

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

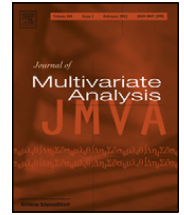
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at SciVerse ScienceDirect

## Journal of Multivariate Analysis

journal homepage: [www.elsevier.com/locate/jmva](http://www.elsevier.com/locate/jmva)

# Bayesian modeling of the dependence in longitudinal data via partial autocorrelations and marginal variances

Y. Wang<sup>a</sup>, M.J. Daniels<sup>b,\*</sup><sup>a</sup> Department of Biostatistics, University of Florida, Dauer Hall, Gainesville, FL 32611, United States<sup>b</sup> Section of Integrative Biology, Division of Statistics & Scientific Computation, University of Texas at Austin, 141MC Patterson Hall, Austin, TX 78712, United States

## ARTICLE INFO

## Article history:

Received 14 May 2012

Available online 19 December 2012

## AMS subject classifications:

62F15

62H12

62P10

## Keywords:

Markov chain Monte Carlo

Generalized linear model

Uniform prior

G-prior

## ABSTRACT

Many parameters and positive-definiteness are two major obstacles in estimating and modeling a correlation matrix for longitudinal data. In addition, when longitudinal data is incomplete, incorrectly modeling the correlation matrix often results in bias in estimating mean regression parameters. In this paper, we introduce a flexible and parsimonious class of regression models for a covariance matrix parameterized using marginal variances and partial autocorrelations. The partial autocorrelations can freely vary in the interval  $(-1, 1)$  while maintaining positive definiteness of the correlation matrix so the regression parameters in these models will have no constraints. We propose a class of priors for the regression coefficients and examine the importance of correctly modeling the correlation structure on estimation of longitudinal (mean) trajectories and the performance of the DIC in choosing the correct correlation model via simulations. The regression approach is illustrated on data from a longitudinal clinical trial.

© 2012 Elsevier Inc. All rights reserved.

## 1. Introduction

Longitudinal data, measurements on the same subject over time, arise in many areas, from clinical trials to environmental studies. In such studies, to draw valid inferences, the covariance between repeated observations on the same individuals needs to be properly modeled. On top of this, in incomplete longitudinal data, mis-modeling the covariance matrix can result in biased estimates of fixed effect mean parameters [22,8]. Two major obstacles for modeling covariance matrices are (1) the number of parameters and (2) positive-definiteness.

Many approaches have been proposed for estimating a covariance matrix more efficiently, whether by shrinking eigenvalues to obtain more stability [34,15] or reducing the dimension via structure [20,4,26,27,14]. Some regression approaches to introduce structure and/or unit-specific covariates [4,17] can result in a difficulty in interpreting regression coefficients due to complex (but unconstrained) parameterizations. This is not our focus here, but the parameterizations in [26] and the one given here do not suffer from this issue. There has also been research on shrinkage to introduce stability in structured ways [9,10,12] or without structure [33,21].

These approaches can often be thought of in terms of specific decompositions of a covariance matrix. Our approach will focus on the variance/correlation decomposition, used recently in [3], which decomposes the covariance matrix  $\Sigma$  as  $\Sigma = DRD$ , where  $R$  is correlation matrix and  $D$  is diagonal matrix of standard deviations. Our approach here will rely on this decomposition and a further decomposition of the correlation matrix  $R$  into partial autocorrelations which we review next.

\* Corresponding author.

E-mail addresses: [yanpin@ufl.edu](mailto:yanpin@ufl.edu) (Y. Wang), [mjdaniels@austin.utexas.edu](mailto:mjdaniels@austin.utexas.edu) (M.J. Daniels).

Consider a  $p \times p$  correlation matrix  $R$  with  $(j, j+k)$ th element, the marginal correlation  $\rho_{jj+k} \equiv \text{Cor}(Y_j, Y_{j+k})$ . The matrix  $R$  can be re-parameterized using partial autocorrelations,  $\pi_{jj+k} \equiv \text{Cor}(Y_j, Y_{j+k} | Y_l, j < l < j+k)$ . To provide an expression for the partial autocorrelations as a function of the marginal correlations, we first define some notation (similar to [18]).

Let  $R^{(j,j+k)}$  be the  $(k+1) \times (k+1)$  submatrix of  $R$  which takes elements from the  $j$ th row to the  $(j+k)$ th row and the  $j$ th column to the  $(j+k)$ th column. Then we partition  $R^{(j,j+k)}$  as follows,

$$\begin{pmatrix} 1 & \mathbf{r}_1(j, k) & \rho_{j,j+k} \\ \mathbf{r}_1^T(j, k) & R_2(j, k) & \mathbf{r}_3^T(j, k) \\ \rho_{j+k,j} & \mathbf{r}_3(j, k) & 1 \end{pmatrix},$$

where  $\mathbf{r}_1(j, k) = (\rho_{j,j+1}, \dots, \rho_{j,j+k-1})$ ,  $\mathbf{r}_3(j, k) = (\rho_{j+k,j+1}, \dots, \rho_{j+k,j+k-1})$ , and  $R_2(j, k)$  contains the middle  $k-1$  rows and columns of  $R^{(j,j+k)}$ .

The partial autocorrelations have the following form as a function of the marginal correlations,

$$\begin{cases} \pi_{j,j+1} = \rho_{j,j+1} \\ \pi_{j,j+k} = \frac{\rho_{j,j+k} - r_{jk}}{B_{jk}} \quad 2 \leq k \leq p-j \end{cases} \quad (1)$$

where  $r_{jk} = \mathbf{r}_1(j, k)R_2^{-1}(j, k)\mathbf{r}_3^T(j, k)$  and  $B_{jk} = [1 - \mathbf{r}_1(j, k)R_2^{-1}(j, k)\mathbf{r}_1^T(j, k)][1 - \mathbf{r}_3(j, k)R_2^{-1}(j, k)\mathbf{r}_3^T(j, k)]^{1/2}$ . The marginal correlations,  $\rho_{j,j+k}$  can also be written as a simple function of the partial autocorrelations,

$$\rho_{j,j+k} = r_{jk} + \pi_{j,j+k}B_{jk}.$$

One of the advantages of this parameterization is that  $\pi_{jk}$  can vary independently in  $(-1, 1)$  while maintaining positive definiteness of  $R$ , unlike  $\rho_{jk}$  (see, e.g. [18]). Based on reparameterizing the marginal correlations into partial autocorrelations, Daniels and Pourahmadi [13] introduce a prior for  $R$  induced by independent uniform priors on the partial autocorrelations, i.e.,  $p(\pi) = 2^{-p(p-1)/2}$ .

After reparameterizing the off-diagonal elements of  $R = (\rho_{jk})$  in terms of the partial autocorrelations  $\{\pi_{jk}\}$ , we transform them to the entire real line using Fisher's  $z$  transformation. Moving from the constrained marginal correlations to the unconstrained transformed partial autocorrelations provides a link function framework similar to the theory of generalized linear models in [23]. The models proposed here will extend recent models from the literature for correlation matrices including the multivariate probit [6] and related models [11].

This article expands on previous work by describing how to do Bayesian inference in the proposed models, including an appropriate choice of priors. Under an improper prior on the mean parameters and monotone (ignorable) missingness, we provide conditions under which the posterior is proper. We also provide simulation results that demonstrate the importance of specifying the correct correlation structure under ignorable missingness and evaluate the ability of the DIC to select the correct model in these situations.

