Supplement to: "AR(1) Latent Class Models for Longitudinal Count Data"

A Transition Function

The transition function $p(y_{ij}|y_{i,j-1}, \mathbf{X}_i, \boldsymbol{\theta}_c) = p_i(y_{ij}|y_{i,j-1}; \boldsymbol{\theta}_c)$ can be expressed as the convolution of a beta-binomial distribution and a negative binomial distribution

$$p_i(y_{ij}|y_{i,j-1}; \boldsymbol{\theta}_c) = \sum_{k=0}^{\min(y_{ij}, y_{i,j-1})} f_i(y_{ij} - k) g_i(k). \tag{1}$$

Above, $f_i(\cdot)$ represents the probability mass function for a beta-binomial distribution with parameters $\Big(y_{ij}, \alpha_c \mu_{i,j-1}^c/\gamma_c, (1-\alpha_c)\mu_{i,j-1}^c/\gamma_c\Big)$

$$f_i(k) = \frac{\Gamma(y_{i,j-1}+1)\Gamma(\eta_{i,j-1}^c)\Gamma(k+\alpha_c)\eta_{i,j-1}^c)\Gamma(y_{i,j-1}+(1-\alpha_c)\eta_{i,j-1}^c)}{\Gamma(k+1)\Gamma(y_{i,j-1}-k+1)\Gamma(\alpha_c\eta_{i,j-1}^c)\Gamma((1-\alpha_c)\eta_{i,j-1}^c)\Gamma(\eta_{i,j-1}^c+y_{i,j-1})} \tag{2}$$

and $g_i(\cdot)$ represents the probability mass function for a negative binomial distribution with mean $\mu_{ij}^c(1-\alpha_c)$ and variance $\mu_{ij}^c(1-\alpha_c)(1+\gamma_c)$

$$g_i(k) = \frac{\Gamma\{k+(1-\alpha_c)\mu_{ij}^c/\gamma_c\}}{\Gamma\{k+1\}\Gamma\{k+(1-\alpha_c)\mu_{ij}^c/\gamma_c\}} \Big(\frac{\gamma_c}{1+\gamma_c}\Big)^{(1-\alpha_c)\mu_{ij}^c/\gamma_c} \Big(\frac{\gamma_c}{1+\gamma_c}\Big)^k. \tag{3}$$

By applying (1), the transition function can then be written as

$$\begin{aligned} p_i(y_{ij}|y_{i,j-1}; \boldsymbol{\theta}_c) &= \frac{\gamma_c^{y_{i,j-1}}\Gamma(y_{i,j-1}+1)\Gamma(\eta_{i,j-1}^c)}{\Gamma(\lambda_{ij}^c)\Gamma(\eta_{i,j-1}^c-\lambda_{ij}^c)\Gamma(\eta_{ij}^c-\lambda_{ij}^c)\Gamma(\eta_{i,j-1}^c+y_{i,j-1})(1+\gamma_c)^{y_{i,j-1}+\eta_{ij}^c-\lambda_{ij}^c}} \\ &\times \sum_{k=0}^{\min(y_{ij}, y_{i,j-1})} \Big(\frac{1+\gamma_c}{\gamma_c}\Big)^k \frac{\Gamma(\lambda_{ij}^c)\Gamma(\eta_{i,j-1}^c-\lambda_{ij}^c+y_{i,j-1}-k)\Gamma(\eta_{ij}^c-\lambda_{ij}^c+y_{ij}-k)}{\Gamma(y_{i,j-1}-k+1)\Gamma(y_{ij}-k+1)\Gamma(k+1)}, \end{aligned}$$

where $\lambda_{ij}^c = \alpha_c\sqrt{\mu_{ij}^c\mu_{i,j-1}^c}/\gamma_c$, $\eta_{ij}^c = \mu_{ij}^c/\gamma_c$, and $\eta_{i,j-1}^c = \mu_{i,j-1}^c/\gamma_c$.

