axioms

MDPI

Article

Longitudinal Data Analysis Based on Bayesian Semiparametric Method

Guimei Jiao [1,*,†], Jiajuan Liang [2,3,*,†], Fanjuan Wang [1,†], Xiaoli Chen [1,†], Shaokang Chen [1], Hao Li [1], Jing Jin [1], Jiali Cai [1] and Fangjie Zhang [1]

[1] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China; wangfj2021@lzu.edu.cn (F.W.); 18393919665@163.com (X.C.); chenshk21@lzu.edu.cn (S.C.); haoli21@lzu.edu.cn (H.L.); jinj21@lzu.edu.cn (J.J.); caijl21@lzu.edu.cn (J.C.); zhangfj21@lzu.edu.cn (F.Z.)
[2] Department of Statistics and Data Science, BNU-HKBU United International College, Zhuhai 519087, China
[3] Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, Zhuhai 519087, China
* Correspondence: gmjiao@lzu.edu.cn (G.J.); jiajuanliang@uic.edu.cn (J.L.)
† These authors contributed equally to this work.

Abstract: A Bayesian semiparametric model framework is proposed to analyze multivariate longitudinal data. The new framework leads to simple explicit posterior distributions of model parameters. It results in easy implementation of the MCMC algorithm for estimation of model parameters and demonstrates fast convergence. The proposed model framework associated with the MCMC algorithm is validated by four covariance structures and a real-life dataset. A simple Monte Carlo study of the model under four covariance structures and an analysis of the real dataset show that the new model framework and its associated Bayesian posterior inferential method through the MCMC algorithm perform fairly well in the sense of easy implementation, fast convergence, and smaller root mean square errors compared with the same model without the specified autoregression structure.

Keywords: Bayesian semiparametric method; covariance structure; Dirichlet process; linear mixed model; longitudinal data; MCMC algorithm

MSC: 62F10; 62F15; 62F40; 65C40

check for updates

Citation: Jiao, G.; Liang, J.; Wang, F.; Chen, X.; Chen, S.; Li, H.; Jin, J.; Cai, J.; Zhang, F. Longitudinal Data Analysis Based on Bayesian Semiparametric Method. *Axioms* **2023**, *12*, 431. https://doi.org/10.3390/axioms12050431

Academic Editor: J. Alberto Conejero

Received: 31 December 2022
Revised: 23 April 2023
Accepted: 24 April 2023
Published: 27 April 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Longitudinal data arise from repeated observations from the same individual or group of individuals at different time points. The structure of longitudinal data is shown in the following Table 1. The basic tasks of analyzing longitudinal data can be summarized as (1) studying the trend in the covariance structure of the observed variables with respect to time; (2) discovering the influence of covariates on the observable variables; and (3) determining the within-group correlation structures [1].

Table 1. Longitudinal data structures.

Observation Object	Number of Observations		
	1	**· · ·**	***k***
1	$t_{11}, x_{1,11}, \ldots, x_{p,11}, y_{11}$	$\cdots$	$t_{1k}, x_{1,1k}, \ldots, x_{p,1k}, y_{1k}$
$\vdots$	$\vdots$		$\vdots$
i	$t_{i1}, x_{1,i1}, \ldots, x_{p,i1}, y_{i1}$	$\cdots$	$t_{ik}, x_{1,ik}, \ldots, x_{p,ik}, y_{ik}$
$\vdots$	$\vdots$		$\vdots$
n	$t_{n1}, x_{1,n1}, \ldots, x_{p,n1}, y_{n1}$	$\cdots$	$t_{nk}, x_{1,nk}, \ldots, x_{p,nk}, y_{nk}$

Note: In Table 1, t_{ij} stands for the j-th observation time of the individual i; $(x_{1,ij}, \ldots, x_{p,ij})$ for the p-dimensional covariate of individual i at t_{ij}; and y_{ij} for the j-th observation of individual i.

Longitudinal data are often highly unbalanced. It is usually difficult to apply traditional multiple regression techniques to analyze highly unbalanced data directly. Statisticians developed various Bayesian statistical inference models for longitudinal data analysis [2]. A parametric assumption may result in modeling bias and may relax the assumption about parametric structure. Nonparametric methods have the characteristics of robustness because they do not require model assumptions. Semiparametric models integrate the characteristics of parametric and nonparametric models and have the characteristics of flexibility and ease of interpretation. Nonparametric statistical methods and semiparametric statistical methods are not only hot spots in current statistical research but also widely used in many practical applications. Xiang et al. [3] summarize some common outcomes of nonparametric regression analysis of longitudinal data. Bayesian methods for parametric linear mixed models have been widely used in different areas. Assuming a normal random effect and using the standard Gibbs sampling to realize some simple posterior inference, Quintana et al. [4] extend the general class of mixed models for longitudinal data by generalizing the GP part to the nonparametric case. The GP is a probabilistic approach to learning nonparametric models. Cheng et al. [5] propose the so-called LonGP, which is a flexible and interpretable nonparametric modeling framework. It provides a versatile software implementation that can solve commonly faced challenges in longitudinal data analysis. It also develops a fully Bayesian, predictive inference for LonGP, which can be employed to carry out model selection. Kleinman and Ibrahim [6] relax the normal assumption by assuming a Dirichlet process prior with a Gaussian measure of zero mean, semiparametrically modeling the random effects distribution [7]. Although a parametric model may have limitations in some applications, it is simpler than a nonparametric model whose scope may be too wide to draw a concise conclusion. Nonparametric regression has a fatal weakness, known as the "curse of dimensionality", which refers to the fact that when the independent variable X is multivariate, the estimation accuracy of the regression function becomes very poor as the dimension of X increases. To solve this problem, the semiparametric model is a good compromise maintaining some excellent characteristics of parametric and nonparametric models [8]. There is a lot of literature on the semiparametric modeling of longitudinal data. Most of the existing literature employs random effects to model within-group correlations [9].

Semiparametric linear mixed models generalize traditional linear mixed models by modeling a covariate effect with a nonparametric function and parametrically modeling other covariate effects. Semiparametric linear mixed models mainly use frequentist methods for statistical inference and assume normal random effects [10]. In this paper, we employ a Bayesian semiparametric framework for linear mixed models given by Quintana et al. [4] by imposing stronger conditions on the random effect to obtain an explicit solution to posterior distributions of the model parameters and fast convergence of the MCMC algorithm in the posterior inference of model parameters. The proposed framework generalizes the default priori of variance components and adjusts the inference of fixed effects associated with nonparametric random effects. The latter includes the extrapolation of nonparametric mean functions over time [11,12]. The stochastic process approach in [4] is a good choice for characterizing the intragroup correlation. The Gaussian process (GP) covariance has an exponential form and is uniquely determined by two parameters. GP can specify autoregressive correlation (the AR covariance structure, [13]). A nonparametric Dirichlet process (DP) priori assigned to the covariance parameter results in an Ornstein–Uhlenbeck process (OUP). A partial Dirichlet process mixture (DPM) is performed on the OUP. By imposing stronger conditions on the random effect and decomposing the OUP into some single Gaussian variables, we can substantially simplify the MCMC sampling in the posterior inference. The framework of our proposed semiparametric autoregression model (SPAR) is demonstrated in Figure 1.

This paper is organized as follows. Section 2 briefly introduces some basic theoretical background and principles. Section 3 introduces the model setup using a partial Dirichlet process mixture in terms of the OUP, and the Bayesian semiparametric autoregressive model

is proposed with a recommended solution. In Section 4, the Dirichlet process and Dirichlet process mixture (DPM) are simply introduced. The formulation of the semiparametric autoregressive model is given in Section 5. Section 6 derives the marginal likelihood, prior determination, and posteriori inference. Section 7 gives a simple Monte Carlo study and a real dataset application based on the proposed model. Some concluding remarks are given in the last section.

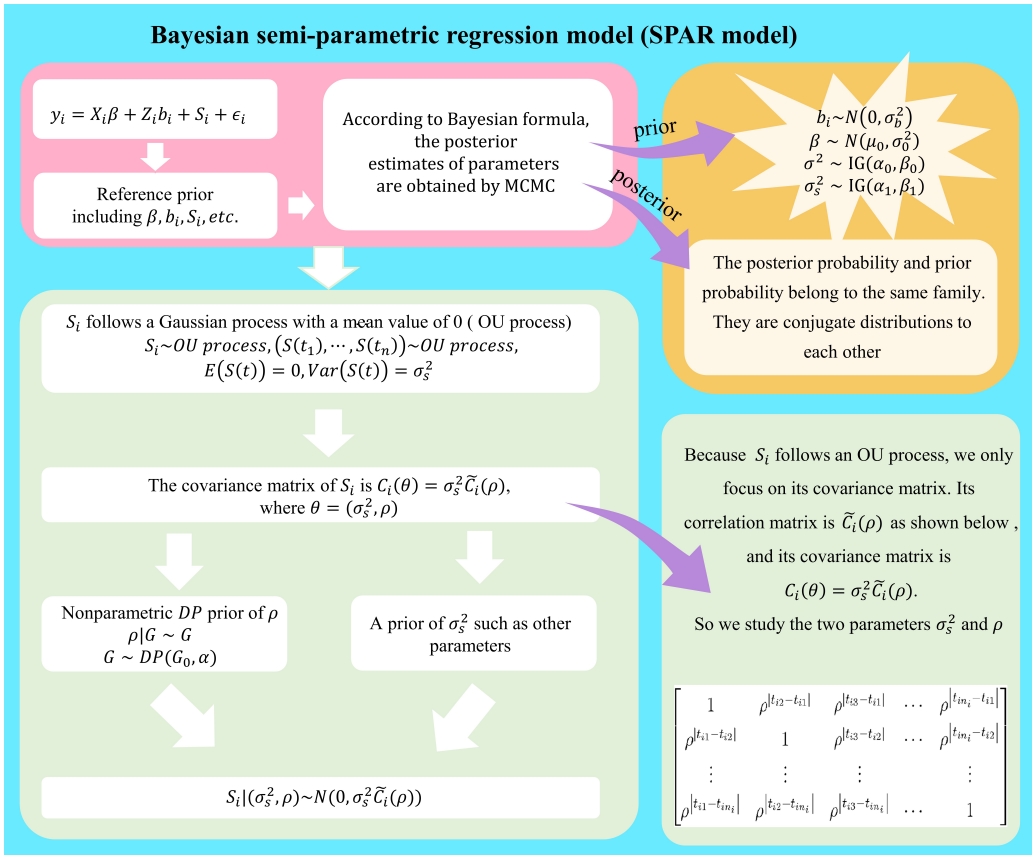


Figure 1. Bayesian semiparametric autoregression model (SPAR model).

2. Theoretical Basis

2.1. A General Linear Hybrid Model Containing an AR Structure

For an observation object i $(i = 1, 2, \ldots, n)$, we denote the observation time points by $\{t_{i1}, t_{i2}, \ldots, t_{in_i}\}$. Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$ $(n_i \times 1)$ be the observation vector from object i. At time point t_{ij}, we consider the possible time-dependent covariate vector $\boldsymbol{x}_{ij}' = (1, x_{i1}(t_{ij}), x_{i2}(t_{ij}), \ldots, x_{ip}(t_{ij}))$. Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$ $(n_i \times 1)$ be the observation vector from object i. At time point t_{ij}, we consider the possible time-dependent covariate vector $\boldsymbol{x}_{ij}' = (1, x_{i1}(t_{ij}), x_{i2}(t_{ij}), \ldots, x_{ip}(t_{ij}))$. Let $E(y_{ij}) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}$. Define the $n_i \times (p+1)$-dimensional fixed-effect design matrix $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{in_i})'$. Assume $E(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$. Define the corresponding $n_i \times q (q \leq p)$-dimensional random effects design matrix $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}, \boldsymbol{z}_{i2}, \ldots, \boldsymbol{z}_{in_i})'$, where $\boldsymbol{z}_{ij}' = (1, z_{i1}(t_{ij}), z_{i2}(t_{ij}), \ldots, z_{iq}(t_{ij}))$. The general linear mixed model containing an AR covariance structure is

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{S}_i + \boldsymbol{\epsilon}_i, \quad & \boldsymbol{S}_i|\boldsymbol{\theta} &\sim N(\boldsymbol{0}, \boldsymbol{C}_i(\boldsymbol{\theta})), \\ \boldsymbol{b}_i|\boldsymbol{\tau} &\sim N(\boldsymbol{0}, \boldsymbol{D}(\boldsymbol{\tau})), & \boldsymbol{\epsilon}_i &\sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n_i}), \end{aligned} \tag{1}$$

where $\boldsymbol{X}_i\boldsymbol{\beta}$ stands for the fixed effect and $\boldsymbol{Z}_i\boldsymbol{b}_i$ for the random effect, $\boldsymbol{S}_i$ is a stochastic process characterizing the correlation among the n_i observations from object i, $\boldsymbol{\epsilon}_i$ is the error term, $i = 1, 2, \ldots, n_i$. $\boldsymbol{C}_i(\theta)$ is an $n_i \times n_i$ structural covariance matrix, and vectors $\boldsymbol{\tau}$ and

$\boldsymbol{\theta}$ contain the variance–covariance parameters of the random vector $\boldsymbol{b}_i$ and the stochastic process $\boldsymbol{S}_i$, respectively. $\boldsymbol{I}_{n_i}$ stands for the identity matrix of dimension $n_i \times n_i$. The AR structure is usually specified by a GP with zero mean. The GP is uniquely determined by a covariance function containing parameter $\boldsymbol{\theta} = (\sigma_s, \rho)$. The vector $\boldsymbol{S}_i$ is a GP corresponding to the i-th object. $\boldsymbol{S}_i$ is generated sequentially by the GP $\{s_i(t) : t > 0\}$ at the observation time points $\{t_{i1}, t_{i2}, \ldots, t_{in_i}\}$, i.e.,

$$\boldsymbol{S}_i = \Big(s_i(t_{i1}), s_i(t_{i2}), \ldots, s_i(t_{in_i})\Big).$$

To specify the AR structure, it is assumed that the above GP possesses stationarity:

$$\mathrm{Cov}(s_i(t_{il}), s_i(t_{ik})) = \sigma_s^2 \rho(|t_{il} - t_{ik}|). \tag{2}$$

When the model has a structured covariance function, the covariance matrix of $\boldsymbol{y}_i$ is

$$\mathrm{Cov}(\boldsymbol{y}_i) = \boldsymbol{Z}_i \boldsymbol{D}(\boldsymbol{\tau}) \boldsymbol{Z}_i' + \boldsymbol{C}_i(\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I}_{n_i}. \tag{3}$$

2.2. The Principle for Bayesian Inference

The Bayesian method assumes a piori distribution on the unknown parameter θ and a joint distribution $p(\boldsymbol{x}, \theta)$ between an observable variable X and the parameter θ. The Bayesian method is based on the posterior distribution $\pi(\theta|\boldsymbol{x})$ of the unknown parameter θ after observed data are available. The joint distribution, the priori distribution, and the posterior distribution are related to each other as follows:

$$p(\boldsymbol{x}, \theta) = \pi(\theta|\boldsymbol{x}) p_X(\boldsymbol{x}),$$
$$p_X(\boldsymbol{x}) = \int_\Theta p(\boldsymbol{x}, \theta) \mathrm{d}\theta = \int_\Theta p(\boldsymbol{x}|\theta) \pi(\theta) \mathrm{d}\theta,$$

where $p_X(\boldsymbol{x})$ stands for the sample marginal density of $\boldsymbol{x}$. We use the posterior distribution $\pi(\theta|\boldsymbol{x})$ to carry out statistical inference for the parameter θ:

$$\pi(\theta|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \theta)}{p_X(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\theta)\pi(\theta)}{\int_\Theta p(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta} \tag{4}$$

Equation (4) is the well-known Bayesian formula. Bayesian statistical inference assumes that the posterior distribution of the unknown parameter θ, $\pi(\theta|\boldsymbol{x})$ contains all the available information. As a result, the point estimation, interval estimation, hypothesis testing, and predicting inference of θ can be implemented as usual. Because

$$\pi(\theta|\boldsymbol{x}) = \frac{\pi(\theta)p(\boldsymbol{x}|\theta)}{\int_\Theta \pi(\theta)p(\boldsymbol{x}|\theta)\mathrm{d}\theta} \propto \pi(\theta)p(\boldsymbol{x}|\theta),$$

we construct the Bayesian estimate for θ after observing data $\boldsymbol{x}$ by their conditional expectation:

$$\hat{\theta} = E(\theta|\boldsymbol{x}) = \frac{\int_\Theta \pi(\theta)p(\boldsymbol{x}|\theta)\theta\mathrm{d}\theta}{\int_\Theta \pi(\theta)p(\boldsymbol{x}|\theta)\mathrm{d}\theta} \propto \int_\Theta \pi(\theta)p(\boldsymbol{x}|\theta)\theta\mathrm{d}\theta. \tag{5}$$

In general, the Bayesian estimate $\hat{g}(\theta)$ for a function of θ, $g(\theta)$ can be obtained as follows:

$$\hat{g}(\theta) = E[g(\theta)|\boldsymbol{x}] = \frac{\int_\Theta \pi(\theta)p(\boldsymbol{x}|\theta)g(\theta)\mathrm{d}\theta}{\int_\Theta \pi(\theta)p(\boldsymbol{x}|\theta)\mathrm{d}\theta} \propto \int_\Theta \pi(\theta)p(\boldsymbol{x}|\theta)g(\theta)\mathrm{d}\theta.$$

Obviously, the estimator $\hat{g}(\theta)$ is the usual expectation of $g(\theta)$ with respect to the posterior distribution $\pi(\theta|\boldsymbol{x})$:

$$\hat{g}(\theta) = \int_\Theta g(\theta)\pi(\theta|\boldsymbol{x})\mathrm{d}\theta = E[g(\theta)|\boldsymbol{x}]. \tag{6}$$

2.3. The MCMC Sampling and Its Convergence

When the posterior distribution $\pi(\theta|\boldsymbol{x})$ in (6) is difficult to compute, estimating multiple parameters given by $g(\theta)$ can be realized by the Markov Chain Monte Carlo (MCMC) method [14] following these steps: (1) establish a Markov chain, the stationary distribution for $\pi(\theta|\boldsymbol{x})$; (2) use the Markov chain for the posterior distribution $\pi(\theta|\boldsymbol{x})$ to carry out sampling to obtain an MCMC sample $\{\theta_0, \theta_1, \ldots, \theta_n\}$; and (3) obtain the MCMC estimator $\hat{g}(\theta)$ for $g(\theta)$ by

$$\hat{g}(\theta) = \bar{g}(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(\theta_i).$$

The MCMC sample can estimate the parameter function $g(\theta)$ more and more effectively when the sample size n becomes larger and larger. The MCMC sample $\{\theta_0, \theta_1, \theta_2, \ldots, \}$ possesses some properties, such as stability, normal recurrence, periodicity, and irreducibility. The choice of the initial value θ_0 has little impact on the estimate of θ_t. Gibbs sampling is one of the MCMC algorithms. It is used to construct a multivariate probability distribution of a random sample. The defect of the standard Gibbs sampling is that it cannot process the nonconjugated distribution. Because the prior distribution of parameter $\boldsymbol{\rho}$ in the model in this paper is nonconjugated, the improved version of the Gibbs algorithm is adopted [15].