This article is arranged as follows. In Section 2, we introduce regression models for the partial autocorrelations and marginal variances. We derive and investigate priors for the regression parameters for the partial autocorrelation and marginal variance parameters in Section 3. We provide details on posterior computations in Section 4. Results of a simulation study to investigate correlation structure misspecification are given in Section 5. Application of the models to a schizophrenia clinical trial is presented in Section 6. Section 7 provides conclusions and extensions.

## 2. Models for the covariance matrix

Let  $Y_i : i = 1, \dots, n$  be a  $p \times 1$  vector of longitudinal responses measured (without loss of generality) at times  $1, \dots, p$  with distribution,

$$Y_i \sim N_p(x_i^T \beta, \Sigma_i), \quad (2)$$

where  $\beta$  is a vector of (mean) regression parameters with dimension  $p_\beta \times 1$ ,  $x_i$  is a  $p_\beta \times p$  covariate matrix, and  $\Sigma_i = D_i R_i D_i^T$ . We build regression models for  $R_i$  via the partial autocorrelations and  $D_i$ , via the marginal standard deviations in the following subsections.

### 2.1. Partial autocorrelations

Consider the following regression model for  $\pi_{i,jk}$ , the  $jk$ -th partial autocorrelation for subject  $i$ ,

$$z(\pi_{i,jk}) = w_{i,jk}^* \gamma, \quad (3)$$

where  $z(\cdot)$  is Fisher's  $z$ -transform,  $z(\pi) = \frac{1}{2} \log \frac{1+\pi}{1-\pi}$  and  $w_{i,jk}^*$  is a  $1 \times q$  vector of covariates to model structure and subject-level covariates;  $\gamma$  is unconstrained in  $q$ -dimensional real space  $\mathbb{R}^q$ . Given that the partial autocorrelations are correlations between longitudinal observations, conditional on intermediate ones, we might expect higher order ones to be zero. For example, we might specify  $w_{i,jk}^* = I(|k-j| = 1)$  corresponding to an AR(1) structure with all lag partial autocorrelations bigger than one equal to zero. The design vector,  $w_{i,jk}^{*T} = (1, |k-j|)$  implies that  $\tilde{\Pi}$ , the corresponding partial autocorrelation matrix, which has 1's on the main diagonal and  $(j, k)$ -th element  $\pi_{jk}$  (for  $j, k \neq 1$ ) has a Toeplitz form with the  $z$ -transform of the elements on each subdiagonal having a linear relationship in lag. The design vector,

$w_{i,jk}^* = (I\{|k-j|=1\}, I\{|k-j|=1\}j, I\{|k-j|>1\}, I\{|k-j|>1\}|k-j|)$  implies that the vector of  $\pi$ 's has nonstationary structure with the  $z$ -transform of the lag one correlations a linear function of time and a stationary form for the rest of the matrix, with the  $z$ -transform of the elements for each lag (after the first one) having a linear relationship in lag. For related structures for the parameters of the modified Choleski decomposition, see [26,28].

## 2.2. Marginal variances

We assume the logarithms of the marginal standard deviations,  $\sigma_{i,j}$  (i.e., the  $j$ th diagonal element of  $D_i$ ) follow the regression models,

$$\log(\sigma_{i,j}) = A_{i,j}\eta, \quad (4)$$

where  $A_{i,j}$  is a  $1 \times q_0$  vector of covariates to model structure and unit-level covariates. For example,  $A_{(i,j)} = (I(j=1), I(j>1))$  induces a structure of equal variance except for time 1;  $A_{(i,j)} = (1, j)$  corresponds to the marginal variances that are log linear in time. Verbyla [31] proposed models for the marginal variances (residual variances) in terms of unit level covariates (i.e., heterogeneity) in the setting of independent responses.

## 3. Priors for $\gamma$

Standard diffuse priors for  $\gamma$  in (3), e.g., an improper uniform prior on  $\mathbb{R}^q$  or a diffuse normal prior, result in most of the mass for the partial autocorrelations,  $\pi_{i,jk}$  being put at  $-1$  and  $+1$ . This happens in many settings with diffuse priors on transformed spaces, e.g., coefficients in logistic regression (see [1]). These are not sensible prior beliefs. In the next two subsections, we will review a prior proposed for the partial autocorrelations from [13] and propose alternative ones that both avoid this behavior for an unstructured vector of  $\pi$ 's. We then propose a way to use these priors, which are both within the class of independent transformed Beta priors, to construct priors for  $\gamma$  and point out their connections to  $g$ -priors. We also construct a similar prior for  $\eta$  in (4).

### 3.1. Review of priors for unstructured partial autocorrelations

Independent uniform priors on the partial autocorrelations can be expressed as independent Beta priors on the interval  $(-1, 1)$ ,  $Beta_{(-1,1)}(a, b)$ , with parameters  $a = 1$  and  $b = 1$ . These priors induce desirable behavior for longitudinal (ordered) data by shrinking higher lag marginal correlations toward zero [13] and a priori favoring serial correlation often seen in longitudinal data [24]. The behavior can be understood by examining the determinant of the Jacobian from  $\rho$  to  $\pi$ ,  $J(\rho \rightarrow \pi)$  [18],

$$J(\rho \rightarrow \pi) = \prod_{j=1}^{p-1} (1 - \pi_{j,j+1})^{-(p-2)/2} \prod_{k=2}^{p-2} \prod_{j=1}^{p-k} (1 - \pi_{j,j+k}^2)^{-(p-1-k)/2}.$$

As lag ( $k$ ) increases, more mass is placed toward zero. This is not surprising since most priors on  $R(\rho)$  do not use information on potential ordering of the responses and would induce this prior form on partial correlations to obtain identical marginal priors for the marginal correlations [3]. However, the  $Beta_{(-1,1)}(1, 1)$  prior does *not* favor positive correlations as we typically see in longitudinal data.

### 3.2. An alternative prior for unstructured partial autocorrelations

Here we introduce a prior on the partial autocorrelations that favors positive correlations over negative correlations. We propose independent priors on  $\pi_{jk}$  with pdf's,  $p(\pi_{jk}) = \frac{1+\pi_{jk}}{2}$ , which is a  $Beta_{(-1,1)}(2, 1)$  distribution; we refer to these as triangular priors given their shape. The implied marginal priors for  $\rho_{jk}$  are given in Fig. 1(a). The priors have decreasing mass close to 1 as lag ( $|j-k|$ ) increases. This is consistent with serial correlation often seen in longitudinal data and places more mass on *positive correlations* than the  $Beta_{(-1,1)}(1, 1)$  priors.

