Web-based Supplementary Materials for "Quantification of the Dose-Response Relationship for a Continuous Treatment in the Presence of Confounding or Informative Non-Compliance" by Erica E. M. Moodie and David A. Stephens.

Appendix A

Estimation in Linear Mixed Models

Suppose, that

$$Y = X\beta + Zu + \epsilon, \tag{A.1}$$

where $u \sim N(0, G)$ and $\epsilon \sim N(0, R)$ with u and ϵ independent. This model can be interpreted as $Y|\beta, u \sim N(X\beta+Zu, R)$, $u \sim N(0, G)$, yielding (on integrating out u) the marginal model $Y|\beta \sim N(X\beta, ZGZ^{\mathsf{T}} + R)$. Let $V = ZGZ^{\mathsf{T}} + R$. Then the maximum penalized likelihood estimates of β and u given G and R are given by

$$\widehat{\theta} = \begin{bmatrix} \widehat{\beta} \\ \widehat{u} \end{bmatrix} = (C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1}y \tag{A.2}$$

where $C = [X \ Z]$ and B is the block diagonal matrix with blocks $\mathbf{0}$ and G^{-1}. The variance of the estimators are given by $\mathrm{Var}[\widehat{\theta}] = (C^{\mathsf{T}}R^{-1}C + B)^{-1}$. Fitted values are obtained routinely as

$$\widehat{y} = C(C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1}y$$

whereas predictions under this model at new design point $c_0 = [x_0 \ z_0]$ are obtained from

$$\widehat{y}_0 = c_0(C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1}y$$

with variance $c_0(C^\mathsf{T}R^{-1}C + B)^{-1}c_0^\mathsf{T}$. The quantities R and G that together define V and B can be estimated using maximum profile (integrated) likelihood

$$l_P(V) = \text{constant} - \frac{1}{2}\left[\log|V| + y^\mathsf{T}V^{-1}(I - X(X^\mathsf{T}V^{-1}X)^{-1}X^\mathsf{T}V^{-1})y\right]$$

obtained from the likelihood plugging in $\widehat{\beta} = (X^\mathsf{T}V^{-1}X)^{-1}X^\mathsf{T}V^{-1}y$, or REML, using the restricted likelihood

$$l_R(V) = l_P(V) - \frac{1}{2}\log|X^\mathsf{T}V^{-1}X|.$$

obtained by first integrating out β from the likelihood $Y \sim N(X\beta, V)$.

Appendix B

An Informative Prior Specification for the SPALM

In the SPALM model, the specification of random effects prior matrix G can be engineered to match prior opinion about the nature (that is, smoothness or curvature) of the modelled function. Consider a single semiparametric component $Y = Zu$, where $u \sim N(0, G)$, so that $Y \sim N(0, ZGZ^\mathsf{T})$, where we require that *a priori* $Y \sim N(0, V_0)$. Then

$$V_0 = ZGZ^\mathsf{T} \Longrightarrow G = (Z^\mathsf{T}Z)^{-1}Z^\mathsf{T}V_0Z(Z^\mathsf{T}Z)^{-1}$$

and G should adopt a data-dependent form, giving a prior that is similar in structure to the "g-prior" (Zellner, 1983). Conditional on knot points $\kappa_1, \ldots, \kappa_M$, we can specify any required prior autocovariance structure. For example, we could specify a prior with high autocorrelation, thereby encouraging smoothness in the semiparametric component. In our analysis, we specify V_0 to be a diagonal matrix such that the prior variation in the semiparametric function is concentrated on the range ± 2. This results in a required prior variance for the

dose component to be specified by

$$G^{-1} = 10^{-3} \begin{pmatrix} 14 & 13 & 12 & 11 & 11 & 10 & 9 & 8 & 6 & 5 & 2 \\ 13 & 12 & 11 & 11 & 10 & 9 & 8 & 7 & 6 & 4 & 2 \\ 12 & 11 & 11 & 10 & 9 & 9 & 8 & 7 & 6 & 4 & 2 \\ 11 & 11 & 10 & 10 & 9 & 8 & 8 & 7 & 6 & 4 & 2 \\ 11 & 10 & 9 & 9 & 9 & 8 & 7 & 6 & 5 & 4 & 2 \\ 10 & 9 & 9 & 8 & 8 & 7 & 7 & 6 & 5 & 4 & 1 \\ 9 & 8 & 8 & 8 & 7 & 7 & 6 & 5 & 5 & 3 & 1 \\ 8 & 7 & 7 & 7 & 6 & 6 & 5 & 5 & 4 & 3 & 1 \\ 6 & 6 & 6 & 6 & 5 & 5 & 5 & 4 & 4 & 3 & 1 \\ 5 & 4 & 4 & 4 & 4 & 4 & 3 & 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

This is a much more precise specification than the noninformative prior we selected. However, it is much less precise than the prior deduced using the empirical Bayes procedure based on ML/REML estimation of the components of G; the parameters $(\widehat{\sigma}_1, \widehat{\sigma}_2, \widehat{\sigma}_3)$ that define the diagonal components of G are (7.589e-08, 2.314e-06, 7.430e-04). Finally, as the prior is design-dependent, this specification is only strictly appropriate for the fixed-knot case, and will change in a straightforward fashion when the knot positions change.