With regard to the convergence of the MCMC sampling, we consider three convergence criteria: (1) Combine the posterior sampling trend graph with the energy graph of the MCMC sampling process for convergence assessment. By observing the posterior sampling trend graph of a random variable, we can determine if the information of the sampling tends to be stable as the number of iterations increases. (2) Compare the overall distribution of energy levels in real data with energy changes between successive samples. If the two distributions are similar to each other, we can conclude that the algorithm converges. (3) Draw the trajectory of the negative evidence lower bound (ELBO, [16]) obtained in the optimization process to verify the convergence of the algorithm. Minimizing the KL (Kullback–Leibler) divergence is equivalent to minimizing the negative ELBO [17,18].

3. The OU Process

The random process $\boldsymbol{S}_i$ in the general linear mixed model (1) is a Gaussian process (GP). A GP can be regarded as an infinite dimensional extension of the multivariate Gaussian distribution, and its probability characteristics are uniquely determined by the mean function and covariance function. In particular, a GP with zero mean is completely determined by its covariance function [18]. To give a simple review of GP and keep the notation $\boldsymbol{S}_i$ specially used in model (1), we return to a general notation $X(t)$ for GP to avoid a mix-up with the model parameter in model (1). $X(t)$ in this section can be considered as a copy of $\boldsymbol{S}_i$ in model (1). If random process $\{X(t), t \in \boldsymbol{T}\}$ is a GP, any finite observation vector $\boldsymbol{x} = \Big(X(t_1), \ldots, X(t_n)\Big)'$ has a multivariate Gaussian distribution. Let $E[X(t)] = m_X(t), \mathrm{var}[X(t)] = \sigma_X^2(t)$. The joint density function of the multivariate Gaussian random vector is

$$f(x(t_1), \ldots x(t_n)) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{C}_X|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m}_X)' \boldsymbol{C}_X^{-1} (\boldsymbol{x} - \boldsymbol{m}_X)\right\},$$

where $\boldsymbol{x} = \Big(x(t_1), \ldots, x(t_n)\Big)'$, mean vector $\boldsymbol{m}_X = (m_X(t_1), \ldots, m_X(t_n))'$, and $\boldsymbol{C}_X$ stands for the covariance matrix. Let $c_{ij} = \mathrm{cov}\Big(X(t_i), X(t_j)\Big) = E\Big[(X(t_i) - m_X(t_i))(X(t_j) - m_X(t_j))\Big]$. When the Gaussian process $\{X_t = X(t), t \in \boldsymbol{T}\}$ has a zero mean and it is smooth, we have $c_{ij} = E[X(t_i)X(t_j)]$ and

$$
\begin{aligned}
E[X(t)] &= 0 \\
\operatorname{var}[X(t)] &= \sigma_X^2 \\
R_X(X_t, X_s) &= E[(X_t)(X_s)] = R_X(t-s),\ t > s, \\
C_X(X_t, X_s) &= \sigma_X^2 R_X(t-s) = C_X(t-s),\ t > s,
\end{aligned}
$$

where $R_X(X_t, X_s)$ and $C_X(X_t, X_s)$ stand for the GP autocorrelation function and the autocovariance function, respectively, which only depend on the time interval $t - s$. Thus, $R_X(t-s) = \rho(|t-s|)$, and $c_{ij} = \operatorname{cov}\big(X(t_i), X(t_j)\big) = \sigma_X^2 \rho(|t_i - t_j|)$. The correlation coefficient between $X(t_i)$ and $X(t_j)$ is denoted by $\rho(|t_i - t_j|)$. It is assumed that $\rho(|t_i - t_j|) = \rho^{|t_i - t_j|}$ in this paper. A zero-mean smooth GP $\{X(t), t \in \boldsymbol{T}\}$ is an Ornstein–Uhlenbeck (OU) process [19]. An OU process can be regarded as a continuous analogy of the discrete-time first-order autoregressive AR(1) process.

To understand some properties of the AR(1) process, we express the AR(1) process as

$$
X_t = \phi_1 X_{t-1} + \epsilon_t,\ t \geq 1, \quad X_0 = \epsilon_0, \tag{7}
$$

where $|\phi_1| < 1$ is a weight parameter to ensure stability, and $\epsilon_0, \epsilon_1, \epsilon_2, \ldots$ are uncorrelated random variables satisfying

$$
E(\epsilon_t) = 0,\ t \geq 0, \quad Var(\epsilon_t) = \begin{cases} \frac{\sigma^2}{1-\phi_1^2}, & t = 0, \\ \sigma^2, & t \geq 1. \end{cases}.
$$

It is assumed that the random error satisfies $\operatorname{cov}(\epsilon_t, \epsilon_{t+k}) = 0, k \neq 0$. Therefore,

$$
\begin{aligned}
& \operatorname{cov}(X_t, X_{t+k}) = \operatorname{cov}\left(\sum_{i=0}^{t} \phi^{t-i} \epsilon_i, \sum_{i=0}^{t+k} \phi^{t+k-i} \epsilon_i \right) \\
= & \sum_{i=0}^{t} \phi^{t-i} \phi^{t+k-i} \operatorname{cov}(\epsilon_i, \epsilon_i) \\
= & \sigma^2 \phi_1^{2t+k} \left(\frac{1}{1-\phi_1^2} + \sum_{i=1}^{t} \phi_1^{-2i} \right) = \frac{\sigma^2 \phi_1^k}{1-\phi_1^2}.
\end{aligned}
$$

The autocorrelation function ρ_k is assumed to follow the first-order difference equation

$$
\rho_k = \phi_1 \rho_{k-1}, \quad k \geq 1, \quad \rho_0 = 1,
$$

so that we have the following solution:

$$
\rho_k = \phi_1^k, \quad k \geq 0.
$$

As shown in Figure 2 below, when $|\phi_1| < 1$, the absolute correlation between $X(t_i)$ and $X(t_j)$ of $X(t)$ at two different time points t_i and t_j approaches 0 when the time is increasing.

The above AR(1) process (7) is a special case of an OU process by taking $\phi_1 = 1 - \theta$ and $\mu = 0$, that is,

$$
X_{t+1} = X_t + \theta(\mu - X_t) + \epsilon_{t+1}, \quad X_0 = \epsilon_0.
$$

Therefore, the exponential covariance matrix constructed by the OU process can be used to describe the correlation in longitudinal data. An OU process is showed in Figure 3. It is obvious that when $X(t)$ moves toward its mean position with the increase in time t, the correlation between $X(t_i)$ and $X(t_j)$ at two different time points becomes weaker and weaker.

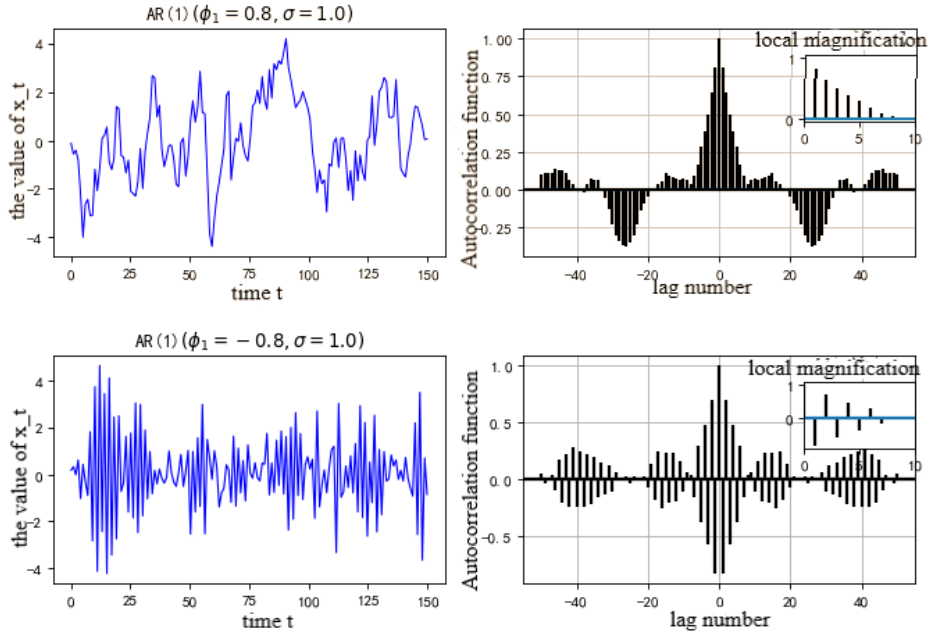


Figure 2. Example of an AR(1) process.

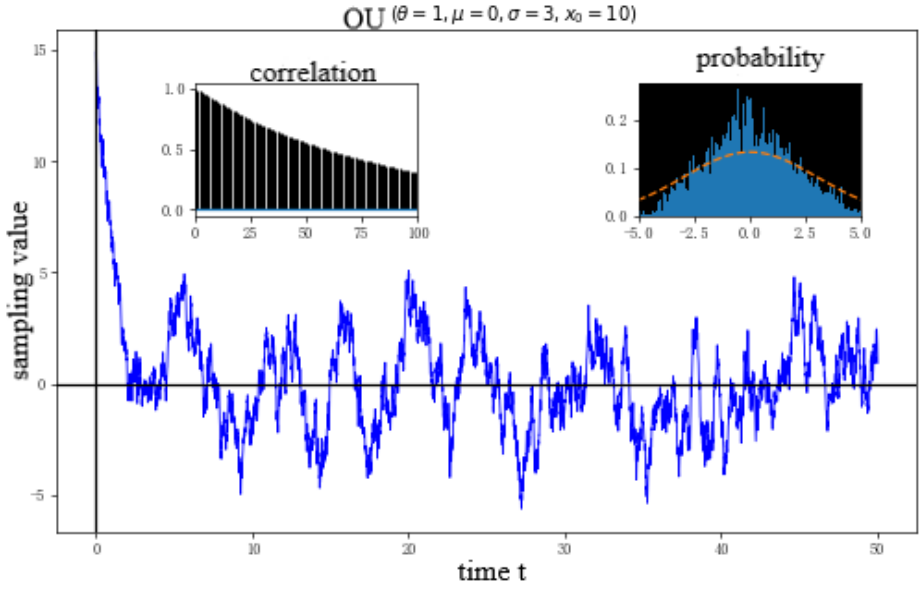


Figure 3. Example of an OU process.

4. Dirichlet Process and Dirichlet Process Mixture

4.1. Dirichlet Process

A Dirichlet process (DP) is a class of stochastic processes whose trace is a probability distribution. DP is often used in Bayesian inference to describe the prior knowledge of random variables.

Definition 1. *Given a measurable set $\mathcal{C}$, a basis distribution G_0, and a positive real number α, a Dirichlet process* $\mathrm{DP}(G_0, \alpha)$ *is a random process whose realization is a probability distribution on $\mathcal{C}$. For any measurable finite partition of $\mathcal{C}$, $\{B_i\}_{i=1}^n$, if $G \sim \mathrm{DP}(G_0, \alpha)$, then*

$$(G(B_1), \ldots, G(B_n)) \sim Dir(\alpha G_0(B_1), \ldots, \alpha G_0(B_n)).$$

where Dir stands for the Dirichlet distribution. A DP is specified by the basis distribution G_0 and the concentration parameter α.

A DP can also be regarded as an infinite dimensional generalization of the n-dimensional Dirichlet distribution. A DP is the conjugate priori of an infinite, nonparametric discrete distribution. An important application of DP is to use it as a prior distribution for infinite mixture models. A statistically equivalent description of DP is based on the stick-breaking process, described as follows. Given a discrete distribution

$$G(\rho) = \sum_{i=1}^{\infty} w_k \delta_{\rho_k}(\rho)$$

where $\delta_{\rho k}$ is an indicator function, namely

$$\delta_{\rho_k} = \begin{cases} 1, & \theta = \rho_k; \\ 0, & \theta \neq \rho_k. \end{cases}$$

$\rho_k \overset{iid}{\sim} G_0, w_k = v_k \prod_{i=1}^{k-1}(1 - v_i), v_i \overset{iid}{\sim} \text{Beta}(1, \alpha)$, the probability distribution G defined in this way is said to obey the Dirichlet process, denoted as $G \sim DP(G_0, \alpha)$. A DP can be constructed by a stochastic process. It possesses some advantages in its application in Bayesian model evaluation, density estimation, and mixed model clustering [20].

4.2. Dirichlet Process Mixture

We consider a set of i.i.d. sample data $\{y_1, \ldots, y_n\}$ as a part of an infinitely exchangeable sequence. y_i follows the probability distribution $F(y; \boldsymbol{\theta})$ with parameter $\boldsymbol{\theta} \in \Theta$, that is, $y_i|\boldsymbol{\theta} \sim F(y; \boldsymbol{\theta})$. It may be assumed that the prior distribution of $\boldsymbol{\theta}$ is an unknown random probability measure G, and G can be constructed by DP, that is, $\boldsymbol{\theta}|G \sim G$, $G \sim DP(G_0, \alpha)$. Thus, a Dirichlet process can be obtained using the hybrid (DPM) model definition:

$$y_i|\boldsymbol{\theta} \sim F(y; \boldsymbol{\theta}), \quad \boldsymbol{\theta}|G \sim G, \quad G \sim \text{DP}(G_0, \alpha),$$

where G_0 and α are the basis distribution and model parameter, respectively. In general, the distributions F and G_0 will depend on additional hyperparameters not mentioned above. Since the trace of DP is discretized with probability 1, the above model can be regarded as a countably infinite mixture. We can integrate the distribution G in the above DPM to obtain the prior distribution of $\boldsymbol{\theta}$:

$$\begin{aligned} \theta_i|(\theta_1, \ldots, \theta_{i-1}) &\sim \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\theta_j} + \frac{\alpha}{\alpha + i - 1} G_0 \\ &= \frac{i-1}{\alpha + i - 1} \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \delta_{\theta_j} \right) + \frac{\alpha}{\alpha + i - 1} G_0, \end{aligned}$$

where δ_{θ_j} $(j = 1, \ldots, i-1, i = 2, 3, \ldots)$ represents the point quality at θ_j. This is a mixture of two distributions with weights $p = (i-1)/(\alpha + i - 1)$ and $1 - p = \alpha/(\alpha + i - 1)$.

Let $\left\{ f(\cdot|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \Re^d \right\}$ be a family of density functions with parameter $\boldsymbol{\theta}$. We assume that the density function of the observed data y_i follows the probability distribution family $\mathcal{F} = \{ f(\cdot|G) : G \in \mathcal{P} \}$ defined by

$$f(y_i|G) = \int f(y_i|\boldsymbol{\theta}) \mathrm{d}G(\boldsymbol{\theta}), \tag{8}$$

which is called the DPM density of y_i, where $G(\boldsymbol{\theta})$ is the distribution of $\boldsymbol{\theta}$. $\mathcal{F}$ is the nonparametric family of mixed distributions. The distribution G can be made random by assuming that G comes from a DP. In practice, to obtain the DPM density functions of $y_1, \ldots, y_n$, it

is necessary to introduce the latent variable θ_i related to y_i. That is, after introducing θ_i, the joint density function of $\{y_i : i = 1, \ldots, n\}$ is $\prod_{i=1}^{n} f(y_i|\theta_i)$, where $\{\theta_i : i = 1, \ldots, n\}$ is a sample from the distribution G. The joint density function of $\{y_i : i = 1, \ldots, n\}$ can be expressed as $\prod_{i=1}^{n} f(y_i|G)$. We assign the prior distribution $DP(G_0, \alpha)$ to the distribution G with $G \sim DP(G_0, \alpha)$. The Bayesian model is set up as follows:

$$\begin{aligned} y_i|(\boldsymbol{\beta}, \boldsymbol{S}_i, \boldsymbol{b}_i, \boldsymbol{\theta}, \sigma) &\sim F(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{S}_i, \boldsymbol{b}_i, \sigma), & w_k &= \nu_k \prod_{i=1}^{k-1}(1-\nu_i), \\ \theta_k|G &\sim G,\ k = 1, 2, \ldots & & \\ G(\cdot) &= \sum_{k=1}^{\infty} w_k \delta_{\theta_k}(\cdot), & \nu_i &\overset{iid}{\sim} Beta(1, \alpha), \\ & & \theta_k &\overset{iid}{\sim} G_0,\ k = 1, 2, \ldots \end{aligned} \tag{9}$$

where $\{y_1, \ldots, y_n\}$ is a set of observations and $\{\theta_1, \ldots, \theta_n\}$ is a set of latent variables. Based on the discretization of DP, we can obtain the DPM density function of S_i by

$$p(S_i|G) = \int p(S_i|\theta) G(\mathrm{d}\theta) = \sum_{k=1}^{\infty} w_k p(S_i|\theta_k).$$

A semiparametric model can be set up by replacing the DP mixture on $\boldsymbol{\theta}$ with the DP mixture on any subset of $\boldsymbol{w} = \{w_k : k \geq 1\}$ (given by (9)), depending on $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_i = (\boldsymbol{w}_i, \boldsymbol{\eta})$ with a prior distribution $\boldsymbol{\eta}$. The above DP mixture is only performed on the parameter $\boldsymbol{w}_i$. The joint density function $\prod_{i=1}^{n} f(y_i|\boldsymbol{\eta}, G)$ of $\{y_i : i = 1, \ldots, n\}$ can be obtained. The prior distribution of $\boldsymbol{\eta}$ and G are prespecified. As a result, the semiparametric model can be determined [21].

5. Formulation of the Semiparametric Autoregressive Model

5.1. The Partial Dirichlet Process Mixture of Stochastic Process

The stochastic process in (1) is semiparameterized to generalize the general linear mixed model for longitudinal data. A nonparametric DP prior is assigned to the covariance parameter (σ_s^2, ρ) of $\boldsymbol{S}_i$. To reduce the number of unknown parameters, a partial Dirichlet process mixture is performed on $\boldsymbol{S}_i$, i.e., only a nonparametric DP prior is assigned to the parameter ρ. The semiparametric process is created as follows. First, we consider the OU process $\boldsymbol{S}_i$ associated with object i to have a covariance matrix $\boldsymbol{C}_i(\boldsymbol{\theta}) = \sigma_s^2 \tilde{\boldsymbol{C}}_i(\rho)$, where $\boldsymbol{\theta} = (\sigma_s^2, \rho)$. The (k, l)-element of the matrix $\tilde{\boldsymbol{C}}_i(\rho)$ is given by $\rho^{|t_{il} - t_{ik}|}$. The matrix $\tilde{\boldsymbol{C}}_i(\rho)$ associated with the random process $\boldsymbol{S}_i$ has the following form:

$$\begin{bmatrix} 1 & \rho^{|t_{i2}-t_{i1}|} & \rho^{|t_{i3}-t_{i1}|} & \ldots & \rho^{\left|t_{in_i}-t_{i1}\right|} \\ \rho^{|t_{i1}-t_{i2}|} & 1 & \rho^{|t_{i3}-t_{i2}|} & \ldots & \rho^{\left|t_{in_i}-t_{i2}\right|} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{\left|t_{i1}-t_{in_i}\right|} & \rho^{\left|t_{i2}-t_{in_i}\right|} & \rho^{\left|t_{i3}-t_{in_i}\right|} & \ldots & 1 \end{bmatrix}; \tag{10}$$

Second, we give the parameter ρ a nonparametric DP prior:

$$\rho|G \sim G, \quad G \sim \mathrm{DP}(G_0, \alpha), \tag{11}$$

where G_0 is a known probability distribution, α is a constant, and the probability distribution G is generated by $\mathrm{DP}(G_0, \alpha)$ satisfying

$$\begin{aligned} G(\cdot) &= \sum_{k=1}^{\infty} w_k \delta_{\phi_k}(\cdot), & \nu_i &\overset{iid}{\sim} Beta(1, \alpha), \\ w_k &= \nu_k \prod_{i=1}^{k-1}(1-\nu_i), & \phi_k &\overset{iid}{\sim} G_0, \end{aligned} \tag{12}$$

where $\delta_{\phi}(\cdot)$ is the point quality of ϕ, and G is a random probability distribution.