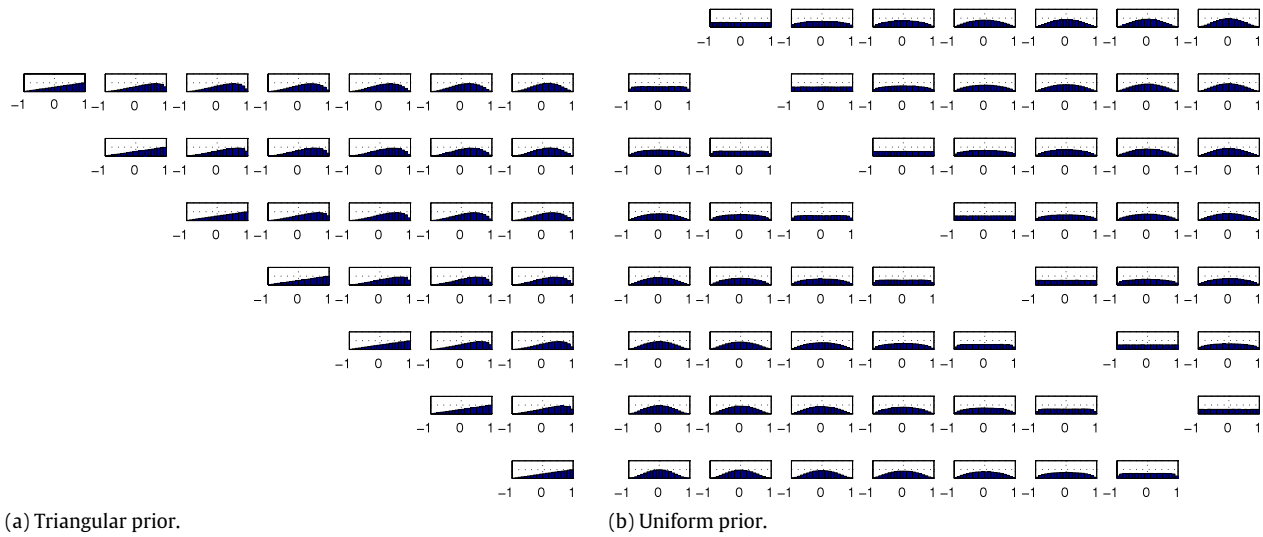
In the following section, we will use these two priors as a starting point to construct a prior for the regression coefficients,  $\gamma$  in (3). In the remainder, all Beta priors will be specified on the interval  $(-1, 1)$ , but we just denote them as  $Beta(a, b)$ .

### 3.3. Proposed prior on $\gamma$

In the following, when it is not necessary we will drop the subscripts on  $\pi_{jk}$ . We start by deriving the distribution of  $z(\pi_{jk})$  when the  $\pi_{jk}$  follow independent  $Beta(1, 1)$  priors. For this prior on  $\pi$ ,  $z(\pi) = z = \frac{1}{2} \log \frac{1+\pi}{1-\pi}$ , with pdf,

$$f(z) = \frac{2e^{2z}}{(1+e^{2z})^2},$$

where  $z \in (-\infty, +\infty)$ . This is the pdf of a logistic distribution,  $z \sim \text{logistic}(0, \frac{1}{2})$  with variance  $\pi/12$ . It is well known that the logistic distribution can be approximated with a  $t$ -distribution [2]. However, the easy to use construction of the multivariate  $t$ -distribution as a gamma mixture of normals has  $t$ -marginals but they are not independent as we required



**Fig. 1.** Marginal priors for the marginal correlations induced by (a) the triangular prior and (b) the uniform prior. In (b), the upper triangles are the marginal priors on the  $\rho$ 's induced by the original priors; the lower triangles are the marginal priors on  $\rho$ 's after the normal approximation.

based on our original specification of independent Betas. As a result, we will use a normal approximation to the logistic distribution, whose multivariate version does have independent marginals,  $z \sim N(0, \frac{\pi}{12})$ ; that is, the random vector  $\mathbf{z} \sim N(0, \frac{\pi}{12} I_{T \times T})$ , where  $T = \frac{p(p-1)}{2}$ .

Fig. 1(b) shows how well the normal prior approximates the original uniform prior on the hypercube (i.e., the independent Beta(1,1) priors) in terms of the marginal correlations. The upper triangular elements represent the marginal priors of  $\rho_{jk}$  from the original uniform prior and the lower triangular elements represent the marginal priors of  $\rho_{jk}$  from the prior based on the normal approximation. The approximate prior appears to behave sufficiently similarly.

Now, we show how this prior can be used to construct a prior for  $\gamma$  in (3). We first focus on the case of  $z(\pi_{i,jk}) = z(\pi_{jk})$  and for ease of notation, let  $z_{jk} = z(\pi_{jk})$  and  $\mathbf{z} = (z_{12}, \dots, z_{1p}, z_{23}, \dots, z_{2p}, \dots, z_{p-1p})^T$ . Consider the full rank linear transformation  $\mathbf{z} = \mathbf{w}\gamma$ , where

$$\mathbf{w} = \begin{pmatrix} \mathbf{w}^* & \mathbf{w}^\perp \end{pmatrix}.$$

$\mathbf{w}^*$  is a  $T \times q$  ( $T \geq q$ ) full column rank matrix corresponding to the regression in (3). The matrix  $\mathbf{w}^\perp$  is a  $T \times (T - q)$  full column rank matrix such that  $(\mathbf{w}^*)^T \times (\mathbf{w}^\perp) = \mathbf{0}_{q \times (T-q)}$  and  $(\mathbf{w}^\perp)^T \times \mathbf{w}^\perp = I_{(T-q) \times (T-q)}$ .

Therefore

$$\gamma = \begin{pmatrix} \gamma^* \\ \gamma^\perp \end{pmatrix} = \begin{pmatrix} \mathbf{w}^* & \mathbf{w}^\perp \end{pmatrix}^{-1} \mathbf{z},$$

with

$$\begin{pmatrix} \mathbf{w}^* & \mathbf{w}^\perp \end{pmatrix}^{-1} = \begin{pmatrix} ((\mathbf{w}^*)^T \mathbf{w}^*)^{-1} (\mathbf{w}^*)^T \\ (\mathbf{w}^\perp)^T \end{pmatrix}.$$

We define  $E(\mathbf{z}) = \mu$  and  $\text{Var}(\mathbf{z}) = \sigma^2$  based on the multivariate normal prior on  $\mathbf{z}$ . Under the Beta(1,1) prior on  $\pi$ ,  $\mu = 0$  and  $\sigma^2 = \pi/12$ ; under the Beta(2,1) prior (triangular prior),  $\mu = \frac{1}{2}$  and  $\sigma^2 = 0.5722$ . The corresponding prior for  $\gamma$  is multivariate normal with mean and variance given below,

$$E(\gamma) = \mu \begin{pmatrix} ((\mathbf{w}^*)^T \mathbf{w}^*)^{-1} (\mathbf{w}^*)^T \\ (\mathbf{w}^\perp)^T \end{pmatrix} \mathbf{1}_{T \times 1}$$

and

$$\text{Var}(\gamma) = \sigma^2 \begin{pmatrix} ((\mathbf{w}^*)^T \mathbf{w}^*)^{-1} & \mathbf{0} \\ \mathbf{0} & I_{(T-q) \times (T-q)} \end{pmatrix}.$$

The resulting prior for  $\gamma^*$  is also multivariate normal with expectation,  $\mu(\mathbf{w}^{*T} \mathbf{w}^*)^{-1} \mathbf{w}^{*T} \mathbf{1}_{T \times 1}$  and variance,  $\sigma^2 (\mathbf{w}^{*T} \mathbf{w}^*)^{-1}$ . The dimension reduction from  $\mathbf{z}$  to  $\gamma^*$  results in the prior variance being too small. To see this note that the variance of  $i$ th component of  $\mathbf{z}$  in  $\mathbf{z} = \mathbf{w}\gamma$  is

$$\begin{aligned} \text{var}(z_{i(\gamma)}) &= \text{var}(w_{i(\cdot)} \gamma) \\ &= \sigma^2 (\mathbf{w}_{i(\cdot)}^* ((\mathbf{w}^*)^T \mathbf{w}^*)^{-1} (\mathbf{w}_{i(\cdot)}^*)^T + \mathbf{w}_{i(\cdot)}^\perp (\mathbf{w}_{i(\cdot)}^\perp)^T). \end{aligned}$$