The results from an analysis using this informative prior (and fixed knots) are depicted in Figure 4, where results for the Non-Informative and Empirical Bayes Priors are also shown for comparison. Overall, results are broadly similar when the two fully Bayesian procedures are used, but the magnitude of the various dose effects are estimated to be much larger than those estimated using the empirical Bayes procedures. We note that the deduced empirical Bayes prior has **extremely** (we argue unreasonably) high precision for several of the components, and prefer the informative specification.

Appendix C

Estimation of the Average Potential Outcome in a frequentist GPS analysis

The approach to estimating the Average Potential Outcome (APO) at dose level d, $\mu(d)$, was described in Hirano and Imbens (2004) and proceeds as follows:

I. Estimate β in the predictive model for D given $X = x$, $f_{D|X}(d|x, \beta)$.

II. Compute the estimated GPS, $\widehat{r}_i = f_{D|X}(d_i|x_i, \widehat{\beta})$.

III. Estimate α in the predictive model for Y given $D = d$ and $R = \widehat{r}$, $f_{Y|D,R}(y|d, \widehat{r}, \alpha)$.

IV. Estimate the APO at dose level d by

$$\widehat{\mu}(d) = \widehat{E}[Y|D = d] = \frac{1}{N}\sum_{i=1}^{N} E_{Y|D,R}[Y_i(d)|D = d, \widehat{r}_i = \widehat{r}(d, x_i), \widehat{\alpha}]$$

for d in a suitable range in $\mathcal{D}$, where $\widehat{r}$ is evaluated at $\beta = \widehat{\beta}$. Then $\widehat{\mu}(d), d \in \mathcal{D}$ is the GPS-adjusted estimated dose-response function.

Several components in this model must be user-specified; the two key conditional models $f_{D|X}(d|x, \beta)$ and $f_{Y|D,R}(y|d, r, \alpha)$ must be selected to reflect the various relationships between the variables. However, the adequacy of both components is testable in a straightforward statistical fashion.

Appendix D

Bayesian Posterior Calculation for Repeated Measures Data

Consider a linear model formulation using the notation introduced earlier, that is where $Y \sim N(X\beta, R)$, where R is a block diagonal error covariance matrix $R = \text{diag}(R_1, \ldots, R_N)$. For example, the components of R can be specified via

the exponential decay, or AR(1) autocorrelation functions. We focus on the former for illustration.

To complete the specification, we use a diffuse (improper uniform) prior specification for β and an improper Jeffreys-type prior on the positive parameters in the exponential autocovariance function; i.e., take $p(\beta, \lambda, \zeta, \nu) = (\lambda\zeta\nu)^{-1}$ and derive the posterior distribution. This factorizes $p(\beta, \lambda, \zeta, \nu|y) = p(\lambda, \zeta, \nu|y)\, p(\beta|y, \lambda, \zeta, \nu)$ where

$$p(\lambda, \zeta, \nu|y) \propto \frac{|\ M_3|^{-1/2}}{\prod_{i=1}^{N} |R_i|^{1/2}} \exp\left\{-\frac{1}{2}\left[M_1 - M_2^{\mathsf{T}} M_3^{-1} M_2\right]\right\} \frac{1}{\lambda\zeta\nu} \tag{D.3}$$

with

$$M_1 = \sum_{i=1}^{N} y_i^{\mathsf{T}} R_i^{-1} y_i \qquad M_2 = \sum_{i=1}^{N} X_i^{\mathsf{T}} R_i^{-1} y_i \qquad M_3 = \sum_{i=1}^{N} X_i^{\mathsf{T}} R_i^{-1} X_i$$

and $\beta|y, \lambda, \zeta, \nu \sim N_p(M_3^{-1} M_2, M_3)$. The posterior distribution in equation (D.3) is not available analytically, but inference may be carried out using Markov chain Monte Carlo (MCMC) on the three parameter joint posterior. We use a Metropolis update on a sweep of the conditionals, reparameterized onto the log scale, and jointly on the block of the three parameters. The conditional posterior for β given (λ, ζ, ν) can be sampled directly.

To extend the mixed model, a further level can be added to the hierarchy in some cases, although this is not sensible for the semiparametric components. For example, fitting a random effects model is straightforward using a Gibbs sampler. Denoting by $\eta = (\eta_1, ..., \eta_N)$ the vector of child-specific random effects (intercepts), the posterior of interest becomes the joint distribution $p\left(\theta, \lambda, \zeta, \nu, \eta, \sigma_\eta^2|y\right)$, where σ_η^2 is the (unknown) random effect error variance, which is included in the MCMC cycle; we might assign an Inverse Gamma

prior with parameters 2.5 and 0.25. Then, conditional on η, the posterior for $(\theta, \lambda, \zeta, \nu)$ is updated as in the fixed effect only model, with datum y_{ij} replaced by $y_{ij} - \eta_i$. Conditional on $(\theta, \lambda, \zeta, \nu)$ and σ_η^2, the posterior for η_i is univariate Gaussian. Finally, conditional on all other parameters, the posterior for σ_η^2 is Inverse Gamma.