B Unbiased Estimating Function

Using the notation of Section 3.2, let $U_i^k(\boldsymbol{\theta}_c; \mathbf{y_i})$ be the k^{th} component of $U_i(\boldsymbol{\theta}_c)$. We can then see that $G(\boldsymbol{\Theta}, \boldsymbol{\pi})$ is an unbiased estimating function because the expectation of the (k, c) component of $\mathbf{V}_i$ is given by

$$E\Big\{W_{ic}(\boldsymbol{\Theta}, \boldsymbol{\pi})U_i^k(\boldsymbol{\theta}_c; \mathbf{y}_i)\Big\} = \pi_c E\Big\{\frac{p_{\theta_c}(\mathbf{y}_i)}{p(\mathbf{y}_i)}U_i^k(\boldsymbol{\theta}_c; \mathbf{y}_i)\Big\} = \pi_c E_{\theta_c}\Big\{U_i^k(\boldsymbol{\theta}_c; \mathbf{y}_i)\Big\} = 0 \ ,$$

and the expectation of the c^{th} element of $\mathbf{b}_i$ is given by

$$E\Big\{W_{ic}(\mathbf{\Theta}, \boldsymbol{\pi}) - \pi_c\Big\} = \pi_c E\Big\{\frac{p_{\theta_c}(\mathbf{y_i})}{p(\mathbf{y_i})}\Big\} - \pi_c = 0.$$

C Parameter Initialization

The parameter initialization procedure is detailed below

1. Choose K "cluster centers" for the regression coefficients $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$. This is done by randomly selecting K subject and fitting a separate Poisson regression for each subject.

2. Assign each subject to one of the K classes through $S_i = \underset{c}{\operatorname{argmin}}\{D(\mathbf{y}_i; \boldsymbol{\beta}_c)\}$. Here, $D(\mathbf{y}_i; \boldsymbol{\beta}_c) = -2 \log L(\boldsymbol{\beta}_c|\mathbf{y}_i)$ is the usual deviance associated with a Poisson regression. Namely, $D(\mathbf{y}_i; \boldsymbol{\beta}_c) = 2\sum_{j=1}^{n_i}\Big(\mu_{ij}^c - y_{ij}\log(\mu_{ij}^c) + \log(y_{ij}!)\Big)$.

3. Using this hard assignment of subjects to clusters, compute a new value of $\boldsymbol{\beta}_c$ by fitting a Poisson regression for the subjects in the set $\mathcal{S}_c = \{i : S_i = c\}$. Do this for each cluster $c = 1, \ldots, K$.

4. Repeat steps (2)-(3) twice.

5. Compute the mixture proportions through $\pi_c = \frac{1}{m}\sum_{i=1}^m \mathbf{1}\{S_i = c\}$, and compute each $\boldsymbol{\theta}_c = (\boldsymbol{\beta}_c, \alpha_c, \gamma_c)$ by solving $\sum_{i\in\mathcal{S}_c} U_i(\boldsymbol{\theta}_c) = \mathbf{0}$, where $U_i(\cdot)$ is the estimating function described in Section 3.1 and is also shown in (4) below.

$$U_i(\boldsymbol{\theta}_c) = \frac{1}{\phi_c}\begin{bmatrix} \mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\mu}_i^c)\mathbf{R}_i^{-1}(\alpha_c)\mathbf{A}_i^{-1/2}(\boldsymbol{\mu}_i^c)(\mathbf{y}_i - \boldsymbol{\mu}_i^c) \\ \frac{2\phi_c\alpha_c(n_i-1)}{1-\alpha_c^2} - (\mathbf{y}_i - \boldsymbol{\mu}_i^c)^T\frac{d\mathbf{R}_i^{-1}(\alpha_c)}{d\alpha_c}(\mathbf{y}_i - \boldsymbol{\mu}_i^c) \\ \frac{1}{\phi_c}(\mathbf{y}_i - \boldsymbol{\mu}_i^c)^T\mathbf{A}_i^{-1/2}(\boldsymbol{\mu}_i^c)\mathbf{R}_i^{-1}(\alpha_c)\mathbf{A}_i^{-1/2}(\boldsymbol{\mu}_i^c)(\mathbf{y}_i - \boldsymbol{\mu}_i^c) - n_i \end{bmatrix}. \quad (4)$$

D Estimation with Sampling Weights

Suppose v_i, $i = 1, \ldots, m$ are sampling weights such that v_i is proportional to the inverse probability that subject i is included in the sample. We compute initial estimates

$(\boldsymbol{\Theta}^{(0)}, \boldsymbol{\pi}^{(0)})$ in the same way as the unweighted case. Given estimates $(\boldsymbol{\Theta}^{(k)}, \boldsymbol{\pi}^{(k)})$, we produce updated estimates $(\boldsymbol{\Theta}^{(k+1)}, \boldsymbol{\pi}^{(k+1)})$ in the $(k+1)^{st}$ step through the following process.

