A Bayesian Approach for Computing a Multivariate Mixed-Effects Prediction Model with Design Space Applications

John J. Peterson
Research Statistics Unit
GlaxoSmithKline Pharmaceuticals

BAYES2010 Conference
May 20th, 2010

OUTLINE

- The Mixed-Effects Seemingly Unrelated Regressions (SUR) Model
- Some Bayesian Approaches to Fitting the Mixed Effect-SUR Model
- A Bayesian Analysis of a Mixed Effect-SUR Model for optimizing one stage of an active Pharmaceutical Ingredient (API)
- An ICH Q8 Design Space Application using the above model.
- Summary
- References

The Mixed Effects SUR Model

(SUR = "seemingly unrelated regressions")

- The basic mixed model form: $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{e}$

- Here, $\boldsymbol{\beta}$ is a vector of fixed effects and $\boldsymbol{\gamma}$ is a vector of random effects.

- This is a multivariate regression model with r response types.

- Here, it is typically assumed that:

  $\gamma \sim N(\mathbf{0}, \boldsymbol{G})$ for some variance-covariance matrix, $\boldsymbol{G}$.

  $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ for some variance-covariance matrix, $\boldsymbol{\Sigma}$.

The Mixed Effects SUR Model

(SUR = "seemingly unrelated regressions")

- The basic mixed model form: $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{e}$
- Here, $\boldsymbol{\beta}$ is a vector of fixed effects and $\boldsymbol{\gamma}$ is a vector of random effects.
- This is a multivariate regression model with r response types.
- A somewhat more detailed format can be:

Observation (vector) 1 →

Observation (vector) n →

$$\begin{bmatrix} \begin{bmatrix} y_{11} \\ \vdots \\ y_{1r} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} y_{n1} \\ \vdots \\ y_{nr} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} x'_{11}\beta_1 \\ \vdots \\ x'_{1r}\beta_r \end{bmatrix} \\ \vdots \\ \begin{bmatrix} x'_{n1}\beta_1 \\ \vdots \\ x'_{nr}\beta_r \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} z'_{11}\gamma_{11} \\ \vdots \\ z'_{1r}\gamma_{r1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} z'_{n1}\gamma_{1J} \\ \vdots \\ z'_{nr}\gamma_{rJ} \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} e_{11} \\ \vdots \\ e_{1r} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} e_{n1} \\ \vdots \\ e_{nr} \end{bmatrix} \end{bmatrix}$$

- Here, $\beta' = (\beta'_1, \ldots, \beta'_r)$, but notationally it is more difficult to describe $\boldsymbol{\gamma}$ in detail. This is because random effects can apply across groups of observation vectors, due to the hierarchical nature of many designs involving random effects.

The Mixed Effects SUR Model

(SUR = "seemingly unrelated regressions")

- The basic mixed model form: $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{e}$

Observation (vector) 1

Observation (vector) n

$$\begin{bmatrix} \begin{bmatrix} y_{11} \\ \vdots \\ y_{1r} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} y_{n1} \\ \vdots \\ y_{nr} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} x'_{11}\beta_1 \\ \vdots \\ x'_{1r}\beta_r \end{bmatrix} \\ \vdots \\ \begin{bmatrix} x'_{n1}\beta_1 \\ \vdots \\ x'_{nr}\beta_r \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} z'_{11}\gamma_1 \\ \vdots \\ z'_{1r}\gamma_1 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} z'_{n1}\gamma_J \\ \vdots \\ z'_{nr}\gamma_J \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} e_{11} \\ \vdots \\ e_{1r} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} e_{n1} \\ \vdots \\ e_{nr} \end{bmatrix} \end{bmatrix}$$

- Note that each response type (1,2,..,r) the fixed-effects model can be of a *different* model form (unlike that for the classical multivariate regression model).

- Note also that for each response type, the random-effects part of the model can be of a *different* model form. Notationally , this needs to be achieved though the $\boldsymbol{Z}$ matrix.

The Mixed Effects SUR Model

(SUR = "seemingly unrelated regressions")

- The basic mixed model form: $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{e}$

- Suppose we have a model with two response types, two factors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and some γ batch effects as follows:

$$\text{Model 1: } y_1 = \boldsymbol{x}_1'\beta_1 + \boldsymbol{z}_1'\gamma_1 + e_1 = \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \gamma_{01} + \gamma_{11}x_1 + \gamma_{21}x_2 + e_1$$

$$\text{Model 2: } y_2 = \boldsymbol{x}_2'\beta_2 + \boldsymbol{z}_2'\gamma_2 + e_2 = \beta_{02} + \beta_{12}x_1 + \gamma_{02} + \gamma_{12}x_1 + e_2 \ ,$$

$$\text{where } \boldsymbol{x}_1 = (1, x_1, x_2)', \boldsymbol{x}_2 = (1, x_1)', \boldsymbol{z}_1 = (1, x_1, x_2)', \text{ and } \boldsymbol{z}_2 = (1, x_1)$$

- For the models above, all batch effects are in **blue**.

