtx} - \mu_{mtx}^{(j)'}(X'X + \tau)\mu_{mtx}^{(j)} \\ = (Y - X\mu_{mtx}^{(j)})'(Y - X\mu_{mtx}^{(j)}) + (\mu_{mtx}^{(j)} - \mu_{mtx})'\tau(\mu_{mtx}^{(j)} - \mu_{mtx}). \end{aligned}$$

Note that the right-hand-side of the claim is exactly the matrix V . Hence, using the claim, we obtain

$$\begin{aligned} P(\Sigma \mid \mathbf{y}_j^*, \mathbf{x}_j^*) \propto |\Sigma|^{-(n_j + \nu + p + 1)/2} \exp \left\{ -\frac{1}{2} \text{trace} [(\mathbf{W} + \Lambda)\Sigma^{-1}] \right\} \\ \times \exp \left\{ -\frac{1}{2} \text{trace} [(\mathbf{V} - \mathbf{W})\Sigma^{-1}] \right\}. \end{aligned}$$

Consequently, $P(\Sigma \mid \mathbf{y}_j^*, \mathbf{x}_j^*)$ is an inverse-Wishart distribution with $n_j + \nu$ degrees of freedom, and scale matrix $V + \Lambda$.

Now we proceed to show the claim. We have

$$\begin{aligned} (Y - X\mu_{mtx}^{(j)})'(Y - X\mu_{mtx}^{(j)}) &= Y'Y - 2Y'X\mu_{mtx}^{(j)} + \mu_{mtx}^{(j)'}X'X\mu_{mtx}^{(j)} \\ (\mu_{mtx}^{(j)} - \mu_{mtx})'\tau(\mu_{mtx}^{(j)} - \mu_{mtx}) &= \mu_{mtx}^{(j)'}\tau\mu_{mtx}^{(j)} - 2\mu_{mtx}^{(j)'}\tau\mu_{mtx} + \mu'_{mtx}\tau\mu_{mtx}. \end{aligned}$$

Hence

$$V = Y'Y + \mu'_{mtx}\tau\mu_{mtx} + \mu_{mtx}^{(j)'}(X'X + \tau)\mu_{mtx}^{(j)} - 2(Y'X\mu_{mtx}^{(j)} + \mu_{mtx}^{(j)'}\tau\mu_{mtx}).$$

But

$$Y'X\mu_{mtx}^{(j)} + \mu_{mtx}^{(j)'}\tau\mu_{mtx} = (X'Y + \tau\mu_{mtx})'\mu_{mtx}^{(j)} = \mu_{mtx}^{(j)'}(X'X + \tau)\mu_{mtx}^{(j)}.$$

Hence

$$V = Y'Y + \mu'_{mtx}\tau\mu_{mtx} - \mu_{mtx}^{(j)'}(X'X + \tau)\mu_{mtx}^{(j)}.$$

This concludes the proof of the claim and of the Lemma. ■

Proof of Lemma 3

We are going to compute the marginal using the identity

$$P(\mathbf{Y} \mid \mathbf{X}) = \frac{P(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}) P(\boldsymbol{\theta}, \boldsymbol{\Sigma})}{P(\boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{Y}, \mathbf{X})}, \quad \text{for all } (\boldsymbol{\theta}, \boldsymbol{\Sigma}).$$

From the calculations done in the proof of Lemma 2

$$\begin{aligned} P(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}) P(\boldsymbol{\theta}, \boldsymbol{\Sigma}) &= \frac{(2\pi)^{-(n_j p + pq)/2} 2^{-\nu p/2}}{\Gamma_p(\nu/2)} |\boldsymbol{\Lambda}|^{\nu/2} |\boldsymbol{\tau}|^{p/2} |\boldsymbol{\Sigma}|^{-(n_j + q + \nu + p + 1)/2} \\ &\times \exp \left\{ -\frac{1}{2} \text{trace} [(\boldsymbol{\Lambda} + \mathbf{V}) \boldsymbol{\Sigma}^{-1}] \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}^{(j)})' [\boldsymbol{\tau}^{(j)}]^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}^{(j)}) \right\}, \quad (\text{A.3}) \end{aligned}$$