The  $i$ th component of  $z = w^* \gamma^*$  has variance

$$\begin{aligned} \text{var}(z_{i(\gamma^*)}) &= \text{var}(w_{(i,:)}^* \gamma^*) \\ &= \sigma^2 (w_{(i,:)}^* ((w^*)^T w^*)^{-1} (w_{(i,:)}^*)^T). \end{aligned}$$

Clearly,  $\text{var}(z_{i(\gamma)}) > \text{var}(z_{i(\gamma^*)})$ . It is easy to adjust for this by noting that the average variance of  $z(\pi_{jk}) = w_{jk}^* \gamma^*$ ,  $\overline{\text{var}(z)}$  is

$$\begin{aligned} \overline{\text{var}(z)} &= \frac{1}{T} \sum_{i=1}^T \text{var}(z_i) \\ &= \frac{q}{T} \sigma^2, \end{aligned}$$

where  $\sigma^2$  is the desired variance. Hence we can inflate  $\text{var}(\gamma^*)$  by a factor of  $\frac{T}{q}$ . The resulting prior for  $\gamma^*$  is

$$\gamma^* \sim N \left( \mu ((w^*)^T w^*)^{-1} (w^*)^T \mathbf{1}_{T \times 1}, \frac{T}{q} \sigma^2 ((w^*)^T w^*)^{-1} \right).$$

See the supplementary materials for additional details on the above derivations.

### 3.4. Extension to unit-specific covariates

We can easily extend this prior to unit-specific covariates. Suppose, for  $i = 1, \dots, n$ ,

$$z(\pi_{i,jk}) = w_{i,jk}^* \gamma.$$

Let  $w_i^*$  be a  $p \times q$  matrix such that  $z_i = w_i^* \gamma$ . We first stack  $z_1, \dots, z_n$  and  $w_1^*, \dots, w_n^*$  together, i.e.,  $z = (z_1^T, z_2^T, \dots, z_n^T)^T$ , and  $w^* = (w_1^{*T}, w_2^{*T}, \dots, w_n^{*T})^T$ . So, we have  $z = w^* \gamma$  and  $w^*$  is  $nT \times q$  full column rank matrix. Similar to the previous case, we obtain

$$\gamma \sim N \left( \mu \left( \sum_{i=1}^n w_i^{*T} w_i^* \right)^{-1} w^{*T} \mathbf{1}_{nT \times 1}, \frac{nT}{q} \sigma^2 \left( \sum_{i=1}^n w_i^{*T} w_i^* \right)^{-1} \right), \quad (5)$$

which is our recommended prior in the general case.

### 3.5. Connection to g-priors

Our priors on  $\gamma$  have similar form to the  $g$ -priors introduced in [35]. However, our derivation begins with a prior on an unconstrained parameter space as opposed to Zellner's construction of a prior based on the posterior distribution of imaginary data  $y_0, y_0 = x^T \beta + \epsilon$  where  $\epsilon \sim N(0, \sigma_0^2 I_n)$  (with independent priors on  $\beta \propto 1$  and  $\sigma_0 \propto \frac{1}{\sigma_0}$ ). The Zellner prior for  $\beta | \sigma_0$  has the form  $\beta \sim N \left( \hat{\beta}_0, \frac{\sigma_0^2}{g} (xx^T)^{-1} \right)$ , where  $\hat{\beta}_0$  is the least squares estimate based on the imaginary data and  $g$  is a penalty parameter; in practice, the mean is typically set to zero so no imaginary data is actually required. Our prior has a similar form but it is based on the projection of  $z(\pi_i)$  on  $w_i^*$  with weights based on the original prior for  $\pi$  on the unconstrained space (here a hypercube). The 'weights' based on the prior in (5) come in through  $\mu$  in the prior mean,  $(\sum_{i=1}^n w_i^{*T} w_i^*)^{-1} w^{*T} \mu \mathbf{1}_{nT \times 1}$  and  $\sigma^2$  in the prior variance,  $\sigma^2 (w^{*T} w^*)^{-1}$ . As a result, with these priors, we do not have to deal with the issue of the choice of  $g$  (for some discussion, see [16,5]).

### 3.6. Prior for $\eta$

The most commonly used prior on the marginal variances  $\sigma^2$  is the inverse gamma prior, which facilitates computations due to conditional conjugacy. Daniels [7] used a uniform prior on the transformed innovation (IV) parameters with or without structure similar to the model (4) in Section 2.2 for the marginal standard deviations. Barnard et al. [3] discussed independent normal priors on logarithmic transformed  $\sigma$ . In particular, they proposed the following prior

$$\log(\sigma_i) \sim N(\xi, \Lambda), \quad (6)$$

with  $\Lambda$  diagonal. We will derive a prior for  $\eta$  similar to that for  $\gamma$  based on Barnard et al.'s prior for the marginal standard deviations. The resulting prior is

$$\eta \sim N \left( \lambda \left( \sum_{i=1}^n A_i^T A_i \right)^{-1} A^T \mathbf{1}_{np \times 1}, \tau^2 \frac{np}{q_0} \left( \sum_{i=1}^n A_i^T A_i \right)^{-1} \right).$$

Note in the derivation, we have assumed  $\xi = \lambda \mathbf{1}_{p \times 1}$  and  $\Lambda = \tau^2 I_{p \times p}$  in (6), where  $\lambda$  and  $\tau$  are fixed a priori.



#### 4. Posterior distribution and computations

The full data likelihood,  $L(\eta, \beta, \gamma | y)$  is proportional to

$$\sum_{i=1}^n |D_i(\eta)R_i(\gamma)D_i(\eta)|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \sum_{i=1}^n R_i(\gamma)^{-1} D_i(\eta)^{-1} (Y_i - x_i \beta) (Y_i - x_i \beta)^T D_i(\eta)^{-1} \right] \right\}.$$

We specify the following priors for  $\beta$ ,  $\eta$ , and  $\gamma$ ,

$$\beta \propto 1 \quad (7)$$

$$\eta \sim N \left( \lambda_0 \left( \sum_{i=1}^n A_i^T A_i \right)^{-1} A^T \mathbf{1}_{np \times 1}, \tau^2 \frac{np}{q_0} \left( \sum_{i=1}^n A_i^T A_i \right)^{-1} \right) \quad (8)$$

$$\gamma \sim N \left( \mu \left( \sum_{i=1}^n w_i^{*T} w_i^* \right)^{-1} w^{*T} \mathbf{1}_{nT \times 1}, \frac{nT}{q} \sigma^2 \left( \sum_{i=1}^n w_i^{*T} w_i^* \right)^{-1} \right) \quad (9)$$

where  $\sum_{i=1}^n A_i^T A_i$  and  $\sum_{i=1}^n w_i^{*T} w_i^*$  are non-singular. In the setting of incomplete longitudinal responses, under an assumption of ignorable dropout (monotone missingness), we only need to specify the full data response model and the likelihood of interest is the observed data likelihood,  $L(\beta, \gamma, \eta | y_{obs}, x)$ , where the observed data response is  $y_{obs}$  [8]; the form of the observed data likelihood is given in the supplementary materials.

