APPENDIX AVAILABLE ON THE HEI WEB SITE

Research Report 183

Development of Statistical Methods for Multipollutant Research

Part 2. Development of Enhanced Statistical Methods for Assessing Health Effects Associated with an Unknown Number of Major Sources of Multiple Air Pollutants

E.S. Park et al.

Appendix C. Full Conditional Distributions for Parameters

Note: Appendices available only on the Web have been reviewed solely for spelling, grammar, and cross-references to the main text. They have not been formatted or fully edited by HEI.

Correspondence may be addressed to Dr. Eun Sug Park, Texas A&M Transportation Institute, The Texas A&M University System, 3135 TAMU, College Station, TX 77843-3135; e-park@tamu.edu.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83467701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

This document was reviewed by the HEI Health Review Committee.

Appendix C: Full Conditional Distributions for Parameters

FULL CONDITIONAL DISTRIBUTIONS FOR PARAMETERS UNDER THE CONTINUOUS HEALTH OUTCOME MODEL

The full conditional posterior distributions are given as:

$$\gamma_t \left| \cdots \sim N_q \left(m_{\gamma_t}, V_{\gamma_t} \right) \right.,$$

$$\sigma_j^{-2} \left| \cdots \sim Gamma\left(a_0 + \tfrac{1}{2} T, b_{0j} + \tfrac{1}{2} d_j \right) \right.,$$

where d_j is the jth diagonal element of $d = \left(X - 1_T \otimes \mu - \Gamma \mathbf{P} \right)' \left(X - 1_T \otimes \mu - \Gamma \mathbf{P} \right)$,

$$\sigma_y^{-2} \left| \cdots \sim Gamma\left(a_0^y + \tfrac{1}{2} T, b_0^y + \tfrac{1}{2} d_y \right) \right.,$$

where $d_y = \left(Y - \alpha 1_T - \Gamma \beta - Z\eta \right)' \left(Y - \alpha 1_T - \Gamma \beta - Z\eta \right)$, $Y = \left(y_1, y_2, \cdots, y_T \right)'$,

$\Omega \sim IW\left(R, T + r_0 \right)$ where $R = \Gamma' \Gamma + R_0$,

$$\mu \left| \cdots \sim N_J \left(m_\mu, V_\mu \right) \right.,$$

$$\alpha \left| \cdots \sim N \left(m_\alpha, V_\alpha \right) \right.,$$

$$\beta \left| \cdots \sim N_q \left(m_\beta, V_\beta \right) \right.,$$

$$\eta \left| \cdots \sim N_I \left(m_\eta, V_\eta \right) \right.,$$

where

$$m_{\gamma_t} = \left\{ \left(X_t - \mu \right) \Sigma^{-1} \mathbf{P}' + \left(y_t - \alpha - Z_t \eta \right) \beta' \sigma_y^{-2} \right\} V_{\gamma_t}, \quad V_{\gamma_t} = \left\{ \Omega^{-1} + P \Sigma^{-1} \mathbf{P}' + \beta \beta' \sigma_y^{-2} \right\}^{-1},$$

$$m_\mu = \left\{ T \left(\bar{X} - \bar{\gamma} \mathbf{P} \right) \Sigma^{-1} + m_0 M_0^{-1} \right\} V_\mu, \quad V_\mu = \left(M_0^{-1} + T \Sigma^{-1} \right)^{-1},$$

$$m_\alpha = \left(\sum_{t=1}^{T} (y_t - \gamma_t \beta - Z_t \eta) \sigma_y^{-2} + U_0^{-1} \alpha_0 \right) V_\alpha , \qquad V_\alpha = \left(T \sigma_y^{-2} + U_0^{-1} \right)^{-1},$$

$$m_\beta = V_\beta \left(\sum_{t=1}^{T} \gamma_t{}' \left(y_t - \alpha - Z_t \eta \right) \sigma_y^{-2} + B_0^{-1} \beta_0 \right), \quad V_\beta = \left(\Gamma^t \Gamma \sigma_y^{-2} + B_0^{-1} \right)^{-1},$$

$$m_\eta = V_\eta \left(\sum_{t=1}^{T} Z_t{}' \left(y_t - \alpha - \gamma_t \beta \right) \sigma_y^{-2} + \Psi_0^{-1} \eta_0 \right), \quad V_\eta = \left(Z^t Z \sigma_y^{-2} + \Psi_0^{-1} \right)^{-1},$$

For the columns of **P** with no preassigned zero elements,

$$P_j \,\middle|\, \cdots \sim N_q \left(c_j, C_j \right) \mathbf{I} \left(P_{kj} \geq 0, k = 1, \cdots, q \right),$$

where $c_j = C_j \left\{ \sigma_j^{-2} \Gamma' \left(X_j - \mu_j 1_T \right) + C_{0j}^{-1} c_{0j} \right\}$, $C_j = \left(\sigma_j^{-2} \Gamma' \Gamma + C_{0j}^{-1} \right)^{-1}$, c_{0j} is a q-dimensional prior mean vector of P_j, C_{0j} is a corresponding submatrix of C_0, and X_j is the jth column of X.