and

$$\begin{aligned} P(\boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{Y}, \mathbf{X}) &= \frac{2^{-(\nu + n_j)p/2}}{\Gamma_p((\nu + n_j)/2)} |\boldsymbol{\Lambda} + \mathbf{V}|^{(\nu + n_j)/2} |\boldsymbol{\Sigma}|^{-(\nu + n_j + p + 1)/2} \\ &\exp \left\{ -\frac{1}{2} \text{trace} [(\boldsymbol{\Lambda} + \mathbf{V}) \boldsymbol{\Sigma}^{-1}] \right\} \times (2\pi)^{-pq/2} |\boldsymbol{\Sigma}|^{-q/2} |\mathbf{X}' \mathbf{X} + \boldsymbol{\tau}|^{p/2} \\ &\times \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}^{(j)})' [\boldsymbol{\tau}^{(j)}]^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}^{(j)}) \right\}. \quad (\text{A.4}) \end{aligned}$$

Then (13) is obtained by computing the ratio between (A.3) and (A.4). ■

Proof of Proposition 1

Note that when $\boldsymbol{\mu}_{\bar{\ell}}^{(p,*)} \approx \boldsymbol{\mu}_{\bar{\ell}}^{(j)}$, one has

$$\mathbf{V}_{\bar{\ell}}(\mathbf{w}, [\rho_n, b_{n+1}]) \approx \mathbf{V}_{\bar{\ell}(\rho_n)} + (\mathbf{w} - \mathbf{x}'_{n+1} \boldsymbol{\mu}_{\bar{\ell}, mtx}^{(j)}) (\mathbf{w} - \mathbf{x}'_{n+1} \boldsymbol{\mu}_{\bar{\ell}, mtx}^{(j)})'.$$

The proposition is based on the fact that for any $p \times p$ matrix $\mathbf{A}$ and vector b of dimension p , the determinant $|\mathbf{A} + bb'| = |\mathbf{A}|(1 + b' \mathbf{A}^{-1} b)$. Therefore,

$$\begin{aligned} &|\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\mathbf{w}, [\rho_n, b_{n+1}])| \\ &\approx |\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}(\rho_n)}| \left(1 + (\mathbf{w} - \mathbf{x}'_{n+1} \boldsymbol{\mu}_{\bar{\ell}, mtx}^{(j)})' [\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}(\rho_n)}]^{-1} (\mathbf{w} - \mathbf{x}'_{n+1} \boldsymbol{\mu}_{\bar{\ell}, mtx}^{(j)}) \right). \end{aligned}$$

Let $\boldsymbol{\theta} = \mathbf{x}'_{n+1} \boldsymbol{\mu}_{\bar{\ell}, mtx}^{(j)}$. We have,

$$\begin{aligned}
& \int_{\mathbf{w}} |\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\mathbf{w}, [\rho_n, b_{n+1}])|^{-(\nu+n_{\bar{\ell}})/2} d\mathbf{w} \\
& \approx \int_{\mathbf{w}} |\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n)|^{-(\nu+n_{\bar{\ell}})/2} \left(1 + (\mathbf{w} - \boldsymbol{\theta})' [\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n)]^{-1} (\mathbf{w} - \boldsymbol{\theta}) \right)^{-(\nu+n_{\bar{\ell}})/2} d\mathbf{w} \\
& = |\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n)|^{-(\nu+n_{\bar{\ell}})/2} \int_{\mathbf{w}} \left(1 + (\mathbf{w} - \boldsymbol{\theta})' [\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n)]^{-1} (\mathbf{w} - \boldsymbol{\theta}) \right)^{-(\nu+n_{\bar{\ell}})/2} d\mathbf{w} \\
& \quad = |\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n)|^{-(\nu+n_{\bar{\ell}}-1)/2} \pi^{p/2} \Gamma\left(\frac{\nu + n_{\bar{\ell}} - p}{2}\right) / \Gamma\left(\frac{\nu + n_{\bar{\ell}}}{2}\right) \\
& \quad \times \int_{\mathbf{w}} \frac{\Gamma\left(\frac{\nu+n_{\bar{\ell}}}{2}\right)}{\Gamma\left(\frac{\nu+n_{\bar{\ell}}-p}{2}\right) \pi^{p/2} |\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n)|^{1/2} (1 + (\mathbf{w} - \boldsymbol{\theta})' [\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n)]^{-1} (\mathbf{w} - \boldsymbol{\theta}))^{(\nu+n_{\bar{\ell}})/2}} d\mathbf{w}.
\end{aligned}$$

