

Web-based Supplementary Materials for “Bayesian Model Selection For Incomplete Data using the Posterior Predictive Distribution” by Michael J. Daniels, Arkendu S. Chatterjee, and Chenguang Wang

## 1 Web Appendix A

### Derivation of PPL for $h(\mathbf{y})$ with full data

Consider the loss function

$$\mathcal{L}\{h(\mathbf{y}_{\text{rep}}), a; \mathbf{y}\} = L\{h(\mathbf{y}_{\text{rep}}), a\} + kL\{h(\mathbf{y}), a\}$$

When  $L\{\cdot\}$  is squared error loss, the posterior predictive expectation of  $\mathcal{L}$  minimized with respect to the action,  $a$ , is

$$\begin{aligned} C_k(h) &= \sum_i^n \min_{a_i} \mathbb{E} \left[ L\{h(\mathbf{y}_{\text{rep},i}), a_i\} + kL\{h(\mathbf{y}_i), a_i\} \right] \\ &= \sum_i^n \min_{a_i} \mathbb{E} \left[ \{h(\mathbf{y}_{\text{rep},i}) - a_i\}^2 + k\{h(\mathbf{y}_i) - a_i\}^2 \mid \mathbf{y} \right] \\ &= \sum_{i=1}^n \text{Var}\{h(\mathbf{y}_{\text{rep},i}) \mid \mathbf{y}\} + \sum_{i=1}^n \min_{a_i} \left[ \mathbb{E}\{h(\mathbf{y}_{\text{rep},i}) \mid \mathbf{y}\} - a_i\}^2 + k\{h(\mathbf{y}_i) - a_i\}^2 \right] \end{aligned} \quad (1)$$

The value of  $a_i$  that minimizes (1) is

$$a_i = \frac{1}{1+k} \left[ \mathbb{E}\{h(\mathbf{y}_{\text{rep},i}) \mid \mathbf{y}\} + kh(\mathbf{y}_i) \right].$$

### Derivation of PPL for $h(\mathbf{y})$ with incomplete data

Consider the loss function

$$\mathcal{L}\{h(\mathbf{y}_{\text{rep}}), a; h(\mathbf{y})\} = L\{h(\mathbf{y}_{\text{rep}}); a\} + kL\{h(\mathbf{y}); a\}$$

When  $L\{\cdot\}$  is squared error loss, the posterior predictive expectation of  $\mathcal{L}$  minimized with respect to the action,  $a$ , is

$$\begin{aligned} C_k(h) &= \sum_{i=1}^n \min_{a_i} \left[ \mathbb{E}\{(h(\mathbf{y}_{\text{rep},i}) - a_i)^2 \mid \mathbf{y}_{\text{obs}}, \mathbf{r}\} + k \mathbb{E}\{(h(\mathbf{y}_i) - a_i)^2 \mid \mathbf{y}_{\text{obs}}, \mathbf{r}\} \right] \\ &= \sum_{i=1}^n \text{Var}\{h(\mathbf{y}_{\text{rep},i}) \mid \mathbf{y}_{\text{obs}}, \mathbf{r}\} + k \sum_{i=1}^n \text{Var}\{h(\mathbf{y}_i) \mid \mathbf{y}_{\text{obs}}, \mathbf{r}\} \\ &\quad + \sum_{i=1}^n \min_{a_i} \left[ \{E(h(\mathbf{y}_{\text{rep},i}) \mid \mathbf{y}_{\text{obs}}) - a_i\}^2 + k\{E(h(\mathbf{y}_i) \mid \mathbf{y}_{\text{obs}}) - a_i\}^2 \right]. \end{aligned} \quad (2)$$

The value of  $a_i$  that minimizes (2) is

$$a_i = \frac{1}{1+k} \left[ \mathbb{E}\{h(\mathbf{y}_{\text{rep},i})|\mathbf{y}_{\text{obs}}, \mathbf{r}\} + k \mathbb{E}\{h(\mathbf{y}_i)|\mathbf{y}_{\text{obs}}, \mathbf{r}\} \right].$$

