Reversible Jump Markov Chain Monte Carlo (RJMCMC)

Tracy Holsclaw

Spring 2008

Overview:

-RJMCMC can be used in problems with the dimensions of the problem change

-model comparison, model averaging, variable selection, and multiple change point problems to name a few

- each model must have a probability, prior, and likelihood associated with it

-there must be a countable number of models

-RJMCMC must satisfy aperiodic, irreducible, and satisfy the detailed balance criteria

-Gibbs steps are not usually possible

-All normalizing constants should be retained

-Proper priors should be used

Auxiliary variables in RJMCMC:

When moving between dimensions of different sizes it is useful to use auxiliary variables to match or make dimensions the same size.

It is easiest to match dimensions when only moving at most one dimension at a time. Algorithms with birth and deaths of steps work well with this idea.

Example: Model Choice- Two Regression Models

M1: a basic linear regression with one independent variable: $(1,x_1)$, the regression coefficients for this model will be Betas.

M2: a linear regression with two independent variables: $(2, z_1, z_2)$, the regression coefficients for this model with be Etas.

Informative priors:

$$p(\boldsymbol{\beta}),\, p(\boldsymbol{\eta}) \sim N(0,\, 10)$$

$$p(\sigma^2) \sim Unif\,(0,2)$$

Model probabilities: favor the smaller model

r_{11} = 0.6, r_{12} =1 - r_{11}= 0.6, r_{22} = 0.6, and r_{21} = 1 - r_{22} = 0.6

a) in M1 and stay in M1, r_{11}=0.6

$$T_{11}(\beta_0, \beta_1, \sigma_1^2) = (\beta_0, \beta_1, \sigma_1^2)$$ is the identity transformation;

therefore $\left|\frac{\partial T_{11}(\beta_0, \beta_1, \sigma_1^2)}{\partial(\beta_0, \beta_1, \sigma_1^2)}\right|$ =1.

$$\alpha_{MH} = \frac{p(y \mid \beta_0^*, \beta_1^*, \sigma_1^{2*})\, p(\sigma_1^{2*})\, p(\beta_0^*)\ p(\beta_1^*)\ (r_{11}) / j(\sigma_1^{2*})}{p(y \mid \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2)\, p(\sigma_{1,i-1}^2)\, p(\beta_{0,i-1})\ p(\beta_{1,i-1})\ (r_{11}) / j(\sigma_{1,i-1}^2)} \left|\frac{\partial T_{11}(\beta_0, \beta_1, \sigma_1^2)}{\partial(\beta_0, \beta_1, \sigma_1^2)}\right|$$

$$= \frac{p(y \mid \beta_0^*, \beta_1^*, \sigma_1^{2*})\ p(\sigma_1^{2*})\, p(\beta_0^*)\ p(\beta_1^*) / j(\sigma_1^{2*})}{p(y \mid \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2)\ p(\sigma_{1,i-1}^2)\, p(\beta_{0,i-1})\ p(\beta_{1,i-1}) / j(\sigma_{1,i-1}^2)}$$

$$j(\sigma^2) = Unif\left(\frac{\sigma^2}{a}, a\sigma^2\right)$$ σ^{2*} is drawn from a non-symmetric proposal

b) in M2 and stay in M2, r_{22}=0.4

$T_{22}(\eta_0,\eta_1,\eta_2,\sigma_2^2)=(\eta_0,\eta_1,\eta_2,\sigma_2^2)$ is the identity transform and

therefore $\left|\dfrac{\partial T_{22}(\eta_0,\eta_1,\eta_2,\sigma_2^2)}{\partial(\eta_0,\eta_1,\eta_2,\sigma_2^2)}\right|=1$

$$\alpha_{MH}=\frac{p(y\,|\,\eta_0^*,\eta_1^*,\eta_2^*,\sigma_2^{2*})\ p(\sigma_2^{2*})\,p(\eta_0^*)\ p(\eta_1^*)\,p(\eta_2^*)\,\left(r_{22}\right)/\,j(\sigma_2^{2*})}{p(y\,|\,\eta_{0,i-1},\eta_{1,i-1},\eta_{2,i-1},\sigma_{2,i-1}^2)\,p(\sigma_{2,i-1}^2)\,p(\eta_{0,i-1})\ p(\eta_{1,i-1})\,p(\eta_{2,i-1})\,\left(r_{22}\right)/\,j(\sigma_{2,i-1}^2)}\left|\frac{\partial T_{22}(\eta_0,\eta_1,\eta_2,\sigma_2^2)}{\partial(\eta_0,\eta_1,\eta_2,\sigma_2^2)}\right|$$

