Bayesian Analysis of First-Order Markov Models for Autocorrelated Binary Responses

Dasom Lee1 · Sujit Ghosh1

Accepted: 4 November 2022 / Published online: 23 November 2022
© Grace Scientific Publishing 2022

Abstract
In many clinical trials, patient outcomes are often binary-valued which are measured asynchronously over time across various dose levels. To account for autocorrelation among such longitudinally observed outcomes, a first-order Markov model for binary data is developed. Moreover, to account for the asynchronously observed time points, nonhomogeneous models for the transition probabilities are proposed. The transition probabilities are modeled using B-spline basis functions after suitable transformations. Additionally, if the underlying dose-response curve is assumed to be non-decreasing, our model allows for the estimation of any underlying non-decreasing curve based on suitably constructed prior distributions. We also extended our model to the mixed effect model to incorporate individual-specific random effects. Numerical comparisons with traditional models are provided based on simulated data sets, and also practical applications are illustrated using real data sets.

Keywords Bayesian inference · Binary longitudinal data · B-spline basis · Nonhomogeneous Markov model

1 Introduction
In many clinical trials, patient outcomes are often binary-valued that are measured asynchronously over time across various dose levels. For example, for patients with toenail infection, a popular measure of clinical efficacy of treatment for the nail is a 4-point scale (cleared, minimal symptoms, slightly improved, failure). A significant improvement is defined as cleared or minimal clinical symptoms at the end of the study. Thus, the efficacy outcome is a binary random variable, with 1 indicating sig-
significant improvement and 0 otherwise. As these binary-valued responses measured over time for each patient are likely to be different for different patients and expected to be correlated over time, a first-order Markov model that allows capturing such autocorrelations is often used to analyze the data sets.

In most of the common modeling approaches, the Markov chain is often considered to be time-homogeneous. That is, the transition probability from one state to another (e.g., a value of 1 to 0 in the case of binary data) depends only on the difference between state values and is independent of the time index. For example, Shoko and Chikobvu [1] used the time-homogeneous Markov process to model HIV/AIDS progression using a cohort study. Similarly, Ma et al. [2] developed a Bayesian approach using the log transformation of $S \times S$ transition probability matrix, where $S$ is the total number of observed outcomes of all subjects, assuming the time-homogeneous Markov process.

However, it is more natural in general to assume that the probability of state-to-state jumps may change as time goes on. For example, the probability of showing improvement would depend on how long the patient has been on treatment. However, modeling a nonhomogeneous Markov chain is relatively difficult as the common stationary property does not hold and the dimension of parameters to estimate is large. Previously, Jackson et al. [3] reviewed various Markov models and their extension to fit a continuous-time panel data, but this approach for the time nonhomogeneous model is restricted to piecewise-constant transition probabilities. Hubbard et al. [4] proposed the time nonhomogeneous models by transforming the time scale of the nonhomogeneous Markov process to the time scale of the homogeneous process. However, these models relied on the assumption that the ratio of transition probabilities stays constant over time. Jones et al. [5] modeled the nonhomogeneous transition probabilities of longitudinal binary data as exponential functions. Zeng and Cook [6] proposed an estimation and inference method for transition probability using Generalized Estimating Equation approaches and logistic regression. However, these models may be inconsistent if the true transition probability functions are different from their assumed models. B-spline basis functions can be used to model smooth curves. For example, Andersen et al. [7] modeled occupancy in an office environment using the nonhomogeneous Markov model where the transition probabilities are estimated using B-splines and exponential smoothing. Similarly, Iversen et al. [8] proposed a non-homogeneous Markov model using logit transformation and the B-spline basis to describe driving patterns of electric vehicles. However, these models lack solving the missingness, which is often the case in clinical outcome data.

The proposed model provides a Bayesian approach for the flexible non-homogeneous Markov model for autocorrelated binary responses. One major disadvantage of the common nonhomogeneous Markov model is that there are a large number of parameters to be estimated. In our model, B-spline basis functions are used to reduce the number and to estimate a smooth transition probability curve without any assumptions about the shape of the curve. Our proposed model, using a Bayesian framework, allows dealing with non-ignorable missingness and unbalance problems with ease based on suitably constructed prior distributions.

This paper is organized as follows: Sect. 2 introduces the proposed non-homogeneous Markov model for asynchronous data. In Sect. 3, numerical results are provided. Sec-
tion 4 illustrates practical applications using toenail clinical trial data and Dow Jones stock market data. Section 5 concludes the paper with some discussion.

2 Non-homogeneous Markov Model for Asynchronous Data

2.1 Constructing Markov Properties

Suppose there are \( i = 1, \ldots, n \) subjects and \( Y^{\text{obs}}_{i0} = Y^{\text{obs}}_{i}(t_{i0}) \) is the initial outcome of a subject \( i \) at initial time \( t_0 \). Also, suppose that each subject \( i \) has been measured at \( J_i \) time points afterward. Let \( t_{ij} \) be the \( j \)th time of subject \( i \) and \( Y^{\text{obs}}_{ij} = Y^{\text{obs}}_{i}(t_{ij}) \) be a binary outcome of the subject \( i \) at time \( t_{ij} \), where \( t_{i0} < t_{i1} < \cdots < t_{iJ_i} \). That is, for subject \( i = 1, \ldots, n \) with \( j = 0, 1, \ldots, J_i \) observed time points,

\[
Y^{\text{obs}}_{ij} = \begin{cases} 
1 & \text{if the subject } i \text{ had an event at time } t_{ij} \\
0 & \text{otherwise}
\end{cases}
\]

Let \( T_A = \sum_{i=1}^{n} J_i \) be the total number of observed time points of \( n \) subjects after initial time \( t_0 \), and let’s assume that among \( T_A \) time points, there are \( T \) distinct time points where \( \max_i(J_i) \leq T \leq T_A \). Also, let \( t = (t_0, t_1, \ldots, t_T) \) be the vector of all distinct ordered time points of length \( T + 1 \). In order to define the Markov property, we reconstruct the binary response outcome based on the all distinct time \( t \). Specifically, for subjects \( i \), the new outcome \( Y_{ik} \), \( k \in t \) is defined as \( Y_{ik} = Y^{\text{obs}}_{ij} \) if \( k = j \) for some \( j = 1, \ldots, J_i \), and missing if the subject \( i \) does not have an outcome at time \( k \).