For the columns of **P** containing preassigned zero elements,

$$P_j^+ \,\middle|\, \cdots \sim N_{q_j^+} \left(c_j^+, C_j^+ \right) \mathbf{I} \left(P_{kj} \geq 0, k = 1, \cdots, q_j^+ \right)$$

where P_j^+ is a column vector consisting of free elements in the jth column of **P** (i.e., a vector that corresponds to the jth column of **P** after deleting prespecified zero elements used for identifiability if there are any), q_j^+ is the length of P_j^+ (i.e., the number of free elements in the jth column of **P**), $c_j^+ = C_j^+ \left\{ \sigma_j^{-2} \Gamma_j^{+\prime} \left(X_j - \mu_j 1_T \right) + \left(C_{0j}^+ \right)^{-1} c_{0j}^+ \right\}$, $C_j^+ = \left\{ \sigma_j^{-2} \Gamma_j^{+\prime} \Gamma_j^+ + \left(C_{0j}^+ \right)^{-1} \right\}^{-1}$, c_{0j}^+ is a q_j^+ -dimensional prior mean vector of P_j^+, C_{0j}^+ is a corresponding submatrix of C_0, and Γ_j^+ consists of the columns of Γ corresponding to free elements of the jth column of **P**.

FULL CONDITIONAL DISTRIBUTIONS FOR PARAMETERS UNDER THE DISCRETE HEALTH OUTCOME MODEL AND THE ALGORITHM FOR SAMPLE GENERATION IN MCMC

Let $H_t\left(\alpha \middle| \gamma_t, \beta, \eta, w_t\right)$ denote the subregion of α such that, for given $\left(\gamma_t, \beta, \eta, w_t\right)$, $\left(\alpha, \gamma_t, \beta, \eta, w_t\right)$ is in $H_t\left(\alpha, \gamma_t, \beta, \eta, w_t\right)$. In the same manner, we denote the subregion of γ_t, β, η, and w_t in $H_t\left(\alpha, \gamma_t, \beta, \eta, w_t\right)$ given others, by $H_t\left(\gamma_t \middle| \alpha, \beta, \eta, w_t\right)$, $H_t\left(\beta \middle| \gamma_t, \alpha, \eta, w_t\right)$, $H_t\left(\eta \middle| \gamma_t, \beta, \alpha, w_t\right)$, and $H_t\left(W_t \middle| \gamma_t, \beta, \eta, \alpha\right)$, respectively. Then the full conditional posterior distributions can be given as:

$$\gamma_t \left| \cdots \sim N_q\left(m_{\gamma_t}, V_{\gamma_t}\right) I\left[H_t\left(\gamma_t \middle| \beta, \eta, \alpha, w_t\right)\right],\right.$$

$$\alpha \left| \cdots \sim N\left(m_\alpha, V_\alpha\right) I\left[\bigcap_{t=1}^{T} H_t\left(\alpha \middle| \beta, \eta, \gamma_t, w_t\right)\right],\right.$$

$$\beta \left| \cdots \sim N_q\left(m_\beta, V_\beta\right) I\left[\bigcap_{t=1}^{T} H_t\left(\beta \middle| \alpha, \eta, \gamma_t, w_t\right)\right],\right.$$

$$\eta \left| \cdots \sim N_I\left(m_\eta, V_\eta\right) I\left[\bigcap_{t=1}^{T} H_t\left(\eta \middle| \alpha, \beta, \gamma_t, w_t\right)\right],\right.$$

$$W_t \left| \cdots \sim N\left(\alpha + \gamma_t \beta + Z_t \eta, \quad 1\right) I\left[H_t\left(W_t \middle| \gamma_t, \beta, \eta, \alpha\right)\right],\right. \qquad \text{(C.1)}$$

$\sigma_j^{-2} \left| \cdots \sim Gamma\left(a_0 + \frac{1}{2}T, b_{0j} + \frac{1}{2}d_j\right)\right.$ where d_j is the jth diagonal element of

$$d = \left(X - 1_T \otimes \mu - \Gamma \mathbf{P}\right)' \left(X - 1_T \otimes \mu - \Gamma \mathbf{P}\right),$$