$$=\frac{p(y\,|\,\eta_0^*,\eta_1^*,\eta_2^*,\sigma_2^{2*})\,p(\sigma_2^{2*})\,p(\eta_0^*)\ p(\eta_1^*)\,p(\eta_2^*)\,/\,j(\sigma_2^{2*})}{p(y\,|\,\eta_{0,i-1},\eta_{1,i-1},\eta_{2,i-1},\sigma_{2,i-1}^2)\ p(\sigma_{2,i-1}^2)\,p(\eta_{0,i-1})\ p(\eta_{1,i-1})\,p(\eta_{2,i-1})\,/\,j(\sigma_{2,i-1}^2)}$$

c) Move from Model 1 to Model 2

-have to switch from two dimensions Betas to three dimension Eta space

-So we will draw an auxiliary variable so that $u \sim N(\eta_{3,i}, .5)$ $\qquad g(u) = \Phi(u)$

-taking the i-1 term from the Etas and using them to propose new Betas has been a problem. Something I tried was: (the sigmas are very similar from one model to the other)

$$\eta_0^* = \beta_{0,i-1} + u - \eta_{3,i-1} + 2 \qquad \eta_1^* = \beta_{1,i-1} + u - \eta_{3,i-1} + .1 \qquad \eta_2^* = u \qquad \sigma_2^{2*} = \sigma_{1,i-1}^2$$

(I am not sure if these are symmetric proposals or not…but for now I am not taking that into account in my acceptance probability)

$$\frac{\partial T_{12}(\sigma_1^2, \beta_0, \beta_1, u)}{\partial(\sigma_1^2, \beta_0, \beta_1, u)} = \begin{vmatrix} \frac{\partial T_{12}(\sigma_1^2)}{\partial \sigma_1^2} & \frac{\partial T_{12}(\sigma_1^2)}{\partial \beta_0} & \frac{\partial T_{12}(\sigma_1^2)}{\partial \beta_1} & \frac{\partial T_{12}(\sigma_1^2)}{\partial u} \\ \frac{\partial T_{12}(\beta_0)}{\partial \sigma_1^2} & \frac{\partial T_{12}(\beta_0)}{\partial \beta_0} & \frac{\partial T_{12}(\beta_0)}{\partial \beta_1} & \frac{\partial T_{12}(\beta_0)}{\partial u} \\ \frac{\partial T_{12}(\beta_1)}{\partial \sigma_1^2} & \frac{\partial T_{12}(\beta_1)}{\partial \beta_0} & \frac{\partial T_{12}(\beta_1)}{\partial \beta_1} & \frac{\partial T_{12}(\beta_1)}{\partial u} \\ \frac{\partial T_{12}(u)}{\partial \sigma_1^2} & \frac{\partial T_{12}(u)}{\partial \beta_0} & \frac{\partial T_{12}(u)}{\partial \beta_1} & \frac{\partial T_{12}(u)}{\partial u} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1$$

$$\alpha_{MH} = \frac{p(y \mid \eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*})\ p(\sigma_2^{2*})\, p(\eta_0^*)\ p(\eta_1^*)\, p(\eta_2^*)\ r_{21}}{p(y \mid \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2)\, p(\sigma_{1,i-1}^2)\, p(\beta_{0,i-1})\ p(\beta_{1,i-1})\ r_{12}\, g(u)} \left| \frac{\partial T_{12}(\sigma_1^2, \beta_0, \beta_1, u)}{\partial(\sigma_1^2, \beta_0, \beta_1, u)} \right|$$

$$= \frac{p(y \mid \eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*})\ p(\sigma_2^{2*})p(\eta_0^*)\ p(\eta_1^*)\, p(\eta_2^*)\ 0.6}{p(y \mid \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2)\ p(\sigma_{1,i-1}^2)p(\beta_{0,i-1})\ p(\beta_{1,i-1})\ 0.4\ \Phi(u)}$$

d) In Model 2 and move to Model 1

Going from a three dimension model to a two dimension model has been just as big a pain to get the new proposed values. (so there are still some issues to work out here)