Figure 1 illustrates the construction of Markov properties on asynchronous data. Suppose that the first patient was observed four times and the corresponding outcome is \( y_1 = (y^{\text{obs}}_{11}, y^{\text{obs}}_{12}, y^{\text{obs}}_{13}, y^{\text{obs}}_{14})^T \). Similarly, the second patient has three outcomes and the third patient has 4 outcomes. The outcomes under the same vertical dashed line imply those were obtained at the same time. Since there are six distinct time points, we reconstruct the new response variable \( Y \) as,

\[
Y = \begin{pmatrix}
Y_1 \\
Y_2 \\
Y_3
\end{pmatrix} = \begin{pmatrix}
y^{\text{obs}}_{11} & y^{\text{obs}}_{12} & y^{\text{obs}}_{13} & y^{\text{obs}}_{14} \\
y^{\text{obs}}_{21} & y^{\text{obs}}_{22} & y^{\text{obs}}_{23} & y^{\text{obs}}_{24} \\
y^{\text{obs}}_{31} & y^{\text{obs}}_{32} & y^{\text{obs}}_{33} & y^{\text{obs}}_{34}
\end{pmatrix} = \begin{pmatrix}
y_{11} & y_{12} & y_{13} & y_{14} & y_{16} \\
y_{21} & y_{22} & y_{25} & y_{26} \\
y_{31} & y_{33} & y_{35} & y_{36}
\end{pmatrix}
\]

When time points are different for every patient but close, we can reconstruct the outcome vector by applying a common practice in clinical trial studies that specifies an appropriate range of time, or visit window, to allow feasible scheduling of visit. For example, a week 4 visit can be defined as 4 weeks post study enrollment \( \pm 1 \) week, and similarly for the week 8 and week 12 visits. In this way, such time points that are different but close can be grouped into a few pre-specified visits. In general, a similar time windowing strategy is applicable based on the research interest and subject matter knowledge.

For the reconstructed the outcome vector \( Y_i = \{ Y_{i0}, Y_{i1}, \ldots, Y_{iT} \}^T \), we assume that \( Y_{i0}, Y_{i1}, \ldots, Y_{iT} \) has Markov properties. By the definition of the first-order Markov
chain that the event at the current time only depends on the event at an immediate past time, two transition probability functions are defined as,

\[
p(Y_{ij} = 1 \mid Y_{i,j-1} = 1) = \theta_i(t_j)
p(Y_{ij} = 1 \mid Y_{i,j-1} = 0) = \theta_i(t_j),
\]

for \( j = 1, \ldots, T \). The initial probability is \( \theta_{i,00} = p(Y_{i0} = 1) \). In general, the transition probabilities may depend on individual \( i \), but for simplicity, we begin with the setting that \( \theta_{il}(t_j) = \theta_l(t_j) \) for \( l = 0, 1 \) and \( j = 1, \ldots, T \). By assuming that each individual is independent of each other, the corresponding generic model for a subject can be expressed by the following hierarchical model,

\[
Y_0 \sim \text{Ber}(\theta_{00})
Y_1 \mid Y_0 \sim \text{Ber}(\theta_1(t_1)Y_0 + \theta_0(t_1)(1 - Y_0))
Y_2 \mid Y_1 \sim \text{Ber}(\theta_1(t_2)Y_1 + \theta_0(t_2)(1 - Y_1))
\]

\[
\vdots
Y_T \mid Y_{T-1} \sim \text{Ber}(\theta_1(t_T)Y_{T-1} + \theta_0(t_T)(1 - Y_{T-1})).
\]

In some applications such as dose-response model, it is natural to assume that the true transition probability function \( \theta_l(t) \), \( l = 0, 1 \) satisfies the non-decreasing condition: \( \theta_l(t_0) \leq \theta_l(t_1) \leq \cdots \) for \( t_0 \leq t_1 \leq t_2 \leq \cdots \). Note that this condition is not necessary for our model development and can be relaxed or modified depending on the background information that we may have from the data.

### 2.2 Modeling Transition Probabilities with B-splines

As the observed data over time can be asynchronous, we would like to estimate the transition probabilities as smooth functions for any time \( t \in [t_0, t_T] \). It is well known (see [9]) that if the underlying true transition probability \( \theta_l(t) \) is smooth, then it can be uniformly approximated by a sequence of B-splines by choosing the appropriately
large number of knots. In practice, the selection of the knots and number of knots are based on observed time points and the sample sizes making it data adaptive.

Let \( \kappa \) be the vector of non-decreasing knots that \( \kappa_0 = \min(t) \leq \kappa_1 \leq \ldots \leq \kappa_m = \max(t) \). The \( k \)-th B-spline basis functions of degree \( q \) with \( m \) internal knots at time \( t \), written as \( B_{k,q}(t) \), are recursive functions that are constructed as

\[
B_{k,0}(t) = \begin{cases} 
1, & t \in [\kappa_k, \kappa_{k+1}) \\
0, & \text{otherwise}
\end{cases}
\]

\[
B_{k,q}(t) = \frac{t - \kappa_k}{\kappa_{k+q} - \kappa_k} B_{k,q-1}(t) + \frac{\kappa_{k+q+1} - t}{\kappa_{k+q+1} - \kappa_{k+1}} B_{k+1,q-1}(t), \quad \text{when} \ q \geq 1,
\]

where \( k = 0, \ldots, M = m + q \) and \( \kappa = (\kappa_0, \ldots, \kappa_M) \) is a vector of knots such that \( \kappa_0 \leq \kappa_1 \leq \ldots \leq \kappa_M \). The first knot \( \kappa_0 \) and the last knot \( \kappa_M \) are boundary knots, and the first B-spline term \( B_{0,q} \) is often regarded as the intercept. Using B-spline approximation, the transition probability function \( \theta_l(t) \) can be estimated as a linear combination of a B-spline basis given by

\[
\theta_l(t) = \sum_{k=0}^{M} \beta_{lk} B_{k,q}(t) = B_M(t)^T \beta_l, \ t \in [\kappa_0, \kappa_m],
\]

where \( B_M(t) = [B_{0,q}(t), \ldots, B_{M,q}(t)]^T \) and \( \beta_l = [\beta_{l0}, \ldots, \beta_{lM}]^T \).

For a fixed number of internal knots \( m \) and a spline degree \( q \), B-spline basis functions form a partition of unity, i.e.,

\[
\sum_{k=0}^{M} B_{k,q}(t) = 1, \text{ for all } t, q.
\]

Therefore, if the coefficients \( \beta_{lk} \)s are restricted to be in the interval \([0, 1]\) then \( \theta_l(t) \in [0, 1] \) for any \( m \geq 1 \) and \( q \geq 1 \). Typically, cubic spline with \( q = 3 \) is used for the most practical applications.