- Update $\boldsymbol{\theta}_c = (\boldsymbol{\beta}_c, \alpha_c, \gamma_c)$ by solving

$$\sum_{i=1}^{m} v_i W_{ic}(\boldsymbol{\Theta}^{(k)}, \boldsymbol{\pi}^{(k)}) U_i(\boldsymbol{\theta}_c) = \mathbf{0}, \tag{5}$$

  where $U_i(\cdot)$ is the estimating function described in Section 3.1 and is also shown in (4).

- Update π_c through

$$\pi_c^{(k+1)} = \frac{\sum_{i=1}^{m} v_i W_{ic}(\boldsymbol{\Theta}^{(k)}, \boldsymbol{\pi}^{(k)})}{\sum_{i=1}^{m} v_i} \tag{6}$$

We determine convergence by stopping when the weighted estimating function $\sum_{i=1}^{m} v_i G_i(\boldsymbol{\Theta}, \boldsymbol{\pi})$ is sufficiently close to zero.

E Simulation Parameter Values

In Scenarios I and II of the INAR simulations, we define the mean curves through $\log(\mu_{ij}^c) = \beta_0^c + \beta_1^c t_{ij}$. The values of (β_0^c, β_1^c) used for Scenarios I and II are given in Table S3.

The Poisson-Normal simulations use four simulation settings. One of these settings (the setting with the largest class separation index) has eight time points with time points $t_{ij} = j/9$ for $j = 1, \ldots, 8$. The other three settings have five time points with $t_{ij} = j/6$ for $j = 1, \ldots, 5$. Each setting utilizes the parameters $(\beta_0^c, \beta_1^c, \sigma_{c0}^2, \sigma_{c1}^2, \pi_c)$ for $c = 1, \ldots, 4$. The values of these parameters for each simulation setting (as indexed by the value of the CSI) are shown in table S4

List of Figures

S1 Empirical Discrimination from INAR(1)-NB Simulations 5
S2 Diagnostics for the overdispersion parameters 6
S3 Marginal residual plot 7