After substituting the value of  $a$ , we obtain

$$\begin{aligned} C_h(k) = & \sum_{i=1}^n \text{Var}\{h(\mathbf{y}_{\text{rep},i})|\mathbf{y}_{\text{obs}}, \mathbf{r}\} + k \sum_{i=1}^n \text{Var}\{h(\mathbf{y}_i)|\mathbf{y}_{\text{obs}}, \mathbf{r}\} \\ & + \frac{k}{1+k} \sum_{i=1}^n \left[ \mathbb{E}\{h(\mathbf{y}_i)|\mathbf{y}_{\text{obs}}, \mathbf{r}\} - \mathbb{E}\{h(\mathbf{y}_{\text{rep},i})|\mathbf{y}_{\text{obs}}, \mathbf{r}\} \right]^2. \end{aligned}$$

### Proof of Theorem I:

To prove Theorem I, we will examine each of the three terms individually. First we note that the condition that  $p(\omega)$  and  $p(\mathbf{y}_{\text{obs}}, \mathbf{r}; \omega_O)$  are the same ensures  $p(\omega|\mathbf{y}_{\text{obs}}, \mathbf{r})$  is the same for both models. For clarity and conciseness in the below, we let  $o = \text{obs}$  and  $m = \text{mis}$ .

*Term 3:* Clearly, under the conditions in the Theorem,  $\mathbb{E}\{h(\mathbf{y})|\mathbf{y}_O, \mathbf{r}\}$  is the same for a degenerate and non-degenerate model. We now show that  $\mathbb{E}\{h(\mathbf{y}_{\text{rep}})|\mathbf{y}_{\text{obs}}, \mathbf{r}\}$  is equal as well.

$$\begin{aligned} \mathbb{E}\{h(\mathbf{y}_{\text{rep}})|\mathbf{y}_O, \mathbf{r}\} &= \int h(\mathbf{y}_{\text{rep}})p(\mathbf{y}_{\text{rep}}|\mathbf{y}_O, \mathbf{r})d\mathbf{y}_{\text{rep}} \\ &= \int \int \int \int h(\mathbf{y}_{\text{rep}})p(\mathbf{y}_{\text{rep},m}|\mathbf{y}_{\text{rep},o}, \mathbf{r}_{\text{rep}}, \omega)p(\mathbf{y}_{\text{rep},o}|\mathbf{r}_{\text{rep}}, \omega)p(\mathbf{r}_{\text{rep}}|\omega) \\ &\quad p(\omega|\mathbf{y}_O, \mathbf{r})d\omega d\mathbf{r}_{\text{rep}}d\mathbf{y}_{\text{rep},o}d\mathbf{y}_{\text{rep},m} \\ &= \int \int \int \mathbb{E}\{h(\mathbf{y}_{\text{rep}})|\mathbf{y}_{\text{rep},o}, \mathbf{r}_{\text{rep}}, \omega\}p(\mathbf{y}_{\text{rep},o}|\mathbf{r}_{\text{rep}}, \omega)p(\mathbf{r}_{\text{rep}}|\omega) \\ &\quad p(\omega|\mathbf{y}_O, \mathbf{r})d\omega d\mathbf{r}_{\text{rep}}d\mathbf{y}_{\text{rep},o} \\ &= \int \int \int \mathbb{E}\left\{h(\mathbf{y}_{\text{rep},o}, \mathbb{E}(\mathbf{y}_{\text{rep},m}|\mathbf{y}_{\text{rep},o}, \mathbf{r}_{\text{rep}}, \omega))|\mathbf{y}_{\text{rep},o}, \mathbf{r}_{\text{rep}}, \omega\right\}p(\mathbf{y}_{\text{rep},o}|\mathbf{r}_{\text{rep}}, \omega) \\ &\quad p(\mathbf{r}_{\text{rep}}|\omega)p(\omega|\mathbf{y}_O, \mathbf{r})d\omega d\mathbf{r}_{\text{rep}}d\mathbf{y}_{\text{rep},o}. \end{aligned}$$

The conditional expectation,  $\mathbb{E}(\mathbf{y}_{\text{rep},m}|\mathbf{y}_{\text{rep},o}, \mathbf{r}_{\text{rep}}, \omega)$  is the same under the degenerate and non-degenerate models and the other terms in the integrand are the same under both models.