Using the partition of unity property, the proposed model with the B-spline approximation is as follows,

\[
Y_{i0} \sim \text{Ber}(\theta_{00}) \text{ where } \theta_{00} \sim \text{Beta}(1, 1)
\]

\[
Y_{ij} \mid Y_{i,j-1} \sim \text{Ber} \left( \theta_1(t_j)Y_{i,j-1} + \theta_0(t_j)(1 - Y_{i,j-1}) \right), \ j = 1, \ldots, T
\]

\[
\theta_l(t_j) = \sum_{k=1}^{M} \beta_{lk} B_{k,q}(t_j), \ l = 0, 1
\]

\[
\beta_{lk} \sim \text{Beta}(\mu_k \tau, (1 - \mu_k) \tau), \ \forall l, k
\]

\[
\mu_k \sim \text{Beta}(1, 1)
\]

\[
\tau \sim \text{Unif}(0.1, 10). \quad (1)
\]
The coefficients $\beta_{lk}$ were given a Beta prior to ensure that the transition probability function $\theta_l(t) \in [0, 1]$ for $l = 0, 1$ and for all $t$. The hyper-parameters $\mu_k, k = 1, \ldots, M$ represent the mean of the prior distribution of $\beta_{lk}$, and $\tau$ controls the variance of the distribution.

Alternatively, the logit transformation for $\theta_l$ can be used considering $\beta_{lk}$ as logistic regression coefficients. The primary advantage of using the logit transformation is that we no longer need any restriction on the coefficients $\beta_{lk}$. This can be modeled as

\begin{align*}
  Y_{i0} & \sim \text{Ber}(\theta_{00}) \text{ where } \theta_{00} \sim \text{Beta}(1, 1) \\
  Y_{it} | Y_{i,j-1} & \sim \text{Ber}(\theta_1(t_j)Y_{i,j-1} + \theta_0(t_j)(1 - Y_{i,j-1})), \quad j = 1, \ldots, T \\
  \theta_l(t_j) & = \left[1 + \exp\left\{-\sum_{k=1}^{M} \beta_{lk} B_{k,q}(t_j)\right\}\right]^{-1}, \quad l = 0, 1 \\
  \beta_{lk} & \sim \text{ind } N(\mu_k, \tau) \quad \forall k, l \\
  \mu_k & \sim \text{iid } N(0, 10^3) \\
  \tau & \sim \text{IG}(2, 1).
\end{align*}

Both models in (1) and (2) allow flexible estimation of the transition probability functions $\theta_l(t)$ for $l = 0, 1$ for autocorrelated and asynchronous binary outcomes. In addition, one can consider probit link model by representing the transition probability functions as

\begin{align*}
  \theta_l(t) & = \Phi \left\{ \sum_{k=1}^{M} \beta_{lk} B_{k,q}(t) \right\}, \quad l = 0, 1,
\end{align*}

where $\Phi(\cdot)$ is a Gaussian CDF function. However, as we are using a nonparametric model, the results will not be sensitive with the choice of the link function because the function space spanned by either the use of logit or probit or any other CDF will be the same.
2.3 Non-decreasing Constraints

In many cases, the true transition probability curves are assumed to be non-decreasing functions. For example, the probability of transitioning from a severe symptom to a mild symptom would likely to be non-decreasing with respect to the time a patient was on a treatment. In such cases, our model allows for the estimation of the underlying non-decreasing curve. Meyer [10] proved that the non-decreasing constraint on $\beta_{lk}$ imposes the non-decreasing shape constraints on $\theta_l(t)$.

The linear link model (1) can easily be extended to incorporate non-decreasing constraints using the reparametrization technique following Shin and Ghosh [11] and Cutis and Ghosh [12]. Even though their method was applied to Bernstein polynomials [13], a similar technique can be applied to the B-spline basis functions. We consider the difference of the two consecutive coefficients, $\gamma_{lk} = \beta_{lk} - \beta_{l,k-1}$ and the partial sum of B-spline functions, $F_{k,q}(t) = \sum_{k=1}^{M} B_{k,q}(t)$ for $t \in [t_{k-1}, t_k]$ and $M = m + q + 1$. Then, we can represent the transition probability function $\theta_l(t)$ with non-decreasing constraints in terms of $\gamma_{lk}$ and $F_{k,q}(t)$ as follows,

$$\theta_l(t) = B_M(t)^T \beta_l = F_M(t)^T \gamma_l \text{ for } l = 0, 1$$

s.t. $\begin{pmatrix} 1^T \\ -1^T \\ M \end{pmatrix} \gamma_l \geq \begin{pmatrix} 0^T \\ -1 \end{pmatrix}$, (3)

where $F_M(t) = \{F_{0,q}(t), ..., F_{M,q}(t)\}$ and $\gamma_l = \{\gamma_{l0}, ..., \gamma_{lM}\}^T$. The constraints in (3) imply $\gamma_{lk} \geq 0$ and $\sum_{k=1}^{M} \gamma_{lk} = \beta_{l,M} \leq 1$ for all $l = 0, 1$ and $k = 1, ..., M$, resulting in the non-decreasing $\beta_{lk}$. The full Bayesian model with this non-decreasing restriction is

$$Y_{i1} \sim \text{Ber}(\theta_{00}) \text{ where } \theta_{00} \sim \text{Beta}(1, 1)$$

$$Y_{ij} | Y_{i,j-1} \sim \text{Ber} \left( \theta_l(t_j) Y_{i,j-1} + \theta_0(t_j) (1 - Y_{i,j-1}) \right), \quad j = 1, ..., T$$

$$\theta_l(t_j) = \sum_{k=1}^{M} \gamma_{lk} F_{k,q}(t_j), \quad l = 0, 1.$$

(4)

For the prior distribution of $\gamma_{lk}$, following Shin and Ghosh [11], we consider the following reparametrization,

$$\gamma_{l1} = \delta_{l1} \text{ and } \gamma_{lk} = \delta_{lk} \prod_{j=1}^{k-1} (1 - \delta_{lj}), \quad k = 2, ..., m,$$

which is similar to the stick-breaking representation. Using this reparametrization, $\gamma_{lk}$s satisfy constraint in (3) as long as $\delta_{lk} \in [0, 1]$. Therefore, for $\delta_{lk}$, Bernoulli-Beta mixture prior were given that

$$\delta_{lk} = \begin{cases} 0 \quad \text{with probability } 1 - \psi_{lk} \\ \psi_{lk} \quad \text{with probability } \psi_{lk} \end{cases},$$

where $\psi_{lk}$ is a parameter for the Bernoulli distribution.
with

$$\psi_{lk} \sim \text{Beta}(0.5, 0.5) \text{ and } \nu_{lk} \sim \text{Beta}(1, 1).$$

This representation allows estimating ‘flat’ transition probability curves as $$\delta_{lk} = 0$$ implies $$\beta_{lk} = \beta_{l,k-1}$$ for $$t \in [t_{k-1}, t_k]$$. Note that even though the primary focus of this paper is on the non-decreasing constraints, the proposed model would also work for non-increasing constraints with minor modifications by simply redefining the $$\gamma_{lk} = \beta_{l,k-1} - \beta_{lk}$$.