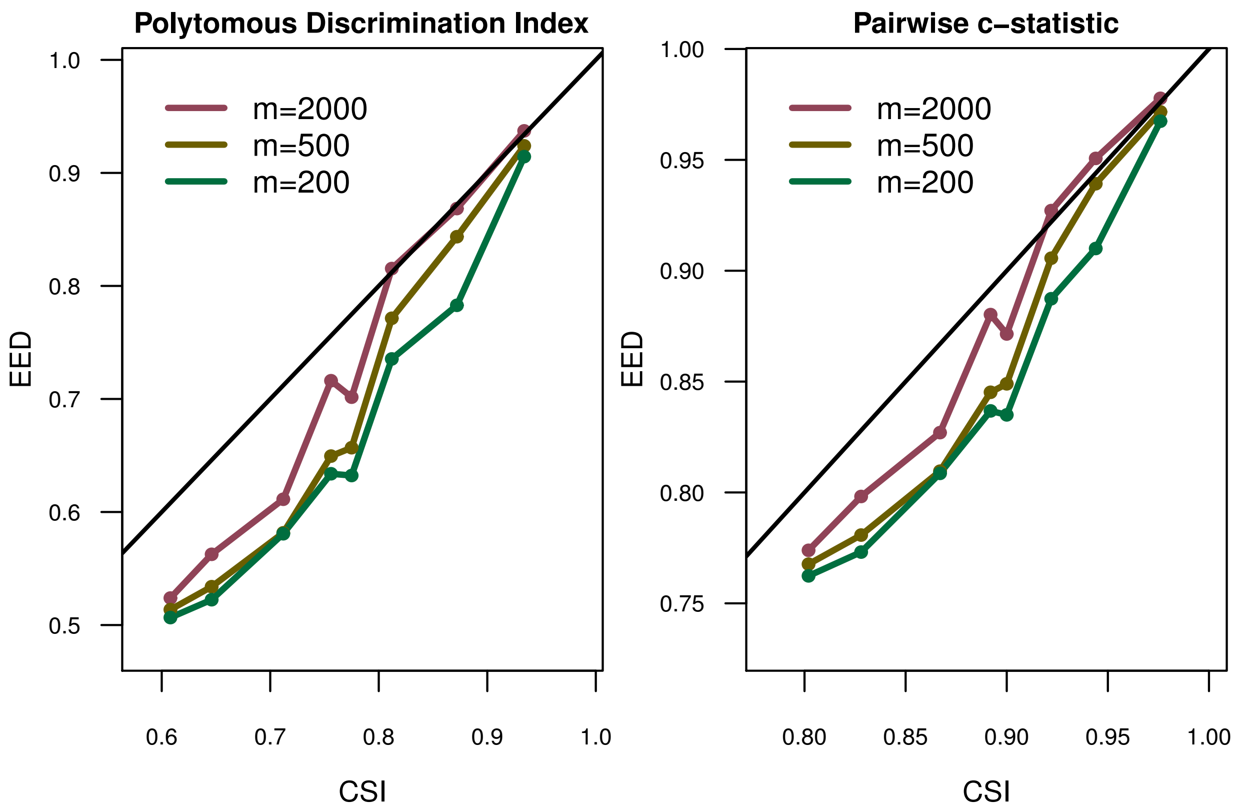

Figure S1: Expected empirical discrimination (EED) and Class Separation Indices (CSI) for Scenarios I and II of the INAR(1)-NB model. Values of the EED are shown when the parameters are estimated from simulations with $m = 200$, $m = 500$, $m = 2,000$ subjects.

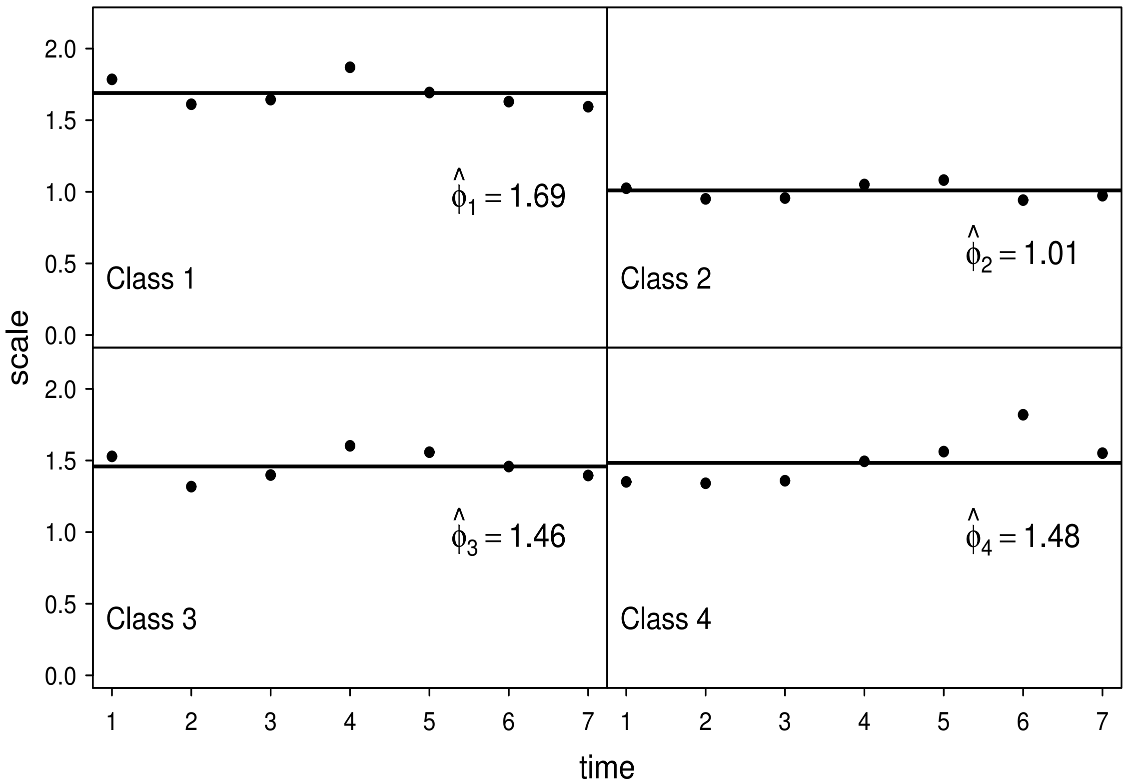

Figure S2: Sample overdispersion (weighted) obtained by using the estimated posterior probabilities of class membership to randomly assign each subject to one of the four latent classes. This random assignment procedure was repeated 1,000 times; the displayed overdispersion values represent the average overdispersion values from these replications.