Using the discretization of DP, for any parameter $\boldsymbol{\phi}$, we assume that $f(\cdot|\boldsymbol{\phi})$ is a density function depending on the parameter $\boldsymbol{\phi}$. Let $\boldsymbol{\phi}|G \sim G$, $G \sim \mathrm{DP}(G_0, \alpha)$. The DPM density function can be obtained:

$$p(\cdot|G) = \int p(\cdot|\boldsymbol{\phi})G(\mathrm{d}\boldsymbol{\phi}) = \sum_{k=1}^{\infty} w_k p(\cdot|\phi_k), \tag{13}$$

where $\phi_k \overset{iid}{\sim} G_0$. After embedding the above conditional prior into the distribution of the random process $\boldsymbol{S}_i$, we formulate the DPM model for $\boldsymbol{S}_i$ as

$$\boldsymbol{S}_i|(\sigma_s^2, \rho) \sim N(\mathbf{0}, \sigma_s^2 \tilde{\boldsymbol{C}}_i(\rho)), \quad \rho|G \sim G, \quad G \sim \mathrm{DP}(G_0, \alpha). \tag{14}$$

Using the discretization of the distribution function G, we obtain the DPM density of the random process $\boldsymbol{S}_i$ as follows.

$$p(\boldsymbol{S}_i|\sigma_s^2, G) = \int N(\boldsymbol{S}_i|\mathbf{0}, \sigma_s^2 \tilde{\boldsymbol{C}}_i(\rho))\mathrm{d}G(\rho) = \sum_{k=1}^{\infty} w_k N\Big(\boldsymbol{S}_i|\mathbf{0}, \sigma_s^2 \tilde{\boldsymbol{C}}_i(\tilde{\rho}_k)\Big). \tag{15}$$

This is an infinitely countable mixture of multivariate Gaussian density functions, where $\tilde{\rho}_k \overset{iid}{\sim} G_0$, $N(\cdot|\mathbf{0}, \sigma_s^2 \tilde{\boldsymbol{C}}_i(\rho))$ represents a multivariate Gaussian density function with a zero-mean vector and a covariance matrix of $\sigma_s^2 \tilde{\boldsymbol{C}}_i(\rho)$. It can be seen that after the semiparametric treatment of the stochastic process $\boldsymbol{S}_i$, its distribution is a mixture of stochastic processes, which is a more general mixture of OU processes.

Since G is discretized with probability 1, it provides an automatic clustering effect on the autocorrelation structure of the objects. After the OU process $\boldsymbol{S}_i$ is semiparameterized, the covariance between any two time points t_{il}, t_{ik} can be obtained by

$$\begin{aligned} &\mathrm{cov}(s_i(t_{il}), s_i(t_{ik})|\sigma_s^2, G) = \int \mathrm{cov}(s_i(t_{il}), s_i(t_{ik})|\sigma_s^2, \rho)\, dG(\rho) \\ = &\int \sigma_s^2 \rho^{|t_{il}-t_{ik}|}\, dG(\rho) = \sum_{k=1}^{\infty} w_k \sigma_s^2 \tilde{\rho}_k^{|t_{il}-t_{ik}|}. \end{aligned} \tag{16}$$

After semiparameterization of the OU process $\boldsymbol{S}_i$, it not only contains an AR structure but also has an automatic clustering effect. If the observations $\{\boldsymbol{y_i} : i = 1, \ldots, n\}$ from model (1) are obtained at equal interval time points, the covariance matrix structure of $\boldsymbol{y_i}$ becomes AR(1).

5.2. The Framework of a Hierarchical Model

We introduce potential parameters $\rho_1, \rho_2, \ldots, \rho_n$ (corresponding to n objects) and semiparameterize the general linear mixed model (1) with observations $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ satisfying the following model:

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{S}_i + \boldsymbol{\epsilon}_i, & \rho_i|G &\sim G, \\ \boldsymbol{b}_i|\boldsymbol{\tau} &\sim N(\mathbf{0}, \boldsymbol{D}(\boldsymbol{\tau})), & G &\sim \mathrm{DP}(G_0, \alpha), \\ \boldsymbol{S}_i|(\sigma_s^2, \rho_i) &\sim N\Big(\mathbf{0}, \sigma_s^2 \tilde{\boldsymbol{C}}_i(\rho_i)\Big), & \boldsymbol{\epsilon}_i &\sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_{n_i}). \end{aligned} \tag{17}$$

This is converted to a hierarchical model as follows:

$$\begin{aligned} \boldsymbol{y}_i|(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{S}_i, \sigma^2) &\overset{ind}{\sim} N(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{S}_i, \sigma^2\boldsymbol{I}), \\ \boldsymbol{S}_i|(\sigma_s^2, \rho_i) &\overset{ind}{\sim} N\left(\boldsymbol{0}, \sigma_s^2\tilde{\boldsymbol{C}}_i(\rho_i)\right), \\ \{\rho_1, \rho_2, \ldots, \rho_n\}|G &\overset{iid}{\sim} G, \\ G &\sim \mathrm{DP}(G_0, \alpha), \\ \boldsymbol{b_i}|\boldsymbol{\tau} &\overset{iid}{\sim} N(0, \boldsymbol{D}(\boldsymbol{\tau})), \\ (\sigma^2, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_s^2) &\sim p(\sigma^2) \times N(\boldsymbol{\beta}_0, \boldsymbol{B}) \times p(\boldsymbol{\tau}) \times p(\sigma_s^2), \end{aligned} \tag{18}$$

where "ind" means independent only, $\boldsymbol{S}_i$ and $\boldsymbol{b}_i (1 \le i \le n)$ are independent of each other, and $p(\sigma^2)$, $N(\boldsymbol{\beta}_0, \boldsymbol{B})$, $p(\boldsymbol{\tau})$, and $p(\sigma_s^2)$ are the prior distributions of parameters σ^2, $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, and σ_s^2, respectively. If b_i is a scalar quantity corresponding to the random intercept of the model, we have $\tau = \sigma_b^2$. This model generalizes the exponential covariance function of the OU process. It realizes the automatic clustering effect between objects through parameter ρ_i. We call model (18) the Bayesian semiparametric autoregressive model, or simply the SPAR (semiparametric autoregressive) model.

Note that the SPAR model (18) is a parallel version of Quintana et al.'s [4] with additional conditions that the random effects $\{S_i : i = 1, \ldots, n\}$ have a common variance component σ_S^2, and the autocorrelation parameters $(\rho_1, \rho_2, \ldots, \rho_n)$ are assumed to be conditional i.i.d. with $DP(G_0, \alpha)$. This is different from model (3) in Quintana et al. [4], which assumes that the random-effect parameters $(\phi_1, \ldots, \phi_n)$ are conditional i.i.d. with $DP(G_0, \alpha)$, where $\phi_i = (\sigma_i^2, \rho_i)$ with $\mathrm{var}(S_i) = \sigma_i^2\tilde{\boldsymbol{C}}_i(\rho_i)$ in our notation. The simpler assumptions help simplify the posterior inference. Quintana et al.'s [4] model assumptions do not lead to explicit posterior distributions of model parameters. Their posterior inference on model parameters may have to be performed by nonparametric MCMC algorithms. Because no explicit expressions related to posterior inference on model parameters are given by [4], we are not able to conclude if Quintana et al.'s [4] Bayesian posterior inference is a complete semiparametric MCMC method or a combination of semiparametric and nonparametric MCMC methods. By imposing simpler assumptions on the parameters in model (18), we are able to obtain the explicit posterior distributions of all model parameters in the subsequent context and conclude that our approach to handling the SPAR model (18) is a complete semiparametric MCMC method.

In the SPAR model (18), the random process part is an OU process mixture. The correlation matrix $\tilde{\boldsymbol{C}}_i(\rho_i)$ possesses the following form:

$$\tilde{\boldsymbol{C}}_i(\rho_i) = \begin{bmatrix} 1 & \rho_i^{|t_{i2}-t_{i1}|} & \rho_i^{|t_{i3}-t_{i1}|} & \cdots & \rho_i^{|t_{il}-t_{i1}|} & \cdots & \rho_i^{|t_{in_i}-t_{i1}|} \\ \rho_i^{|t_{i1}-t_{i2}|} & 1 & \rho_i^{|t_{i3}-t_{i2}|} & \cdots & \rho_i^{|t_{il}-t_{i2}|} & \cdots & \rho_i^{|t_{in_i}-t_{i2}|} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \rho_i^{|t_{i1}-t_{ik}|} & \rho_i^{|t_{i2}-t_{ik}|} & \rho_i^{|t_{k3}-t_{ik}|} & \cdots & \rho_i^{|t_{il}-t_{ik}|} & \cdots & \rho_i^{|t_{in_i}-t_{ik}|} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \rho_i^{|t_{i1}-t_{in_i}|} & \rho_i^{|t_{i2}-t_{in_i}|} & \rho_i^{|t_{i3}-t_{in_i}|} & \cdots & \rho_i^{|t_{il}-t_{in_i}|} & \cdots & 1 \end{bmatrix}. \tag{19}$$

Using the property of the OU process structure, $\tilde{\boldsymbol{C}}_i(\rho_i)$ can be analyzed backwards. The inverse matrix of $\tilde{\boldsymbol{C}}_i(\rho_i)$ is a tridiagonal. Denote by $r_{ik} = \rho_i^{|t_{i,k+1}-t_{ik}|}$ $(k = 1, 2, \ldots, n_i - 1)$. We have the (k, l)-th element of $\tilde{\boldsymbol{C}}_i(\rho_i)$ given by

$$\left\{\tilde{\boldsymbol{C}}_i(\rho_i)\right\}_{kl} = \rho_i^{|t_{il}-t_{ik}|} = \rho_i^{|t_{i,k+1}-t_{ik}|+|t_{i,k+2}-t_{i,k+1}|+\cdots+|t_{il}-t_{i,l-1}|} = r_{ik}r_{i,k+1}\ldots r_{i,l-1}. \tag{20}$$

So, $\tilde{\boldsymbol{C}}_i(\rho_i)$ can be rewritten as

$$\begin{bmatrix} 1 & r_{i1} & r_{i1}r_{i2} & \cdots & r_{i1}\ldots r_{i,l-1} & \cdots & r_{i1}\ldots r_{i,n_i-1} \\ r_{i1} & 1 & r_{i2} & \cdots & r_{i2}\ldots r_{i,l-1} & \cdots & r_{i2}\ldots r_{i,n_i-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ r_{i1}\ldots r_{i,k-1} & r_{i2}\ldots r_{i,k-1} & r_{i3}\ldots r_{i,k-1} & \cdots & r_{ik}\ldots r_{i,l-1} & \cdots & r_{ik}\ldots r_{i,n_i-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ r_{i1}\ldots r_{i,n_i-1} & r_{i2}\ldots r_{i,n_i-1} & r_{i3}\ldots r_{i,n_i-1} & \cdots & r_{il}\ldots r_{i,n_i-1} & \cdots & 1 \end{bmatrix}.$$

Using the correlation theory of the anticorrelation random variables [22], we can compute the elements of the inverse matrix $\tilde{\boldsymbol{C}}_i^{-1}(\rho_i)$ of $\tilde{\boldsymbol{C}}_i(\rho_i)$ as follows:

$$\begin{aligned} \left\{\tilde{\boldsymbol{C}}_i^{-1}(\rho_i)\right\}_{11} &= \frac{1}{1-r_{i1}^2}, \\ \left\{\tilde{\boldsymbol{C}}_i^{-1}(\rho_i)\right\}_{kk} &= \frac{1-r_{i,k-1}^2 r_{ik}^2}{1-r_{i,k-1}^2-r_{ik}^2+r_{i,k-1}^2 r_{ik}^2},\ k=2,3,\ldots,n_i-1, \\ \left\{\tilde{\boldsymbol{C}}_i^{-1}(\rho_i)\right\}_{kk+1} &= -\frac{r_{ik}}{1-r_{ik}^2},\ k=1,2,\ldots,n_i-1, \\ \left\{\tilde{\boldsymbol{C}}_i^{-1}(\rho_i)\right\}_{n_in_i} &= \frac{1}{1-r_{i,n_i-1}^2}. \end{aligned} \tag{21}$$

$\tilde{\boldsymbol{C}}_i^{-1}(\rho_i)$ turns out to be a tridiagonal matrix.

The random process corresponding to the i-th object $\boldsymbol{S}_i = (s_{i1}, s_{i2}, \ldots, s_{in_i})'$ satisfies the conditional distribution:

$$\boldsymbol{S}_i|\sigma_s^2, \rho_i \sim N_{n_i}\left(\boldsymbol{0}, \sigma_s^2\tilde{\boldsymbol{C}}_i(\rho_i)\right).$$

So, the conditional density function is

$$f(\boldsymbol{S}_i|\sigma_s^2,\rho_i) = f(s_{i1},\ldots,s_{in_i}|\sigma_s^2,\rho_i) = f(s_{i1}|\sigma_s^2,\rho_i)\cdots f(s_{in_i}|s_{i1},\ldots,s_{i,n_i-1},\sigma_s^2,\rho_i). \tag{22}$$

When the inverse matrix of the covariance matrix $\tilde{\boldsymbol{C}}_i(\rho_i)$ of the random process $\boldsymbol{S}_i$ has the above tridiagonal form, based on the property of the multivariate normal distribution, the conditional density function $f(\boldsymbol{S}_i|\sigma_s^2,\rho_i)$ of $\boldsymbol{S}_i$ can be decomposed into the product of n_i univariate Gaussian density functions as follows:

$$\begin{aligned} s_{i1} &\sim N(0,\sigma_s^2), \\ s_{i2}|(s_{i1}=\tilde{s}_1) &\sim N\left(\tilde{s}_1 r_{i1}, \sigma_s^2(1-r_{i1}^2)\right), \\ &\vdots \\ s_{ik}|(s_{i1}=\tilde{s}_1,\ldots,s_{i,k-1}=\tilde{s}_{k-1}) &\sim N\left(\tilde{s}_{k-1} r_{i,k-1}, \sigma_s^2(1-r_{i,k-1}^2)\right), \\ &\vdots \\ s_{in_i}|(s_{i1}=\tilde{s}_1\ldots,s_{i,n_i-1}=\tilde{s}_{n_i-1}) &\sim N\left(\tilde{s}_{n_i-1} r_{i,n_i-1}, \sigma_s^2(1-r_{i,n_i-1}^2)\right). \end{aligned} \tag{23}$$

It is known that $r_{ik} = \rho_i^{|t_{i,k+1}-t_{ik}|}, k = 1,2,\ldots,n_i-1$. Hence,

$$
\begin{aligned}
s_{i1} &\sim N(0, \sigma_s^2), \\
s_{i2}|(s_{i1} = \tilde{s}_1) &\sim N\left(\tilde{s}_1 \rho^{|t_{i2}-t_{i1}|}, \sigma_s^2\left(1-\rho^{2|t_{i2}-t_{i1}|}\right)\right), \\
&\vdots \\
s_{ik}|(s_{i,k-1} = \tilde{s}_{k-1}) &\sim N\left(\tilde{s}_{k-1} \rho_i^{|t_{ik}-t_{i,k-1}|}, \sigma_s^2\left(1-\rho_i^{2|t_{ik}-t_{i,k-1}|}\right)\right), \\
&\vdots \\
s_{in_i}|(s_{i,n_i-1} = \tilde{s}_{n_i-1}) &\sim N\left(\tilde{s}_{n_i-1} \rho_i^{|t_{in_i}-t_{i,n_i-1}|}, \sigma_s^2\left(1-\rho_i^{2|t_{in_i}-t_{i,n_i-1}|}\right)\right).
\end{aligned} \tag{24}
$$

As a result, the conditional density function of $\boldsymbol{S}_i$ in the random process $f(\boldsymbol{S}_i|\sigma_s^2, \rho_i)$ can be decomposed as follows:

$$
\begin{aligned}
f(\boldsymbol{S}_i\Big|\sigma_s^2, \rho_i) =& N\left(s_{i1}\Big|0, \sigma_s^2\right) \cdots N\left(s_{ik}\Big|\tilde{s}_{k-1}\rho_i^{|t_{ik}-t_{i,k-1}|}, \sigma_s^2\left(1-\rho_i^{2|t_{ik}-t_{i,k-1}|}\right)\right) \cdots \\
& N\left(s_{in_i}\Big|\tilde{s}_{n_i-1}\rho_i^{|t_{in_i}-t_{i,n_i-1}|}, \sigma_s^2\left(1-\rho_i^{2|t_{in_i}-t_{i,n_i-1}|}\right)\right).
\end{aligned} \tag{25}
$$

The above decomposition greatly simplifies the computation of the distribution of the random process $\boldsymbol{S}_i$, which can be obtained by direct computation of the distribution of a single Gaussian variable.

6. The Marginal Likelihood, Prior Determination, and Posteriori Inference

Let $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)$, $\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$, $\boldsymbol{S} = (\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_n)$, and $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_n)$. Denote by $\boldsymbol{\Psi} = (\boldsymbol{\beta}, \sigma^2, \sigma_s^2, \boldsymbol{\rho}, \boldsymbol{\tau})$, which represents the parameter vector in model (18). A random process that specifies the AR structure in the SPAR model (18) is

$$
\boldsymbol{S}_i|\sigma_s^2, \rho_i \sim N\left(\boldsymbol{0}, \sigma_s^2 \tilde{\boldsymbol{C}}_i(\rho_i)\right),\ \rho_i|G \sim G_0,\ G \sim \mathrm{DP}(G_0, \alpha). \tag{26}
$$

The marginal distribution of the covariance parameter ρ_i can be obtained from the joint distribution of all terms in (26) as follows:

$$
\rho_1 \sim G_0, \quad \rho_i|(\rho_1, \rho_2, \ldots, \rho_{i-1}) \sim \frac{\alpha}{\alpha+i-1} G_0 + \frac{1}{\alpha+i-1} \sum_{k=1}^{i-1} \delta_{\rho_k} \tag{27}
$$

for $i = 2, \ldots, n$. We assume that $\{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_n\}$ is a sample from the multivariate Gaussian distribution. It can be regarded as a part of an exchangeable sequence. By using the exchangeable property, the corresponding order i of the observation $\boldsymbol{y}_i$ can be considered as the last one in all n observations from n objects. Then, $\boldsymbol{y}_i$ is the corresponding vector S_i. Or we can consider it as the last one that gives us $\rho_{(-i)}$. The conditional prior score of ρ_i is

$$
\rho_i|\rho_{(-i)} \sim \frac{\alpha}{\alpha+i-1} G_0 + \frac{1}{\alpha+i-1} \sum_{k \neq i} \delta_{\rho_k}, \tag{28}
$$

where $\rho_{(-i)}$ represents all other ρ's except ρ_i. According to the product formula, we have

$$
p(\boldsymbol{\rho}) = p(\rho_1, \rho_2, \ldots, \rho_n) = p(\rho_1) p(\rho_2|\rho_1) \cdots p(\rho_n|\rho_1, \rho_2, \ldots, \rho_{n-1}).
$$

Based on the above conditional prior distributions, we can obtain the prior distribution $p(\boldsymbol{\rho})$ of parameter $\boldsymbol{\rho}$. Compute the likelihood function of the SPAR model (18) as follows:

$$\begin{aligned} \boldsymbol{L}(\boldsymbol{\Psi}|\boldsymbol{y},\boldsymbol{b},\boldsymbol{S}) &= f(\boldsymbol{y},\boldsymbol{b},\boldsymbol{S}|\boldsymbol{\Psi}) \\ &= f(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{S},\boldsymbol{\Psi})f(\boldsymbol{b}|\boldsymbol{\Psi})f(\boldsymbol{S}|\boldsymbol{\Psi}) \\ &= f(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{S},\boldsymbol{\beta},\sigma^2)f(\boldsymbol{b}|\boldsymbol{\tau})f(\boldsymbol{S}|\sigma_s^2,\boldsymbol{\rho}). \end{aligned} \tag{29}$$

Let

$$p(\boldsymbol{\Psi}) = p(\boldsymbol{\beta},\sigma^2,\sigma_s^2,\boldsymbol{\rho},\boldsymbol{\tau}) = p(\boldsymbol{\beta})p(\sigma^2)p(\sigma_s^2)p(\boldsymbol{\rho})p(\boldsymbol{\tau}),$$

This implies that the prior distributions of each parameter are independent of each other. The joint posterior of the SPAR model is computed as follows:

$$\begin{aligned} p(\boldsymbol{\Psi},\boldsymbol{b},\boldsymbol{S}|\boldsymbol{y}) &\propto p(\boldsymbol{y},\boldsymbol{b},\boldsymbol{S}|\boldsymbol{\Psi})p(\boldsymbol{\Psi}) \\ &= \left\{\prod_{i=1}^{n} f(\boldsymbol{y}_i,\boldsymbol{b}_i,\boldsymbol{S}_i|\boldsymbol{\Psi})\right\}p(\boldsymbol{\Psi}) \\ &= \left\{\prod_{i=1}^{n} f\left(\boldsymbol{y}_i|\boldsymbol{\beta},\boldsymbol{S}_i,\boldsymbol{b}_i,\sigma^2\right)f(\boldsymbol{b}_i|\boldsymbol{\tau})f\left(\boldsymbol{S}_i|\sigma_s^2,\rho_i\right)\right\}p(\boldsymbol{\Psi}). \end{aligned} \tag{30}$$

That is, the joint posterior distribution is the product of the likelihood function and the prior distribution. The first term of the likelihood function is the density function of the estimated n-dimensional Gaussian distribution $N(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{b}_i+\boldsymbol{S}_i,\sigma^2\boldsymbol{I})$ at $\boldsymbol{y}_i$. The second term is in $\boldsymbol{b}_i$, the density function of the estimated q-dimensional Gaussian distribution $N(\boldsymbol{0},\boldsymbol{D}(\boldsymbol{\tau}))$. The third term is the product of the n_i univariate Gaussian distribution density functions.