### 2.4 Penalized B-Splines

For B-spline basis models, choosing the appropriate number of knots and selecting their locations are not trivial problems. Eilers and Marx [14] proposed penalized B-splines, also known as P-splines, which is a combination of B-spline basis functions and penalization to optimize the number and location of knots. Lang and Brezger [15] extended the P-splines in a Bayesian framework.

Currie and Durban [16] proposed a mixed model approach to estimate the P-spline coefficients, and Kauermann et al. [17] proved the asymptotic properties of P-splines when the number of bases increases with the sample size. Large sample consistencies of the proposed spline-based estimators of the transition probabilities can be established by applying some of the above mentioned methods under both frequentist and Bayesian framework, which we omit in this work and rather our focus is in the finite sample performances of the proposed methodologies using extensive simulation studies that we present later.

Following [16], under a frequentist framework with $$E(\theta_l) = g(B\beta_l)$$, a general P-spline approach is to find $$\beta_l$$ such that

$$\min \left( \theta_l - g(B\beta_l) \right)^T \left( \theta_l - g(B\beta_l) \right) + \lambda_l \beta_l^T D_v^T D_v \beta_l$$

where $$g(x)$$ is a link function and $$D_v \in \mathbb{R}^{(M-v) \times M}$$ denotes the $$v$$th order difference matrix. For example, the first-order and the second-order difference matrix are constructed as

$$D_1 = \begin{pmatrix}
-1 & 1 & 0 & \cdots & 0 & 0 \\
0 & -1 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -1 & 1
\end{pmatrix}_{(M-1) \times M}, \quad D_2 = \begin{pmatrix}
1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1
\end{pmatrix}_{(M-2) \times M}.$$

In our context of Bayesian Markov models, we extend the logit link model specified in (2) to incorporate P-splines as mixed models, similar to [16]. We represent $$\beta_l = Gb_l + Zu_l$$, where $$Z = D_v^T (D_v^T)^{-1} \in \mathbb{R}^{M \times (M-v)}$$ and $$u_l \sim N(0, \sigma^2_{u_l} I)$$. A common choice of $$G$$ is $$G = \{1, g, ..., g^{v-1}\}$$ where $$g = \{1, ..., M\}$$. Since it is natural to assume that $$b_l$$ are independent, we give a prior distribution $$b_l \sim N(0, \sigma^2_b)$$ with large
The smoothing parameter $\lambda_l$ is $\lambda_l = 1/\sigma^2_{u,l}$, and we assign a conjugate prior $\sigma^2_{u,l} \sim IG(a_u, b_u)$. A small $\lambda_l$ indicates undersmoothing, and a large $\lambda_l$ indicates oversmoothing.

The proposed P-splines model with logit link is

$$Y_{i0} \sim \text{Ber}(\theta_{i0}) \text{ where } \theta_{i0} \sim \text{Beta}(1, 1)$$

$$Y_{ij} \mid Y_{i,j-1} \sim \text{Ber} \left( \theta_l(t_j) Y_{i,j-1} + \theta_0(t_j)(1 - Y_{i,j-1}) \right), \quad j = 1, ..., T$$

$$\theta_l(t_j) = \left[1 + \exp \left\{ - \sum_{k=1}^{M} \beta_{lk} B_{k,q}(t_j) \right\} \right]^{-1}, \quad l = 0, 1$$

$$\beta_l = \{ \beta_1, ..., \beta_M \} = Gb_l + Zu_l. \quad (5)$$

The linear link model specified in (1) can also be extended to the P-spline model. However, it is less straightforward than that of the logit link model as the coefficients $\beta_{lk}$ are restricted to be in the interval [0, 1] for (1).

Kauermann et al. [17] proved a general consistency of P-splines for the generalized linear model (GLM) that when the dimension of the spline basis increases as the order of $O(n^{1/(2q+3)})$ and the smoothing parameter $\sigma^{-2}_{u,l} = O(n^r)$, $r \leq 2/(2q + 3)$, under mild regularity conditions, $\hat{\theta}_l(t_j)$ are consistent estimators of $\theta_l(t_j)$, for $l = 0, 1, j = 1, ..., T$. Yoshida and Naito [18] discussed asymptotic properties of P-splines in generalized additive models; their approach could be used to establish consistency by utilizing the GLM framework of our proposed model.

### 2.5 Subject Specific B-Spline Model

In addition to the mean transition probability, one could be interested in estimating the transition probability of a set of specific subjects. Therefore, it is natural to extend the previously proposed models (1)–(5) with an additional hierarchy, assuming that each subject has his/her own transition probability functions which deviate from the mean transition probability. This relationship is analogous to that of population-mean models and subject-specific models from several longitudinal studies [e.g., 19, 20]. This extension allows us to estimate individual-specific transition probabilities as well as overall mean transition probabilities.

Here we only consider the extension of the linear link B-spline model in (1), but other models such as the logit link model or P-spline model can easily be extended. We assume that at time $t_0$, the initial probability of having an event for subject $i$ is $\theta_{i,00}$ such that

$$Y_{i0} \sim \text{Ber}(\theta_{i,00}) \text{ where } \theta_{i,00} \sim \text{Beta}(\mu_{00} \tau_0, (1 - \mu_{00}) \tau_0)$$

$$\mu_{00} \sim \text{Beta}(1, 1)$$

$$\tau_0 \sim \text{Unif}(0.1, 10), \quad (6)$$

where $\mu_{00}$ is the mean of $\theta_{i,00}$, which implies the overall initial probability of having an event across subjects.
For a subject $i$, let $\theta_{i,1}(t_j)$ denote the true transition probability from 1 to 1 and $\theta_{i,0}(t_j)$ denote the true transition probability from 0 to 1 for $j = 1, \ldots, T$. Then the subject-specific transition probability curve $\theta_{i,l}(t_j), l = 0, 1$ can be written as a linear combination of B-splines basis functions.

$$Y_{ij} \mid Y_{i,j-1} \sim \text{Ber} \left( \theta_{i,1}(t_j)Y_{i,j-1} + \theta_{i,0}(t_j)(1-Y_{i,j-1}) \right), \quad j = 1, \ldots, T$$

$$\theta_{i,l}(t_j) = \sum_{k=1}^{m_l+q+1} \beta_{i,lk} B_{k,q}(t_j), \quad l = 0, 1, i = 1, \ldots, n$$

$$\beta_{i,lk} \sim \text{Beta}(\mu_{i,k} \tau_i, (1 - \mu_{i,k}) \tau_i), \quad \forall i, l, k$$

$$\mu_{i,k} \sim \text{Beta}(a_k b_k, (1 - a_k) b_k)$$

$$\tau_i \sim \text{Unif}(0.1, 10)$$

$$a_k \sim \text{Beta}(1, 1)$$

$$b_k \sim \text{Unif}(0.1, 10).$$

(7)

Similar to (1), the hyper-parameter $\mu_{i,k}$s represent the mean of the subject-specific prior distributions of $\beta_{i,lk}$, and $\tau_i$s control the variance of the distributions. Note that $\mu_{i,k}$ deviates from the common $a_k$ and $b_k$, for all $i = 1, \ldots, n$ and $k = 1, \ldots, M$.