$\Omega \sim IW\left(R, T + r_0\right)$ where $R = \Gamma'\Gamma + R_0$, $\mu \left| \cdots \sim N_J\left(m_\mu, V_\mu\right)\right.$,

where

$$m_{\gamma_t} = \left\{\left(W_t - \alpha - Z_t\eta\right)\beta' + \left(X_t - \mu\right)\Sigma^{-1}\mathbf{P}'\right\} V_{\gamma_t},\ V_{\gamma_t} = \left\{\Omega^{-1} + P\Sigma^{-1}\mathbf{P}' + \beta\beta'\right\}^{-1},$$

$$m_\alpha = \left(\sum_{t=1}^{T} \left(W_t - \gamma_t \beta - Z_t \eta \right) + U_0^{-1} \alpha_0 \right) V_\alpha , \qquad V_\alpha = \left(T + U_0^{-1} \right)^{-1} ,$$

$$m_\beta = V_\beta \left(\sum_{t=1}^{T} \left(W_t - \alpha - Z_t \eta \right) \gamma_t' + B_0^{-1} \beta_0 \right), \qquad V_\beta = \left(\sum_{t=1}^{T} \gamma_t' \gamma_t + B_0^{-1} \right)^{-1} ,$$

$$m_\eta = V_\eta \left(\sum_{t=1}^{T} \left(W_t - \alpha - \gamma_t \beta \right) Z_t' + \Psi_0^{-1} \eta_0 \right), \qquad V_\eta = \left(\sum_{t=1}^{T} Z_t' Z_t + \Psi_0^{-1} \right)^{-1} ,$$

$$m_\mu = \left(T \left(\bar{X} - \bar{\gamma} \mathbf{P} \right) \Sigma^{-1} + m_0 M_0^{-1} \right) V_\mu , \quad V_\mu = \left(M_0^{-1} + T \Sigma^{-1} \right)^{-1} .$$

where $c_j = C_j \left\{ \sigma_j^{-2} \Gamma' \left(X_j - \mu_j 1_T \right) + C_{0j}^{-1} c_{0j} \right\}$, $C_j = \left(\sigma_j^{-2} \Gamma' \Gamma + C_{0j}^{-1} \right)^{-1}$, c_{0j} is a q-dimensional prior mean vector of P_j, C_{0j} is a corresponding submatrix of C_0, and X_j is the jth column of X.

For the columns of **P** containing preassigned zero elements,

$$P_j^+ \Big| \cdots \sim N_{q_j^+} \left(c_j^+, C_j^+ \right) \mathbf{I} \left(P_{kj} \geq 0, k = 1, \cdots, q_j^+ \right),$$

where P_j^+ is a column vector consisting of free elements in the jth column of **P** (i.e., a vector that corresponds to the jth column of **P** after deleting prespecified zero elements used for identifiability if there are any), q_j^+ is the length of P_j^+ (i.e., the number of free elements in the jth column of **P**), $c_j^+ = C_j^+ \left\{ \sigma_j^{-2} \Gamma_j^{+\prime} \left(X_j - \mu_j 1_T \right) + \left(C_{0j}^+ \right)^{-1} c_{0j}^+ \right\}$, $C_j^+ = \left\{ \sigma_j^{-2} \Gamma_j^{+\prime} \Gamma_j^+ + \left(C_{0j}^+ \right)^{-1} \right\}^{-1}$, c_{0j}^+ is a q_j^+ -dimensional prior mean vector of P_j^+, C_{0j}^+ is a corresponding submatrix of C_0, and Γ_j^+ consists of the columns of Γ corresponding to free elements of the jth column of **P**, and $\bigcap$ denotes the intersection of intervals.

Using the idea of Oh and Park (2002) that approximates the Poisson cdf (F) by the standard normal cdf with appropriate transformations (F^*),

$$F^*\left(y_t \middle| \delta_t\right) = \Phi\left(-3\sqrt{y_t+1}\left[\frac{\left(e^{\delta_t}\right)^{1/3}}{\left(y_t+1\right)^{1/3}} - 1 + \frac{1}{9\left(y_t+1\right)}\right]\right),$$

we can use the following approximation:

$$\Phi^{-1}F\left(y_t \middle| \delta_t\right) \approx -3\sqrt{y_t+1}\left[\frac{\left(e^{\delta_t}\right)^{1/3}}{\left(y_t+1\right)^{1/3}} - 1 + \frac{1}{9\left(y_t+1\right)}\right] \equiv b\left(\delta_t, y_t\right),$$

and the restriction $H_t\left(\alpha, \gamma_t, \beta, \eta, w_t\right)$ can be replaced by

$$H_t^*\left(\alpha, \gamma_t, \beta, \eta, w_t\right) = H_t^*\left(\delta_t, w_t\right) = \left\{\left(\delta_t, w_t\right); h\left(y_t - 1, \delta_t, w_t\right) < 0 \le h\left(y_t, \delta_t, w_t\right)\right\}$$

where

$$h\left(y_t, \delta_t, w_t\right) = b\left(\delta_t, y_t\right) - w_t + \delta_t.$$