*Term 1:*

Given that we showed for Term 3 that  $\mathbb{E}\{h(\mathbf{y}_{\text{rep}})|\mathbf{y}_{\text{obs}}, \mathbf{r}\}$  is the same for both cases, we just need to show that  $\mathbb{E}\{h^2(\mathbf{y}_{\text{rep}})|\mathbf{y}_{\text{obs}}, \mathbf{r}\}$  is the same or larger in the non-degenerate (ND) model as the degenerate model. Using a similar development to Term 3, for the degenerate (D) model we have the following expectation (taken with respect to  $p(\mathbf{y}_{\text{rep},m}|\mathbf{y}_{\text{rep},o}, \mathbf{r}_{\text{rep}}, \omega)$ ),

$$\begin{aligned}
E_D\{h^2(\mathbf{y}_{\text{rep}})|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\} &= E_D\{h^2(\mathbf{y}_{\text{rep,o}}, \mathbf{y}_{\text{rep,m}})|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\} \\
&= E_D\left\{h^2(\mathbf{y}_{\text{rep,o}}, E(\mathbf{y}_{\text{rep,m}}|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega))|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\right\} \\
&= \text{Var}_D\left\{h(\mathbf{y}_{\text{rep,o}}, E(\mathbf{y}_{\text{rep,m}}|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega))|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\right\} \\
&\quad + \left[E_D\left\{h(\mathbf{y}_{\text{rep,o}}, E(\mathbf{y}_{\text{rep,m}}|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega))|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\right\}\right]^2 \\
&= 0 + \left[E_D\left\{h(\mathbf{y}_{\text{rep,o}}, E(\mathbf{y}_{\text{rep,m}}|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega))|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\right\}\right]^2 \\
&\leq \text{Var}_{ND}\{h(\mathbf{y}_{\text{rep,o}}, \mathbf{y}_{\text{rep,m}})|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\} \\
&\quad + \left[E_D\{h(\mathbf{y}_{\text{rep,o}}, E(\mathbf{y}_{\text{rep,m}}|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega))|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\}\right]^2 \\
&= E_{ND}\{h^2(\mathbf{y}_{\text{rep}})|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\}
\end{aligned}$$

The expression following the last inequality holds given the equality of the expectations

$$E\{h(\mathbf{y}_{\text{rep}})|\mathbf{y}_{\text{rep,o}}, \mathbf{r}_{\text{rep}}, \omega\}$$

between the degenerate and non-degenerate models.

*Term 2:* Clearly,  $\text{Var}\{h(\mathbf{y})|\mathbf{y}_{\text{obs}}, \mathbf{r}\}$  is equal to zero if  $p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r})$  is degenerate. If  $p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r})$  is not degenerate, then this term will be positive.

## 2 Web Appendix B

**Table S.1: Growth Hormone Trial**

Treatment	$s$	$n_s$	Month		
			0	6	12
EG	1	12	58 (26)		
	2	4	57 (15)	68 (26)	
	3	22	78 (24)	90 (32)	88 (32)
	All	38	69 (25)	87 (32)	88 (32)
EP	1	7	65 (32)		
	2	2	87 (52)	86 (51)	
	3	31	65 (24)	81 (25)	73 (21)
	All	40	66 (26)	82 (26)	73 (21)

Table S.1: Growth hormone trial: quadriceps strength means (standard deviations) for treatment groups Exercise + Growth Hormone (EG) and Exercise + Placebo (EP) at months 0, 6 and 12. Missing pattern  $s = \{1, 2, 3\}$  corresponds to the last observation obtained at month 0, 6, or 12, respectively.  $n_s$  is the number of participants with missing pattern  $s$ .