To estimate the joint posterior distribution of the SPAR model (18), it is necessary to use the Bayesian theorem to obtain the conditional distribution of parameters in the model:

$$\begin{aligned} &p(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{b},\boldsymbol{S},\sigma^2,\sigma_s^2,\boldsymbol{\rho},\boldsymbol{\tau}), \\ &p(\boldsymbol{b}_i|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{b}_{-i},\boldsymbol{S},\sigma^2,\sigma_s^2,\boldsymbol{\rho},\boldsymbol{\tau}), \\ &p(\boldsymbol{S}_i|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{S}_{-i},\sigma^2,\sigma_s^2,\boldsymbol{\rho},\boldsymbol{\tau}), \end{aligned} \quad \begin{aligned} &p(\sigma^2|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{S},\sigma_s^2,\boldsymbol{\rho},\boldsymbol{\tau}), \\ &p(\sigma_s^2|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{S},\sigma^2,\boldsymbol{\rho},\boldsymbol{\tau}), \\ &p(\rho_i|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{S},\sigma^2,\sigma_s^2,\rho_{-i},\boldsymbol{\tau}), \\ &p(\boldsymbol{\tau}|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{S},\sigma^2,\sigma_s^2,\boldsymbol{\rho}). \end{aligned} \tag{31}$$

The MCMC algorithm is employed to estimate these conditional distributions. The conditional distribution of a parameters is denoted by $p(\cdot|*)$ in the subsequent context.

7. A Monte Carlo Study

7.1. Simulation Design

To verify that the SPAR model (18) can effectively simulate the correlation structure in the longitudinal data, the empirical sample data were generated under the four situations of zero mean and covariance structure being compound symmetric (CS), autoregressive (AR), mixed CS and AR, and nonstructured, respectively. The MCMC method was employed to estimate the covariance matrix and the correlation matrix in the four different cases, respectively, and compared with the traditional Bayesian inverse-Wishart estimation method.

Consider a brief form of the SPAR model (18):

$$\boldsymbol{y}_i = (\beta + b_i)e + \boldsymbol{S}_i + \boldsymbol{\epsilon}_i, \tag{32}$$

where β represents a fixed intercept, e is an $n_i \times 1$ vector of ones, $b_i \sim N(0,\sigma_b^2)$ is a random intercept, and $\boldsymbol{S}_i$ is an OU process corresponding to $\boldsymbol{y}_i$. Convert the above model into a hierarchical model:

$$\begin{aligned} \boldsymbol{y}_i|(\beta, b_i, \boldsymbol{S}_i, \sigma^2) &\overset{ind}{\sim} N\Big((\beta + b_i)e + \boldsymbol{S}_i, \sigma^2\boldsymbol{I}\Big), & b_i &\overset{iid}{\sim} N(0, \sigma_b^2), \\ \boldsymbol{S}_i|(\sigma_s^2, \rho_i) &\overset{ind}{\sim} N\Big(\boldsymbol{0}, \sigma_s^2\tilde{\boldsymbol{C}}(\rho_i)\Big), & \beta &\sim N(\mu_0, \sigma_0^2), \\ \{\rho_1, \rho_2, \ldots, \rho_n\}|G &\overset{iid}{\sim} G, & \sigma^2 &\sim IG(\alpha_0, \beta_0), \\ G &\sim \mathrm{DP}(G_0, \alpha), & \sigma_s^2 &\sim IG(\alpha_1, \beta_1). \end{aligned} \tag{33}$$

For the special case of balanced sample design with $n_i = m$ $(i = 1, 2, \ldots, n)$ and $t_{ij} = t_j$ for $i = 1, 2, \ldots, n$, the joint posterior distribution of model (18) can be easily computed by

$$\begin{aligned} p(\beta, \sigma^2, \sigma_s^2, \boldsymbol{\rho}, \boldsymbol{b}, \boldsymbol{S}|\boldsymbol{y}) &= \frac{p(\boldsymbol{y}|\beta, \boldsymbol{b}, \boldsymbol{S}, \sigma^2)p(\boldsymbol{b})p(\boldsymbol{S}|\sigma_s^2, \boldsymbol{\rho})p(\beta, \sigma^2, \sigma_s^2, \boldsymbol{\rho})}{p(\boldsymbol{y})} \\ &\propto p(\boldsymbol{y}|\beta, \boldsymbol{b}, \boldsymbol{S}, \sigma^2)p(\boldsymbol{b})p(\boldsymbol{S}|\sigma_s^2, \boldsymbol{\rho})p(\beta, \sigma^2, \sigma_s^2, \boldsymbol{\rho}) \\ &= \left\{\prod_{i=1}^{n} f\Big(\boldsymbol{y}_i|\beta, b_i, \boldsymbol{S}_i, \sigma^2\Big)f(b_i)f\Big(\boldsymbol{S}_i|\sigma_s^2, \rho_i\Big)\right\}p(\beta)p(\sigma^2)p(\sigma_s^2)p(\boldsymbol{\rho}). \end{aligned} \tag{34}$$

To realize the Bayesian inference of the model, it is necessary to use the Bayesian theorem to obtain the conditional distribution of each parameter. The lengthy derivations of all conditional probability distributions are given in the Appendix A at the end of the paper.

Performing a posteriori inference on the covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ $(m \times m)$ is equivalent to estimating the posteriori mean of $\boldsymbol{\Sigma}$:

$$\hat{\boldsymbol{\Sigma}} = E\{\boldsymbol{\Sigma}|\boldsymbol{y}\} = E\Big\{\sigma_b^2\boldsymbol{A} + \sigma_s^2\tilde{\boldsymbol{C}}(\rho) + \sigma^2\boldsymbol{I}\Big|\boldsymbol{y}\Big\}, \quad m \times m \tag{35}$$

where $\boldsymbol{A} = ee'$. The posteriori estimate for the correlation matrix $\boldsymbol{R}$ can be obtained by using $\hat{\boldsymbol{\Sigma}}$ to calculate the estimated $\hat{\boldsymbol{R}}$ of the correlation matrix $\boldsymbol{R}$:

$$\hat{\boldsymbol{R}} = \boldsymbol{R}_0\hat{\boldsymbol{\Sigma}}\boldsymbol{R}_0, \quad \boldsymbol{R}_0 = \mathrm{diag}\Big(\hat{\sigma}_{11}^{-\frac{1}{2}}, \ldots, \hat{\sigma}_{mm}^{-\frac{1}{2}}\Big), \quad \hat{\boldsymbol{\Sigma}} = \Big(\hat{\sigma}_{ij}\Big) : \ m \times m. \tag{36}$$

The mean square error loss function is used to evaluate the performance of the posterior mean:

$$MSE(\boldsymbol{\Sigma}) = \left(\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{\sigma}_{ij} - \sigma_{ij})^2\right)^{\frac{1}{2}}. \tag{37}$$

Compared with the Bayesian estimates of the covariance matrix and correlation matrix under other loss functions [23], the most common one is the entropy loss function defined by

$$L_1(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - m. \tag{38}$$

The quadratic loss function is given by

$$L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - I)^2. \tag{39}$$

The Bayesian estimates of the covariance matrices based on the loss functions $L_1(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ and $L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ are given by

$$\hat{\boldsymbol{\Sigma}} = \Big[E\Big(\boldsymbol{\Sigma}^{-1}|\boldsymbol{y}\Big)\Big]^{-1} \quad \text{and} \quad \mathrm{vec}(\hat{\boldsymbol{\Sigma}}) = \Big[E\Big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}|\boldsymbol{y}\Big)\Big]^{-1}\mathrm{vec}\Big[E\Big(\boldsymbol{\Sigma}^{-1}|\boldsymbol{y}\Big)\Big], \tag{40}$$

respectively, where vec stands for the column vectorization of a matrix, and "$\otimes$" denotes the Kronecker product of matrices. Similarly, the Bayesian estimate of correlation moment $\boldsymbol{R}$ under various loss functions can be obtained.

7.2. Simulation Specification and Display of Empirical Results

In the simulation, we firstly set up the four different covariance structures as in the following Equations (41) and the priors as given in the following Equations (43)–(46). Then, we run the MCMC training trial 2000 times. After seeing the convergence trend approach relative stability after 2000 training trials, we generated 20 datasets consisting of 100 sample points of length 6 for each object. This is equivalent to the longitudinal data structure in Table 1, with $n = 100$, $p = 6$, and $k = 1$ for each generated longitudinal dataset.

The four covariance matrices are designed as follows:

$$\begin{aligned} \mathbf{\Sigma}_1(i,j) &= 10I(i=j) + 7I(i \neq j) \\ \mathbf{\Sigma}_2(i,j) &= 10 \times 0.4^{|i-j|} \\ \mathbf{\Sigma}_3(i,j) &= 0.3\mathbf{\Sigma}_1(i,j) + 0.7\mathbf{\Sigma}_2(i,j) \\ \mathbf{\Sigma}_4(i,j) &= \frac{10}{\sqrt{1+|i-j|}}. \end{aligned} \tag{41}$$

The root mean square error (RMSE) for estimating $\mathbf{\Sigma}$ is computed by

$$RMSE = \frac{1}{20}\sum_{k=1}^{20} \text{MSE}, \quad MSE = \sqrt{\frac{1}{36}\sum_{i=1}^{6}\sum_{j=1}^{6}(\mathbf{\Sigma}_{ij} - \hat{\mathbf{\Sigma}}_{ij})^2}. \tag{42}$$

The RMSE for estimating the correlation matrix $\boldsymbol{R}$ is computed similarly.

For each of the four different covariance structures, we used the R-package (called Pandas, available upon request) to perform a preliminary analysis on the simulated data with a specified prior distribution, respectively. Then, we employed the MCMC algorithm to perform the sampling estimation on each parameter in the model. We used the three methods mentioned before to perform the convergence assessment on the sampling results. Finally, we obtained the estimates for the four types of covariance and correlation matrices. The estimation error was computed based on three types of loss functions. The inverse-Wishart estimation error was also obtained for comparison.

For the case of the CS covariance structure , the prior distribution of the parameter in the model is specified as follows:

$$\begin{aligned} G_0 = N(0,10), \alpha = 0.75, \quad & \beta \sim N(0,25), \\ b_i \overset{iid}{\sim} N(0,7), \quad & \sigma^2 \sim IG(3,2), \\ & \sigma_s^2 \sim IG(6,10). \end{aligned} \tag{43}$$

Each parameter in the model is sampled and estimated using the MCMC method. The last 1000 iterations were taken to draw the posterior sampling trend diagram, as shown in Figure 4.

The negative ELBO loss histogram and energy graph estimated by the model with the CS structure are shown in Figures 5 and 6, respectively, where the energy graph in Figure 6 was generated by the Python package PyMC3 (https://www.pymc.io/projects/docs/en/v3/pymc-examples/examples/getting_started.html) (accessed on 26 April 2023), which displays two simulated density curves: the blue one stands for the energy value at each MCMC sample point subtracted by the average energy (like conducting data centerization); the green one stands for the difference of the energy function (like deriving the derivative of a differential function). A normal energy distribution from an MCMC sampling indicates the sampling process tends to a stable point. It implies convergence of the MCMC sampling. More details on energy computation in MCMC sampling by PyMC3 can be found on this website (https://www.pymc.io/projects/docs/en/v3/api/inference.html#module-pymc3.sampling). Figures 5 and 6 incorporate both popular methods for evaluating the convergence of the MCMC sampling in our Monte Carlo study. All energy graphs in the

subsequent context have the same interpretation as they do here. We skip the tedious interpretations for all other energy graphs to save space.

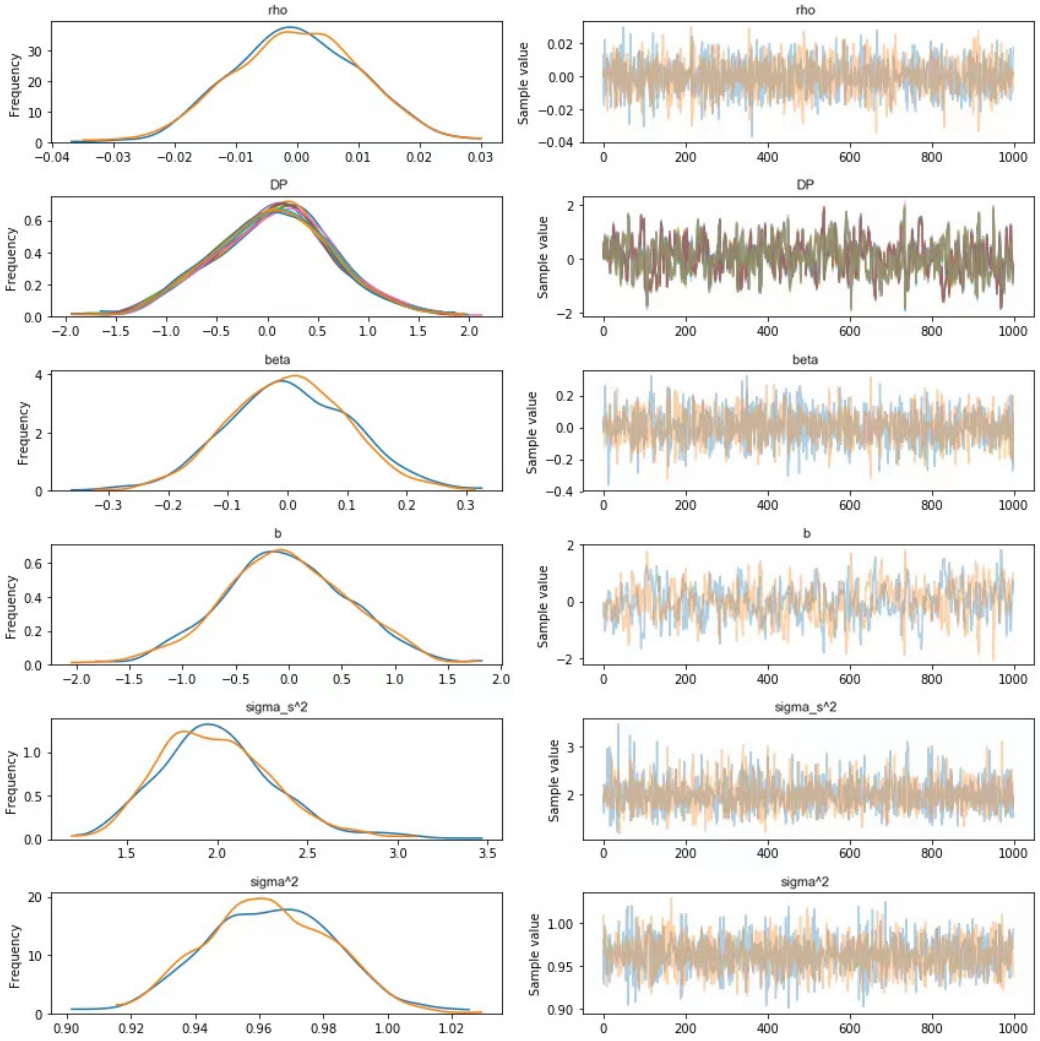


Figure 4. Sampling results under the CS structure: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed for $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the CS structure and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (43)). The orange line stands for sampling distribution under the SPAR model (18), and the blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (orange) and the inverse-Wishart model (blue), respectively, as well as the Markov chain for each individual DP (different color for each). Each graph in the right column shows the relationship between each estimated parameter (the ordinate) versus the number of random samples in the abscissa, ranging from 1 to 1000. Each graph in the left column presents the kernel-estimated density function of the parameter from the last 1000 samples.

As can be seen from Figures 5 and 6, after several iterations of the MCMC algorithm, the negative ELBO loss is stable between 0 and 25, and the sample energy conversion distribution is basically consistent with the true energy distribution. Based on the sampling distribution and the trend graph, we can conclude that the MCMC algorithm is convergent.

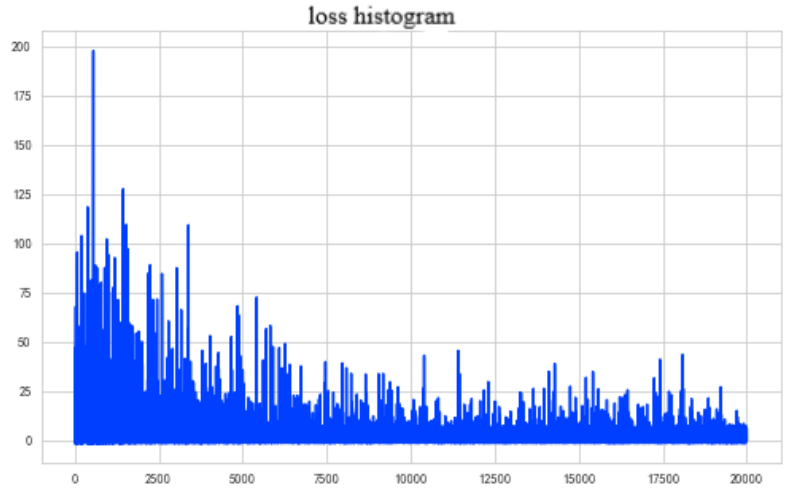


Figure 5. Negative ELBO loss histogram: in Figure 5, the horizontal axis stands for the number of iterations in the MCMC sampling with size $n = 100$, the vertical axis for the negative ELBO loss.