Since we specify an improper prior on  $\beta$ , we need to prove the posterior distribution of  $(\beta, \gamma, \eta)$  is proper. In the next section, we provide a theorem which gives simple sufficient conditions under which the posterior is proper. The supplementary materials contain details on the MCMC algorithm to sample from the posterior distribution.

##### 4.1. Posterior propriety

In the following theorem, we state conditions that are sufficient for the posterior to be proper. First, we need to introduce some notation. Suppose full-data  $Y_i : i = 1, \dots, n$  are independently distributed random variables with distribution  $Y_i \sim N_p(x_i^T \beta, \Sigma_i)$ , where  $x_i$  is a  $p_\beta \times p$  covariate matrix,  $\beta$  is a  $p_\beta \times 1$  (mean) regression parameter vector,  $\Sigma_i = D_i^{\frac{1}{2}}(\eta)R_i(\gamma)D_i^{\frac{1}{2}}(\eta)$ , and  $D_i(\eta) = \text{diag}(\sigma_{i\eta}^2)$  with  $\sigma_{i\eta}^2 = (\sigma_{i1}^2, \sigma_{i2}^2, \dots, \sigma_{ip}^2)$  specified by (4) and  $R_i(\gamma)$  specified by (1) and (3); define  $\mathfrak{S}_\beta, \mathfrak{S}_\gamma, \mathfrak{S}_\eta$  to be sample spaces of  $\beta, \gamma, \eta$ , respectively. Let  $(Q_{i1}, \dots, Q_{ip})^T$  be a vector of observed data indicators, where  $Q_{ik} = 1$  if  $Y_{ik}$  is observed (0 otherwise). Let  $Y_i^{k_i} = \{Y_{ij}, j = 1, \dots, k_i \text{ where } k_i : Q_{ik_i} = 1, Q_{i,k_i+1} = 0\}$  be the vector of observed data (of dimension  $k_i$ ) for subject  $i$  and  $S_k = \{i, Q_{ik_i} = 1 \text{ and } Q_{i,k_i+1} = 0, \text{ and } k_i = k, \text{ where } 1 \leq k \leq p-1; i = 1, \dots, n\}$  be the set of subjects with observed data of dimension  $k$ .

**Theorem 1.** We assume the observed data distribution for the  $i$ th subject ( $i = 1, \dots, n$ ) is  $Y_i^{k_i} \sim N_k(x_i^{k_iT} \beta, \Sigma_i^{k_i})$ , where  $x_i^{k_i} = x_{i[1:k_i]}$  is a  $p_\beta \times k_i$  submatrix of  $x_i$  and  $\Sigma_i^{k_i} = \Sigma_{i[1:k_i, 1:k_i]}$  is the  $k_i \times k_i$  principal submatrix of  $\Sigma_i$ . We also assume the priors on the parameters are given by (7)–(9) and missingness is monotone and ignorable. Then the posterior of  $(\beta, \gamma, \eta)$  will be proper under the following three (easy to check) conditions:

1.  $\sum_{i \in S_k} x_i^{k_i} x_i^{k_iT}$  is non-singular for all  $k \in \{1, 2, \dots, p-1\}$ .
2.  $\sum A_i A_i^T$  is non-singular.
3.  $\sum w_i^* w_i^{*T}$  is non-singular.

The proof is given in the supplementary materials. Note that the three conditions are conditions for the three design matrices in our model (for the mean, the variance, and the correlations, respectively); the latter two guarantee that the priors (8) and (9) are proper.

#### 5. Simulation

To assess the importance of the correlation structure on estimating the (mean) longitudinal trajectories in incomplete data, we conducted a simulation study. The true model was (2) with  $p = 6$ . For each individual, the rows of the mean design matrix were specified as a second order orthogonal polynomial. We set  $\beta = (27, -2.3, 0.50)^T$ . We considered three sample sizes (30, 100, and 400). For each scenario we simulated 200 data sets.

The true models for marginal variances and partial autocorrelation coefficients were given by

$$z(\pi_{jk}) = \gamma_1 I(|k-j| = 1 \cap j = 1) + \gamma_2 I(|k-j| = 2) + \gamma_3 I(|k-j| = 1 \cap j > 1) \quad (10)$$

**Table 1**

Posterior means of  $\beta$ : The values in first row, second row, and third row correspond to posterior means of  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ , respectively. The true value for  $\beta$  is (27.0, −2.0, 0.50). ‘Unstr’, ‘True’, ‘AR(1)’, and ‘Indep.’ represent unstructured model, true model, AR(1) model, and independence model, respectively.

Sample Size 30				Sample Size 100				Sample Size 400			
Unstr	True	AR(1)	Indep.	Unstr	True	AR(1)	Indep.	Unstr	True	AR(1)	Indep.
26.5	26.9	26.7	25.3	26.7	26.9	26.6	25.3	26.9	27.0	26.7	25.4
−2.1	−2.0	−2.1	−2.5	−2.0	−2.0	−2.1	−2.5	−2.0	−2.0	−2.1	−2.5
0.52	0.51	0.50	0.58	0.51	0.51	0.50	0.58	0.50	0.50	0.49	0.57

**Table 2**

Summary measures from the simulation: The values in rows correspond to sample size 30, 100, and 400, respectively. ‘Unstr’, ‘True’, ‘AR(1)’, and ‘Indep.’ represent unstructured model, true model, AR(1) model, and independence model, respectively. ‘Total MSE’ and ‘Change from Baseline’ correspond to the mean square errors of  $\beta$  and the change of the mean responses from the beginning to the end of study (True change is 12.0), respectively.

Total MSE				Change from Baseline			
Unstr	True	AR(1)	Indep.	Unstr	True	AR(1)	Indep.
6.9	6.5	6.6	10.1	12.6	12.0	12.7	15.1
1.8	1.7	2.0	4.8	12.2	12.0	12.7	14.9
0.42	0.41	0.52	3.4	12.1	12.0	12.8	15.0

and

$$\log \sigma_{jj} = \eta_1 I(j = 1) + \eta_2 I(j > 1),$$

with  $\gamma = (0.65, 0.21, 0.85)^T$ , and  $\eta = (150, 200)^T$ . The structure on  $\pi$  represents a second order model with the lag one partial autocorrelations constant except for time 1, the lag two partial autocorrelations constant over time, and higher lag partial autocorrelations equal to zero. The structure on the variances corresponds to a constant variance over time after time one.

After simulating the complete data, we induce ignorable missingness via the following missing data mechanism,

$$\text{logit}\{P(Q_{ik} = 1 | Q_{i,k-1} = 1, y_{obs})\} = \alpha_1 + \alpha_2 y_{k-1},$$

where  $Q_{jk} = I\{Y_{jk} \text{ is observed}\}$  and  $\alpha = (3.86, -0.05)$ .

We fit four models to the simulated data. For each model, we use the same true mean and marginal variance models, but different partial autocorrelation models. Our objective is to evaluate the impact of mis-specifying the partial autocorrelation model on inference on the marginal mean regression coefficients,  $\beta$ . Specifically, the models we compare are:

- (1) True structured model for  $\pi$  given in (10)
- (2) Independence model,  $\pi = 0$
- (3) AR(1) model:  $z(\pi_{jk}) = \gamma_1 I\{|k - j| = 1\}$
- (4) Unstructured model (no structure on  $\pi$ ).