Model	$p_D$	$\overline{\text{Deviance}}$	$\text{DIC}_o$
SM	23.4	5.6	29.1
MM1	30.7	10.6	41.3
MM2	25.4	26.4	51.8

Table S.2: DIC based on observed data likelihood ( $\text{DIC}_o$ ), posterior mean deviance ( $\overline{\text{Deviance}}$ ), and effective number of parameters ( $p_D$ ) for the growth hormone data analysis in Section 4 for Selection Model (SM), Mixture Model 1 (MM1), and Mixture Model 2 (MM2).

### 3 Web Appendix C

Closed form expectations for the mixture model in Section 4 for several choices of  $T(\cdot)$

Recall  $S = \sum_{j=1}^3 r_j$ . First, we derive the expectations for  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_3 y_3 - r_1 y_1$ .

$$\begin{aligned} \mathbb{E}_{\mathbf{y}_{\text{rep}}, \mathbf{r}_{\text{rep}} | \omega} (r_3 y_3 - r_1 y_1) &= \mathbb{E}_{\mathbf{r}_{\text{rep}} | \omega} (r_3 \mathbb{E}_{\mathbf{y}_{\text{rep}} | \mathbf{r}_{\text{rep}}, \omega} y_3 - r_1 \mathbb{E}_{\mathbf{y}_{\text{rep}} | \mathbf{r}_{\text{rep}}, \omega} y_1) \\ &= \mathbb{E}_{\mathbf{r}_{\text{rep}} | \omega} (r_3 \mu_3^{(s)} - r_1 \mu_1^{(s)}) \\ &= \eta_3 \mu_3^{(3)} - \mu_1 \end{aligned}$$

where,  $\mu_1 = \sum_{s=1}^3 \eta_s \mu_1^{(s)}$  and

$$\begin{aligned} \mu_3^{(3)} &= \mathbb{E}(Y_3 | S = 3) \\ &= \alpha_3 + \alpha_2 \phi_{32} + (\phi_{31} + \phi_{32} \phi_{21}) \mu_1^{(3)}. \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{y}_{\text{rep}}, \mathbf{r}_{\text{rep}} | \omega} \{(r_3 y_3 - r_1 y_1)^2\} &= \mathbb{E}_{\mathbf{r}_{\text{rep}} | \omega} \mathbb{E}_{\mathbf{y}_{\text{rep}} | \mathbf{r}_{\text{rep}}, \omega} \{(r_3 y_3 - r_1 y_1)^2\} \\ &= \mathbb{E}_{\mathbf{r}_{\text{rep}} | \omega} \mathbb{E}_{\mathbf{y}_{\text{rep}} | \mathbf{r}_{\text{rep}}, \omega} \{(r_3^2 y_3^2) + (r_1^2 y_1^2) - 2r_1 r_3 y_1 y_3\} \\ &= \mathbb{E}_{\mathbf{r}_{\text{rep}} | \omega} \left\{ r_3^2 \mathbb{E}_{\mathbf{y}_{\text{rep}} | \mathbf{r}_{\text{rep}}, \omega} (y_3^2) + r_1^2 \mathbb{E}_{\mathbf{y}_{\text{rep}} | \mathbf{r}_{\text{rep}}, \omega} (y_1^2) \right. \\ &\quad \left. - 2r_1 r_3 \mathbb{E}_{\mathbf{y}_{\text{rep}} | \mathbf{r}_{\text{rep}}, \omega} (y_1 y_3) \right\} \\ &= \eta_3 (V_{33} + \mu_3^{(3)2}) + \sum_{s=1}^3 \eta_s (\sigma_1^{(s)2} + \mu_1^{(s)2}) \\ &\quad - 2\eta_3 \{(\phi_{31} + \phi_{32} \phi_{21}) \sigma_1^{(3)2} + \mu_1^{(3)} \mu_3^{(3)}\} \end{aligned}$$

where  $V_{33} = \text{Var}(Y_3 | S = 3) = (\phi_{31}^2 + \phi_{32}^2 \phi_{21}^2 + 2\phi_{31} \phi_{32} \phi_{21}) \sigma_1^{(3)2} + \phi_{32}^2 \tau_2^2 + \tau_3^2$ .