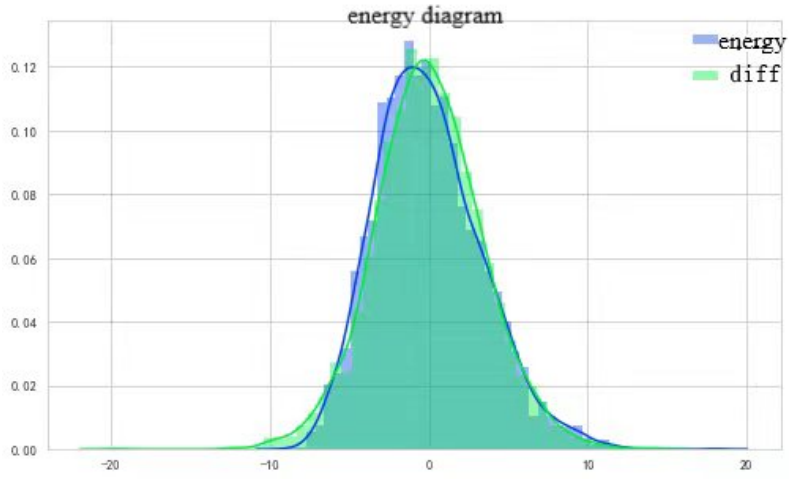


Figure 6. Energy graph: in Figure 6, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

For the case of the AR covariance structure, the prior distribution of each parameter in the model is specified as follows:

$$\begin{array}{ll} G_0 = N(0.4, 10), \alpha = 0.75, & \beta \sim N(0, 10) \\ b_i \overset{iid}{\sim} N(0, 0.01), & \sigma^2 \sim IG(1.25, 0.01) \\ & \sigma_s^2 \sim IG(3.76, 9.566) \end{array} \tag{44}$$

Each parameter in the model is sampled and estimated using the MCMC method. The last 1000 iterations were taken to draw the posterior sampling trend diagram, as shown in Figure 7. Note that the DP-estimated density curves show different central locations from those in Figure 4, because they are generated from different prior distributions with different covariance structures (see the difference between Equations (43) and (44)).

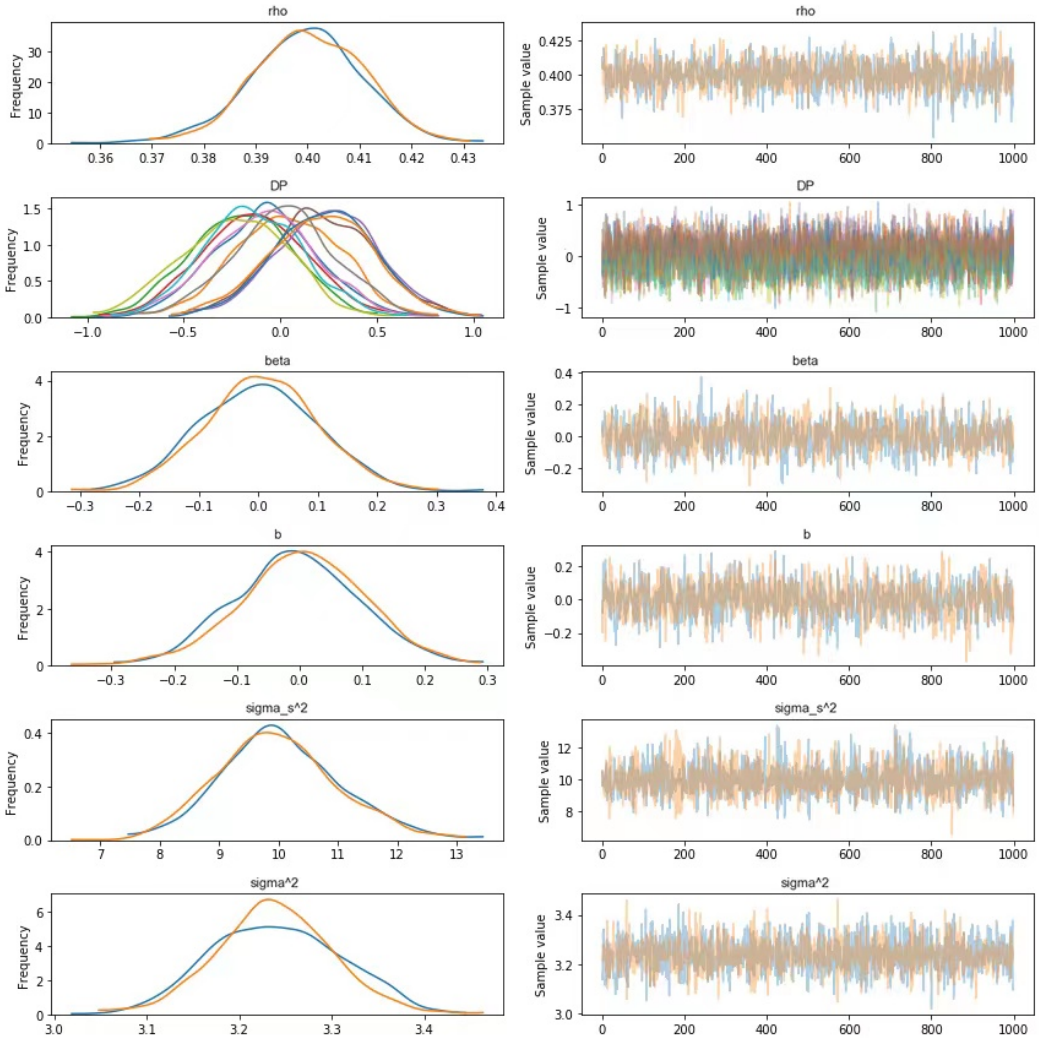


Figure 7. Sampling results under the AR structure: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the AR structure and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (44)). The orange line stands for sampling distribution under the SPAR model (18) and the blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (orange) and the inverse-Wishart model (blue), respectively, as well as the Markov chain for each individual DP (different color for each). Figures 4 and 7 have the same structure in both axes for the two columns of graphs.

The histogram of negative ELBO loss and the energy graph for when the AR structure is estimated are shown in Figures 8 and 9, respectively. They show that the histogram of the negative ELBO loss tends to be stable after several iterations, which is basically below 50, and the sample energy conversion distribution is basically consistent with the true energy distribution. Based on the posteriori sampling and the trend graph, it can be concluded that the algorithm converges quickly.

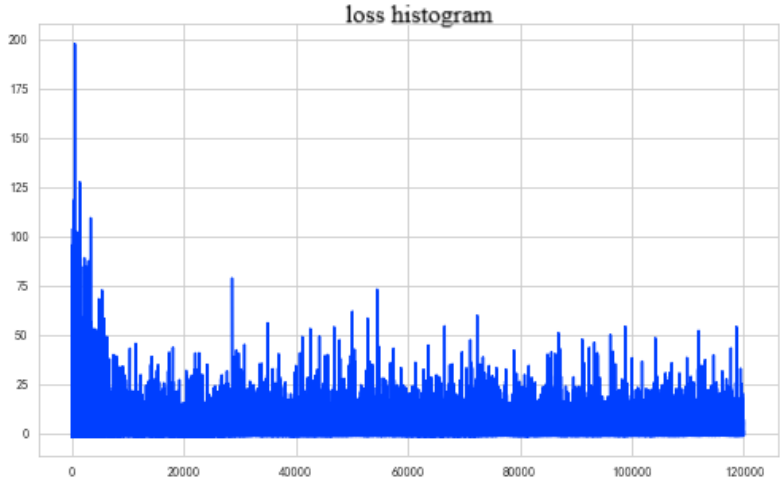


Figure 8. Negative ELBO loss histogram: in Figure 8, the horizontal axis stands for the number of iterations in the MCMC sampling with size $n = 100$, the vertical axis for the negative ELBO loss.

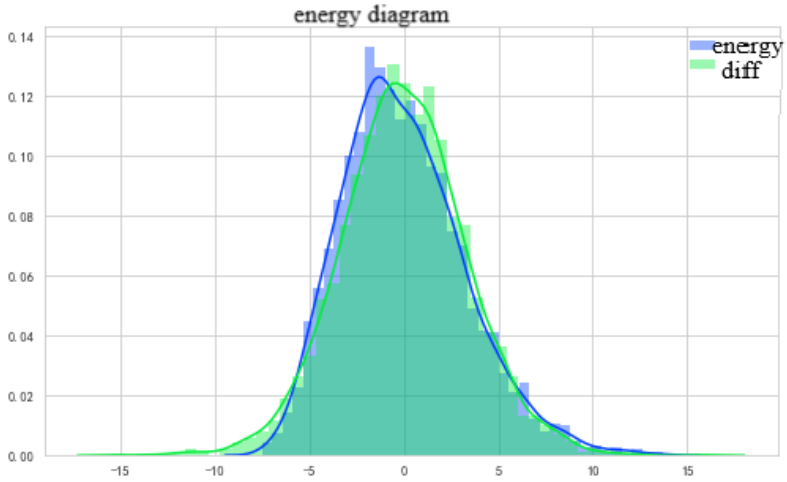


Figure 9. Energy graph: In Figure 9, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

For the covariance of the mixed structure of CS and AR, the prior distribution of each parameter in the model is specified as follows:

$$\begin{aligned} G_0 = N(0.4, 10),\ \alpha = 0.75, \quad & \beta \sim N(0, 10), \\ b_i \overset{iid}{\sim} N(0, 2.1), \quad & \sigma^2 \sim IG(15.2, 14), \\ & \sigma_s^2 \sim IG(3.82, 6.867). \end{aligned} \tag{45}$$

Each parameter in the model is sampled and estimated using the MCMC method. The last 1000 iterations were taken to draw the posterior sampling trend graph, as shown in Figure 10. The negative ELBO loss histogram and the model energy graph are shown in Figures 11 and 12, respectively.

As can be seen from the histogram of the negative ELBO loss in Figure 11, the negative ELBO loss is basically under control between 0 and 30, after several iterations of the algorithm, and the sample energy conversion distribution is basically consistent with the true energy distribution. Based on the posteriori sampling and the trend graph, we can conclude that the algorithm is convergent.

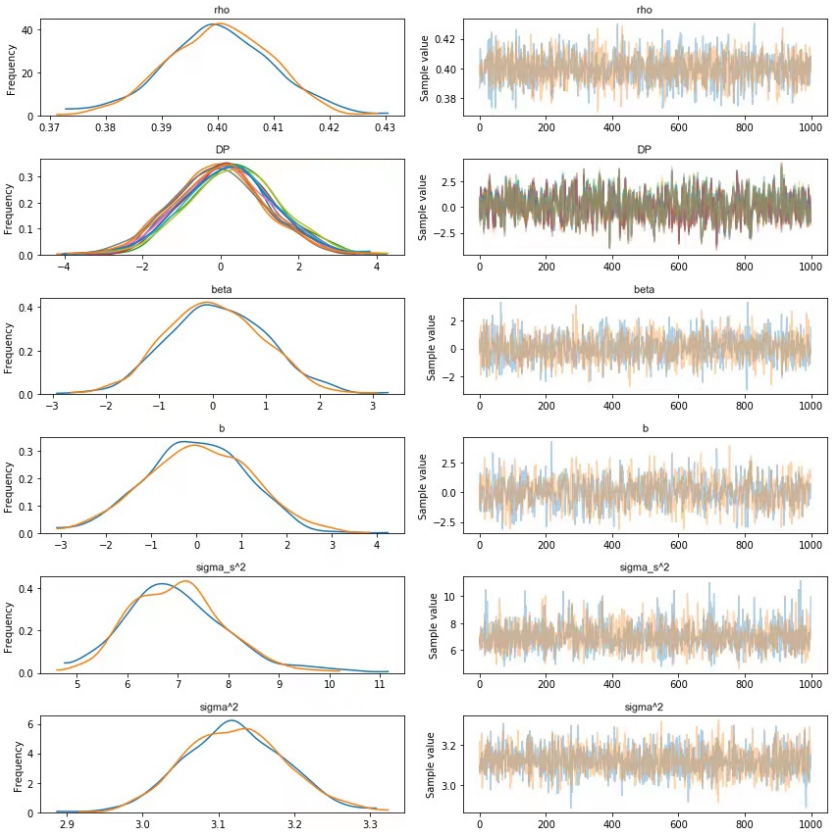


Figure 10. Sampling results under the mixed structure of CS and AR: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the mixed structure of CS and AR and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (45)). The orange line stands for sampling distribution under the SPAR model (18) and the blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (orange) and the inverse-Wishart model (blue), respectively, as well as the Markov chain for each individual DP (different color for each). Figures 4 and 10 have the same structure in both axes for the two columns of graphs.

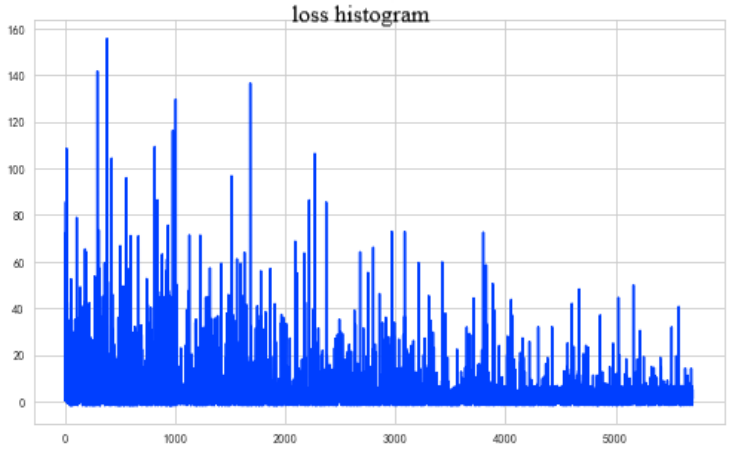


Figure 11. Negative ELBO loss histogram: in Figure 11, the horizontal axis stands for the number of iterations in the MCMC sampling with size $n = 100$, the vertical axis for the negative ELBO loss.

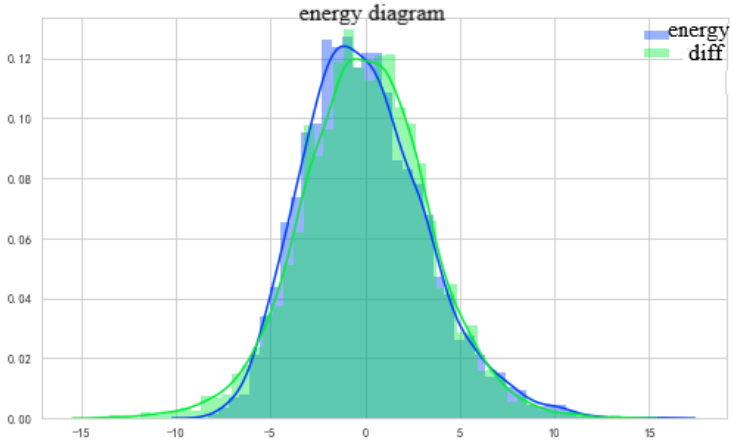


Figure 12. Energy graph: in Figure 12, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

For the case of independent structure covariance, the prior distributions of the parameters in the model are specified as follows:

$$\begin{aligned} G_0 = N(0.672, 10),\ \alpha = 0.75, \quad & \beta \sim N(0, 20), \\ b_i \overset{iid}{\sim} N(0, 1.106), \quad & \sigma^2 \sim IG(2.22, 7.48), \\ & \sigma_s^2 \sim IG(3.20, 3.443). \end{aligned} \tag{46}$$

The MCMC method was used to sample and estimate each parameter in the model. The last 1000 iterations were taken to draw the posterior sampling trend graph shown in Figure 13.

The negative ELBO loss histogram and the model energy graph are shown in Figures 14 and 15, respectively. It can be seen that the negative ELBO loss approaches 0 quickly. Based on the posterior sampling and the trend diagram, we can conclude that the algorithm is convergent.

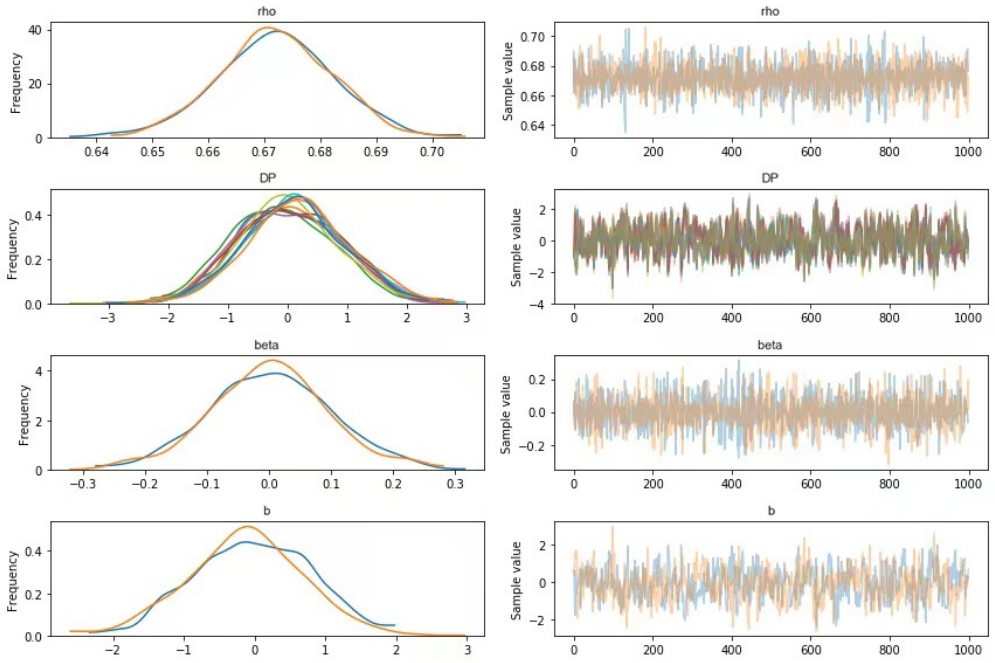


Figure 13. *Cont.*

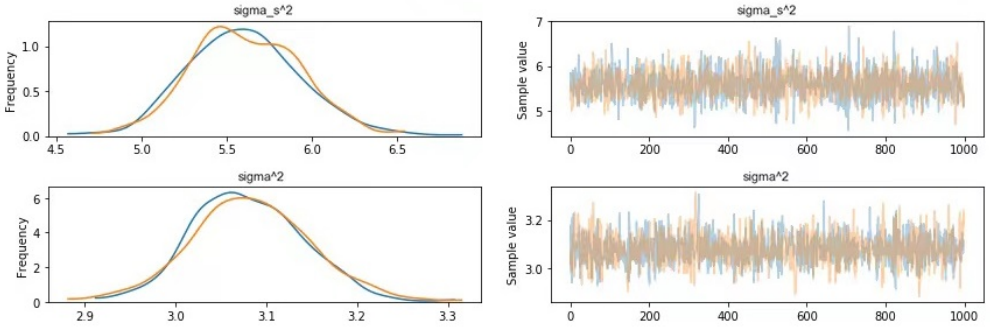


Figure 13. Sampling results under the independent structure: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the independent structure and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (46)). The orange line stands for sampling distribution under the SPAR model (18) and the blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (orange) and the inverse-Wishart model (blue), respectively, as well as the Markov chain for each individual DP (different color for each). Figures 4 and 13 have the same structure in both axes for the two columns of graphs.