### 3 Simulation Study

In this section, we conduct multiple sets of simulation studies to evaluate the finite-sample performance of the proposed models. Section 3.1 compares the performance of the B-spline models and that with monotone increasing constraints when the true data were generated from the exponential decay, log-logistic and S/U-shaped functions. In Sect. 3.2, we present the performance of P-spline models with penalties of different orders.

#### 3.1 B-Spline Models

Monte Carlo simulations were conducted to assess the performance of the proposed nonparametric B-spline models and a B-spline model with monotone increasing constraints with two common parametric models, i.e., log-logistic model and exponential decay model. The log-logistic model, which is also known as the Fisk distribution in Economics, is widely used in survival analysis. The transition probability under the log-logistic model is

$$\theta_l(t) = (1 + \exp(\alpha_l - \beta_l \log(t)))^{-1},$$

where $\alpha_l \in \mathbb{R}$ and $\beta_l \geq 0, \forall l = 0, 1$. Another popular parametric model is the exponential decay model which is often used to model the decay of a quantity when it decreases at a rate that is proportional to the amount at time $t$, such as the chemical compound decay. The corresponding transition probability can be written in the form of
\[ \theta_l(t) = \alpha_l(1 - \exp(-\beta_l(t))), \]

where \( \alpha_l \in [0, 1] \) and \( \beta_l \geq 0, \forall l = 0, 1. \)

In this simulation, we consider the exponential decay model and log-logistic model as the true transition probability curves. In addition to these concave models, we also consider S-shaped and U-shaped functions to assess the flexibility of our proposed model. We choose the S/U-shaped curves because they resemble the Toenail data that we will analyze in Sect. 4.1. The three sets of underlying true transition probability curves are as follows,

- **Exponential decay:**

  \[
  \begin{align*}
  \theta_0(t) &= 0.65(1 - \exp(-0.1t)) + 0.1 \\
  \theta_1(t) &= 0.85(1 - \exp(-0.05t)) + 0.1
  \end{align*}
  \]

- **Log-logistic:**

  \[
  \begin{align*}
  \theta_0(t) &= (1 + \exp(1.5 - 0.6 \log(t)))^{-1} + 0.1 \\
  \theta_1(t) &= (1 + \exp(1 - 0.75 \log(t)))^{-1} + 0.1
  \end{align*}
  \]

- **S/U-shaped:**

  \[
  \begin{align*}
  \theta_0(t) &= 0.04t - 6 \cdot 10^{-5} \cdot t^2 + 0.1 \\
  \theta_1(t) &= 0.4(0.25\sqrt{t} - 0.9)^3 + 0.3916
  \end{align*}
  \]

For all three curves, the true intercept is set as \( \theta_{00} = 0.1 \). The illustration of the true data-generating curves is provided in Appendix A.1.

For each true transition probability curve, we fit five different models, i.e., linear link model, logit link model, linear link model with non-decreasing constraints, parametric exponential decay model, and parametric log-logistic model. All models were implemented using the Bayesian MCMC algorithm using R-Jags. All simulations are based on 500 Monte Carlo replications. In each replication, we ran two chains for each model, discarded the first 2,000 iterations of each chain as burn-in, and used the next 8,000 iterations as posterior samples. In order to generate realistic data, we choose the number of subjects \( n = 100 \) and 12 observed time points after the initial time point, \( \mathbf{t} = (0, 2, 4, 8, 12, 20, 24, 36, 44, 48, 52, 56, 60)^T \), which is motivated from the Toenail data that we will present in Sect. 4.1. For simplicity, all subjects have the same time points. For B-splines, we choose \( \lfloor \sqrt{T} \rfloor = 3 \) internal knots and cubic splines.

To compare the model fit, the root mean integrated squared error (RMISE), Watanabe–Akaike information criterion [WAIC;21], and Leave-One-Out information criterion [LOOIC;22] were computed. The RMISE of an estimate \( \hat{\theta}(i) \) is defined as

\[
\text{RMISE} = \sqrt{\frac{1}{T} \int_0^T \left( \theta(s) - \hat{\theta}(s) \right)^2 ds} \approx \sqrt{\frac{1}{T} \sum_{i=1}^h \left( \theta(s_i) - \hat{\theta}(s_i) \right)^2 \Delta s_i}.
\]
For our simulation, $t_T = 60$, and we choose $h = 100$ and equidistant $\Delta s_i$.

WAIC is an extension of the Akaike Information Criterion (AIC) and an improvement of Deviance Information Criterion (DIC) in a fully Bayesian approach. WAIC is obtained as $\text{WAIC} = -2(\text{lpdp} - p_{\text{waic}})$ where $\text{lpdp}$ refers to log pointwise predictive density which measures the predictive accuracy of the fitted model [23]. The $\text{lpdp}$ which is originally defined by the posterior distribution $p(\theta \mid y)$ can be computed using posterior samples $\theta^{(s)}, s = 1, ..., S$, by $\text{lpdp} = \sum_{i=1}^n \log \int p(y_i \mid \theta) p(\theta \mid y) d\theta \approx \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S \text{Pr}(y_i \mid \theta^{(s)}) \right)$, where $n$ is the number of data points. Moreover, the quantity $p_{\text{waic}}$ represents the ‘effective’ number of parameters, which measures the complexity of the model. Specifically, as the parameters are random within a Bayesian model, the $p_{\text{waic}}$ approximates the number of unconstrained parameters in the model; each parameter counts as 1 if it is estimated with no constraints or weakly informative prior, 0 if fully constrained or prior distributions are informative, or an intermediate value if in between them [24]. The quantity is calculated as $p_{\text{waic}} = \sum_{i=1}^n \text{Var}_y(\log(\text{Pr}(y_i \mid \theta^{(s)})))$, with $\text{Pr}(y_i \mid \theta^{(s)})$ representing the posterior predictive likelihood. WAIC is a more robust criterion compared to AIC and DIC, as it averages over the posterior distribution. Alternative to WAIC, Vehtari et al., [22] approximate Leave-One-Out CV (LOOCV) using importance weights. Even though WAIC is asymptotically equal to Leave-One-Out CV, according to [22], the LOOCV approach could be more robust. Note that the model with a small value of the RMISE, WAIC, or LOO is considered to be better in terms of model fit.