- Note that we have batch-by-factor interactions , but for response type 2 we cannot have a batch-by-$\boldsymbol{x}_2$ interaction effect because model 2 does not have an $\boldsymbol{x}_2$ factor .

Two Gibbs Sampling Algorithms

(using conditionally conjugate priors)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{e}, \text{ where } \gamma \sim N\left(\boldsymbol{0}, \boldsymbol{G}\right) \text{ and } \boldsymbol{e} \sim N\left(\boldsymbol{0}, \Sigma\right)$$

Conditionally Conjugate Priors:

$$\beta \sim N\left(\boldsymbol{b}, \boldsymbol{B}\right), \boldsymbol{G}^{-1} \sim W\left(\eta, \Gamma\right), \Sigma^{-1} \sim W\left(\nu, \boldsymbol{R}\right)$$

- A useful theorem from Lindley and Smith (1972):

Suppose $\boldsymbol{Y} \sim N\left(A_1\theta_1, \boldsymbol{C}_1\right)$, and $\theta_1 \sim N\left(A_2\theta_2, \boldsymbol{C}_2\right)$, then $\theta_1 \mid \boldsymbol{y} \sim N\left(\boldsymbol{D}\boldsymbol{d}, \boldsymbol{D}\right)$

where $\boldsymbol{D} = \left[A_1'\boldsymbol{C}_1^{-1}A_1 + \boldsymbol{C}_2^{-1}\right]^{-1}$ and $\boldsymbol{d} = A_1'C_1^{-1}\boldsymbol{y} + \boldsymbol{C}_2^{-1}A_2\theta_2$

- This theorem is useful for obtaining posterior distributions for $\boldsymbol{\beta}$ and γ (given $\boldsymbol{y}$ and other parameters).

Two Gibbs Sampling Algorithms

(using conditionally conjugate priors)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{e}, \text{ where } \gamma \sim N\left(\boldsymbol{0}, \boldsymbol{G}\right) \text{ and } \boldsymbol{e} \sim N\left(\boldsymbol{0}, \Sigma\right)$$

Conditionally Conjugate Priors:

$$\beta \sim N\left(\boldsymbol{b}, \boldsymbol{B}\right), \boldsymbol{G}^{-1} \sim W\left(\eta, \Gamma\right), \Sigma^{-1} \sim W\left(\nu, \boldsymbol{R}\right)$$

Algorithm #1:

$$\beta \mid \gamma, \boldsymbol{G}, \Sigma, \text{data}$$

$$\gamma \mid \beta, \boldsymbol{G}, \Sigma, \text{data}$$

$$\boldsymbol{G} \mid \gamma$$

$$\Sigma \mid \beta, \gamma, \text{data}$$

Algorithm #2:

$$\beta, \gamma \mid \boldsymbol{G}, \Sigma, \text{data}$$

$$\begin{cases} \beta \mid \boldsymbol{G}, \Sigma, \text{data} \\ \gamma \mid \beta, \boldsymbol{G}, \Sigma, \text{data} \end{cases}$$

$$\boldsymbol{G} \mid \gamma$$

$$\Sigma \mid \beta, \gamma, \text{data}$$

Two Gibbs Sampling Algorithms

(using conditionally conjugate priors)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{e}, \text{ where } \gamma \sim N\left(\boldsymbol{0}, \boldsymbol{G}\right) \text{and } \boldsymbol{e} \sim N\left(\boldsymbol{0}, \Sigma\right)$$

Conditionally Conjugate Priors:

$$\beta \sim N\left(\boldsymbol{b}, \boldsymbol{B}\right), \boldsymbol{G}^{-1} \sim W\left(\eta, \Gamma\right), \Sigma^{-1} \sim W\left(\nu, \boldsymbol{R}\right)$$

- How do we choose specific values for $\boldsymbol{b}, \boldsymbol{B}, \eta, \Gamma, \nu$, and $\boldsymbol{R}$ for the priors?

- Recommendations from Shafer and Yucel (2002)
  In the absence of informative prior information do the following:

  - Set $\boldsymbol{b}=\boldsymbol{0}$, and $\boldsymbol{B}^{-1}=\boldsymbol{0}$ (i.e. set the precision matrix equal to the zero matrix)

  - Set η = dimension of $\boldsymbol{G}$ and ν = dimension of $\boldsymbol{\Sigma}$.
  - Set $\Gamma = \hat{E}\left(\boldsymbol{G}^{-1}\right) = \eta^{-1}\hat{\boldsymbol{G}}^{-1}$ and $\boldsymbol{R} = \hat{E}\left(\Sigma^{-1}\right) = \nu^{-1}\hat{\Sigma}^{-1}$, where $\hat{\boldsymbol{G}}$ and $\hat{\Sigma}$ are estimates.

- But I have some suspicions about possible biases these priors could induce.

Use of a scaled inverted-Wishart prior

• The Wishart prior for Σ^{-1} *or* $\boldsymbol{G}^{-1}$ can be somewhat restrictive with regard to specification of information about the precisions (or equivalently the variances).