$$\beta_0^* = c_1\eta_{0,i-1} \qquad \beta_1^* = c_2\eta_{1,i-1} \qquad \sigma_1^{2*} = \sigma_{2,i-1}^2$$

$$\left|\frac{\partial T_{21}(\sigma_2^2,\eta_0,\eta_1,\eta_2)}{\partial(\sigma_2^2,\eta_0,\eta_1,\eta_2)}\right| = \begin{vmatrix} \frac{\partial T_{21}(\sigma_2^2)}{\partial\sigma_2^2} & \frac{\partial T_{21}(\sigma_2^2)}{\partial\eta_0} & \frac{\partial T_{21}(\sigma_2^2)}{\partial\eta_1} & \frac{\partial T_{21}(\sigma_2^2)}{\partial\eta_2} \\ \frac{\partial T_{21}(c_1\eta_0)}{\partial\sigma_2^2} & \frac{\partial T_{21}(c_1\eta_0)}{\partial\eta_0} & \frac{\partial T_{21}(c_1\eta_0)}{\partial\eta_1} & \frac{\partial T_{21}(c_1\eta_0)}{\partial\eta_2} \\ \frac{\partial T_{21}(c_2\eta_1)}{\partial\sigma_2^2} & \frac{\partial T_{21}(c_2\eta_1)}{\partial\eta_0} & \frac{\partial T_{21}(c_2\eta_1)}{\partial\eta_1} & \frac{\partial T_{21}(c_2\eta_1)}{\partial\eta_2} \\ \frac{\partial T_{21}(\eta_2)}{\partial\sigma_2^2} & \frac{\partial T_{21}(\eta_2)}{\partial\eta_0} & \frac{\partial T_{21}(\eta_2)}{\partial\eta_1} & \frac{\partial T_{21}(\eta_2)}{\partial\eta_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 \\ 0 & 0 & c_2 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix} = 0$$

$$\alpha_{MH} = \frac{p(y \mid \beta_0^*,\beta_1^*,\sigma_1^{2*})\,p(\sigma_1^{2*})\,p(\beta_0^*)\ p(\beta_1^*)\ r_{12}}{p(y \mid \eta_{0,i-1},\eta_{1,i-1},\eta_{2,i-1},\sigma_{2,i-1}^2)\ p(\sigma_{2,i-1}^2)\,p(\eta_{0,i-1})\ p(\eta_{1,i-1})\,p(\eta_{2,i-1})\,r_{21}}\left|\frac{\partial T_{21}(\sigma_2^2,\eta_0,\eta_1,\eta_2)}{\partial(\sigma_2^2,\eta_0,\eta_1,\eta_2)}\right|$$

$$= \frac{p(y \mid \beta_0^*,\beta_1^*,\sigma_1^{2*})\,p(\sigma_1^{2*})\,p(\beta_0^*)\ p(\beta_1^*)\ 0.4}{p(y \mid \eta_{0,i-1},\eta_{1,i-1},\eta_{2,i-1},\sigma_{2,i-1}^2)\,p(\sigma_{2,i-1}^2)\,p(\eta_{0,i-1})\ p(\eta_{1,i-1})\,p(\eta_{2,i-1})\,0.6}\,0$$

Big Picture:

a) From M1 to M1

propose $\beta_0^*, \beta_1^*, \text{ and } \sigma_1^{2*}$ $\text{ from } \beta_{0,i-1}, \beta_{1,i-1}, \text{and } \sigma_{1,i-1}^2$

If accept – keep $\beta_0^*, \beta_1^*, \text{ and } \sigma_1^{2*}$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

If reject –keep $\beta_{0,i-1}, \beta_{1,i-1}, \text{and } \sigma_{1,i-1}^2$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

b) From M2 to M2

propose $\eta_0^*, \eta_1^*, \eta_2^*, \text{ and } \sigma_2^{2*}$ $\text{ from } \eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \text{and } \sigma_{2,i-1}^2$

If accept – keep $\eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*}$ then we can randomly move to b) $r_{22} = 0.4$ or d) $r_{21} = 0.6$

If reject –keep $\eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \sigma_{2,i-1}^2$ then we can randomly move to b) $r_{22} = 0.4$ or d) $r_{21} = 0.6$

c) From M1 to M2

propose $\eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*}$ $\text{ from } \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2, u$

If accept – keep $\eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*}$ then we can randomly move to a) $r_{21} = 0.6$ or c) $r_{22} = 0.4$