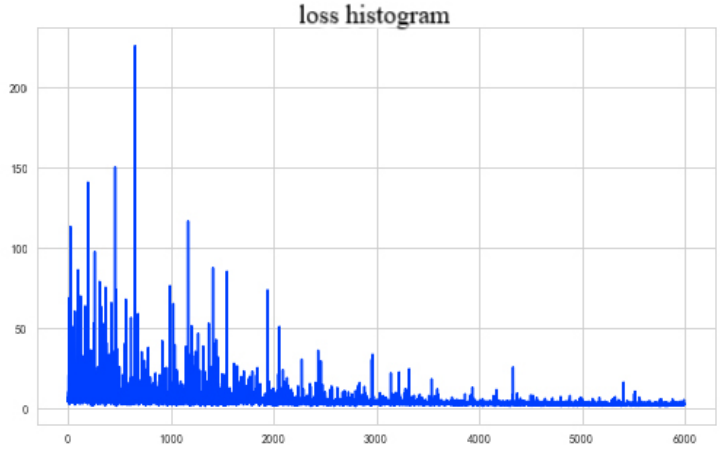


Figure 14. Negative ELBO loss histogram: in Figure 14, the horizontal axis stands for the number of iterations in the MCMC sampling with size $n = 100$, the vertical axis for the negative ELBO loss.

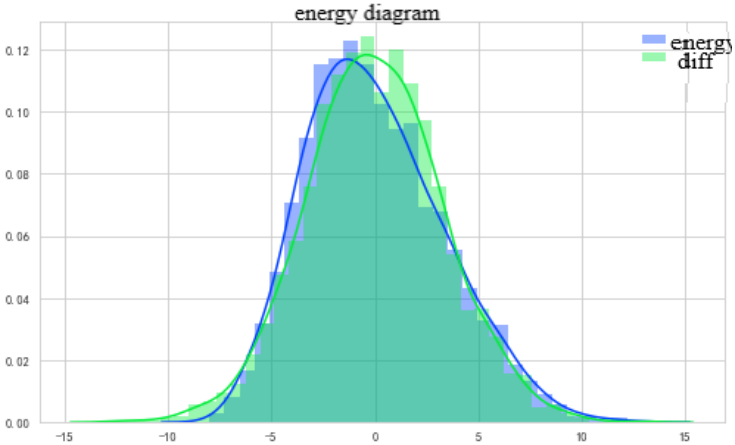


Figure 15. Energy graph: in Figure 15, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

The above model is applied to the estimation of covariance matrices, correlation matrices, and model errors. We compare it with with the inverse-Wishart method. The outcomes of the estimation errors are shown in Table 2:

Table 2. Estimated RMSE, L_1, and L_2 based on covariance matrix $\boldsymbol{\Sigma}$.

Loss Function	Model	C1	C2	C3	C4
RMSE	SPAR	0.9960	0.1037	0.8995	1.5358
	Inv-W	2.9691	2.9352	3.4804	3.3285
L_1	SPAR	0.1810	0.1201	1.3217	0.9558
	Inv-W	1.8389	1.4632	3.0516	1.2926
L_2	SPAR	0.1289	0.1345	1.7575	0.6044
	Inv-W	0.2405	0.9576	2.8083	1.7565

C1 represents the covariance of the complex symmetry (CS) structure; C2 represents the AR structure covariance; C3 represents the mixed structure covariance of CS and AR; C4 represents independent structure covariance; and Inv-W = inverse-Wishart.

Similarly, the estimation results of the four corresponding correlation matrices are shown in Table 3:

Table 3. Estimated RMSE, L_1, and L_2 based on correlation matrix $\boldsymbol{R}$.

Loss Function	Model	C1	C2	C3	C4
RMSE	SPAR	0.1628	1.1417	0.3775	0.1542
	Inv-W	0.2308	2.4152	0.9115	0.2771
L_1	SPAR	0.0956	0.5482	0.0631	1.2354
	Inv-W	0.1519	1.1424	0.1756	1.9137
L_2	SPAR	0.1013	0.9809	0.4937	0.5854
	Inv-W	0.1238	1.1935	0.5205	0.6484

C1 represents the covariance of the complex symmetry (CS) structure; C2 represents the AR structure covariance; C3 represents the mixed structure covariance of CS and AR; and C4 represents independent structure covariance.

In Tables 2 and 3, the covariance models are compared with each other based on three types of loss functions. In estimating the four covariance structures, the SPAR model performs better than the traditional inverse-Wishart method for each covariance structure. The estimation error of the SPAR model is much smaller than that of the inverse-Wishart method. When estimating the correlation matrix, except for the relatively poor SPAR performance under the strict AR structure, all other models perform roughly the same based on the quadratic loss function $L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$. Based on the comparison of the estimation errors in Tables 2 and 3, the SPAR model shows fairly good performance in estimating the covariance matrix and correlation matrix.

7.3. Analysis of a Real Wind Speed Dataset

To verify the effectiveness of the SPAR model built in this paper in practical application, we employ the Hongming data, which contain the ground meteorological data of Dingxin Station, Jinta County, Jiuquan City, Gansu Province, China. We use the SPAR model and the MCMC method to estimate the covariance matrix and correlation matrix under four different covariance structures (CS, AR, the mixture of CS and AR, and unstructured) and compare the estimates with those from the traditional Bayesian estimation by the inverse-Wishart method. The covariance structure of the four cases is shown in Equation (41), and the mean square error is shown in Equation (42). The real data are arranged into a longitudinal data structure, as shown in Table 1 with $n = 100$, $p = 6$, and $k = 1$. The following graphs are plotted in the same way as in the corresponding graphs in the Monte Carlo study in Section 7.2 but with real data input in the same Python program.

For the case of CS covariance structure , each parameter in the model is sampled and estimated using the MCMC method. The last 1000 iterations are taken to draw the posterior sampling trend diagram, as shown in Figure 16.

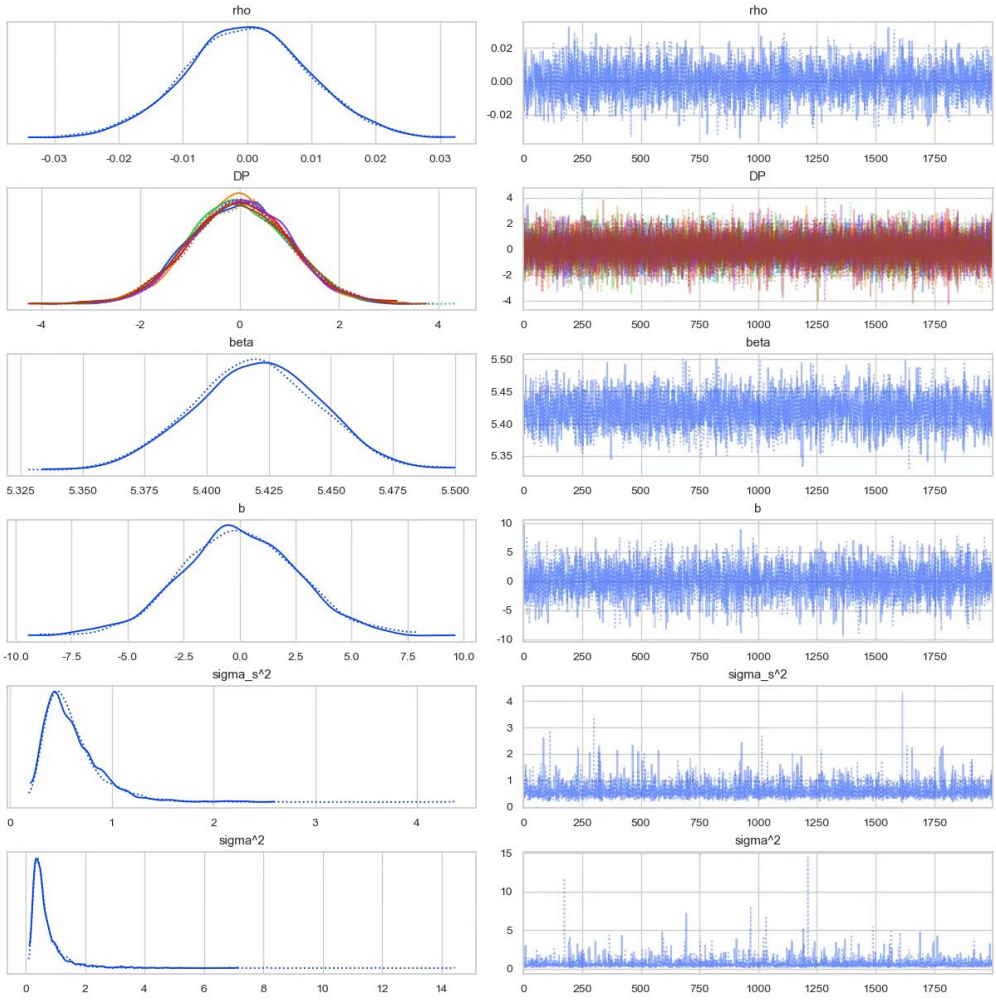


Figure 16. Sampling results under the CS structure: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the CS structure and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (43)). The real blue line stands for sampling distribution under the SPAR model (18) and the dotted blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (real line) and the inverse-Wishart model (dotted line), respectively, as well as the Markov chain for each individual DP (different color for each). Figures 4 and 16 have the same structure in both axes for the two columns of graphs.

The negative ELBO loss histogram and the model energy graph are shown in Figures 17 and 18, respectively. It can be seen that the negative ELBO loss approaches 0 quickly, and the sample energy conversion distribution is basically consistent with the true energy distribution. Based on the posterior sampling and the trend diagram, we can conclude that the algorithm is convergent.

For the case of the AR covariance structure, each parameter in the model is sampled and estimated using the MCMC method. The last 1000 iterations are taken to draw the posterior sampling trend diagram, as shown in Figure 19:

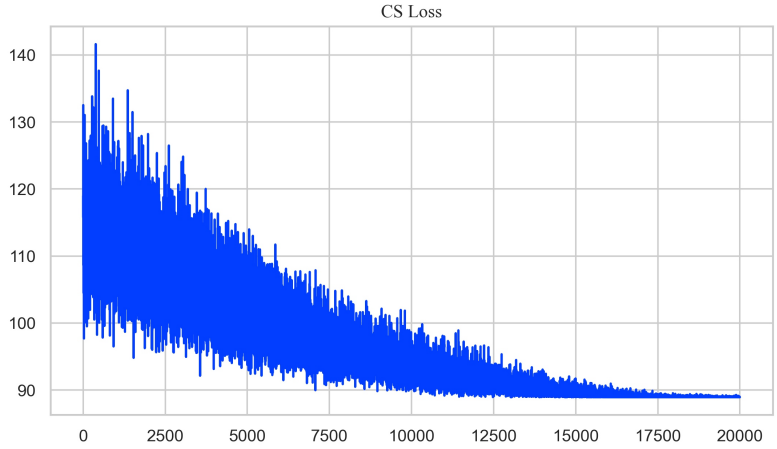


Figure 17. Negative ELBO loss histogram: in Figure 17, the horizontal axis stands for the number of iterations in the MCMC algorithm with size $n = 100$, the vertical axis for the negative ELBO loss.

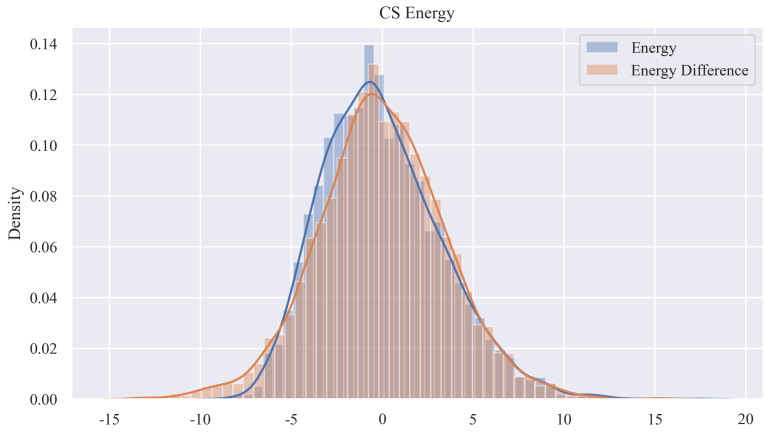


Figure 18. Energy graph: In Figure 18, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

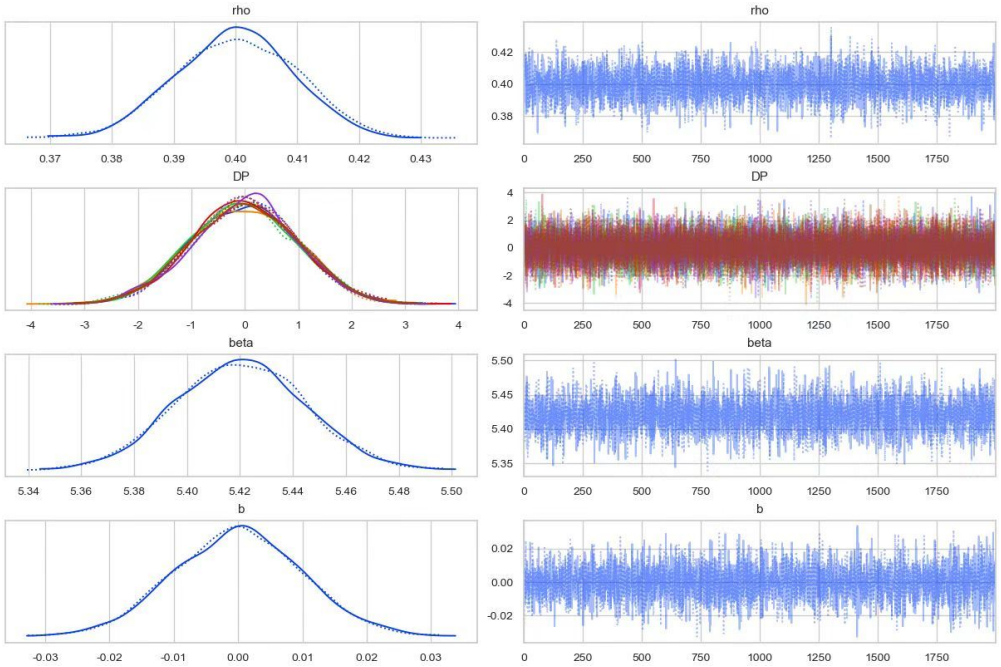


Figure 19. *Cont.*

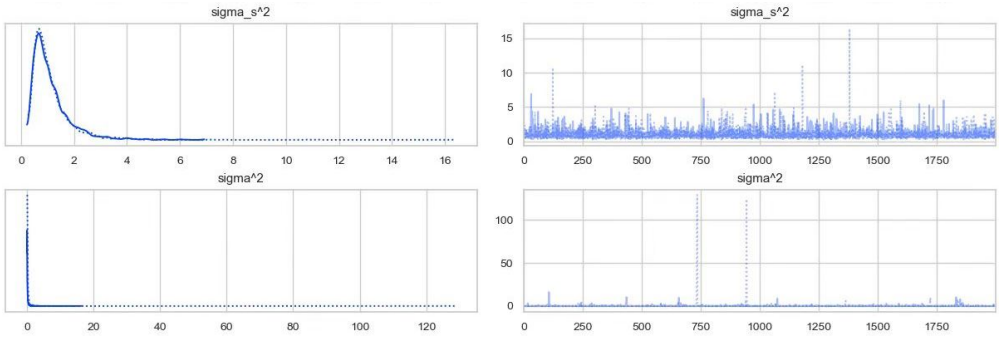


Figure 19. Sampling results under the AR structure: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the AR structure and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (44)). The real blue line stands for sampling distribution under the SPAR model (18) and the dotted blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (real line) and the inverse-Wishart model (dotted line), respectively, as well as the Markov chain for each individual DP (different color for each). Figures 4 and 19 have the same structure in both axes for the two columns of graphs.

The negative ELBO loss histogram and the energy graph estimated by the model under the AR structure are shown in Figures 20 and 21, respectively. It shows that the histogram of the negative ELBO loss approaches 0 quickly. Based on the posteriori sampling and the trend graph, it can be concluded that the algorithm converges quickly.

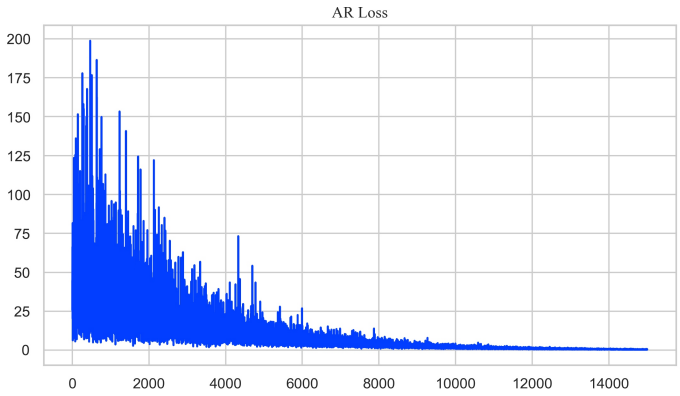


Figure 20. Negative ELBO loss histogram: in Figure 20, the horizontal axis stands for the number of iterations in the MCMC algorithm with size $n = 100$, the vertical axis for the negative ELBO loss.

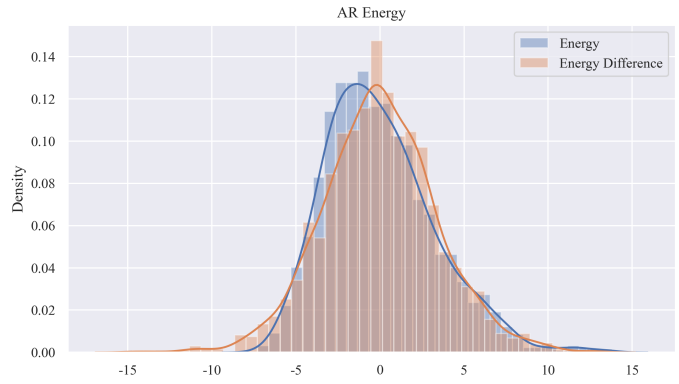


Figure 21. Energy graph: in Figure 21, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

For the covariance of the mixed structure of CS and AR, each parameter in the model is sampled and estimated using the MCMC method. The last 1000 iterations are taken to draw the posterior sampling trend graph, as shown in Figure 22. The negative ELBO loss histogram and the model energy graph are shown in Figures 23 and 24, respectively.