• An approach recommended by Gelman and Hill (2007) is to use a scaled inverted Wishart prior for the variance-covariance matrix).

• Consider $\Sigma = \boldsymbol{DQD}$, where $\boldsymbol{Q}$ is an inverted-Wishart rv with $(r+1)$ df and scale matrix equal to an rxr identity matrix. Here, $\boldsymbol{D}$ is a diagonal matrix of (positive) scaling variables.

• So we have:

$$\boldsymbol{\Sigma = DQD} = \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & & \\ & & \ddots & \\ & & & \xi_r \end{bmatrix} \boldsymbol{Q} \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & & \\ & & \ddots & \\ & & & \xi_r \end{bmatrix} = \begin{bmatrix} \xi_1^2 q_{11} & \xi_1\xi_2 q_{12} & \cdots & \xi_1\xi_r q_{1r} \\ \xi_1\xi_2 q_{12} & \xi_2^2 q_{22} & \cdots & \xi_2\xi_r q_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1\xi_r q_{1r} & \xi_2\xi_r q_{2r} & \cdots & \xi_r^2 q_{rr} \end{bmatrix}$$

Use of a scaled inverted-Wishart prior

- So we have:

$$\boldsymbol{\Sigma} = \boldsymbol{DQD} = \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & & \\ & & \ddots & \\ & & & \xi_r \end{bmatrix} \boldsymbol{Q} \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & & \\ & & \ddots & \\ & & & \xi_r \end{bmatrix} = \begin{bmatrix} \xi_1^2 q_{11} & \xi_1\xi_2 q_{12} & \cdots & \xi_1\xi_r q_{1r} \\ \xi_1\xi_2 q_{12} & \xi_2^2 q_{22} & \cdots & \xi_2\xi_r q_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1\xi_r q_{1r} & \xi_2\xi_r q_{2r} & \cdots & \xi_r^2 q_{rr} \end{bmatrix}$$

- Recall that $\boldsymbol{Q} \sim IW_{(r+1)}(\boldsymbol{I})$. Gelman and Hill (2007) state that the correlation parameters associated with $\boldsymbol{Q}$ have a (marginal) uniform distribution on (-1,1).

- For $\boldsymbol{Q}$, these correlation parameters are: $\dfrac{q_{ij}}{\sqrt{q_{ii}}\sqrt{q_{jj}}}$

- Note also that the correlation parameters for $\boldsymbol{\Sigma}$ are: $\dfrac{\sigma_{ij}^2}{\sigma_i\sigma_j} = \dfrac{\xi_i\xi_j q_{ij}}{\sqrt{\xi_i^2 q_{ii}}\sqrt{\xi_j^2 q_{jj}}} = \dfrac{q_{ij}}{\sqrt{q_{ii}}\sqrt{q_{jj}}}$

- Hence, we can use $\boldsymbol{Q}$ to build a weak prior for the correlation parameters for $\boldsymbol{\Sigma}$.

- However, the diagonal elements of $\boldsymbol{Q}$ alone may be inadequate to model the variances in $\boldsymbol{\Sigma}$.

Use of a scaled inverted-Wishart prior

- So we have:

$$\boldsymbol{\Sigma} = \boldsymbol{DQD} = \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & & \\ & & \ddots & \\ & & & \xi_r \end{bmatrix} \boldsymbol{Q} \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & & \\ & & \ddots & \\ & & & \xi_r \end{bmatrix} = \begin{bmatrix} \xi_1^2 q_{11} & \xi_1\xi_2 q_{12} & \cdots & \xi_1\xi_r q_{1r} \\ \xi_1\xi_2 q_{12} & \xi_2^2 q_{22} & \cdots & \xi_2\xi_r q_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1\xi_r q_{1r} & \xi_2\xi_r q_{2r} & \cdots & \xi_r^2 q_{rr} \end{bmatrix}$$

- Recall that $\boldsymbol{Q} \sim IW_{(r+1)}(\boldsymbol{I})$. Gelman and Hill (2007) state that the correlation parameters associated with $\boldsymbol{Q}$ have a uniform distribution on (-1,1).

- However, the diagonal elements of $\boldsymbol{Q}$ alone are inadequate to model the variances in $\boldsymbol{\Sigma}$.
- But since $\xi_i^2 q_{ii} = \sigma_i^2 \ (i = 1,...,r)$, the ξ_i's can be used to rescale the q_{ii}'s so that we can get better priors for the σ_i^2 terms in $\boldsymbol{\Sigma}$.

Use of a scaled-inverted-Wishart prior

• But since $\xi_i^2 q_{ii} = \sigma_i^2 \ (i = 1,...,r)$, the ξ_i's can be used to rescale the q_{ii}'s so that we can get better priors for the σ_i^2 terms in $\boldsymbol{\Sigma}$.

• Gelman and Hill (2007) recommend uniform or half-*t* priors for the ξ_i terms.