If reject –keep $\beta_{0,i-1}, \beta_{1,i-1}, \text{and } \sigma_{1,i-1}^2$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

d) From M2 to M1

propose $\beta_0^*, \beta_1^*, \text{ and } \sigma_1^{2*}$ $\text{ from } \eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \text{and } \sigma_{2,i-1}^2$

If accept – keep $\beta_0^*, \beta_1^*, \text{ and } \sigma_1^{2*}$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

If reject –keep $\eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \sigma_{2,i-1}^2$ then we can randomly move to a) $r_{21} = 0.6$ or c) $r_{22} = 0.4$

Results:
Model 1 – Quality~Flavor
Model 2 – Quality~Flavor + Oakiness

Both models should fit well but I used r=0.6 for M1, the smaller model

a)	Acceptance rate= 0.2304	M1	pass through a) 5832 times
b)	Acceptance rate= 0.1875	M2	pass through b) 96 times
c)	Acceptance rate= 0.0258	M2	pass through c) 3881 times
d)	Acceptance rate= 0.5316	M1	pass through d) 190 times

Overall: M1 is chosen 97.15% of the time and M2 2.86%
(as a cross check with regular regression we would say both models are significant but Oakiness is barely worth adding to the model)

M1-

β_0= 4.3337

β_1= 1.6933

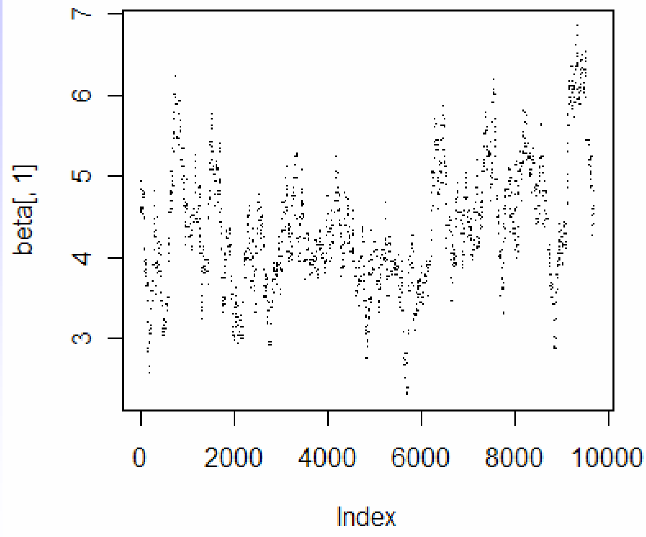
η_0= 5.9238

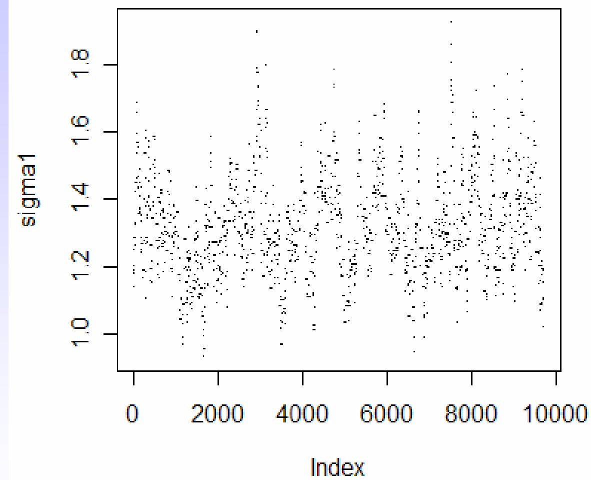
η_1= 1.7355

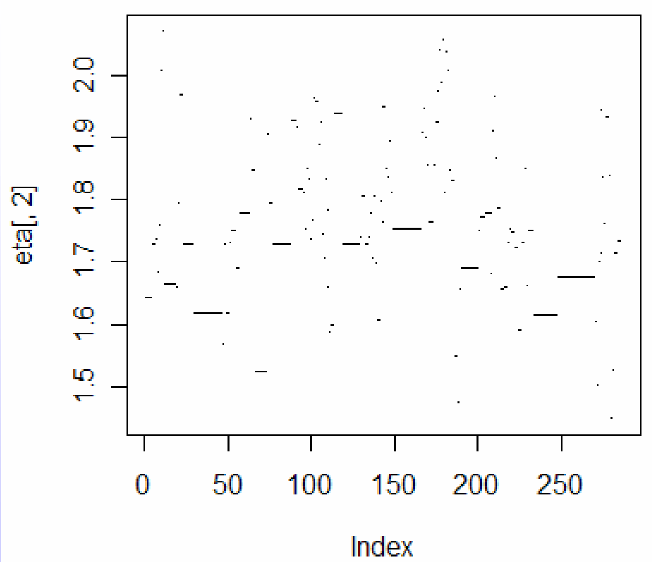
η_2 = -0.4271

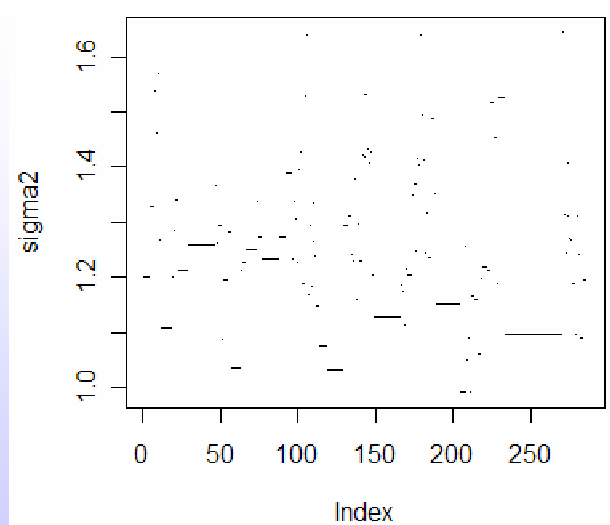
σ^2_1=1.3152

σ^2_2=1.2140









Slide 9 c) Move from Model 1 to Model 2

-have to switch from two dimensions Betas to three dimension Eta space

-So we will draw an auxiliary variable so that $u \sim N(\eta_{3,i}, 0.05)$ $g(u) = \Phi(u)$