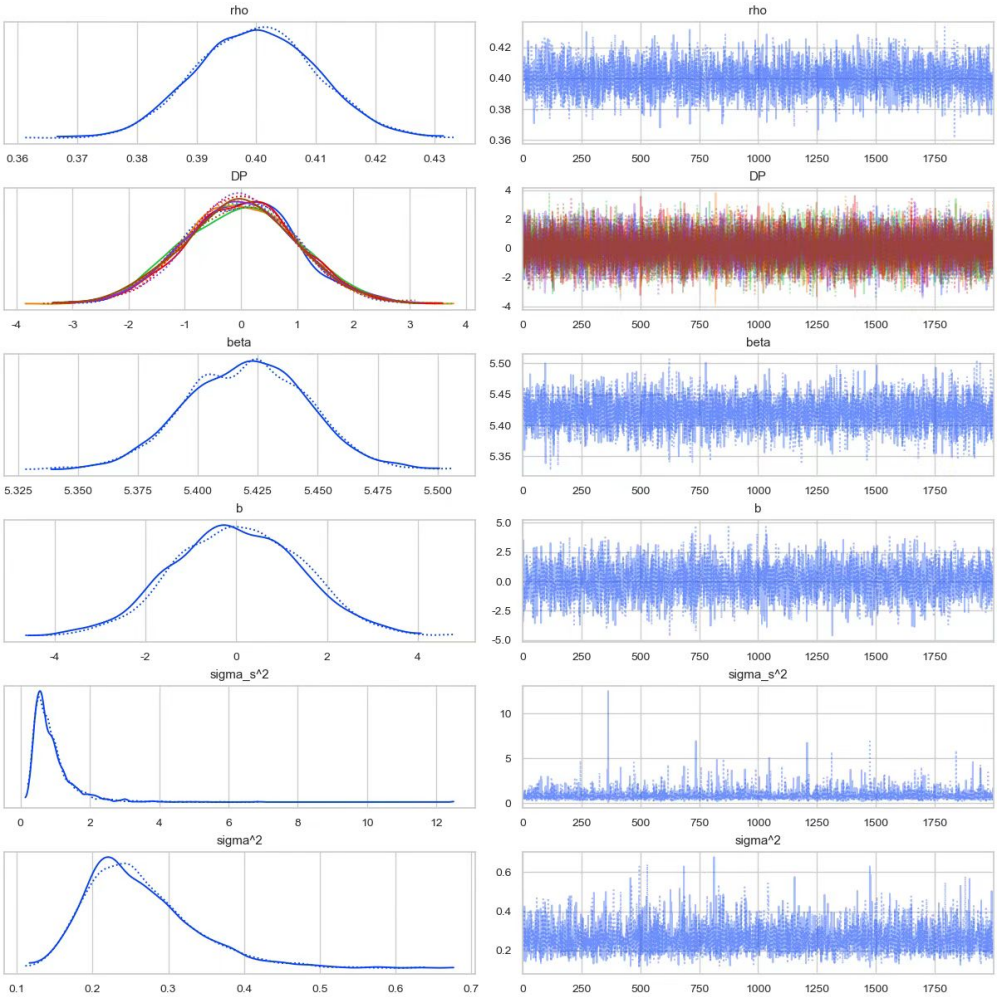


Figure 22. Sampling results under the mixed structure of CS and AR: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the mixed structure of CS and AR and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (45)). The real blue line stands for sampling distribution under the SPAR model (18) and the dotted blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (real line) and the inverse-Wishart model (dotted line), respectively, as well as the Markov chain for each individual DP (different color for each). Figures 4 and 22 have the same structure in both axes for the two columns of graphs.

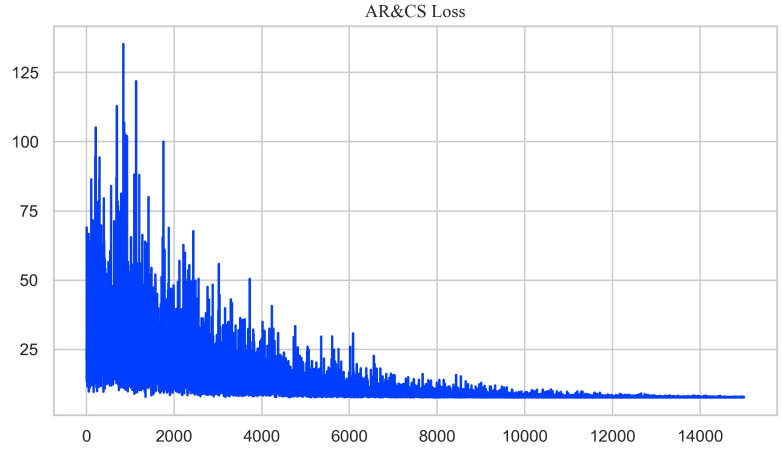


Figure 23. Negative ELBO loss histogram: in Figure 23,the horizontal axis stands for the number of iterations in the MCMC algorithm with size $n = 100$, the vertical axis for the negative ELBO loss.

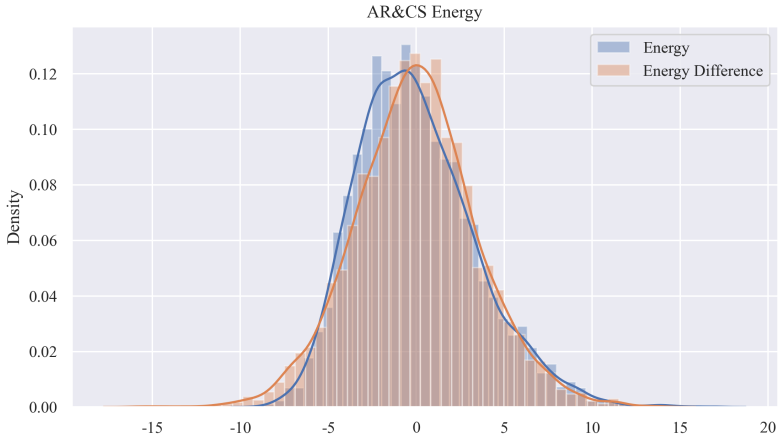


Figure 24. Energy graph: in Figure 24, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

As can be seen from the histogram of the negative ELBO loss in Figure 23, the negative ELBO loss approaches 0 quickly after several iterations of the algorithm. Based on the posteriori sampling and the trend graph, we can conclude that the algorithm is convergent.

For the case of independent structure covariance, the MCMC method is used to sample and estimate each parameter in the model. The last 1000 iterations are taken to draw the posterior sampling trend graph shown in Figure 25.

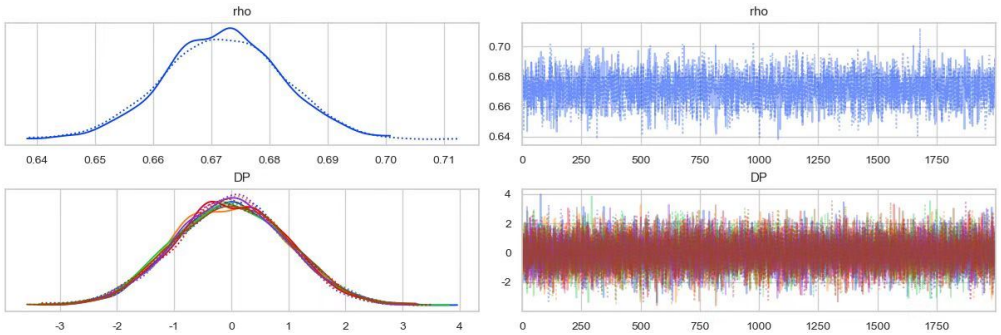


Figure 25. *Cont*.

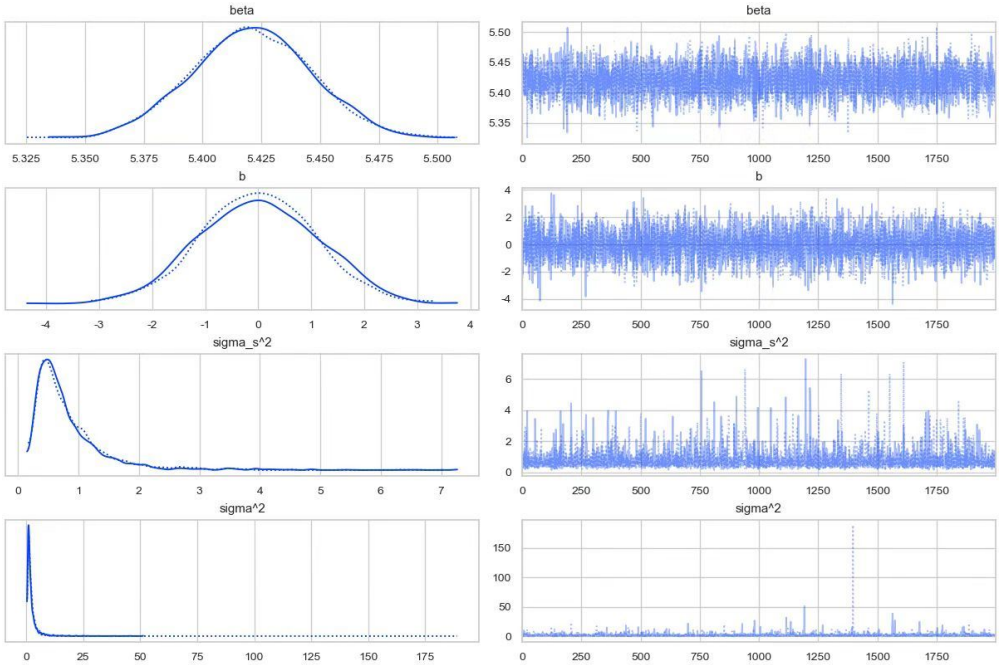


Figure 25. Sampling results under an independent structure: The left panel gives the sampling distributions (smoothed by the kernel density estimation) for the parameters and the DP, which contain six pairs of DPs (each subject is observed $p = 6$ variables in the simulation), each pair consisting of a DP generated from model (18) with the independent structure and a DP generated from the traditional model with the inverse-Wishart structure (sampling population specified by Equations (23), (24) and (46)). The real blue line stands for sampling distribution under the SPAR model (18) and the dotted blue line for the inverse-Wishart model; the right panel gives the the graphs of two Markov chains under the SPAR model (real line) and the inverse-Wishart model (dotted line), respectively, as well as the Markov chain for each individual DP (different color for each). Figures 4 and 25 have the same structure in both axes for the two columns of graphs.

The negative ELBO loss histogram and the model energy graph are shown in Figures 26 and 27, respectively. It can be seen that the negative ELBO loss approaches 0 quickly. Based on the posterior sampling and trend diagram, we can conclude that the algorithm is convergent.

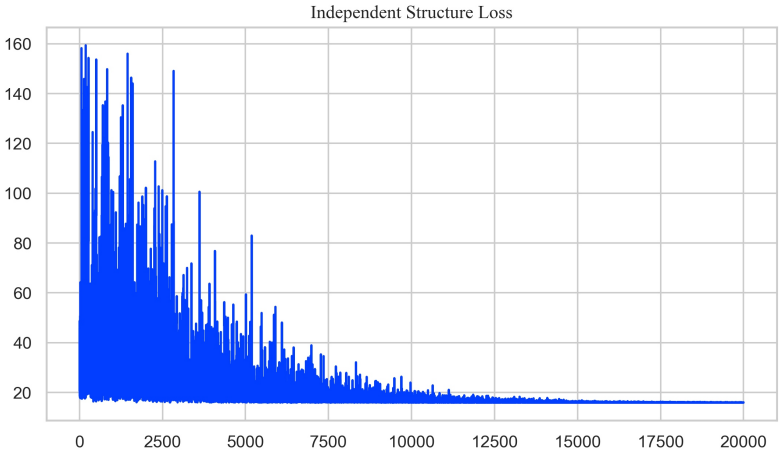


Figure 26. Negative ELBO loss histogram: in Figure 26, the horizontal axis stands for the number of iterations in the MCMC algorithm with size $n = 100$, the vertical axis for the negative ELBO loss.

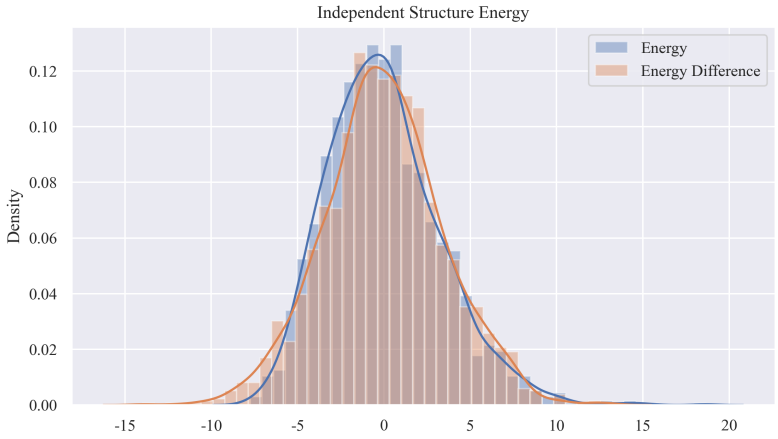


Figure 27. Energy graph: in Figure 27, the estimated distribution of energy is based on 1000 samples with size $n = 100$.

In Tables 4 and 5, the covariance models are compared with each other based on three types of loss functions, RMSE, L_1, and L_2. When using the four covariance structures based on either the covariance matrix or the correlation matrix, the SPAR model always performs better than the traditional inverse-Wishart method—it always has a smaller value of RMSE, L_1, or L_2 when comparing the SPAR model with the Inv-W model under the same covariance structure C1, C2, C3, or C4 in Tables 4 and 5.

Table 4. Estimated RMSE, L_1, and L_2 based on covariance matrix $\boldsymbol{\Sigma}$.

Loss Function	Model	C1	C2	C3	C4
RMSE	SPAR	0.0831	0.0962	0.08837	0.0981
	Inv-W	0.4120	0.6826	0.3181	0.4064
L_1	SPAR	0.1360	0.1996	0.1465	0.2508
	Inv-W	46.6743	16.8967	19.0260	14.3543
L_2	SPAR	0.0870	0.0826	0.2084	0.3517
	Inv-W	661.4643	142.3157	450.6686	340.6994

C1 represents the covariance of the complex symmetry (CS) structure; C2 represents the AR structure covariance; C3 represents the mixed structure covariance of CS and AR; C4 represents independent structure covariance; and Inv-W = inverse-Wishart.

Table 5. Estimated RMSE, L_1, and L_2 based on correlation matrix $\boldsymbol{R}$.

Loss Function	Model	C1	C2	C3	C4
RMSE	SPAR	0.0187	0.5359	0.1976	0.1985
	Inv-W	0.6585	0.6486	0.6515	0.7580
L_1	SPAR	0.0066	0.1670	0.0662	0.1846
	Inv-W	0.0968	4.8245	0.3174	0.8035
L_2	SPAR	0.0036	0.0132	0.0080	0.0939
	Inv-W	0.0079	1.3683	0.1202	0.3488

C1 represents the covariance of the complex symmetry (CS) structure; C2 represents the AR structure covariance; C3 represents the mixed structure covariance of CS and AR; and C4 represents independent structure covariance.

8. Concluding Remarks

The SPAR model (18) provides an explicitly complete semiparametric solution to the estimation of model parameters through the MCMC algorithm. Compared with the model formulation in Quintana et al. [4], the MCMC algorithm for posterior inference on model (18) is easier to implement and may converge faster because of the explicit simple posterior distributions of the model parameters. An effective and fast-converging MCMC algorithm plays an important role in Bayesian statistical inference. The SPAR model (18) gains some trade-off in easy implementation and fast convergence in the MCMC algorithm by imposing simpler assumptions on model parameters.

With regard to the option of choosing the initial values for estimating model parameters by the MCMC algorithm, we recommend using a numerical optimization method such as the maximum posteriori (MAP) estimation to obtain an estimator as the initial value of a parameter. It is likely to speed up the convergence speed of the sampling parameter. We employ the Gibbs sampling algorithm when estimating the parameters in the model. The convergence diagnosis of Markov chains generated by the MCMC algorithm is assessed by the posterior sampling trend plot, negative ELBO histogram, and the energy graph, which show the observation of fast convergence of the MCMC sampling process. By applying the SPAR model to four different covariance structures through both Monte Carlo study and a real dataset, we illustrate its effectiveness in handling nonstationary forms of covariance structures and its domination over the traditional inverse-Wishart method.

It should be pointed out that the effectiveness and fast convergence of the MCMC algorithm depend on both model assumption and the priors of model parameters. Our Monte Carlo study was carried out by choosing normal priors for the model parameters and the inverse Gamma distribution for the variance components. This choice led to the easy implementation of the MCMC algorithm. It will be an interesting future research direction to develop some meaningful criteria for model and algorithm comparison in the area of Bayesian nonparametric longitudinal data analysis. The main purpose of our paper is to give an easily implementable approach to this area with a practical illustration. We can conclude that the complete semiparametric approach to Bayesian longitudinal data analysis in this paper is a significant complement to the area studied by some influential peers, such as Mukhopadhyay and Gelfand (1997, [21]), Quintana et al. (2016, [4]), and others.

Author Contributions: G.J. developed the Bayesian semiparametric method for longitudinal data. J.L. helped validate the methodology and finalize the manuscript. F.W. conducted major data analysis. X.C. conducted initial research in the Bayesian semiparametric modeling study on longitudinal data analysis under the guidance of G.J. in her master's thesis [24]. H.L. helped with data analysis. J.C. found data and reference resources. S.C. and F.Z. finished the simulation study and also helped with data analysis and edited all figures. G.J. and J.J. wrote the initial draft. G.J. and F.W. completed the final writing—review and editing. All authors have read and agreed to the final version of the manuscript.

Funding: The research was supported by National Natural Science Foundation of China Under Grant No.41271 038, Jiajuan Liang's research is supported by a UIC New Faculty Start-up Research Fund R72021106, and in part by the Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College (UIC), project code 2022B1212010006.