I have used independent uniform priors for each ξ_i

• If we can find an informative upper bound for each σ_i, that may help with setting a prior for the scaling parameter ξ_i.

• Because the correlations do not depend upon the ξ_i's, we can take the ξ_i's to have any positive values.

• We could take $\xi_i = u_i / \sqrt{q_{ii}}$ for a uniform rv u_i which would result in $\sigma_i^2 = \xi_i^2 q_{ii} = u_i^2$.

I have tried with bugs and it crashes. ☹

Fitting a ME-SUR model for an API process

• This is an example for fitting a mixed-effect multivariate model for an early phase synthetic chemistry process. (Acknowledgements: Gregory Stockdale, GSK Stat. Sciences.)

• There are four response types, each modeled by a different (linear) model form.

• The experiment used four factors in a central composite design with 30 runs.
  - Each run represents a batch run of a chemical reaction.
  - Within each batch, three measurements were made of each of the four response types.

• So for each response type, we have a within-batch (measurement error) random effect and a random batch effect.

• So if we denote the within-batch (measurement error) random effect by $\boldsymbol{e}$, we have

$$\boldsymbol{e} = (e_1, e_2, e_3, e_4)', \text{ with } \operatorname{var}(\boldsymbol{e}) = \Sigma.$$

• Likewise, if we denote the random batch effect by γ, we have

$$\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)', \text{ with } \operatorname{var}(\gamma) = \boldsymbol{G}.$$

Fitting a ME-SUR model for an API process

- The factors chosen for the experimental design were:

x_1="temperature", x_2="pressure", x_3="catalyst loading", x_4="reaction-time".

- The response variables measured (and specifications) are:

Y_1="% starting material isomer" (<0.15%), Y_2="% product isomer" (<2.0%), Y_3="% impurity #1"(<3.5%), and Y_4="% API purity"(>95.0%).

- The linear models used for optimization are:

$$Y_1 = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + \beta_6 x_2^2 + \beta_7 x_3^2 + \beta_8 x_4^2 + \gamma_1 + e_1$$
$$Y_2 = \beta_9 + \beta_{10} x_1 + \beta_{11} x_3 + \gamma_2 + e_2$$
$$Y_3 = \beta_{12} + \beta_{13} x_1 + \beta_{14} x_2 + \beta_{15} x_3 + \beta_{16} x_1 x_3 + \gamma_3 + e_3$$
$$Y_4 = \beta_{17} + \beta_{18} x_3 + \gamma_4 + e_4$$

$\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)'$, with $\mathrm{var}(\gamma) = \boldsymbol{G}$. $\boldsymbol{e} = (e_1, e_2, e_3, e_4)'$, with $\mathrm{var}(\boldsymbol{e}) = \boldsymbol{\Sigma}$.

- More compactly we have: $\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij}\beta + \gamma_j + \boldsymbol{e}_{ij}$, where $\boldsymbol{Y}_{ij}$ is the i^{th} measurement within the j^{th} batch.

Fitting a ME-SUR model for an API process

Priors Used:

- $\beta = (\beta_1, ..., \beta_{17})'$ $\beta \sim N(\hat{\beta}, \boldsymbol{B})$, where $\hat{\beta}$ is the MLE and $\boldsymbol{B}^{-1}$ is a diagonal precision matrix with diagonal values of 0.00001 .

- For $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$, I used scaled inverted-Wishart priors.
  - For the ξ_i scaling parameters I used independent uniform distributions with support on $(0, \sigma_i^{upper\,bound})$
  - Here, $\sigma_i^{upper\,bound}$ values were obtained from a 99% (one-sided) Bonferroni-adjusted confidence interval (adjusted for the eight variance parameters).

Software Used:

R and the R package BRugs which calls bugs.

Fitting a ME-SUR model for an API process

Results:

Compared to MLE's from SAS Proc Mixed

MPE="median posterior estimate"

Priors:

- weak beta prior
- **scaled inverse-Wishart priors** for $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$.
- scaling parameters uniform on $\left(0, \sigma_i^{upper\ bound}\right)$

	beta values			G values			Sigma values	
	MLE	MPE		MLE	MPE		MLE	MPE
b1	-7.074	-7.087	G11	0.287	0.279	S11	0.036	0.037
b2	1.214	1.211	G12	0.036	0.041	S12	-0.015	-0.016
b3	-0.262	-0.251	G13	0.0098	0.0090	S13	0.00096	0.00092
b4	-0.430	-0.433	G14	-0.018	-0.014	S14	0.00016	0.00021
b5	-0.200	-0.194	G22	1.957	1.933	S22	0.173	0.174
b6	0.850	0.883	G23	0.094	0.086	S23	0.0017	0.0016
b7	0.741	0.766	G24	-0.127	-0.116	S24	-0.0021	-0.0019
b8	-2.080	-2.117	G33	0.035	0.035	S33	0.0029	0.0029
b9	-6.777	-6.779	G34	-0.038	-0.036	S34	0.0000	0.0000
b10	-1.462	-1.468	G44	0.048	0.048	S44	0.0029	0.0028
b11	-2.610	-2.600						
b12	-3.300	-3.299						
b13	0.201	0.200						
b14	0.041	0.038						
b15	-0.090	-0.095						
b16	3.076	3.076						
b17	0.214	0.219						