Now we derive the expectations for  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_3(r_3 y_3 - r_1 y_1)$ .

$$\mathbb{E}_{\mathbf{y}_{\text{rep}}, \mathbf{r}_{\text{rep}} | \omega} \{r_3(r_3 y_3 - r_1 y_1)\} = \eta_3 (\mu_3^{(3)} - \mu_1^{(3)})$$

$$\begin{aligned} \mathbb{E}_{\mathbf{y}_{\text{rep}}, \mathbf{r}_{\text{rep}} | \omega} \{r_3(y_3 - y_1)\}^2 &= \mathbb{E}_{\mathbf{y}_{\text{rep}}, \mathbf{r}_{\text{rep}} | \omega} (r_3 y_3^2 + r_3 y_1^2 - 2r_3 y_3 y_1) \\ &= \eta_3 (V_{33} + \mu_3^{(3)2}) + \eta_3 (\sigma_1^{(3)2} + \mu_1^{(3)2}) \\ &\quad - 2\eta_3 \{(\phi_{31} + \phi_{32} \phi_{21}) \sigma_1^{(3)2} + \mu_1^{(3)} \mu_3^{(3)}\}. \end{aligned}$$

### 4 Web Appendix D

#### Simulation Study Results

Model	$p_D$	$\overline{\text{Deviance}}$	$\text{DIC}_o$
True Model: SM0, Sample Size: 50			
SM0	23.6	1245.8	1269.4
MM1	42.0	1253.5	1295.6
MM2	35.8	1285.8	1321.6
True Model: SM0, Sample Size: 100			
SM0	23.8	2493.9	2517.7
MM1	34.4	2501.6	2536.0
MM2	28.7	2567.8	2596.5
True Model: SM0, Sample Size: 2000			
SM0	24.0	49801.6	49825.6
MM1	34.4	49944.7	49979.1
MM2	29.4	51238.4	51267.8
True Model: MM1, Sample Size: 50			
SM0	22.3	1372.0	1394.3
MM1	39.9	1261.0	1300.9
MM2	35.5	1575.3	1610.8
True Model: MM1, Sample Size: 100			
SM0	23.2	2745.4	2768.6
MM1	34.0	2516.5	2550.4
MM2	28.6	3146.6	3175.2
True Model: MM1, Sample Size: 2000			
SM0	23.9	54889.6	54913.5
MM1	34.1	50274.5	50308.6
MM2	29.3	62820.4	62849.7
True Model: MM2, Sample Size: 50			
SM0	22.2	1380.5	1402.7
MM1	40.2	1262.6	1302.8
MM2	33.6	1265.1	1298.7
True Model: MM2, Sample Size: 100			
SM0	23.2	2760.9	2784.1
MM1	33.9	2517.0	2550.9
MM2	28.2	2520.1	2548.3
True Model: MM2, Sample Size: 2000			
SM0	24.0	55232.5	55256.4
MM1	34.2	50253.8	50287.9
MM2	29.4	50256.9	50286.2

Table S.3: For simulating (true) model being MAR Selection model (SM0), Mixture model 1 (MM1) and Mixture model 2 (MM2) and sample size being 50, 100 and 20000, average (over 200 replications) DIC based on observed data likelihood ( $\text{DIC}_o$ ), posterior mean deviance ( $\overline{\text{Deviance}}$ ), and effective number of parameters ( $p_D$ ). 6