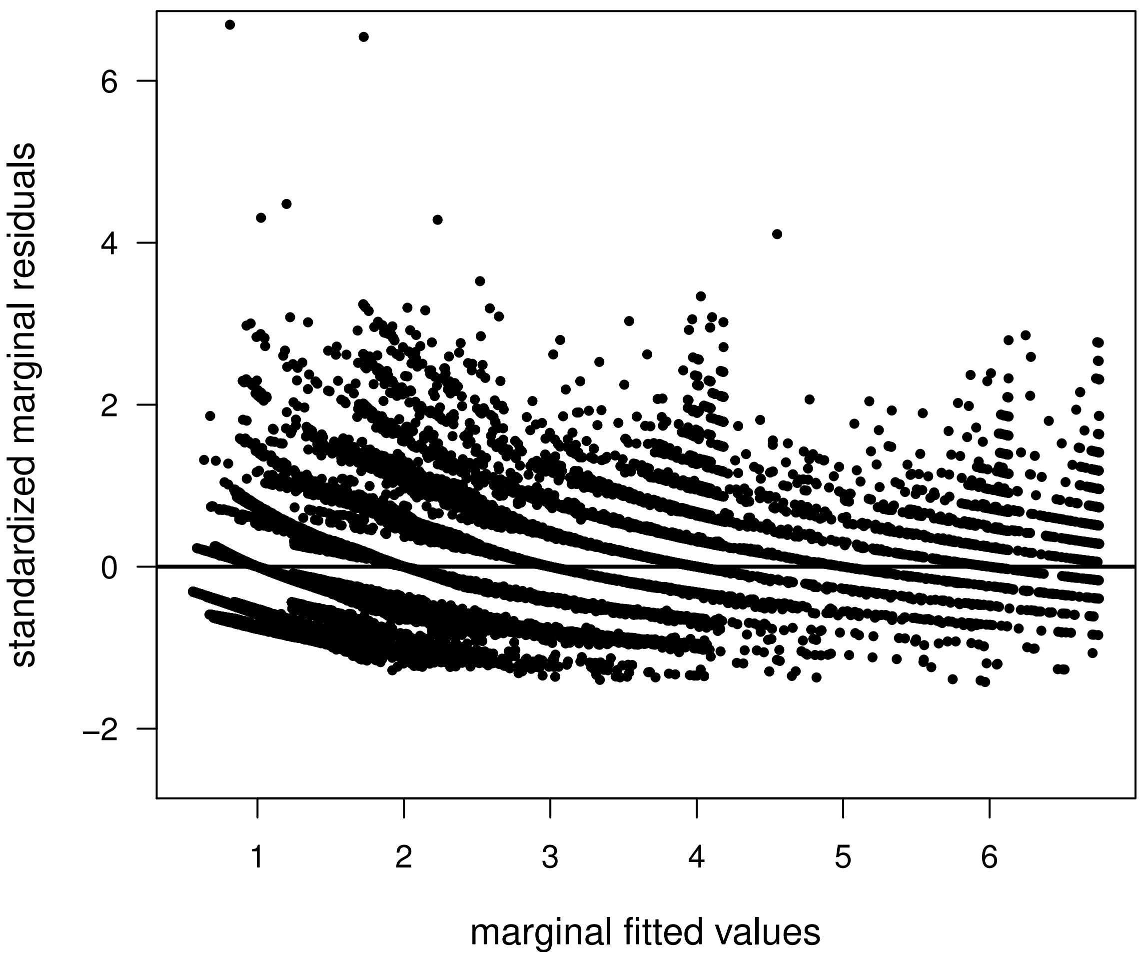

Figure S3: Marginal residuals vs. marginal fitted values obtained from fitting the four-class model to the CNLSY data. The marginal fitted values $\hat{Y}_{ij}$ are defined as $\hat{Y}_{ij} = \sum_{c=1}^{C} \hat{\mu}_{ij}^{c} \hat{p}_{i}^{c}$ where $\hat{p}_{i}^{c}$ is the estimated posterior probability that subject i belongs to class c. The marginal residuals r_{ij} are defined as $r_{ij} = (Y_{ij} - \hat{Y}_{ij})/\sqrt{\widehat{Var}(Y_{ij})}$. Here, the estimated marginal variance is defined as $\widehat{Var}(Y_{ij}) = \sum_{c=1}^{C} \phi_c \hat{\mu}_{ij}^{c} \hat{p}_{i}^{c} + \sum_{c=1}^{C} \{\hat{\mu}_{ij}^{c} - \bar{\mu}_{ij}\}^2 \hat{p}_{i}^{c}$, where the term $\bar{\mu}_{ij}$ is defined as $\bar{\mu}_{ij} = C^{-1} \sum_{c=1}^{C} \hat{\mu}_{ij}^{c}$.