Fitting a ME-SUR model for an API process

MPE="median posterior estimate"

ESS ="effective sample size"

Priors:

- weak beta prior
- **scaled inverse-Wishart priors** for $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$.
- scaling parameters uniform on $\left(0, \sigma_i^{upper\,bound}\right)$

	beta values			G values			Sigma values	
	MPE	ESS		MPE	ESS		MPE	ESS
b1	-7.087	896	G11	0.279	877	S11	0.037	1000
b2	1.211	1000	G12	0.041	1000	S12	-0.016	863
b3	-0.251	1000	G13	0.0090	1000	S13	0.00092	1000
b4	-0.433	966	G14	-0.014	1000	S14	0.00021	1000
b5	-0.194	1000	G22	1.933	925	S22	0.174	1000
b6	0.883	952	G23	0.086	1000	S23	0.0016	1000
b7	0.766	962	G24	-0.116	1000	S24	-0.0019	1000
b8	-2.117	1000	G33	0.035	1000	S33	0.0029	1000
b9	-6.779	564	G34	-0.036	1000	S34	0.0000	1000
b10	-1.468	1000	G44	0.048	1000	S44	0.0028	883
b11	-2.600	992						
b12	-3.299	580						
b13	0.200	909						
b14	0.038	659						
b15	-0.095	1000						
b16	3.076	613						
b17	0.219	621						

Quick MCMC Overview:

- Five independent chains used
- 10,000 burin-in iterations per chain
- thinned each chain by 100
- 1,000 total MCMC samples kept
- All Gelman-Rubin stats < 1.1

Fitting a ME-SUR model for an API process

Results:

Compared to MLE's from SAS Proc Mixed

MPE="median posterior estimate"

Priors:

- weak beta prior
- **Schafer-Yucel Wishart priors** for $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$.
- MLE's for $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$ used for the Schafer-Yucel priors.

beta values			G values			Sigma values		
	MLE	MPE		MLE	MPE		MLE	MPE
b1	-7.074	-7.084	G11	0.287	0.271	S11	0.036	0.159
b2	1.214	1.201	G12	0.036	0.049	S12	-0.015	-0.003
b3	-0.262	-0.231	G13	0.0098	0.050	S13	0.00096	-0.043
b4	-0.430	-0.422	G14	-0.018	0.052	S14	0.00016	-0.0060
b5	-0.200	-0.147	G22	1.957	1.900	S22	0.173	0.197
b6	0.850	0.935	G23	0.094	0.089	S23	0.0017	-0.011
b7	0.741	0.788	G24	-0.127	-0.076	S24	-0.0021	0.0098
b8	-2.080	-2.209	G33	0.035	2.082	S33	0.0029	1.276
b9	-6.777	-6.785	G34	-0.038	1.537	S34	0.0000	0.081
b10	-1.462	-1.555	G44	0.048	1.691	S44	0.0029	1.272
b11	-2.610	-2.592						
b12	-3.300	-3.292						
b13	0.201	0.143						
b14	0.041	0.0692						
b15	-0.090	-0.199						
b16	3.076	3.083						
b17	0.214	0.233						

Fitting a ME-SUR model for an API process

MPE="median posterior estimate"

ESS ="effective sample size"

Priors:

- weak beta prior
- Schafer-Yucel Wishart priors for $\boldsymbol{G}$ *and* $\boldsymbol{\Sigma}$.
- MLE's for $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$ used for the Schafer-Yucel priors.

	beta values			G values			Sigma values	
	MPE	ESS		MPE	ESS		MPE	ESS
b1	-7.087	889	G11	0.279	1000	S11	0.037	1000
b2	1.211	937	G12	0.041	1000	S12	-0.016	1000
b3	-0.251	940	G13	0.0090	1000	S13	0.00092	1000
b4	-0.433	1000	G14	-0.014	1000	S14	0.00021	1000
b5	-0.194	1000	G22	1.933	1000	S22	0.174	974
b6	0.883	1000	G23	0.086	932	S23	0.0016	1000
b7	0.766	1000	G24	-0.116	973	S24	-0.0019	1000
b8	-2.117	1000	G33	0.035	957	S33	0.0029	1000
b9	-6.779	1000	G34	-0.036	1000	S34	0.0000	1000
b10	-1.468	988	G44	0.048	1000	S44	0.0028	986
b11	-2.600	1000						
b12	-3.299	1000						
b13	0.200	1000						
b14	0.038	1000						
b15	-0.095	1000						
b16	3.076	1000						
b17	0.219	1000						

Quick MCMC Overview:

- Five independent chains used
- 10,000 burin-in iterations per chain
- thinned each chain by 100
- 1,000 total MCMC samples kept
- All Gelman-Rubin stats < 1.1

ICH Q8 Definition of Design Space

The ICH Q8 FDA Guidance for Industry defines "Design Space" as:

"The multidimensional combination and interaction of input variables (e.g. material attributes) and process parameters that have been demonstrated to provide assurance of quality."