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	41.2(4.7)	41.6(4.2)	82.8(8.7)
MM1	41.2(4.7)	49.6(6.8)	90.8(10.0)
MM2	41.8(4.8)	53.8(7.8)	95.6(11.2)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	8.4(1.3)	8.9(1.3)	17.3(2.5)
MM1	8.4(1.3)	10.0(1.5)	18.5(2.7)
MM2	9.0(1.6)	10.5(1.8)	19.6(3.3)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T I\{r_j = 1, r_{j+1} = 0\} r_j y_j - I\{r_2 = 1\} r_1 y_1$			
SM0	30.6(5.0)	31.0(4.8)	61.6(9.6)
MM1	30.6(5.0)	37.0(5.1)	67.5(9.6)
MM2	31.3(5.1)	35.7(4.7)	67.0(9.0)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T I\{r_j = 1, r_{j+1} = 0\} r_j y_j - I\{r_2 = 1\} r_1 y_1 \right]^2$			
SM0	2151(660)	2162(527)	4313(1104)
MM1	2179(652)	7038(12083)	9217(12052)
MM2	2185(648)	8370(18109)	10555(18059)

Table S.4: Simulating (true) model MAR Selection model (SM0) and sample size 50: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	41.4(3.7)	41.7(3.1)	83.2(6.5)
MM1	41.5(3.7)	43.2(3.5)	84.7(7.1)
MM2	41.9(3.7)	47.6(4.2)	89.5(7.6)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	8.7(0.9)	8.9(0.9)	17.6(1.7)
MM1	8.7(0.9)	9.4(0.9)	18.1(1.8)
MM2	9.2(1.1)	10.0(1.1)	19.2(2.1)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T I\{r_j = 1, r_{j+1} = 0\} r_j y_j - I\{r_2 = 1\} r_1 y_1$			
SM0	30.5(4.0)	30.8(3.6)	61.4(7.5)
MM1	30.5(4.0)	33.1(3.9)	63.6(7.9)
MM2	31.1(4.0)	31.7(3.6)	62.8(7.5)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T I\{r_j = 1, r_{j+1} = 0\} r_j y_j - I\{r_2 = 1\} r_1 y_1 \right]^2$			
SM0	2122(537)	2140(397)	4263(877)
MM1	2124(537)	2566(607)	4690(1132)
MM2	2127(536)	2611(616)	4739(1139)

Table S.5: Simulating (true) model MAR Selection model (SM0) and sample size 100: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.



Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	41.7(0.8)	41.8(0.7)	83.4(1.5)
MM1	41.7(0.8)	41.0(0.9)	82.7(1.7)
MM2	42.0(0.8)	45.2(1.0)	87.2(1.8)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	8.7(0.2)	8.8(0.2)	17.5(0.4)
MM1	8.8(0.2)	8.8(0.2)	17.6(0.4)
MM2	9.0(0.3)	9.3(0.2)	18.4(0.4)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T I\{r_j = 1, r_{j+1} = 0\} r_j y_j - I\{r_2 = 1\} r_1 y_1$			
SM0	30.8(1.0)	30.8(0.9)	61.6(1.8)
MM1	30.8(1.0)	31.5(1.0)	62.4(1.9)
MM2	31.2(1.0)	30.1(1.0)	61.3(1.9)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T \{I(r_j = 1, r_{j+1} = 0) r_j y_j\} - I(r_2 = 1) r_1 y_1 \right]^2$			
SM0	2116(127)	2126(103)	4242(216)
MM1	2117(127)	2145(130)	4262(254)
MM2	2120(127)	2186(133)	4305(257)

Table S.6: Simulating (true) model MAR Selection model (SM0) and sample size 2000: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	1095.9(67.9)	1097.9(63.2)	2193.8(129.6)
MM1	1094.4(68.1)	1105.7(67.4)	2200.1(135.2)
MM2	1662.4(243.7)	1852.6(657.9)	3515.0(690.0)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	266.1(19.1)	289.9(23.8)	556.0(42.2)
MM1	265.7(19.0)	271.6(18.8)	537.3(37.7)
MM2	833.8(251.4)	1071.7(314.9)	1905.6(480.9)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T \{\mathbf{I}(r_j = 1, r_{j+1} = 0) r_j y_j\} - \mathbf{I}(r_2 = 1) r_1 y_1$			
SM0	413.0(47.6)	443.2(48.0)	856.2(94.1)
MM1	411.3(47.3)	421.6(45.4)	832.9(92.1)
MM2	1184.5(312.0)	1388.6(270.3)	2573.2(460.6)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T \{\mathbf{I}(r_j = 1, r_{j+1} = 0) r_j y_j\} - \mathbf{I}(r_2 = 1) r_1 y_1 \right]^2$			
SM0	296118(32710)	356876(49052)	652994(78363)
MM1	294008(32332)	328074(73249)	622082(93555)
MM2	1722024(1986146)	5537388(4892633)	7259412(6795196)