-taking the i-1 term from the Etas and using them to propose new Betas has been a problem. So I just drew new Etas from their previous chain.

$$\eta_0^* \sim N(\eta_{0,i-1}, 0.1) \qquad \eta_1^* = N(\eta_{1,i-1}, 0.05) \qquad \eta_2^* = u \qquad \sigma_2^{2*} \sim Unif(\sigma_{2,i-1}^2 / 1.1, 1.1\sigma_{2,i-1}^2)$$

$$\frac{\partial T_{12}(\sigma_1^2, \beta_0, \beta_1, u)}{\partial(\sigma_1^2, \beta_0, \beta_1, u)} = unity = 1$$

$$\alpha_{MH} = \frac{p(y \mid \eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*})\ p(\sigma_2^{2*}) p(\eta_0^*)\ p(\eta_1^*) p(\eta_2^*)\ r_{21} / j(\sigma_2^{2*})}{p(y \mid \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2)\ p(\sigma_{1,i-1}^2)\ p(\beta_{0,i-1})\ p(\beta_{1,i-1})\ r_{12} g(u)} \left| \frac{\partial T_{12}(\sigma_1^2, \beta_0, \beta_1, u)}{\partial(\sigma_1^2, \beta_0, \beta_1, u)} \right|$$

$$= \frac{p(y \mid \eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*})\ p(\sigma_2^{2*}) p(\eta_0^*)\ p(\eta_1^*) p(\eta_2^*)\ 0.6 / j(\sigma_2^{2*})}{p(y \mid \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2)\ p(\sigma_{1,i-1}^2) p(\beta_{0,i-1})\ p(\beta_{1,i-1})\ 0.4\, \Phi(u)}$$

d) In Model 2 and move to Model 1
Going from a three dimension model to a two dimension model has been just as big a pain to get the new proposed values. (so there are still some issues to work out here)

$$\beta_0^* = N(\beta_{0,i-1}, 0.2) \qquad \beta_1^* = N(\beta_{1,i-1}, 0.2) \qquad \sigma_1^{2*} = \sigma_{2,i-1}^2$$

$$\left| \frac{\partial T_{21}(\sigma_2^2, \eta_0, \eta_1, \eta_2)}{\partial(\sigma_2^2, \eta_0, \eta_1, \eta_2)} \right| = unity = 1$$