Further more….

"Working within the Design Space is not considered as a change. Movement out of the Design Space is considered to be a change and would normally initiate a post regulatory approval change process. Design Space is proposed by the applicant and is subject to regulatory assessment and approval".

ICH Q8 Definition of Design Space

- The ICH Q8 FDA Guidance for Industry defines "Design Space" as:

"The multidimensional combination and interaction of input variables (e.g. material attributes) and process parameters that have been demonstrated to provide assurance of quality."

- From a Bayesian perspective, one could define the ICH Q8 design space as

$$\left\{ \boldsymbol{x} \in \chi \,|\, \Pr\left(\boldsymbol{Y} \in A \,|\, \boldsymbol{x}, \text{data} \right) \geq R \right\}$$

where $\boldsymbol{x}$ is a vector of process factors, χ is the experimental region, $\boldsymbol{Y}$ is a vector of response-types, A is an acceptance region, and R is a reliability value for "assurance" of quality.

- The above probability measure is based upon the posterior predictive distribution for $\boldsymbol{Y}$ given the process factors.

Application of the Bayesian Predictive Model to a Design Space Computation

(Here, I am using the scaled inverse-Wishart priors)

- For a future batch, we have the model:

$\boldsymbol{Y}_{i,new} = \boldsymbol{X}_{i,new}\beta + \gamma_{new} + \boldsymbol{e}_{i,new}$, where $\boldsymbol{Y}_{i,new}$ is the i^{th} measurement for the new batch.

Here, i =1,2,3.

- For a design space, we may wish to represent our batch response by

$\boldsymbol{Y}^{*}_{new} = \boldsymbol{X}_{new}\beta + \gamma_{new}$, where the measurement error is removed.

- So we could define a Bayesian ICH Q8 design space as:

$$\left\{\boldsymbol{x} \in \chi \,|\, \Pr\left(\boldsymbol{Y}^{*} \in A \,|\, \boldsymbol{x}, \text{data}\right) \geq R\right\}$$

for some experimental region, χ, some acceptance region, A, and reliability level, R.

- Here, $A = (0, 0.0015) \times (0, 0.02) \times (0, 0.035) \times (0.95, 1)$

- This predictive model was fit using a scaled inverse-Wishart prior and normal distributions for the γ's and $\boldsymbol{e}$'s.

Application of the Bayesian Predictive Model to a Design Space Computation

(Here, I am using the scaled inverse-Wishart priors)

- For a design space, we may wish to represent our batch response by $\boldsymbol{Y}^{*}_{new} = \boldsymbol{X}_{new}\beta + \gamma_{new}$, where the measurement error is removed.

- Steps involved:

  1. Sample from the joint posterior of (β, γ).
  2. Compute $\boldsymbol{Y}^{*} = \boldsymbol{X}\beta + \gamma$. (Here $\boldsymbol{X}$ contains the elements of $\boldsymbol{x}$)
  3. Compute $I\left(\boldsymbol{Y}^{*} \in A\right)$
  4. Do Steps 1-3 *N* times (for large *N*, 1,000 say) to compute $\hat{p}\left(\boldsymbol{x}\right)$ as a Monte Carlo estimate of $\Pr\left(\boldsymbol{Y}^{*} \in A \mid \boldsymbol{x}, \text{data}\right)$
  5. Do Steps 1-4 over a grid of $\boldsymbol{x}$'s to map out the design space:

$$\left\{\boldsymbol{x} \in \chi \mid \Pr\left(\boldsymbol{Y}^{*} \in A \mid \boldsymbol{x}, \text{data}\right) \geq R\right\}$$

Here, χ is the experimental region.