The integral in the right-hand-side corresponds to the integral of a multivariate Student-t distribution density with $\nu_0 = \nu + n_{\bar{\ell}} - p$ degrees of freedom, mean $\boldsymbol{\theta}$, scale matrix $(\boldsymbol{\Lambda} + \mathbf{V}_{\bar{\ell}}(\rho_n))/\nu_0$. Thus, this integral is equal to 1.0, which concludes the proof. $\blacksquare$

B. Details for Simulated Data Used in Scenario II

This section gives further details on the Five-copula cluster dataset described in Scenario II of the main document. Table B.1 gives the parameters used to generate 1000 samples from each copula. The data were generated with the *R copula* package (Kojadinovic and Yan, 2010; Hofert et al., 2014). The corresponding two-dimensional response variables were generated according to the regression coefficients and variance-covariance matrices shown in Table B.2.

Marginals		Copulas				
		1 ($\rho = 0.5$)	2 ($\rho = 0.8$)	3 ($\rho = 0.5$)	4 ($\rho = 0.7$)	5 ($\rho = 0.5$)
Normal	Mean	-2.0	2.0	-2.0	2.0	4.0
	Std. Dev.	0.1	0.2	0.2	0.1	0.5
Exponential	Rate	0.2	4.7	1.7	4.7	0.2
Gamma	Shape	6.0	10.0	4.0	2.0	6.0
	Rate	10.0	6.0	10.0	4.0	8.0
Normal	Mean	3.0	0.0	4.0	4.0	-3.0
	Std. Dev.	0.5	0.2	0.5	0.1	0.1
Gamma	Shape	6.0	10.0	4.0	2.0	6.0
	Rate	10.0	6.0	10.0	4.0	8.0
Exponential	Rate	0.2	4.7	1.7	4.7	0.2

TABLE B.1

The parameters used to generate the five-copula dataset. The parameter ρ is the correlation defining the normal copula with compound symmetry.

Copula		Regression coefficients							Covariance		
		β_0	β_1	β_2	β_3	β_4	β_5	β_6	σ_1^2	σ_2^2	σ_{12}
1	Y_1	-1	3	-4	0	-6	-1.5	-2.5	0.028	0.067	0.033
	Y_2	1	5.5	-4.5	-5.5	5.5	4	-0.5			
2	Y_1	-0.5	-3	2	-4	-5.5	4.5	0	0.072	0.089	0.046
	Y_2	5	-5	-1.5	2	-3	1	-3			
3	Y_1	1	2.5	3	-1	5.5	1.5	-5.5	0.122	0.017	0.040
	Y_2	4	2	4	4	5.5	1	0.5			
4	Y_1	-5.5	-4	5	-5	-3	5	-4	0.042	0.038	-0.009
	Y_2	4	-0.5	-2.5	2	4.5	5.5	-1			
5	Y_1	3.5	-1.5	-0.5	3	-0.5	4.5	-6	0.060	0.010	0.006
	Y_2	4	3	2.5	5.5	-4	6	-3.5			

TABLE B.2

The regression coefficients giving rise to the response variable associated with the five-copula dataset (columns 3 to 9) and the response covariance matrix (last three columns) for each copula component. The symbols σ_1^2 , σ_2^2 , and σ_{12} denote the corresponding variances of Y_1 , Y_2 and the covariance between Y_1 and Y_2 .