A fully Bayesian analysis is also possible for the SPALM model. The posterior distribution of the covariance parameters is identical to that in (D.3), but with

$$M_1 = \sum_{i=1}^{N} y_i^\mathsf{T} R_i^{-1} y_i \qquad M_2 = \sum_{i=1}^{N} C_i^\mathsf{T} R_i^{-1} y_i \qquad M_3 = \sum_{i=1}^{N} C_i^\mathsf{T} R_i^{-1} C_i = C^\mathsf{T} R^{-1} C.$$

where $R \equiv R(\lambda, \zeta, \nu)$, and the posterior distribution for $\theta = [\beta \; u]^\mathsf{T}$ is multivariate normal (dimension $p + MK$) with mean and variance

$$\mu = (C^\mathsf{T} R^{-1} C + B)^{-1} C^\mathsf{T} R^{-1} y \qquad \Sigma = (C^\mathsf{T} R^{-1} C + B)^{-1}$$

respectively. In addition, rather than using an improper uniform prior for β, an informative prior can also be specified. In this case, the calculation proceeds as before, with the marginal posterior for (λ, ζ, ν) sampled using Metropolis-Hastings, and the conditional posterior for β (or (β, u)) multivariate normal.

Appendix E

MCMC for the Bayesian Generalized Propensity Score

To sample the posterior distribution for (α, β, γ), given the data (y, d, x), iterate around the following cycle with parameter updating: at iteration m, let the current values of the parameters be $(\alpha^{(m)}, \beta^{(m)}, \gamma^{(m)})$, and let $r_1^{(m)}, \ldots, r_N^{(m)}$ be defined by

$$r_i^{(m)} = f_{D|X}(d_i | x_i, \beta^{(m)}, \gamma^{(m)}).$$

Let $p(\alpha, \beta, \gamma)$ be the joint prior distribution for the three parameters, and let $p(\alpha)$ and $p(\beta, \gamma)$ be the corresponding marginal priors. Then

1. Sample $\alpha^{(m+1)}$ from full conditional $p(\alpha|y, d, x, \beta^{(m)}, \gamma^{(m)})$ using Metropolis-Hastings, where

$$p(\alpha|y, d, x, \beta, \gamma) \propto f_{Y|D,R}(y|d, r, \alpha)p(\alpha)$$

proposing from some appropriate distribution and accepting/rejecting in the usual fashion.

2. Sample $(\beta^{(m+1)}, \gamma^{(m+1)})$ from full conditional $p(\beta, \gamma|y, d, x, \alpha^{(m+1)})$ using Metropolis-Hastings as follows

   (i) Propose candidate values $(\beta^{(new)}, \gamma^{(new)})$ from some appropriate distribution, q, possibly functionally dependent on the current values $(\beta^{(m)}, \gamma^{(m)})$.

   (ii) Compute for $(\beta^{(new)}, \gamma^{(new)})$

$$r_i^{(new)} = f_{D|X}(d_i|x_i, \beta^{(new)}, \gamma^{(new)}).$$

   (iii) For brevity, let

$$L(\beta, \gamma; \alpha) = f_{Y|D,R}(y|d, r, \alpha)f_{D|X}(d|x, \beta, \gamma)$$

   where the first term depends on β and γ through r. Define Λ as the minimum of 1 and

$$\frac{L(\beta^{(new}\gamma^{(new)}; \alpha^{(m+1)})}{L(\beta^{(m)}, \gamma^{(m)}; \alpha^{(m+1)})} \frac{p(\beta^{(new)}, \gamma^{(new)})}{p(\beta^{(m)}, \gamma^{(m)})} \frac{q(\beta^{(m)}, \gamma^{(m)}|\beta^{(new)}, \gamma^{(new)})}{q(\beta^{(new)}, \gamma^{(new)}|\beta^{(m)}, \gamma^{(m)})}.$$

   where the L terms depend on $r^{(new)}$ and $r^{(m)}$) numerator and denominator respectively.

(iv) With probability Λ, set new values for parameters $(\beta^{(m+1)}, \gamma^{(m+1)})$ equal to $(\beta^{(new)}, \gamma^{(new)})$, else set them equal to $(\beta^{(m)}, \gamma^{(m)})$.

After a sufficient number of iterations of this scheme, the required sample from the joint posterior distribution is obtained.

References

Hirano, K. and Imbens, G. W. (2004). The propensity score with continuous treatments. In Gelman, A. and Meng, X.-L., editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, New York. John Wiley.

Zellner, A. (1983). Applications of Bayesian analysis and econometrics. *The Statistician* **132**, 23–34.