All of the previously mentioned criteria, however, do not take into account the auto-correlative characteristics of the Markov models. Therefore, we consider another criterion suggested by Geweke and Amisano [25] that evaluates the predictive performance of the sequential prediction. Specifically, the one-step-ahead predictive likelihood at time $t$ is defined as

$$\text{PL}(t) = p(y_{t+1} \mid y_{1:t}) = \int p(y_{t+1} \mid y_{1:t}, \theta) p(\theta \mid y_{1:t}) d\theta \approx \frac{1}{S} \sum_{s=1}^S p(y_t \mid y_{1:t}, \theta^{(s)}).$$

Consequently, we define the one-step-ahead negative cumulative log predictive likelihood (NCLPL) as

$$\text{NCLPL} = - \sum_{t=H+1}^{T-1} \log \text{PL}(t),$$

where $H$ is a pre-specified time point. A model with the smaller NCLPL better explains the data.

Moreover, since we make predictions of binary outcomes, we compare the accuracy of our prediction which is the proportion of correct prediction out of total prediction. By averaging by the number of subjects and the number of MCMC samples $S$, the accuracy is calculated as

$$\text{Accuracy} = \frac{1}{nS(T-H)} \sum_{s=1}^S \sum_{i=1}^n \sum_{t=H+1}^T I(y_{it} = \hat{y}_{it}^{(s)}),$$

\text{Springer}
where $y_{it}$ denotes the observed outcome of subject $i$ at time $t$, and $\hat{y}^{(s)}_{it}$ denotes the predicted value of the $s$th MCMC sample. In order to be consistent with other model fit criteria that the smaller value implies the better fit, we calculate the error rate defined as 1-Accuracy. For the NCLPL and error rate, we set $H = T - 4$.

Figure 2 displays the simulation results with boxplots based on 500 Monte Carlo repetitions; the numerical results are provided in Appendix A.1 The LOOIC and WAIC show very similar results. When the true transition probability curves are exponential decay or log-logistic, then parametric models show the smaller WAIC, LOOIC, RMISEs, NCLPL, and error rate, as expected. In fact, exponential decay or log-logistic have similar concave shapes, and we found that the log-logistic model performs well under both exponential decay and log-logistic true curves. The proposed nonparametric models showed comparable performance. Particularly, the linear link model with non-decreasing constraints showed better performance than models without the constraints and sometimes even better than parametric models.

On the other hand, when the true transition probability curves are S-shaped or U-shaped, i.e., if the true curves do not have concave shapes, the performance of the parametric models is poor. The nonparametric model with non-decreasing constraints estimates $\theta_1$ well but did not for $\theta_0$, as the former has non-decreasing increasing shapes whereas the latter does not. These simulation results show the robustness of our nonparametric models regardless of the shape of true functions, whereas parametric models showed poor performance when the model is misspecified. Also, the nonparametric model with non-decreasing constraints shows the best performance in general under all three scenarios where the true probability curves are non-decreasing.

### 3.2 Knot Selection

In this section, we compare the model fit of the proposed P-spline models to assess the appropriate number of internal knots and orders of penalty. The data sets are generated from the following transition probability curves when $t \in (0, 100]$,

- $\theta_0(t) = 0.9 - 0.8 \left[ 1 + \exp\{-1.5 \cos(3\pi t/50)\} \right]^{-1}
- \theta_1(t) = 1 - \left[ 1 + \exp\{1.5 \sin(4\pi \times (t + 5)/50)\} \right]^{-1},

and $\theta_{00} = \theta_0(0) = \theta_1(0) = 0.2$. The illustration of the true transition probability curves is provided in Appendix A.2. Given the number of subjects $n = 100$ and the time range $t \in [0, 100]$, we choose the number of distinct time points $T = 25$ where $t = (t_1, \ldots, t_T)$ are equidistant in $[1, 100]$, and $t_0 = 0$.

We compare the model fit of the P-spline model with different orders of penalty, i.e., no penalty and penalties of orders 1, 2, and 4. For each penalty level, we consider internal knots with four different sizes, 5, 10, 15, and 20. We use the same five model comparison criteria in Sect. 3.1 based on 500 Monte Carlo repetitions. In addition to the one-step-ahead PL, we consider d-step-ahead prediction and compare the corresponding NCLPL and error rate. The d-step-ahead predictive likelihood at time $t$ is calculated as
Fig. 2  Boxplots of various model evaluation criteria under three different true models; a Exponential Decay, b Log-Logistic, c S/U shaped

\[
P L_d(t) = p(y_{t+1:t+d} \mid y_{1:t}) \approx \frac{1}{S} \sum_{s=1}^{S} p\left(y_{t+1:t+d} \mid y_{1:t}, \theta^{(s)} \right).
\]

and similarly,
We set $H = T - 4$ and $d = 4$. The average error rate at each predicted time point $H + 1, \ldots, H + d$ is also calculated.

Figures 3 and 4 display the results with boxplots; numerical results are provided in Appendix A.2. Under the small number of internal knots, the model with no penalty performs the best, and under the large number of internal knots, the model with a high order of penalty performs well in general. The performance of all four models under the internal knots of size 5 is relatively poor compared to the higher number of internal knots. However, all models show similar performance when the number of internal knots is greater than 10. Under each of the different internal knot settings, the model with penalty order 2 performs the best in general. Figure 4 illustrates the
4-step-ahead error rate at each time point under the different number of internal knots. As expected, the error rate increases as we predict the later time point, and the model with the penalty order of 2 shows the smallest increase in general. Therefore, from the simulation, the number of the internal knot around 10 was found to be sufficient, and the model with the penalty order of 2 was found to be the best in general.

### 4 Real Data Analysis

In this section, we apply the proposed models to analyze two data sets, the toenail clinical trial data and the Dow Jones data, a stock market index.

#### 4.1 Toenail Data

We first analyze a randomized clinical trial comparing two competing oral treatments, Terbinafine and Itraconazole, for toenail infection. Details of the study are described in De Backer et al. [26], and the dataset is available from R-package HSAUR3. The dichotomous outcome variable is the severity of toenail infection, with 0 denoting a moderate or severe symptom and 1 denoting a none or mild symptom. Patients had oral treatments for 12 weeks, and the follow-up visits were scheduled at weeks 4, 8, 12, 24, 36, and 48. Among 294 enrolled patients, 148 patients were on Terbinafine and 146 patients were on Itraconazole; the data are not complete as 31 patients on Terbinafine and 29 patients on Itraconazole did not complete all seven scheduled visits.