For models 1, 3, and 4, we use a Beta(1,1) distribution in constructing the prior in (5); we note that the simulation results were similar when using a Beta(2,1). For the prior on  $\eta$ , we set  $(\lambda, \tau^2) = (0, 100)$ . For each model, we compute the DIC [30] based on the observed data likelihood [32].

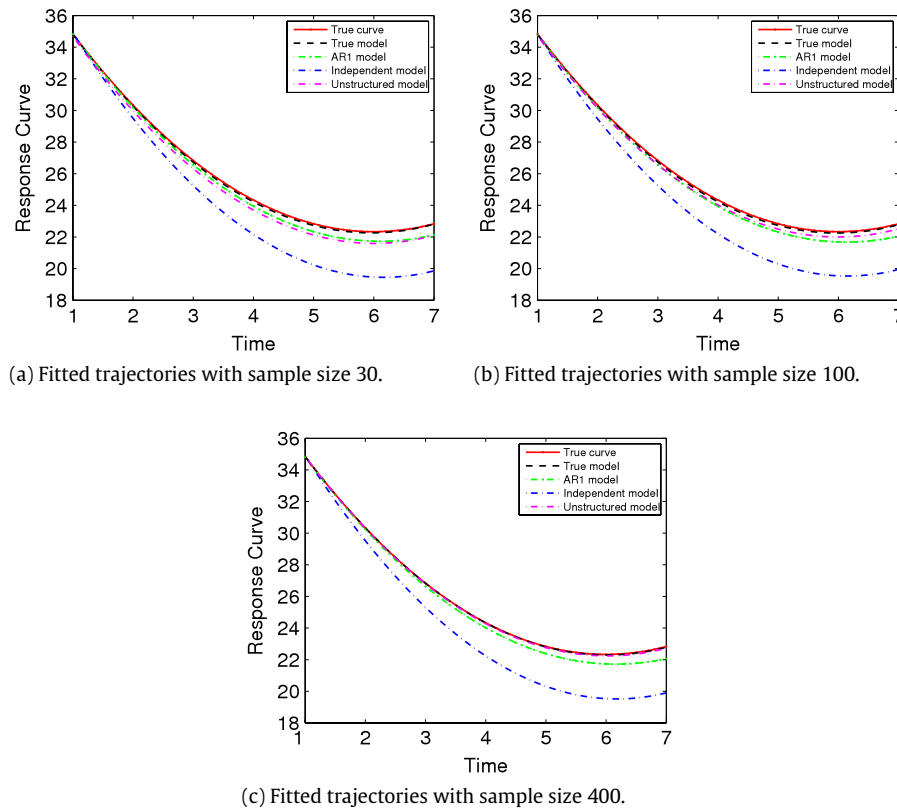
For each of the simulated datasets, we ran 20,000 iterations for each of the four models. For each dataset, we compute the DIC for the four correlation models and rank them (1 = best to 4 = worst) based on their fit (as measured by the DIC). To compare inference on the mean under all four models, we computed the following two quantities: (1) Total MSE, sum of mean squared error of the components of  $\beta$  and (2) Change from Baseline, change of estimated mean responses from time one to time six. We also compare the mean trajectories graphically.

## 5.1. Results

The simulation results are given in Tables 1 and 2 and Fig. 2. As the sample size increases, the estimates for  $\beta$  quickly approach the true value for the true structured correlation model, more slowly for the unstructured correlation model and to the wrong values for the AR(1) and independence correlation models (with the latter with considerable bias) (Table 1). The results in Table 2 are similar with bias in the estimate of change from baseline and larger MSE's for the estimates of the  $\beta$ 's for the incorrect models. Graphically, the fitted trajectories can be seen in Fig. 2 and illustrate the bias in the fitted trajectory when the correlation structure is incorrect.

The DIC chose the true model with high probability, with this probability generally increasing with sample size (see Table 3). For example, for  $n = 30, 100, 400$ , the true structured model is chosen by the DIC with probabilities 0.66, 0.99, and 0.95, respectively. It is interesting to note that the main competitor of the true model in the smaller sample sizes (30, 100) is the (parsimonious) AR(1) structure, while for the larger sample size (400) it is the unstructured model. This is the reason that the probability of the true structured model decreases between  $n = 100$  and  $n = 400$ . For the larger sample sizes, AR(1) is no longer a reasonable competitor, but the unstructured is (though it is only chosen with probability 0.06 for  $n = 400$ ). And note, of course, the unstructured model is correct, but it has more free parameters than needed.





**Fig. 2.** Posterior mean of the trajectories for the unstructured model, true model, AR(1) model, and independence model with sample size 30, 100, and 400.

**Table 3**

Percentage of times each model is chosen as best (row 1), second best (row 2), third best (row 3), and worst (row 4). 'AR(1)', 'Indep', 'True', and 'Unstr' correspond to the true model, AR(1) model, independence model, and the unstructured model, respectively.

Sample Size 30				Sample Size 100				Sample Size 400			
AR(1)	Ind	True	Unstr	AR(1)	Ind	True	Unstr	AR(1)	Ind	True	Unstr
0.35	0.00	0.66	0.00	0.01	0.00	0.99	0.00	0.00	0.00	0.95	0.06
0.66	0.00	0.35	0.00	0.73	0.00	0.01	0.26	0.00	0.00	0.06	0.95
0.00	0.15	0.00	0.85	0.26	0.00	0.00	0.74	1.00	0.00	0.00	0.00
0.00	0.85	0.00	0.15	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00

## 6. Data example: Schizophrenia trial

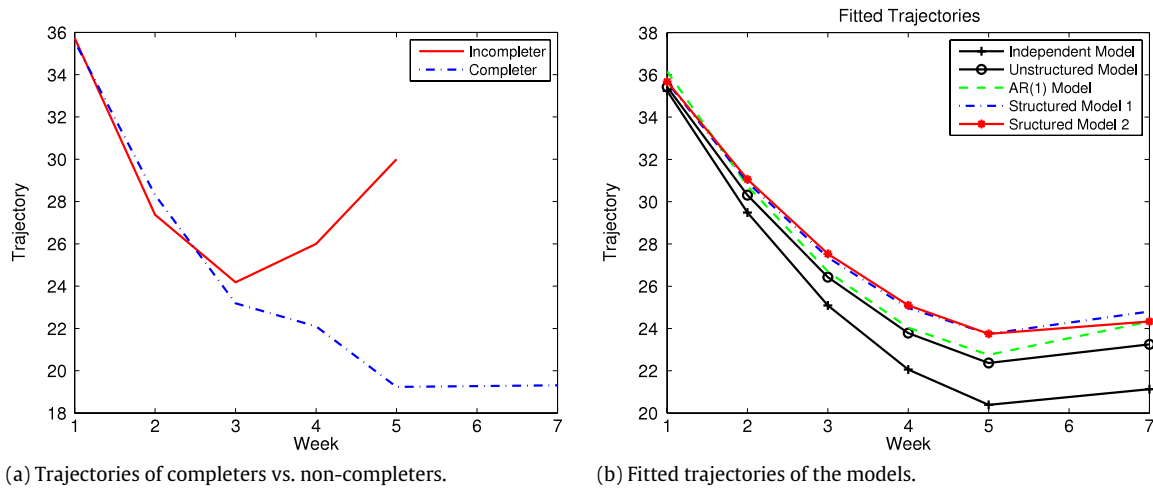
The data were collected as part of a randomized, double-blind clinical trial for a “new” pharmacologic treatment of schizophrenia [19]. The trial compared three doses of the “new” treatment (low, medium, high) to the standard dose of haloperidol, an effective antipsychotic that had known side effects. At the time of the study, the trial was designed to find the appropriate dosing level since the experimental therapy was thought to have similar antipsychotic effectiveness with fewer side effects. Two hundred forty-five patients were enrolled and randomized to one of the four treatment arms. The intended length of follow-up was 6 weeks, with measures taken weekly except for week 5. Schizophrenia severity was assessed using the Brief Psychiatric Rating Scale (BPRS) a sum of scores of 18 items that reflect behaviors, mode, and feelings. The scores ranged from 0 to 108 with higher scores indicating higher severity. To enter the study, the BPRS score had to be no less than 20. We will illustrate our approach using only the medium dose arm. Of main inferential interest is the change in BPRS from the beginning to the end of the study.