Table S.7: Simulating (true) model Mixture model 1 (MM1) and sample size 50: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	1104.2(50.7)	1105.8(46.2)	2210.0(96.2)
MM1	1102.8(50.9)	1107.0(50.1)	2209.9(100.9)
MM2	1607.3(207.1)	1734.7(459.2)	3342.0(456.6)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	267.6(13.3)	290.0(15.7)	557.5(28.7)
MM1	267.3(13.3)	270.1(12.9)	537.4(26.1)
MM2	771.0(204.6)	963.8(205.7)	1734.8(330.4)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T \{\mathbf{I}(r_j = 1, r_{j+1} = 0) r_j y_j\} - \mathbf{I}(r_2 = 1) r_1 y_1$			
SM0	415.0(32.8)	444.1(33.7)	859.1(65.3)
MM1	413.4(32.5)	417.5(32.3)	830.9(64.5)
MM2	1092.3(270.7)	1214.1(233.7)	2306.4(410.1)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T \{\mathbf{I}(r_j = 1, r_{j+1} = 0) r_j y_j\} - \mathbf{I}(r_2 = 1) r_1 y_1 \right]^2$			
SM0	297395(23936)	352048(31864)	649442(53299)
MM1	295418(23730)	303460(22823)	598878(46048)
MM2	1151272(1048539)	3579020(2478676)	4730292(3481461)

Table S.8: Simulating (true) model Mixture model 1 (MM1) and sample size 100: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	1114.3(11.4)	1114.1(10.4)	2228.4(21.4)
MM1	1113.2(11.4)	1113.7(11.6)	2226.8(22.7)
MM2	1601.5(195.4)	1644.6(374.1)	3246.1(183.6)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	270.3(3.0)	291.6(3.5)	561.9(6.4)
MM1	270.0(3.0)	270.2(3.1)	540.2(5.9)
MM2	758.6(194.6)	873.7(60.4)	1632.3(143.2)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T \{\mathbf{I}(r_j = 1, r_{j+1} = 0) r_j y_j\} - \mathbf{I}(r_2 = 1) r_1 y_1$			
SM0	418.9(7.6)	445.6(7.7)	864.5(14.8)
MM1	417.5(7.5)	417.7(7.6)	835.2(14.9)
MM2	1074.9(267.2)	1354.4(114.8)	2429.3(160.3)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T \{\mathbf{I}(r_j = 1, r_{j+1} = 0) r_j y_j\} - \mathbf{I}(r_2 = 1) r_1 y_1 \right]^2$			
SM0	299940(5617)	350146(7537)	650086(12618)
MM1	298273(5586)	298677(5512)	596950(10869)
MM2	1019003(348422)	2908665(176399)	3927667(480104)