We are interested in analyzing the probability that the toenail infection is mitigated after treatment begins for each treatment. Let’s define $\theta_{ab}(t)$ be the transition probability from status $a$ to status $b$ at time $t$. Among the four possible transition probabilities, the most interesting one is $\theta_{01}(t)$, which indicates the transition probability
**Table 1** Model fit comparisons for toenail dataset; bold numbers imply the best model fit, i.e., the smaller the better

|                  | WAIC   | LOOIC  | NCLPL  | Error rate |
|------------------|--------|--------|--------|------------|
| Linear link      | 559.7  | 607.0  | 56.95  | 0.070      |
| Terbinafine      |        |        |        |            |
| Monotone         | 562.0  | 609.8  | 60.43  | 0.077      |
| Subject-specific | 551.5  | 614.0  | 57.03  | 0.082      |
| Linear link      | 681.5  | 703.7  | 90.88  | 0.118      |
| Itraconazole     |        |        |        |            |
| Monotone         | 686.7  | 707.4  | 93.87  | 0.121      |
| Subject-specific | 657.5  | 713.7  | 83.55  | 0.116      |

from moderate or severe to none or mild at time $t$. In addition, $\theta_{11}(t)$, the probability of continuing to have none or mild symptom, is also our interest.

We used the proposed B-spline model with the linear link, the B-spline model with non-decreasing restriction, and the subject-specific B-spline model to fit the data. The monotone increasing assumption is reasonable because the probability of getting better would increase as patients kept receiving treatments. For both treatments, cubic splines with $m$ internal knots are used, where $m = \lfloor \sqrt{6} \rfloor = 2$. Two chains were used, and for each chain, 25,000 MCMC samples were used for posterior inference after discarding the first 5,000 samples.

For the model fit comparison, note that RMISEs cannot be computed for real data as we do not have true parameter values unless we use cross-validation by holding out data. On the other hand, we provide several predictive criteria to evaluate the model performance that can be computed based on real data. Similar to the simulation studies in Sect. 3, we compared model fits using the WAIC, LOOIC, NCLPL, and error rate. Due to the small number of visits, we conducted one-step-ahead predictions from visit 5 (week 24) to visit 7 (week 48). Table 1 shows the model fit comparisons for the toenail data. The error rate is the average of three one-step-ahead predictions. For Terbinafine, the linear link B-spline model was found to be the best model, whereas for Itraconazole, the subject-specific model was the best. Figure 5 shows the estimated transition probability curves from moderate or severe to none or mild and continuing to have none or mild symptom, using the selected model, i.e., the linear link model for Terbinafine and subject-specific model for Itraconazole. For the transition probability from moderate or severe to none or mild, both treatments show increasing trends at first and then decreasing. These trends are reasonable because the patients were on treatments for 12 weeks, and their symptoms were measured during and afterward. On the other hand, the transition probabilities of continuing to have none or mild symptom are very high for both treatments. In both cases, it can be seen from the wide overlapping credible intervals that the two treatments are comparable.

### 4.2 Dow Jones Index

In Econometrics, a Markov model has been widely applied to analyze stock price index data. Hamilton [27] initially proposed an auto-regressive framework using a first-order
Markov chain and extended it with an autoregressive conditional heteroskedasticity [28]. De Angelis and Pass [29] proposed a framework to investigate the dynamic patterns of stock markets by exploiting the latent Markov model and applied this framework to the U.S. stock market indices. Similarly, Sasikumar and Abdullah [30] used Hidden Markov Model to make inferences regarding stock market trends, and Cheng et al. [31] proposed a state-varying endogenous regime-switching model and analyzed Chinese stock market returns.

One of the most commonly used stock indices is the Dow Jones Industrial Average which measures the stock performance of 30 blue-chip companies in the USA. It is one of the most-watched stock indexes in the world, containing companies such as Apple, Boeing, Microsoft, etc. [32]. Our analysis is based on daily log returns for the 30 companies in the Dow Jones Index, defined as $lr(t) = \log(p_t/p_{t-1})$, where $p_t$ denotes the price of the stock on day $t$. The data set is obtained from get.hist.quote

Fig. 5 Estimated transition probabilities curves of two treatments; a transition probabilities from moderate or severe to none or mild, b probability of continuing to have none or mild symptom
function in R-package \texttt{tseries} for the period from July 1, 2019 to November 15, 2020, which includes $T = 348$ time points. Figure 6 illustrates the trend of daily log returns of each stock included in the Dow Jones Index, which shows high volatility that makes it difficult to assess the stock trend.

We applied the proposed Markov models to investigate the behavior of the 30 blue-chip companies included in the Dow Jones Index. In particular, we are interested in the behavior in terms of “safety” during the recent COVID-19 pandemic. In this regard, two states of log return are defined as loss - “negative or zero log return” or gain - “positive log return”. The zero return can be considered the negative return because it is the loss of money considering that one could earn at least some interest instead if one had deposited the money in the bank account.

Using R-jags, we fitted four different models; the linear link model, the subject-specific linear link model, and the penalized B-spline model with penalties of orders 1 and 2. We ran two chains for each model, discarded the first 5,000 iterations of each chain as burn-in, and used the next 10,000 iterations as the posterior sample. For the B-spline models (the first and the second models), we used $\lfloor \sqrt{T} \rfloor = \lfloor \sqrt{348} \rfloor = 18$ internal knots, and for the P-spline models (the third and the fourth), we used 35 internal knots. Similarly to Sect. 3, we use the WAIC, LOOIC, NCLPL, and error rates to compare the model fit. For NCLPL, in addition to the NCLPL 1 with one-step-ahead predictions, we conduct five-step-ahead predictions and calculate the NCLPL 5, since the week ahead prediction is one of the common interests in the stock market. With $H = T - 10$, the NCLPL 1 is the sum of 10 one-step-ahead predictions, and the NCLPL 5 is the sum of 6 five-step-ahead predictions. For error rates, we use $H = T - 10$.

Table 2 compares the model fits of the four models. All five criteria suggest that the P-spline model with the order of 1 is the best model to fit the Dow Jones data. Figure 7 illustrates estimated transition probability curves and 95% credible intervals from the P-spline with order 1 model. The solid black lines indicate the transition probabilities of the overall Dow Jones Index, and the dashed lines represent the 95%
Table 2  Model fit comparisons for the Dow Jones Index data; bold numbers imply the best model fit, i.e., the smaller the better

| Model            | WAIC    | LOOIC   | NCLPL | NCLPL5 | Error rate |
|------------------|---------|---------|-------|--------|------------|
| Naive            | 14181.15| 14181.27| 207.3 | 857.3  | 0.442      |
| Subject-specific | 14362.98| 14366.16| 210.3 | 850.6  | 0.436      |
| P-spline order 1 | **14021.81** | **14022.07** | **200.2** | **756.2** | **0.382**  |
| P-spline order 2 | 14098.21| 14098.40| 201.5 | 813.1  | 0.425      |

Fig. 7  Estimated transition probability curves and 95% credible intervals from the P-spline model with order 1

upper and lower credible bounds. Blue areas indicate the periods when the estimated curves are greater than 0.5 with 95% probability, and red areas indicate the period when the estimated curves are less than 0.5 with at least 95% probability.