Now it remains to solve the inequalities

$$h\left(y_t - 1, \delta_t, w_t\right) < 0 \le h\left(y_t, \delta_t, w_t\right) \qquad \text{(C.2)}$$

for δ_t given y_t and w_t, where $\delta_t = \alpha^* + \gamma_t\beta + Z_t\eta.$

Using the fact that the function h is increasing in y_t and concave in δ_t given w_t, it can be shown that there exist two distinct solutions of δ_t for

$$h\left(y_t, \delta_t, w_t\right) = 0$$

and Equation C.2 is equivalent to

$$\left\{c_{1t} < \delta_t < c_{2t}\right\}$$

or

$$\left\{c_{1t} < \delta_t < d_{1t}\right\} \cup \left\{d_{2t} < \delta_t < c_{2t}\right\},$$

where c_{1t} and c_{2t} are two distinct solutions of $h\left(y_t, \delta_t, w_t\right) = 0$, d_{1t} and d_{2t} are two distinct solutions of $h(y_t - 1, \delta_t, w_t) = 0$, and $\cup$ denotes the union of intervals. (Note that c_{1t}, c_{2t}, d_{1t}, and d_{2t} depend on w_t and y_t.) Thus, given w_t, the region of $\delta_t = \alpha + \gamma_t \beta + Z_t \eta.$ satisfying the restriction H_t^* is given as a fixed interval, hence, given w_t and all the other parameters, the region of each element of α, γ_t, β, and η is given as a fixed interval. From Equation C.1 with H_t replaced by H_t^*, the full conditional distributions of elements of W_t, α, γ_t, β, and η are given as normal distributions restricted to a fixed interval.

The full conditional posterior distribution of W_t is a univariate normal distribution restricted to a fixed interval from which sample generation is easy:

$$W_t | \cdots \sim N(\delta_t, \;\; 1) I\left[b(\delta_t, y_t - 1) < W_t - \delta_t \le b(\delta_t, y_t) \right]$$

The full conditional posterior distribution of γ_t can be given as

$$\gamma_t | \cdots \sim N_q\left(m_{\gamma_t}, V_{\gamma_t}\right) I\left[c_{1t} - \alpha - Z_t\eta < \gamma_t \beta < c_{2t} - \alpha - Z_t \eta\right]$$

or

$$\sim N_q\left(m_{\gamma_t}, V_{\gamma_t}\right) I\left[\{c_{1t} - \alpha - Z_t\eta < \gamma_t\beta < d_{1t} - \alpha - Z_t\eta\} \cup \{d_{2t} - \alpha - Z_t\eta < \gamma_t\beta < c_{2t} - \alpha - Z_t\eta\}\right].$$

for which sample generation of γ_t (for the first case with c_{1t} and c_{2t}) can be done elementwise as follows:

$$\gamma_{tk} | \cdots \sim \begin{cases} N\left(m_{\gamma_{tk}}, \upsilon^2_{\gamma_{tk}}\right) I\left[\dfrac{1}{\beta_k}\left(l_t - \sum_{i \ne k} \gamma_{ti}\beta_i\right) < \gamma_{tk} < \dfrac{1}{\beta_k}\left(u_t - \sum_{i\ne k}\gamma_{ti}\beta_i\right)\right] & \text{if } \beta_k > 0 \\ N\left(m_{\gamma_{tk}}, \upsilon^2_{\gamma_{tk}}\right) I\left[\dfrac{1}{\beta_k}\left(u_t - \sum_{i \ne k} \gamma_{ti}\beta_i\right) < \gamma_{tk} < \dfrac{1}{\beta_k}\left(l_t - \sum_{i\ne k}\gamma_{ti}\beta_i\right)\right] & \text{if } \beta_k < 0 \\ N\left(m_{\gamma_{tk}}, \upsilon^2_{\gamma_{tk}}\right) & \text{if } \beta_k = 0 \end{cases}$$

where $m_{\gamma_{tk}}$ and $\upsilon^2_{\gamma_{tk}}$ are full conditional mean and variance of γ_{tk}, respectively, and

$l_t = c_{1t} - \alpha - Z_t\eta$, $u_t = c_{2t} - \alpha - Z_t\eta$. The second case can be handled similarly.