C. Details for the Potts Regression and PPMx Comparison

We consider now the third scenario from Section 4 of the main document. This scenario is given by the survival data analysis application described in Müller, Quintana and Rosner (2011). Therefore we change the likelihood model to just a univariate normal distribution that does not involve covariate dependence. The corresponding parameters are just the mean and variance, i.e. $\psi_i = (\mu_i, \sigma_i^2)$. In other words, we have $p(y_i \mid \mathbf{x}_i, \psi_i) \equiv N(\mu_i, \sigma_i^2)$, disregarding any available covariate vector $\mathbf{x}_i$. To compare Potts regression with PPMx, we use the same simulated data described in Table 1 of Müller, Quintana and Rosner (2011). The competing models are defined as the univariate likelihood model described above, and either the Potts regression or PPMx priors. The data contain three covariates, one continuous (x_1) and two of binary type (x_2 and x_3). We use both models to independently estimate $E(y_i \mid x_1, x_2, x_3)$ for twelve combinations of the three covariates, as indicated in Table C.1. We randomly generated 100 subsamples of size $n = 200$ from the data. We computed the RMSE over the 100 subsamples with respect to the known true values for the estimated expectations under each of the twelve combinations. All computations are carried out using the same subsamples. Note that the covariate combinations $(x_1, x_2, x_3) = (1, 0, 0)$ and $(1, 1, 0)$ were not supplied to fit the models, and hence the associated results correspond to prediction results. Table C.1 summarizes the results. We again consider the two estimate types discussed earlier (direct means or density-based). The table also includes average RMSE for each of the 12 covariate combinations, as well as the average RMSE for each of the 100 subsamples (last row). We find that PPMx has a very slightly better performance, and that density-based estimation of the mean performs somewhat worse than the other alternatives. However, as in the previous scenarios, the Potts regression model yields better results for the prediction cases $(x_1, x_2, x_3) = (1, 0, 0)$ and $(1, 1, 0)$.

Covariates			Mean				Truth	RMSE		
x_1	x_2	x_3	PPMx	Potts Mean	Potts Density	Empirical		PPMX	Potts Mean	Potts Density
-1	0	0	39	43	45	47	45	9.1	6.4	7.9
0	0	0	33	25	31	28	32	5.3	9.5	7.8
1	0	0	30	23	28	21	26	5.9	5.7	5.8
-1	1	0	42	45	47	45	45	5.8	8.0	13.3
0	1	0	33	28	33	29	32	5.5	7.5	6.5
1	1	0	30	29	34	26	26	6.8	6.5	9.2
-1	0	1	42	41	43	42	49	8.6	9.4	9.6
0	0	1	36	38	41	39	40	5.2	6.4	5.4
1	0	1	29	28	33	30	30	3.9	6.5	7.8
-1	1	1	48	52	57	56	59	11.8	9.4	7.3
0	1	1	47	56	58	58	56	9.6	6.2	5.9
1	1	1	44	45	47	48	49	6.6	6.7	7.0
Subsample average RMSE								7.18	7.35	7.79

TABLE C.1

Summary of PPMx versus Potts Regression for simulated with covariate-independent sampling model. The specific combinations are given by the values in the first three columns. The next four columns indicate means for predictions under each of the models, together with the empirical mean over datasets. The next column is the true mean to be estimated. The final set of columns indicate root mean square errors under each method.

D. Details from the analysis of the Health Outcomes Data of Section 5

The Potts regression model favored seven clusters. The three largest ones contain 45 (Cluster 4), 38 (Cluster 2) and 32 (Cluster 3) countries, respectively. The other clusters are small, containing 8 (Cluster 5), 5 (Cluster 6) and 3 (Clusters 1 and 7) countries. Cluster 4 contains most of the richest countries in the world which show moderate to high *Adjusted Life Expectancy* values, and very low *Child Mortality* values. Cluster 2 contains the poorest countries. These presented very high *Child Mortality* values at the turn of the century. Cluster 3 contains most of the developing countries in Latin America and Asia. These presented mid-range values for *Child Mortality* and high *Adjusted Life Expectancy* values. The list of countries forming the clusters are shown in Table D.1.