Application of the Bayesian Predictive Model to a Design Space Computation

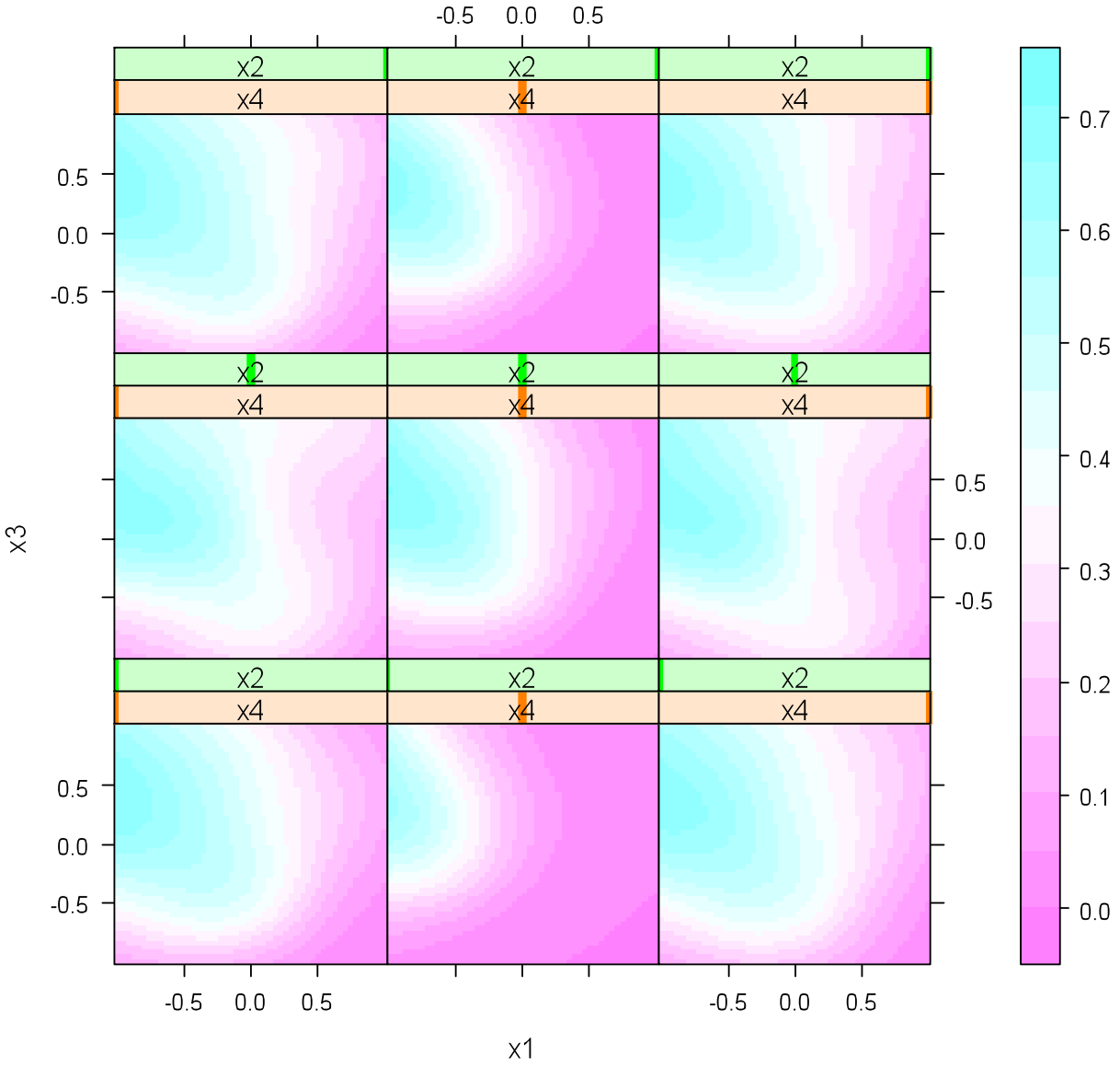
- Do Steps 1-4 over a grid of $\boldsymbol{x}$'s to map out the design space: $\left\{\boldsymbol{x}\in\chi \mid \hat{p}\left(\boldsymbol{x}\right)\geq R\right\}$ where $\hat{p}\left(\boldsymbol{x}\right)$ is the Monte Carlo estimate of $p\left(\boldsymbol{x}\right)=\Pr\left(\boldsymbol{Y}^{*}\in A \mid \boldsymbol{x},\text{data}\right)$
- **Note:** We could get smoother estimates of $\hat{p}\left(\boldsymbol{x}\right)$, and perhaps use a less fine grid over $\boldsymbol{x}$, if we fit a (meta-model) surface to the $\hat{p}\left(\boldsymbol{x}\right)$ values, e.g. using a spline-based regression method. This gives a *closed-form* function, $\tilde{p}\left(\boldsymbol{x}\right)$, for the reliability surface.
- For this study I used a Bayes(!) estimate for $p\left(\boldsymbol{x}\right)$ which is $\hat{p}\left(\boldsymbol{x}\right)=\dfrac{count+1}{N+2}$

where $count$= # of times $\boldsymbol{Y}^{*}$ is in A and N is the number of MCMC values used.

- I then do a logistic transform on the $\hat{p}\left(\boldsymbol{x}\right)$ values and fit them to a generalized additive model using x_1, x_2, x_3, x_4 as the predictor variables.