Table S.9: Simulating (true) model Mixture model 1 (MM1) and sample size 2000: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	1626.4(91.9)	1657.1(85.7)	3283.5(176.2)
MM1	1624.2(92.3)	1635.7(90.7)	3259.9(182.8)
MM2	1624.8(92.2)	1635.3(91.0)	3260.1(182.5)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	499.7(26.4)	541.2(32.2)	1040.9(57.7)
MM1	499.1(26.3)	505.8(25.8)	1004.9(51.9)
MM2	499.7(26.2)	505.1(25.4)	1004.8(50.9)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T \{I(r_j = 1, r_{j+1} = 0)r_j y_j\} - I(r_2 = 1)r_1 y_1$			
SM0	394.5(44.0)	448.6(45.2)	843.1(86.7)
MM1	393.9(43.8)	403.3(43.7)	797.2(86.8)
MM2	394.6(43.8)	402.7(43.5)	797.3(86.3)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T \{I(r_j = 1, r_{j+1} = 0)r_j y_j\} - I(r_2 = 1)r_1 y_1 \right]^2$			
SM0	484041(46637)	585615(69484)	1069657(111360)
MM1	480156(46125)	515598(55962)	995754(99057)
MM2	482642(46208)	512871(56387)	995514(97830)

Table S.10: Simulating (true) model Mixture model 2 (MM2) and sample size 50: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	1655.1(66.9)	1683.5(59.6)	3338.6(125.5)
MM1	1653.6(67.3)	1657.3(66.8)	3310.9(133.9)
MM2	1653.8(67.3)	1656.6(66.8)	3310.4(133.9)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	506.1(18.9)	546.2(22.2)	1052.3(40.5)
MM1	505.7(18.8)	508.6(18.4)	1014.3(37.0)
MM2	506.0(18.9)	507.9(18.4)	1014.0(37.0)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T \{I(r_j = 1, r_{j+1} = 0)r_j y_j\} - I(r_2 = 1)r_1 y_1$			
SM0	408.7(32.3)	460.2(30.3)	868.8(61.2)
MM1	408.2(32.1)	411.3(31.2)	819.5(62.9)
MM2	408.5(32.2)	410.9(31.4)	819.5(63.2)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T \{I(r_j = 1, r_{j+1} = 0)r_j y_j\} - I(r_2 = 1)r_1 y_1 \right]^2$			
SM0	495264(33281)	585498(43762)	1080763(74380)
MM1	491650(32949)	504644(32542)	996295(64911)
MM2	492923(33312)	502432(33131)	995355(65514)

Table S.11: Simulating (true) model Mixture model 2 (MM2) and sample size 100: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.

Model	GOF	Complexity	$C_\infty$
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T y_T - r_1 y_1$			
SM0	1669.6(14.0)	1699.4(13.5)	3369.0(27.1)
MM1	1668.4(14.0)	1668.3(15.1)	3336.7(28.7)
MM2	1668.4(14.0)	1668.5(14.8)	3337.0(28.5)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = r_T(r_T y_T - r_1 y_1)$			
SM0	511.4(4.0)	552.0(5.0)	1063.3(8.8)
MM1	511.0(4.0)	511.2(4.3)	1022.3(8.1)
MM2	511.1(4.0)	511.2(4.2)	1022.3(8.0)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \sum_{j=1}^T \{I(r_j = 1, r_{j+1} = 0)r_j y_j\} - I(r_2 = 1)r_1 y_1$			
SM0	409.3(6.6)	462.0(7.3)	871.3(13.4)
MM1	409.1(6.5)	409.1(7.0)	818.3(13.2)
MM2	409.1(6.5)	409.3(6.7)	818.4(13.0)
$T(\mathbf{r}, \mathbf{r} \circ \mathbf{y}) = \left[ \sum_{j=1}^T \{I(r_j = 1, r_{j+1} = 0)r_j y_j\} - I(r_2 = 1)r_1 y_1 \right]^2$			
SM0	497319(7482)	580996(10900)	1078315(17886)
MM1	494065(7360)	494774(7606)	988839(14647)
MM2	494143(7369)	494568(7702)	988711(14752)

Table S.12: Simulating (true) model Mixture model 2 (MM2) and sample size 2000: average PPL criteria and SD (in parenthesis) over 200 replications for four choices of  $T(\mathbf{r}, \mathbf{r} \circ \mathbf{y})$  for models MAR Selection model (SM0), Mixture model 1 (MM1), and Mixture Model 2 (MM2).  $C_\infty$  is the PPL criterion with  $k = \infty$  and GOF is the goodness of fit component of the criterion.