The dropout rate on the medium dose arm was high, with only 40 out of 61 (about 66%) participants having a measurement at week 7 (the sixth measurement time). Reasons for dropout included adverse events (e.g., side effects), lack of treatment effect, and withdrawal for unspecified reasons. The trajectories of completers vs. non-completers is shown in Fig. 3(a). Clearly those dropping out were doing worse prior to dropping out (higher BPRS).

### 6.1. Models

Let the longitudinal vector of outcomes for subject  $i$  be  $Y_i = (Y_{i1}, \dots, Y_{i6})^T$ , measured at weeks  $\mathbf{t} = (t_1, \dots, t_6) = (1, 2, 3, 4, 5, 7)$ . We assume  $Y_i$  follows (2) with mean.

$$E(Y_{ij}) = \beta_0 + \beta_1 x_1^j + \beta_2 x_2^j,$$



**Fig. 3.** Trajectories of the observed data and posterior mean of the trajectories for the models considered.

**Table 4**

MLE of variances (on main diagonal), partial autocorrelations (in upper triangle), and Fisher's z-transformation of partial autocorrelations (lower triangle).

126.25	0.6578	−0.0738	0.0804	−0.0253	−0.5230
0.7889	210.35	0.8543	−0.0593	−0.3328	0.0292
−0.0740	1.2718	224.42	0.8559	0.2648	0.4375
0.0806	−0.0594	1.2779	240.84	0.8961	0.3506
−0.0253	−0.3460	0.2713	1.4522	221.98	0.8433
−0.5805	0.0292	0.4692	0.3661	1.2325	243.08

where  $x_1^j = (t_j - \bar{t})$  and  $x_2^j = (t_j - \bar{t})^2 - \frac{\sum_{k=1}^6 (t_k - \bar{t})^3}{\sum_{k=1}^6 (t_k - \bar{t})^2} (t_j - \bar{t}) - \frac{\sum_{k=1}^6 (t_k - \bar{t})^2}{6}$ , i.e., an orthogonal quadratic polynomial. We assume missingness is ignorable.

We fit the partial autocorrelation models given below:

Independence Model:  $z(\pi_{jj+k}) = 0$ ,  $\log(\sigma_j) = I(j = 1)\eta_1 + I(j > 1)\eta_2$ .

AR(1) Model:  $z(\pi_{jj+k}) = I(k = 1)\gamma_1$ ,  $\log(\sigma_j) = I(j = 1)\eta_1 + I(j > 1)\eta_2$ .

Unstructured Covariance Model:  $z(\pi_{jj+k}) = \gamma_{jj+k}^{(+)}$  (with  $\gamma = (\gamma_{12}^{(+)}, \gamma_{13}^{(+)}, \gamma_{p-1,p}^{(+)})$ ),  $\log(\sigma_j) = \eta_j$ .

Structured Model 1:

$$z(\pi_{jj+k}) = I(k = 1 \cap j < 2)\gamma_1 + I(k = 1 \cap j > 1)\gamma_2 + I(k = 2)\gamma_3,$$

$$\log(\sigma_j) = I(j = 1)\eta_1 + I(j > 1)\eta_2.$$

Structured Model 2:

$$z(\pi_{jj+k}) = I(k = 1 \cap j < 2)\gamma_1 + I(k = 1 \cap j > 1)\gamma_2 + I(k = 2 \cap j < 3)\gamma_3 \\ + I(k = 2 \cap j > 2)\gamma_4 + I(k = 3)\gamma_5 + I(k = 4)\gamma_6 + I(k = 5)\gamma_7,$$

$$\log(\sigma_j) = I(j = 1)\eta_1 + I(j > 1)\eta_2.$$

Structured Model 1 is the same model as the one considered in the simulation. Structured Model 2 is more flexible than Structured Model 1 for the partial autocorrelations allowing nonstationary lag one and lag two autocorrelations and stationary lag three, four and five (with no structural zeros). The structure on the variances is the same as Structured Model 1. The marginal variance structure for all the models and the partial autocorrelation structures for Structured Models 1 and 2 were chosen after examining the unstructured covariance matrix in Table 4.

We use priors specified in (7)–(9) for  $\beta$ ,  $\eta$ ,  $\gamma$ , respectively. For (9), we consider both the  $Beta(1, 1)$  (uniform) and  $Beta(2, 1)$  (triangular) for the specification of  $(\mu, \sigma^2)$ . For (8), we set  $\lambda = 0$  and  $\tau^2 = 100$ .

## 6.2. Results

For all models, we ran 200,000 iterations with minimal burn-in since the chains converged after a few iterations. The plot of all fitted mean trajectories is given in Fig. 3(b) (only shown under the triangular prior). The mean BPRS initially decreased but started to go back up by week 5. This is related to those dropping out doing more poorly than those staying in the study. Table 5 contains the posterior mean of  $\beta$ , the change from baseline to week 7, their 95% credible intervals, and the DIC based on the observed data likelihood. This table shows that models under uniform and triangular prior fit similarly. The changes from baseline in all models were negative with 95% credible interval excluding 0, showing that Medium-dose reduced the

**Table 5**  
Posterior summaries of the models for the schizophrenia trial.

	$(\beta_0, \beta_1, \beta_2)$	Changes from Baseline (95% CI)	DIC
Independent	(25.6, −2.35, 0.68)	−14.1 (−18.7, −9.4)	1924.6
Unstructured	(26.9, −2.03, 0.62)	−12.2 (−16.3, −7.9)	1663.0
AR(1) (Uniform)	(27.5, −1.97, 0.69)	−11.8 (−15.8, −7.9)	1681.3
AR(1) (Triangular)	(27.5, −1.97, 0.69)	−11.8 (−15.7, −7.9)	1680.6
Structured Model 1(Uniform)	(27.9, −1.80, 0.58)	−10.8 (−14.6, −7.0)	1669.6
Structured Model 1(Triangular)	(27.9, −1.81, 0.58)	−10.8 (−14.7, −7.0)	1670.4
Structured Model 2(Uniform)	(27.9, −1.89, 0.54)	−11.3 (−15.3, −7.3)	1660.4
Structured Model 2(Triangular)	(27.9, −1.89, 0.55)	−11.3 (−15.4, −7.3)	1658.1

BPRS score significantly, which agrees with earlier analysis done in [8]. The changes from baseline varied from −14 to −11 based on the covariance model chosen. According to the DIC, Structured Model 2 with the triangular prior provided the best fit. The change from baseline in Structured Model 2 was almost a full point different from the unstructured model.