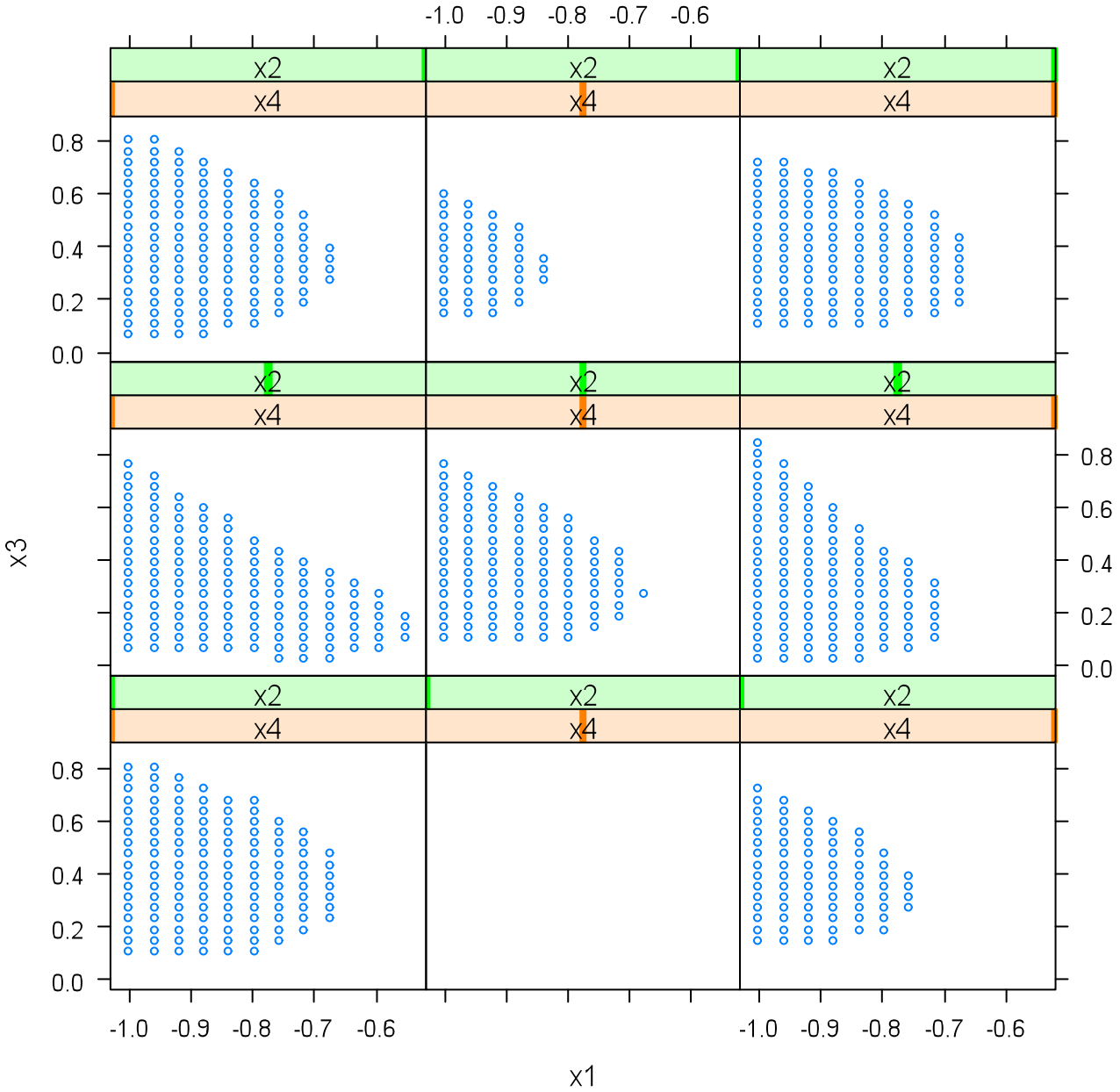
- Back-transforming after fitting produces the reliability surface based upon the fitted meta-model.

Reliability surface from fitted meta-model



Design Space from fitted meta-model

R=0.65



Summary

- Hierarchical models are a very important class of models for many applications: pharmaceutical discovery and development, industrial optimization, natural and social sciences, etc.

- However, statistical inference for multiple-response models with variance component matrices is both statistically and computationally difficult.

- Maximum likelihood estimation often produces estimates for the variance components matrices that are not positive definite.

- The Bayesian approach can avoid this problem, but good priors for the variance component matrices can still be tricky to develop.

Some Bayesian & Design Space References

1. Del Castillo, E. (2007), *Process Optimization: A Statistical Approach*, Springer, New York, NY.

2. Gelman, A. and Hill, J., (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models,* Cambridge University Press, New York, NY.

3. Peterson, J. J. (2008), "A Bayesian Approach to the ICH Q8 Definition of Design Space", *Journal of Biopharmaceutical Statistics*, 18, 959-975.

4. Peterson, J. J. and Lief, K. (2010) "The ICH Q8 Definition of Design Space: A Comparison of the Overlapping Means and the Bayesian Predictive Approaches", *Statistics in Biopharmaceutical Research* (available on-line)

5. Schafer, J. L. and Yucel, R. M. (2002), "Computational Strategies for Multivariate Linear Mixed-Effects Models with Missing Values", *Journal of Computational and Graphical Statistics*", 11(2), 437-457.

6. Stockdale, G. and Cheng, Aili (2009), "Finding Design Space and a Reliable Operating Region using a Multivariate Bayesian Approach with Experimental Design", *Quality Technology and Quantitative Management* **6**(4),391-408 .