Data Availability Statement: The real data and Python code presented in the present study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Derivations of Conditional Probability Distributions in Section 7.1

$$
\begin{aligned}
p(\beta|*) \propto p(\boldsymbol{y}|\beta, \boldsymbol{b}, \boldsymbol{S}, \sigma^2)p(\beta) &= \prod_{i=1}^{n}\prod_{j=1}^{m} f(y_{ij}|\beta, b_i, s_{ij}, \sigma^2)p(\beta) \\
= \quad \exp\left\{ -\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij})^2 - \tfrac{(\beta-\mu_0)^2}{2\sigma_0^2} \right\} &\cdot (2\pi)^{-\frac{(nm+1)}{2}}\sigma^{-nm}\sigma_0^{-1} \qquad \text{(A1)} \\
\propto \quad \exp\left\{ -\frac{1}{2}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(y_{ij}-\beta-b_i-s_{ij})^2}{\sigma^2} + \frac{(\beta-\mu_0)^2}{2\sigma_0^2} \right] \right\}&,
\end{aligned}
$$

where

$$\begin{aligned} & \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(y_{ij}-\beta-b_i-s_{ij})^2}{\sigma^2}+\frac{(\beta-\mu_0)^2}{\sigma_0^2} \\ = & \left(\frac{nm}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\beta^2-2\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-b_i-s_{ij})+\frac{\mu_0}{\sigma_0^2}\right]\beta \\ & +\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-b_i-s_{ij})^2-\frac{2\beta\mu_0}{\sigma_0^2}. \end{aligned} \tag{A2}$$

Combining (A1) with (A2), we have

$$p(\beta|*)\propto\exp\left\{-\frac{1}{2}\left[\left(\frac{nm}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\beta^2-2\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-b_i-s_{ij})+\frac{\mu_0}{\sigma_0^2}\right)\beta\right]\right\}. \tag{A3}$$

Let

$$\sigma_\beta^{-2}=\frac{nm}{\sigma^2}+\frac{1}{\sigma_0^2},\quad \mu_\beta=\sigma_\beta^2\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-b_i-s_{ij})+\frac{\mu_0}{\sigma_0^2}\right). \tag{A4}$$

The complete posterior distribution of β can be expressed as

$$\beta\sim N(\mu_\beta,\sigma_\beta^2). \tag{A5}$$

The conditional distribution of σ^2 is computed as follows:

$$\begin{aligned} & p(\sigma^2|*)\propto p(\boldsymbol{y}|\beta,\boldsymbol{b},\boldsymbol{S},\sigma^2)p(\sigma^2) \\ = & \exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij})^2+\beta_0\right]\right\}\cdot \\ & (2\pi)^{-\frac{nm}{2}}\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}(\sigma^2)^{-\frac{nm+2\alpha_0+2}{2}} \\ \propto & (\sigma^2)^{-\frac{nm+2\alpha_0+2}{2}}\exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}m(y_{ij}-\beta-b_i-s_{ij})^2+\beta_0\right]\right\}. \end{aligned} \tag{A6}$$

This shows that the posterior distribution of σ^2 is an inverse gamma distribution. We use the following notation:

$$\sigma^2\sim IG\left(\frac{nm}{2}+\alpha_0,\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij})^2+\beta_0\right). \tag{A7}$$

The conditional distribution of b_i is computed as follows:

$$\begin{aligned} & p(b_i|*)\propto f(\boldsymbol{y}_i|\beta,b_i,\boldsymbol{S}_i,\sigma^2)p(b_i) \\ = & (2\pi)^{-\frac{m+1}{2}}\sigma^{-m}\sigma_b^{-1}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij})^2-\frac{b_i^2}{2\sigma_b^2}\right\} \\ \propto & \exp\left\{-\frac{1}{2}\left[\sum_{j=1}^{m}\frac{(y_{ij}-\beta-b_i-s_{ij})^2}{\sigma^2}+\frac{b_i^2}{\sigma_b^2}\right]\right\}, \end{aligned} \tag{A8}$$

where

$$\begin{aligned} & \sum_{j=1}^{m} \frac{(y_{ij}-\beta-b_i-s_{ij})^2}{\sigma^2} + \frac{b_i^2}{\sigma_b^2} \\ = \; & \tfrac{1}{\sigma^2}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij})^2 + \tfrac{mb_i^2}{\sigma^2} - \tfrac{2b_i}{\sigma^2}\sum_{j=1}^{m}(y_{ij}-\beta-s_{ij}) + \frac{b_i^2}{\sigma_b^2} \\ = \; & \left(\tfrac{m}{\sigma^2}+\tfrac{1}{\sigma_b^2}\right)b_i^2 - \tfrac{2}{\sigma^2}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij})b_i. \end{aligned} \tag{A9}$$

Combining (A8) with (A9), we have

$$\begin{aligned} & \exp\left\{-\tfrac{1}{2}\left[\textstyle\sum_{j=1}^{m}\frac{(y_{ij}-\beta-b_i-s_{ij})^2}{\sigma^2} + \frac{b_i^2}{\sigma_b^2}\right]\right\} \\ \propto \; & \exp\left\{-\frac{1}{2}\left[\left(\frac{m}{\sigma^2}+\frac{1}{\sigma_b^2}\right)b_i^2 - 2\frac{1}{\sigma^2}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij})b_i\right]\right\}. \end{aligned} \tag{A10}$$

Let

$$\sigma_{b_i}^{-2} = \frac{m}{\sigma^2}+\frac{1}{\sigma_b^2}, \quad \mu_b = \frac{\sigma_b^2}{\sigma^2}\sum_{j=1}^{m}(y_{ij}-\beta-b_i-s_{ij}). \tag{A.11}$$

The posterior distribution of b_i is given by

$$b_i \sim N(\mu_b, \sigma_b^2). \tag{A12}$$

/The conditional distribution of $\boldsymbol{S}_i = (s_{i1}, \ldots, s_{im})'$ is obtained by

$$p(\boldsymbol{S}_i|*) \propto p(\boldsymbol{y}_i|\beta, b_i, \boldsymbol{S}_i, \sigma^2)p(\boldsymbol{S}_i|\sigma_s^2, \rho_i). \tag{A13}$$

It is necessary to compute the conditional distribution of each component of $\boldsymbol{S}_i$ separately:

$$\begin{aligned} & p(s_{i1}|\boldsymbol{y}, \beta, \boldsymbol{b}, \sigma^2, \sigma_s^2, \boldsymbol{\rho}) \\ = \; & \frac{p(\boldsymbol{y}|\beta, \boldsymbol{b}, s_{i1}, \sigma^2)p(\boldsymbol{b})p(s_{i1}|\sigma_s^2, \boldsymbol{\rho})p(\beta, \sigma^2, \sigma_s^2, \boldsymbol{\rho})}{p(\boldsymbol{y}|\beta, \boldsymbol{b}, \sigma^2)p(\boldsymbol{b})p(\beta, \sigma^2, \sigma_s^2, \boldsymbol{\rho})} \\ = \; & \textstyle\prod_{i=1}^{n} f(y_{i1}|\beta, \boldsymbol{b}, s_{i1}, \sigma^2)p(s_{i1}|\sigma_s^2, \boldsymbol{\rho}) \\ = \; & (2\pi)^{-\frac{n+1}{2}}\sigma^{-n}\sigma_s^{-1}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{i1}-\beta-b_i-s_{i1})^2 - \frac{s_{i1}^2}{2\sigma_s^2}\right\} \\ \propto \; & \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\frac{(y_{i1}-\beta-b_i-s_{i1})^2}{\sigma^2} + \frac{s_{i1}^2}{2\sigma_s^2}\right]\right\} \\ \propto \; & \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_s^2}\right)s_{i1}^2 - \frac{2}{\sigma^2}\sum_{i=1}^{n}(y_{i1}-\beta-b_i)s_{i1}\right]\right\}. \end{aligned} \tag{A14}$$

Let

$$\sigma_{s_{i1}}^{-2} = \frac{n}{\sigma^2}+\frac{1}{\sigma_s^2}, \quad \mu_{s_{i1}} = \frac{\sigma_{s_{i1}}^2}{\sigma^2}\sum_{i=1}^{n}(y_{i1}-\beta-b_i). \tag{A15}$$

We obtain the posteriori distribution for s_{i1}:

$$s_{i1} \sim N(\mu_{s_{i1}}, \sigma_{s_{i1}}^2). \tag{A16}$$

Similarly, the posterior distribution of s_{i2} is obtained by

$$\begin{aligned} & p(s_{i2}|*) \propto \prod_{i=1}^{n} p\left(y_{i2}|\beta, \boldsymbol{b}, s_{i2}, \sigma^2\right) p\left(s_{i2}|\hat{s}_{i1}, \sigma_s^2, \boldsymbol{\rho}\right) \\ = & \exp\left\{-\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{i2}-\beta-b_i-s_{i2})^2\right\} \exp\left\{-\frac{\left(s_{i2}-\hat{s}_{i1}\rho_i^{|t_{i2}-t_{i1}|}\right)^2}{2\sigma_s^2\left(1-\rho_i^{2|t_{i2}-t_{i1}|}\right)}\right\}\cdot \\ & (2\pi)^{-\frac{n+1}{2}}\sigma^{-n}\sigma_s^{-1}\left(1-\rho_i^{|t_{i2}-t_{i1}|}\right)^{-1} \\ \propto & \exp\left\{-\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{i2}-\beta-b_i-s_{i2})^2\right\} \exp\left\{-\frac{\left(s_{i2}-\hat{s}_{i1}\rho_i^{|t_{i2}-t_{i1}|}\right)^2}{2\sigma_s^2\left(1-\rho_i^{2|t_{i2}-t_{i1}|}\right)}\right\} \\ = & \exp\left\{-\tfrac{1}{2}\left[\tfrac{1}{\sigma^2}\sum_{i=1}^{n}(y_{i2}-\beta-b_i)^2+\tfrac{ns_{i2}^2}{\sigma^2}-\tfrac{2s_{i2}}{\sigma^2}\sum_{i=1}^{n}(y_{i2}-\beta-b_i)\right]\right\}\cdot \\ & \exp\left\{-\tfrac{1}{2}\left[\frac{s_{i2}^2+\left(\hat{s}_{i1}\rho_i^{|t_{i2}-t_{i1}|}\right)^2-2s_{i2}\hat{s}_{i1}\rho_i^{|t_{i2}-t_{i1}|}}{\sigma_s^2\left(1-\rho_i^{2|t_{i2}-t_{i1}|}\right)}\right]\right\}. \end{aligned} \tag{A17}$$

After removing the constant term, the posterior distribution of s_{i2} is proportional to

$$\exp\left\{-\tfrac{1}{2}\left[\left(\tfrac{n}{\sigma^2}+\frac{1}{\sigma_s^2\left(1-\rho_i^{2|t_{i2}-t_{i1}|}\right)}\right)s_{i2}^2-2\left[\tfrac{1}{\sigma^2}\sum_{i=1}^{n}(y_{i2}-\beta-b_i)+\frac{\hat{s}_{i1}\rho_i^{|t_{i2}-t_{i1}|}}{\sigma_s^2\left(1-\rho_i^{2|t_{i2}-t_{i1}|}\right)}\right]s_{i2}\right]\right\}.$$

Let

$$\begin{aligned} \sigma_{s_{i2}}^{-2} &= \frac{n}{\sigma^2}+\frac{1}{\sigma_s^2\left(1-\rho_i^{2|t_{i2}-t_{i1}|}\right)} \\ \mu_{s_{i2}} &= \sigma_{s_{i2}}^2\left[\frac{\sigma_s^2}{\sigma^2}\sum_{i=1}^{n}(y_{i2}-\beta-b_i)+\frac{\hat{s}_{i1}\rho_i^{|t_{i2}-t_{i1}|}}{\sigma_s^2\left(1-\rho_i^{2|t_{i2}-t_{i1}|}\right)}\right]. \end{aligned} \tag{A18}$$

We obtain the posteriori distribution for s_{i2}:

$$s_{i2} \sim N(\mu_{s_{i2}}, \sigma_{s_{i2}}^2). \tag{A19}$$

Similarly, the posteriori distribution of $s_{ij}(j=1,\ldots,m)$ can be obtained:

$$\begin{aligned} \mu_{s_{ij}} &= \sigma_{s_{ij}}^2\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_{ij}-\beta-b_i)+\frac{\hat{s}_{i,j-1}\rho_i^{|t_{ij}-t_{i,j-1}|}}{\sigma_s^2\left(1-\rho_i^{2|t_{ij}-t_{i,j-1}|}\right)}\right] \\ \sigma_{s_{ij}}^{-2} &= \frac{n}{\sigma^2}+\frac{1}{\sigma_s^2\left(1-\rho_i^{2|t_{ij}-t_{i,j-1}|}\right)} \\ s_{ij} &\sim N(\mu_{s_{ij}}, \sigma_{s_{ij}}^2). \end{aligned} \tag{A20}$$

Combining the above derivations, we can obtain the conditional distribution of σ_s^2 as follows:

$$\begin{aligned}
& p(\sigma_s^2|*) \propto f(\boldsymbol{S}|\sigma_s^2,\boldsymbol{\rho})p(\sigma_s^2) \\
= & \prod_{i=1}^{n} f(s_{i1})f(s_{i2}|s_{i1})\cdots f(s_{im}|s_{i1},s_{i2},\ldots,s_{i,m-1}) \\
= & (2\pi)^{-\frac{m}{2}}\sigma_s^{-m}\prod_{j=2}^{m}\sqrt{1-\rho_i^{2|t_{ij}-t_{i,j-1}|}}\,\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}(\sigma_s^2)^{-\alpha_1-1}\exp\left\{-\frac{\beta_1}{\sigma_s^2}\right\}\cdot \\
& \exp\left\{-\frac{1}{2\sigma_s^2}\left[s_{i1}^2+\sum_{j=2}^{m}\frac{\left(s_{ij}-\hat{s}_{i,j-1}\rho_i^{|t_{ij}-t_{i,j-1}|}\right)^2}{1-\rho_i^{2|t_{ij}-t_{i,j-1}|}}\right]\right\} \\
= & \exp\left\{-\frac{1}{\sigma_s^2}\left[\frac{1}{2}\left(s_{i1}^2+\sum_{j=2}^{m}\frac{\left(s_{ij}-\hat{s}_{i,j-1}\rho_i^{|t_{ij}-t_{i,j-1}|}\right)^2}{1-\rho_i^{2|t_{ij}-t_{i,j-1}|}}\right)+\beta_1\right]\right\}\cdot \\
& (2\pi)^{-\frac{m}{2}}\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\prod_{j=2}^{m}\sqrt{1-\rho_i^{2|t_{ij}-t_{i,j-1}|}}\,(\sigma_s^2)^{-\frac{m}{2}-\alpha_1-1} \\
\propto & (\sigma_s^2)^{-\frac{m}{2}-\alpha_1-1}\exp\left\{-\frac{1}{\sigma_s^2}\left[\frac{1}{2}\left(s_{i1}^2+\sum_{j=2}^{m}\frac{\left(s_{ij}-\hat{s}_{i,j-1}\rho_i^{|t_{ij}-t_{i,j-1}|}\right)^2}{1-\rho_i^{2|t_{ij}-t_{i,j-1}|}}\right)+\beta_1\right]\right\}.
\end{aligned} \tag{A21}$$

The posterior distribution of σ_s^2 is actually an inverse gamma distribution:

$$\sigma_s^2 \sim IG\left(\frac{m}{2}+\alpha_1,\frac{1}{2}\left(s_{i1}^2+\sum_{j=2}^{m}\frac{\left(s_{ij}-\hat{s}_{i,j-1}\rho_i^{|t_{ij}-t_{i,j-1}|}\right)^2}{1-\rho_i^{2|t_{ij}-t_{i,j-1}|}}\right)+\beta_1\right). \tag{A22}$$

The posterior distribution of ρ_i is obtained as follows:

$$\begin{aligned}
& p(\rho_i|*) = p(\rho_i|\boldsymbol{S},\sigma_s^2,\rho_{-i}) = p(\rho_i|\boldsymbol{S}_i,\sigma_s^2,\rho_{-i}) \\
= & \frac{p(\boldsymbol{S}_i|\sigma_s^2,\rho_i)p(\rho_i|\rho_{-i})p(\sigma_s^2,\rho_{-i})}{p(\boldsymbol{S}_i|\sigma_s^2)p(\sigma_s^2,\rho_{-i})} \propto f(\boldsymbol{S}_i|\sigma_s^2,\rho_i)p(\rho_i|\rho_{-i}),
\end{aligned} \tag{A23}$$

which can be expressed as follows:

$$\begin{aligned}
& p(\rho_i|\boldsymbol{S}_i,\sigma_s^2,\rho_{-i}) \sim \sum_{j\neq i} q_{ij}\delta_{\rho_j} + r_i H_i, \quad q_{ij} = bf(\boldsymbol{S}_i|\sigma_s^2,\rho_j) \\
& r_i = b\alpha\int f(\boldsymbol{S}_i|\sigma_s^2,\boldsymbol{\rho})g_0(\boldsymbol{\rho})\mathrm{d}\boldsymbol{\rho}, \quad H_i \sim \frac{f(\boldsymbol{S}_i|\sigma_s^2,\boldsymbol{\rho})g_0(\boldsymbol{\rho})}{\int f(\boldsymbol{S}_i|\sigma_s^2,\boldsymbol{\rho})g_0(\boldsymbol{\rho})\mathrm{d}\boldsymbol{\rho}},
\end{aligned} \tag{A24}$$

where g_0 is the probability density function of the base distribution G_0, b is the constant satisfying the equation $\sum_{j\neq i} q_{ij} + r_i = 1$, and H_i is the marginal distribution of the parameter $\boldsymbol{\rho}$ based on the prior G_0 and the variable $\boldsymbol{S}_i$.

References

1. Pullenayegum, E.M.; Lim, L.S. Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *Statist. Methods Med. Res.* **2016**, *25*, 2992–3014. [CrossRef] [PubMed]
2. Chakraborty, R.; Banerjee, M.; Vemuri , B.C. Statistics on the space of trajectories for longitudinal data analysis. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia, 18–21 April 2017; pp. 999–1002.

3. Xiang, D.; Qiu, P.; Pu, X. Nonparametric regession analysis of multivariate longitudinal data. *Stat. Sin.* **2013**, *23*, 769–789.
4. Quintana, F.A.; Johnson, W.O.; Waetjen, L.E.; BGold, E. Bayesian nonparametric longitudinal data analysis. *J. Am. Statist. Assoc.* **2016**, *111*, 1168–1181. [CrossRef]
5. Cheng, L.; Ramchandran, S.; Vatanen, T.; Lietzén, N.; Lahesmaa, R.; Vehtari, A.; Lähdesmäki, H. An additive Gaussian process regression model for interpretable non-parametric analysis of longitudinal data. *Nat. Commun.* **2019**, *10*, 1798. [CrossRef]
6. Kleinman, K.P.; Ibrahim, J.G. A semiparametric Bayesian approach to the random effects model. *Biometrics* **1998**, *54*, 921–938. [CrossRef] [PubMed]
7. Gao, F.; Zeng, D.; Couper, D.; Lin, D.Y. Semiparametric regression analysis of multiple right- and interval-censored events. *J. Am. Statist. Assoc.* **2019**, *114*, 1232–1240. [CrossRef] [PubMed]
8. Lee, Y. Semiparametric regression. *J. Am. Statist. Assoc.* **2006**, *101*, 1722–1723. [CrossRef]
9. Sun, Y.; Sun, L.; Zhou, J. Profile local linear estimation of generalized semiparametric regression model for longitudinal data. *Lifetime Data Anal.* **2013**, *19*, 317–349. [CrossRef] [PubMed]
10. Zeger, S.L.; Diggle, P.J. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **1994**, *50*, 689–699. [CrossRef] [PubMed]
11. Li, Y.; Lin, X.; Müller, P. Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics* **2010**, *66*, 70–78. [CrossRef] [PubMed]
12. Hunag, Y.X. Quantile regression-based Bayesian semiparametric mixed-effects models for longitudinal data with non-normal, missing and mismeasured covariate. *J. Statist. Comput. Simul.* **2016**, *86*, 1183–1202. [CrossRef]
13. Li, J.; Zhou, J.; Zhang, B.; Li, X.R. Estimation of high dimensional covariance matrices by shrinkage algorithms. In Proceedings of the 2017 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017; pp. 955–962.
14. Doss, H.; Park, Y. An MCMC approach to empirical Bayes inference and Bayesian sensitivity analysis via empirical processes. *Ann. Statist.* **2018**, *46*, 1630–1663. [CrossRef]
15. Neal, R.M. Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.* **2000**, *9*, 249–265.
16. Kingma, D.P.; Welling, M. Auto-encoding variational Bayes. In Proceedings of the International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
17. Csiszar, I. I-Divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **1975**, *3*, 146–158. [CrossRef]
18. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; pp. 63–71.
19. Barndorff-Nielsen, O.E.; Shephard, N. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. Roy. Statist. Soc. (Ser. B)* **2001**, *63*, 167–241. [CrossRef]
20. Griffin, J.E.; Steel, M.F.J. Order-based dependent Dirichlet processes. *J. Am. Statist. Assoc.* **2006**, *101*, 179–194. [CrossRef]
21. Mukhopadhyay, S.; Gelfand, A.E. Dirichlet process mixed generalized linear models. *J. Am. Statist. Assoc.* **1997**, *92*, 633–639. [CrossRef]
22. Zimmerman, D.L.; Núñez-Antón, V.A. *Antedependence Models for Longitudinal Data*; CRC Press: Boca Raton, FL, USA, 2009.
23. Hsu, C.W.; Sinay, M.S.; Hsu, J.S. Bayesian estimation of a covariance matrix with flexible prior specification. *Ann. Inst. Statist. Math.* **2012**, *64*, 319–342. [CrossRef]
24. Chen, X. Longitudinal Data Analysis Based on Bayesian Semiparametric Method. Master's Thesis, Lanzhou University, Lanzhou, China, 2019. (In Chinese)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.