## 7. Discussion

In this paper, we first extended the priors in [13] for partial autocorrelations for the unstructured case by introducing a set of (triangular) priors which favor positive marginal correlations. Using Fisher's z-transformation on the partial autocorrelations, we introduced a regression framework to induce structure and/or unit-specific covariates in the correlation matrix. Based on priors proposed for the partial autocorrelations in the non-regression setting, we introduced a prior for the coefficients in the partial autocorrelation regressions (and for the coefficients of the marginal variance regressions). We conducted simulations that illustrated the importance of correct specification of the correlation structure in the setting of ignorable missingness in longitudinal data and show the ability of the DIC to choose the true correlation model. We also fit the models to data from a longitudinal schizophrenia clinical trial.

There are a variety of extensions to the modeling proposed here. Clearly, it can be difficult to 'find' a good parametric model that imposes structure on the correlation matrix. Thus extending approaches developed under different parameterizations [29,33] to our setting is an important extension. Correlation matrices (instead of covariance matrices) arise commonly in models for longitudinal data modeled using Gaussian copulas [25]; efficient computations using the partial autocorrelation in these settings will be a challenging problem due to the lack of conjugacy. However, the partial autocorrelation models provide an opportunity for flexible dependence in longitudinal categorical data via multivariate probit models. To offer some robustness to a selected model for the correlation structure, an alternative would be to shrink the partial autocorrelations to the structure using independent Beta priors as has been done previously using normal priors on other parameterizations of a covariance matrix [10,12]. Finally, we are considering extensions to irregular longitudinal data using a partial autocorrelation *function* and to time-dependent covariates.

## Acknowledgment

This work was partially supported by NIH grant CA-85295.

## Appendix. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jmva.2012.11.010>.

## References

- [1] A. Agresti, D.B. Hitchcock, Bayesian inference for categorical data analysis, *Statistical Methods and Applications* 14 (2005) 297–330.
- [2] J.H. Albert, S. Chib, Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association* 88 (1993) 669–679.
- [3] J. Barnard, R. McCulloch, X.L. Meng, Modeling covariance matrices in terms of standard deviations and correlations, *Statistica Sinica* 10 (2000) 1281–1312.
- [4] T.Y.M. Chiu, T. Leonard, K.W. Tsui, The matrix-logarithmic covariance model, *Journal of the American Statistical Association* 91 (1996) 198–210.
- [5] M. Clyde, E.I. George, Flexible empirical Bayes estimation for wavelets, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 62 (2000) 681–698.
- [6] C. Czado, Multivariate regression analysis of panel data with binary outcomes applied to unemployment data, *Statistical Papers* 41 (2000) 281–304.
- [7] M.J. Daniels, Bayesian modeling of several covariance matrices and some results on propriety of the posterior for linear regression with correlated and/or heterogeneous errors, *Journal of Multivariate Analysis* 97 (2006) 1185–1207.
- [8] M.J. Daniels, J.W. Hogan, *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Chapman and Hall/CRC, 2008.
- [9] M.J. Daniels, R.E. Kass, Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models, *Journal of the American Statistical Association* 94 (1999) 1254–1263.
- [10] M.J. Daniels, R.E. Kass, Shrinkage estimators for covariance matrices, *Biometrics* 57 (2001) 1173–1184.
- [11] M.J. Daniels, S.L. Normand, Longitudinal profiling and health care units based on mixed multivariate patient outcomes, *Biometrics* 7 (2006) 1–15.
- [12] M.J. Daniels, M. Pourahmadi, Bayesian analysis of covariance matrices and dynamic models for longitudinal data, *Biometrika* 89 (2002) 553–566.
- [13] M.J. Daniels, M. Pourahmadi, Modeling covariance matrices via partial autocorrelations, *Journal of Multivariate Analysis* 100 (2009) 2352–2363.

- [14] M.J. Daniels, Y.D. Zhao, Modeling the random effects covariance matrix in longitudinal data, *Statistics in Medicine* 22 (2003) 1631–1647.
- [15] B. Efron, C. Morris, Multivariate empirical Bayes and estimation of covariance matrices, *The Annals of Statistics* 4 (1976) 22–32.
- [16] E.I. George, D.P. Foster, Calibration and empirical Bayes variable selection, *Biometrika* 87 (2000) 731–747.
- [17] P.D. Hoff, X. Niu, A covariance regression model, *Statistica Sinica* 22 (2012) 729–753.
- [18] H. Joe, Generating random correlation matrices based on partial correlations, *Journal of Multivariate Analysis* 97 (2006) 2177–2189.
- [19] Y.D. Lapierre, N.P.V. Nair, G. Chouinard, A.G. Awad, B. Saxena, B. Jones, D.J. McClure, D. Bakish, P. Max, R. Manchanda, P. Beaudry, D. Bloom, E. Rotstein, R. Ancill, P. Sandor, N. Sladen-Dew, C. Durand, R. Chandrasena, E. Horn, D. Elliot, M. Das, A. Ravindran, G. Matsos, A controlled dose-ranging study of remoxipride and haloperidol in schizophrenia – a Canadian multicentre trial, *Acta Psychiatrica Scandinavica* 82 (1990) 72–77.
- [20] T. Leonard, J.S.J. Hsu, Bayesian inference for a covariance matrix, *The Annals of Statistics* 20 (1992) 1669–1696.
- [21] J.C. Liechty, M.W. Liechty, P. Muller, Bayesian correlation estimation, *Biometrika* 91 (2004) 1–14.
- [22] R.J.A. Little, D.B. Rubin, *Statistical Analysis with Missing Data*, John Wiley, New York, 2002.
- [23] P. McCullagh, J.A. Nelder, *Generalized Linear Models*, second ed., Chapman and Hall, 1989.
- [24] A. Munoz, V. Carey, J.P. Schouten, M. Segal, B. Rosner, A parametric family of correlation structures for the analysis of longitudinal data, *Biometrics* 48 (1992) 733–742.
- [25] R.B. Nelsen, *An Introduction to Copulas*, Springer, 1999.
- [26] M. Pourahmadi, Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation, *Biometrika* 86 (1999) 677–690.
- [27] M. Pourahmadi, Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix, *Biometrika* 87 (2000) 425–435.
- [28] M. Pourahmadi, M.J. Daniels, Dynamic conditionally linear mixed models for longitudinal data, *Biometrics* 58 (2002) 225–231.
- [29] M. Smith, R. Kohn, Parsimonious covariance matrix estimation for longitudinal data, *Journal of the American Statistical Association* 97 (2002) 1141–1153.
- [30] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, A. Linde, Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64 (2002) 583–639.
- [31] A.P. Verbyla, Modelling variance heterogeneity: residual maximum likelihood and diagnostics, *Journal of the Royal Statistical Society. Series B* 55 (1993) 493–508.
- [32] C. Wang, M.J. Daniels, A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data (correction, vol. 68, p. 994), *Biometrics* 67 (2011) 810–818.
- [33] F. Wong, C.K. Carter, R. Kohn, Efficient estimation of covariance selection models, *Biometrika* 90 (2003) 809–830.
- [34] R. Yang, J.O. Berger, Estimation of a covariance matrix using the reference prior, *The Annals of Statistics* 22 (1994) 1195–1211.
- [35] A. Zellner, On assessing prior distributions and Bayesian regression analysis with g-prior distributions, in: *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, 1986, pp. 233–243.