$$\alpha_{MH} = \frac{p(y \mid \beta_0^*, \beta_1^*, \sigma_1^{2*})\, p(\sigma_1^{2*})\, p(\beta_0^*)\ p(\beta_1^*)\ r_{12}}{p(y \mid \eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \sigma_{2,i-1}^2)\ p(\sigma_{2,i-1}^2)\, p(\eta_{0,i-1})\ p(\eta_{1,i-1})\, p(\eta_{2,i-1})\, r_{21}} \left| \frac{\partial T_{21}(\sigma_2^2, \eta_0, \eta_1, \eta_2)}{\partial(\sigma_2^2, \eta_0, \eta_1, \eta_2)} \right|$$

$$= \frac{p(y \mid \beta_0^*, \beta_1^*, \sigma_1^{2*})\, p(\sigma_1^{2*})\, p(\beta_0^*)\ p(\beta_1^*)\ 0.4}{p(y \mid \eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \sigma_{2,i-1}^2)\, p(\sigma_{2,i-1}^2)\, p(\eta_{0,i-1})\ p(\eta_{1,i-1})\, p(\eta_{2,i-1})\, 0.6}$$

Big Picture:

a) From M1 to M1

propose $\beta_0^*, \beta_1^*, \textit{ and } \sigma_1^{2*} \textit{ from } \beta_{0,i-1}, \beta_{1,i-1}, \textit{and } \sigma_{1,i-1}^2$

If accept – keep $\beta_0^*, \beta_1^*, \textit{ and } \sigma_1^{2*}$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

If reject –keep $\beta_{0,i-1}, \beta_{1,i-1}, \textit{and } \sigma_{1,i-1}^2$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

b) From M2 to M2

propose $\eta_0^*, \eta_1^*, \eta_2^*, \textit{ and } \sigma_2^{2*} \textit{ from } \eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \textit{and } \sigma_{2,i-1}^2$

If accept – keep $\eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*}$ then we can randomly move to b) $r_{22} = 0.4$ or d) $r_{21} = 0.6$

If reject –keep $\eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \sigma_{2,i-1}^2$ then we can randomly move to b) $r_{22} = 0.4$ or d) $r_{21} = 0.6$

c) From M1 to M2

propose $\eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*} \textit{ from } \beta_{0,i-1}, \beta_{1,i-1}, \sigma_{1,i-1}^2, u$

If accept – keep $\eta_0^*, \eta_1^*, \eta_2^*, \sigma_2^{2*}$ then we can randomly move to a) $r_{21} = 0.6$ or c) $r_{22} = 0.4$

If reject –keep $\beta_{0,i-1}, \beta_{1,i-1}, \textit{and } \sigma_{1,i-1}^2$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

d) From M2 to M1

propose $\beta_0^*, \beta_1^*, \textit{ and } \sigma_1^{2*} \textit{ from } \eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \textit{and } \sigma_{2,i-1}^2$

If accept – keep $\beta_0^*, \beta_1^*, \textit{ and } \sigma_1^{2*}$ then we can randomly move to a) $r_{11} = 0.6$ or c) $r_{12} = 0.4$

If reject –keep $\eta_{0,i-1}, \eta_{1,i-1}, \eta_{2,i-1}, \sigma_{2,i-1}^2$ then we can randomly move to a) $r_{21} = 0.6$ or c) $r_{22} = 0.4$

Results:

Model 1 – Quality~Flavor

Model 2 – Quality~Flavor + Oakiness

Both models should fit well but I used r=0.6 for M1, the smaller model

a)	Acceptance rate=	0.224	M1	pass through a) 8312 times
b)	Acceptance rate=	0.189	M2	pass through b) 2433 times
c)	Acceptance rate=	0.171	M2	pass through c) 5603 times
d)	Acceptance rate=	0.263	M1	pass through d) 3650 times

Overall: M1 is chosen 69.6% of the time and M2 30.4%

(as a cross check with regular regression we would say both models are significant but Oakiness is barely worth adding to the model)