List of Tables

S1 Class Separation Indices for the Main Simulation Scenarios 9
S2 Summary Statistics from the CNLSY Data 10
S3 Values of Regression Coefficients in the Main Simulation Scenarios 11
S4 Parameter Values used for the Poisson-Normal Simulations. 12

Table S1: Class separation indices for each of the two central scenarios with several different values of the autocorrelation and scale parameters. The class separation indices are computed using both the all-pairwise c-statistic (APC_C) and the polytomous discrimination index (PDI_C) as measures of classification performance.

	Scenario I				Scenario II			
	$\phi = 1.25$		$\phi = 3.0$		$\phi = 1.25$		$\phi = 3.0$	
	$\alpha = 0.1$	$\alpha = 0.4$	$\alpha = 0.1$	$\alpha = 0.4$	$\alpha = 0.1$	$\alpha = 0.4$	$\alpha = 0.1$	$\alpha = 0.4$
APC_C	0.976	0.944	0.922	0.892	0.900	0.867	0.828	0.802
PDI_C	0.934	0.872	0.812	0.756	0.775	0.712	0.646	0.608

Table S2: Summary statistics from the CNLSY data. In total, these data contain $9,626$ subjects each of which was surveyed biennially over the ages 4 to 16 (or 5 to 17). For the age groups 4-5, 6-7, and 8-9, the counts are solely from the behavioral problems index (BPI). For age groups 10-11 and 12-13, the counts represent the sum of the mother-reported BPI and the child-reported number of delinquent acts. For age groups 14-15 and 16-17, the counts are solely from the self-reported number of delinquent acts.

Child Ages	Mean	10%	25%	50%	75%	90%	Max
4-5	2.35	0	1	2	4	6	14
6-7	2.12	0	0	2	3	5	14
8-9	2.15	0	0	2	3	5	14
10-11	3.46	0	1	3	5	8	19
12-13	3.74	0	1	3	5	8	20
14-15	1.36	0	0	1	2	4	7
16-17	1.37	0	0	1	2	4	7

Table S3: Values of the regression coefficients in Scenarios I and II.

	Scenario I				Scenario II			
	c=1	c=2	c=3	c=4	c=1	c=2	c=3	c=4
β_0^c	-0.4	1.5	0.0	1.4	-0.4	1.4	0.0	1.2
β_1^c	-0.1	-0.7	0.65	0.0	-0.1	-1.0	0.9	0.0

Table S4: Parameter values used for the Poisson-Normal simulations. The Class Separation Index (CSI) shown for each simulation setting was computed with the all-pairwise c-statistic.

	CSI=0.983	CSI=0.934	CSI=0.896	CSI = 0.820
β_0^1	-0.90	-0.90	-0.90	-9.00
β_1^1	-0.35	-0.35	-0.35	-0.35
σ_{10}^2	0.30	0.40	0.85	1.50
σ_{11}^2	0.125	0.20	0.50	0.70
π_1	0.50	0.50	0.50	0.50
β_0^2	1.55	1.55	1.55	1.3
β_1^2	-2.10	-2.10	-2.10	-2.10
σ_{20}^2	0.08	0.15	0.35	1.00
σ_{21}^2	0.05	0.075	0.25	0.50
π_2	0.25	0.25	0.25	0.25
β_0^3	-0.40	-0.65	-0.65	-0.7
β_1^3	1.90	2.00	2.00	1.75
σ_{30}^2	0.10	0.20	0.60	1.25
σ_{31}^2	0.06	0.075	0.25	0.55
π_3	0.15	0.15	0.15	0.15
β_0^4	1.40	1.25	1.25	1.00
β_1^4	-0.05	-0.05	-0.05	-0.05
σ_{40}^2	0.06	0.10	0.35	1.00
σ_{41}^2	0.04	0.04	0.30	0.45
π_4	0.10	0.10	0.10	0.10