The full conditional posterior distribution of β can be given as

$$\beta|\cdots \sim N_q\left(m_\beta, V_\beta\right)\times$$

$$I\left(\bigcap_t\left[\begin{cases} c_{1t} - \alpha - Z_t\eta < \gamma_t\beta < c_{2t} - \alpha - Z_t\eta \\ or \\ \left(c_{1t} - \alpha - Z_t\eta < \gamma_t\beta < d_{1t} - \alpha - Z_t\eta\right)\cup\left(d_{2t} - \alpha - Z_t\eta < \gamma_t\beta < c_{2t} - \alpha - Z_t\eta\right)\end{cases}\right]\right)$$

for which the sample generation can be done element-wise as follows:

1) For each t,

let $l_t = c_{1t} - \alpha - Z_t\eta$, $u_t = c_{2t} - \alpha - Z_t\eta$ if the number of intervals is 1, and

let $l_{t1} = c_{1t} - \alpha - Z_t\eta$, $u_{t1} = d_{1t} - \alpha - Z_t\eta$, $l_{t2} = d_{2t} - \alpha - Z_t\eta$, $u_{t2} = c_{2t} - \alpha - Z_t\eta$ if the number of intervals is 2.

2) Then for each k ($k = 1,\ldots,q$) and t, we have

$$l_{tk}^* = \begin{cases} \frac{1}{\gamma_{tk}}\left(l_t - \sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk} > 0 \\ \frac{1}{\gamma_{tk}}\left(u_t - \sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk} < 0 \\ -\infty & \text{if } \gamma_{tk} = 0 \end{cases}, \quad u_{tk}^* = \begin{cases} \frac{1}{\gamma_{tk}}\left(u_t - \sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk} > 0 \\ \frac{1}{\gamma_{tk}}\left(l_t - \sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk} < 0 \\ \infty & \text{if } \gamma_{tk} = 0 \end{cases}$$

when the number of intervals is 1,

or

$$l_{t1k}^{*}=\begin{cases}\dfrac{1}{\gamma_{tk}}\left(l_{t1}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}>0\\ \dfrac{1}{\gamma_{tk}}\left(u_{t1}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}<0\\ \text{-}\infty & \text{if } \gamma_{tk}=0\end{cases},\quad u_{t1k}^{*}=\begin{cases}\dfrac{1}{\gamma_{tk}}\left(u_{t1}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}>0\\ \dfrac{1}{\gamma_{tk}}\left(l_{t1}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}<0\\ \infty & \text{if } \gamma_{tk}=0\end{cases}$$

$$l_{t2k}^{*}=\begin{cases}\dfrac{1}{\gamma_{tk}}\left(l_{t2}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}>0\\ \dfrac{1}{\gamma_{tk}}\left(u_{t2}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}<0\\ \text{-}\infty & \text{if } \gamma_{tk}=0\end{cases},\quad u_{t2k}^{*}=\begin{cases}\dfrac{1}{\gamma_{tk}}\left(u_{t2}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}>0\\ \dfrac{1}{\gamma_{tk}}\left(l_{t2}-\sum_{i\neq k}\gamma_{ti}\beta_i\right) & \text{if } \gamma_{tk}<0\\ \infty & \text{if } \gamma_{tk}=0\end{cases}$$

when the number of intervals is 2.

3) Find the intersection of intervals corresponding to each t over $t = 1,\ldots,T$. Let I_k be the number of final disjoint intervals for β_k, and $\left(l_1^k, u_1^k\right),\cdots,\left(l_{I_k}^k, u_{I_k}^k\right)$ be the disjoint intervals obtained as a result of intersection.

4) The kth element of β can be generated as

$$\beta_k \mid \cdots \sim N\left(m_{\beta_k}, \upsilon_{\beta_k}^2\right) I\left[\bigcup_{i=1}^{I_k}\left\{l_i^k < \beta_k < u_i^k\right\}\right]$$

where m_{β_k} and $\upsilon_{\beta_k}^2$ are full conditional mean and variance of β_k, respectively.

Note that the inequalities in Equation C.2 need to be solved for each t within each iteration of MCMC. Also, the intersection of intervals over $t = 1,\ldots, T$ need to be found within each iteration of MCMC.

The sample generation from the full conditional distributions of α and η can also be handled similarly.