M1-

β_0= 4.62

β_1= 1.63

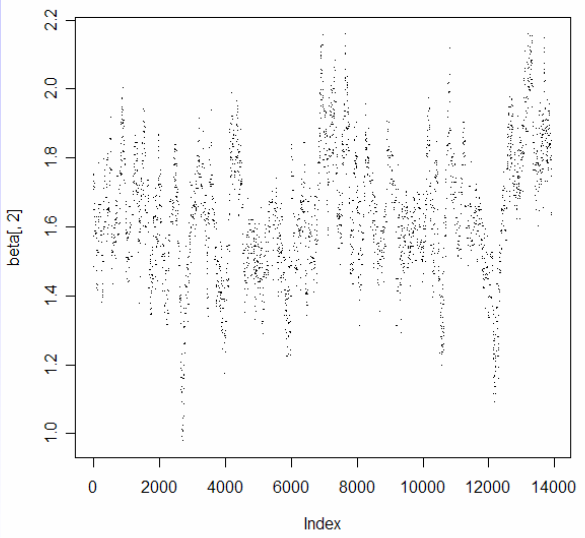
η_0= 6.74

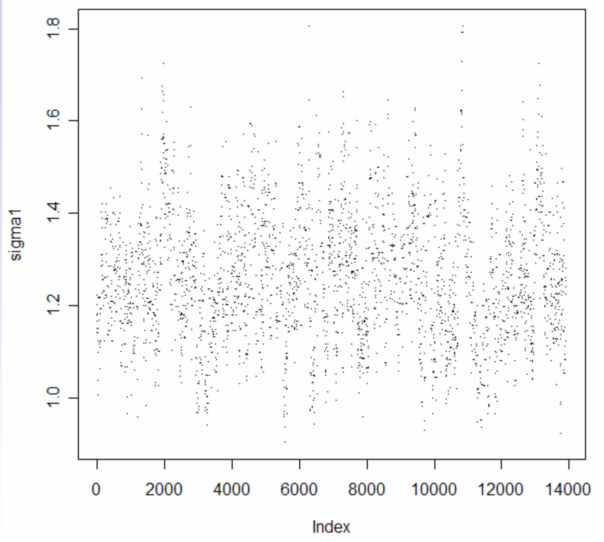
η_1= 1.66

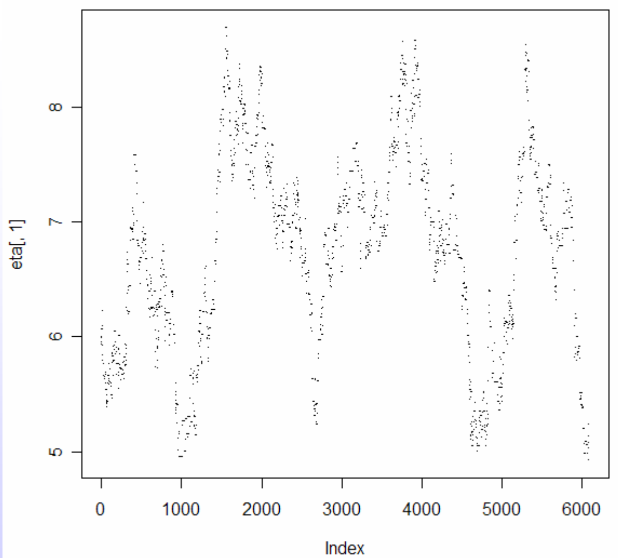
η_2 = -0.52

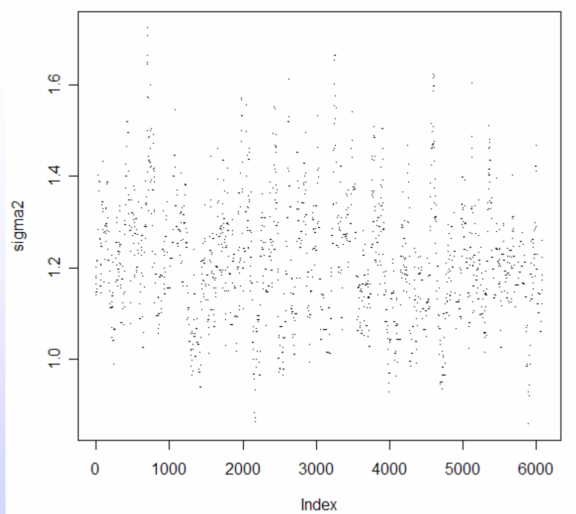
σ^2_1=1.25

σ^2_2=1.20









-It would be good to go on and compare more models than just two.

-This can be done with birth and death of dimensions

-Basically, one can add and remove variables from the regression and see how it would effect the result.

-One example would be to add higher order terms to the regression (or remove them on a death step) and see which order polynomial fit the best.