Figure D.1 displays contours of the corresponding posterior predictive density associated with countries in Cluster 3, that is, the cluster with countries in the mid range of health outcomes. This time the different levels of the variables *Birth Rate* and *Human Resources* correspond to every combination of the observed quantiles 0.1, 0.25, 0.50, 0.75, and 0.90 within Cluster 3 of the *Birth Rate* and *Human Resources* variables. These levels are “realistic” levels of change within this cluster. These are countries interesting to watch since their health outcomes could go either way, for the worse (approaching Cluster 2) or

for the best (approaching Cluster 4). This figure indicates that just moving along a single variable, *Birth Rate* or *Human Resources*, would not make much of a difference in health outcomes. The movement has to be done simultaneously in both variables to obtain a noticeable change in the health outcomes. This implies an interaction effect between these variables, suggesting that policy makers should not concentrate in single factors, but on several of them simultaneously.

CLUSTER 1	Algeria	Iran, Islamic Rep.	Morocco	
CLUSTER 2	Angola	Bangladesh	Benin	Burkina Faso
	Cambodia	Cameroon	Chad	Comoros
	Cote d'Ivoire	Equatorial Guinea	Eritrea	Ghana
	Guatemala	Guinea	Guinea-Bissau	Honduras
	Iraq	Kenya	Lesotho	Madagascar
	Mali	Mauritania	Myanmar	Nepal
	Niger	Nigeria	Oman	Pakistan
	Papua New Guinea	Saudi Arabia	Senegal	Sierra Leone
	Syrian Arab Republic	Tanzania	Togo	Yemen, Rep.
	Zambia	Zimbabwe		
CLUSTER 3	Argentina	Brazil	Chile	Colombia
	Costa Rica	Dominican Republic	Ecuador	Egypt, Arab Rep.
	El Salvador	Guyana	Indonesia	Jordan
	Kyrgyz Republic	Libya	Malaysia	Mauritius
	Mexico	Mongolia	Panama	Peru
	Philippines	South Africa	Sri Lanka	Tajikistan
	Thailand	Trinidad and Tobago	Tunisia	Turkey
	Turkmenistan	Uzbekistan	Venezuela, RB	Vietnam
CLUSTER 4	Armenia	Australia	Austria	Belarus
	Belgium	Bulgaria	Canada	Croatia
	Cuba	Cyprus	Czech Republic	Denmark
	Estonia	Finland	France	Georgia
	Germany	Greece	Hungary	Iceland
	Ireland	Israel	Italy	Japan
	Kazakhstan	South Korea	Latvia	Lithuania
	Moldova	Netherlands	New Zealand	Norway
	Poland	Portugal	Romania	Russian Federation
	Slovak Republic	Slovenia	Spain	Sweden
	Switzerland	Ukraine	United Kingdom	United States
	Uruguay			
CLUSTER 5	Azerbaijan	Bahrain	China	Kuwait
	Macedonia, FYR	Qatar	Singapore	United Arab Emirates
CLUSTER 6	Bolivia	Botswana	Namibia	Nicaragua
	Paraguay			
CLUSTER 7	Haiti	India	Jamaica	

TABLE D.1
The seven clusters favored by the Potts regression model.

FIG D.1. Health data. Contours of the posterior predictive bivariate densities of health outcomes associated with countries in Cluster 3 for different scenarios of Birth Rate (BR) and Human Resources (HR). The horizontal axis corresponds to Adjusted Life Expectancy, whilst the vertical one, corresponds to Child Mortality.

References

- HOFERT, M., KOJADINOVIC, I., MÄCHLER, M. and YAN, J. (2014). Package copula. <http://nacopula.r-forge.r-project.org/>
- KOJADINOVIC, I. and YAN, J. (2010). Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. *Journal of Statistical Software* **34** 1–20.
- MÜLLER, P., QUINTANA, F. A. and ROSNER, G. L. (2011). A Product Partition Model with Regression on Covariates. *Journal of Computational and Graphical Statistics* **20** 260–278.