The upper plot presents the estimated transition probability curves from the *loss* stage to *gain* stage. In general, the overall Dow Jones Index shows a high transition probability from the *loss* to *gain*, since the estimated transition probability is around or greater than 0.5 in general. The periods with high transition probability from loss to
Fig. 8 Estimated overall mean vs. individual stock transition probability curves using the subject-specific model; a transition probability from loss to gain, b probability of being gain

gain include ‘2019-08-12 - 2019-09-11’, ‘2020-03-16 - 2020-04-20’, ‘2020-06-26 - 2020-07-17’, and ‘2020-09-25 - 2020-10-15’. The bottom plot represents the estimated probability curve of being gain stage. The overall trend shows that Dow Jones Index is a safe index since the probability is around 0.5 or greater. However, in the period of ‘2020-02-13 - 2020-04-09’, the probability of being gain was less than 0.5, with the lowest probability of 0.13 (95% CI: 0.09 - 0.19) on ‘2020-03-09’. This could be due to the anxiety about the pandemic since COVID-19 started widely spreading in this period.

If one is interested in estimating the trend of an individual stock, the proposed subject-specific model allows estimating the individual transition probability curves as well as the overall transition probability curves. Figure 8 shows the estimated overall mean vs. individual stock transition probability curves using the subject-specific model. The colored curves are the transition probabilities of individual stocks, and the black curves are the overall mean probability curves. While the behavior of 30 stocks is similar to the overall mean in general, these individual stock transition probability
curves enable us to identify which stocks are safer and which are riskier than other stocks.

5 Discussion

In this paper, we have proposed a flexible Bayesian nonparametric approach for nonhomogeneous Markov model when the response variable is autocorrelated and dichotomous. We use B-spline basis functions to reduce the number of parameters to estimate, and also to estimate smooth transition probability curves. Our model allows imposing non-decreasing restrictions or estimating individual-specific curves. Simulation studies show the robustness of our models under the various shapes of the true transition probability functions.

Within a frequentist framework, our proposed models can possibly be estimated using some optimization methods, e.g., the popular quadratic programming method as we have illustrated (see Sect. 2.4). However, as is well known the frequentist framework is particularly complicated when faced with missing (or censored) values in the data. Whereas within a Bayesian framework, under the usual assumption of missing at random (MAR), the posterior predictive distribution of the missing (or censored) values conditional only on the observed values can be sampled easily to do the imputation on the fly while drawing samples from the posterior distribution. Moreover, within a Bayesian framework the quantification of uncertainty (e.g., standard errors) of the parameter estimates can be obtained rather easily; and in fact the entire posterior or predictive distribution can be estimated using the samples generated by the MCMC methods.

Our proposed models are based on the nonhomogeneous Markov assumption of the binary response $Y_t$. It is natural to extend these models under the hidden Markov assumption [see, e.g.,33] that the Markov process is not observable. Also, under the assumption of the non-decreasing transition probability curves, we impose monotone increasing shape restrictions. To better estimate the true curves, other shape constraints such as convex, concave, S-shaped, and U-shaped, can also be imposed, [e.g.,34], based on subject matter knowledge. As mentioned earlier, the primary focus of this work is to explore the finite sample performances of the proposed dynamic models for transition probabilities, but it would be desirable to establish theoretical large sample properties of the proposed estimators under some technical regularity conditions by applying some of the existing theory for penalized splines models. However, verification of such technical conditions is often not possible in practice, especially those that are based on realistic scenarios.

Acknowledgements The authors would like to thanks the editors and the two anonymous reviewers for their thoughtful suggestions and constructive criticism which have lead to more well-organized version of the manuscript.
Declarations

Conflict of interest  On behalf of all authors, the corresponding author states that there is no conflict of interest.

A Details of Simulation Studies

This section provides additional details of the simulation studies in Sect. 3, including the illustration of the true data generating functions and the numerical results of the simulation studies.

A.1 Details of the Simulation of the B-Spline Models in 3.1

Figure 9 illustrates the three underlying true transition probability curves for simulation studies of the B-spline models in section 3.1. Solid dots indicate the 13 observed time points.

Table 3 summarizes the simulation results based on the model fit criteria described in 3.1. The average WAIC, LOOIC, RMISE, NCLPL, and error rate are summarized with corresponding Monte Carlo standard errors in the parenthesis. Similar to Fig. 2, parametric models show better model fits when the true transition probability curves are exponential decay or log-logistic. However, the proposed B-spline basis models

![Fig. 9](image-url)  
**Fig. 9** True transition probability curves under three different scenarios; exponential decay, log-logistic, and S/U shaped
Table 3  Simulation results under three sets of true transition curves; bold numbers imply the best model fit, i.e., the smaller the better

| True curve     | Model          | RMISE $\theta_0$ | RMISE $\theta_1$ | WAIC   | LOOIC   | NCLPL   | Error rate |
|----------------|----------------|------------------|------------------|--------|---------|---------|------------|
| Exp. Decay     | Linear link    | 0.045 (6E-4)     | 0.031 (5E-4)     | 1322.9 | 1323.0  | 150.3   | 0.222 (1.2E-3) |
|                | Logit link     | 0.047 (7E-4)     | 0.040 (6E-4)     | 1323.9 | 1323.9  | 150.0   | 0.223 (1.1E-3) |
|                | Linear monotone| 0.038 (7E-4)     | 0.032 (6E-4)     | 1321.6 | 1321.6  | 150.0   | 0.219 (1.1E-3) |
| Exp. decay     | 0.030 (7E-4)   | 0.018 (5E-4)     | 1318.2 (1.82)    | 1318.2 | 150.0   | 0.217   | 0.223 (1.1E-3) |
| Log-logistic   | 0.045 (5E-4)   | 0.022 (5E-4)     | 1320.9 (1.83)    | 1320.9 | 150.5   | 0.217   | 0.223 (1.1E-3) |
| Log-logistic   | Log-logistic   | 0.059 (1.0E-3)   | 0.027 (5E-4)     | 922.9  | 922.9   | 44.61   | 0.052 (6E-4) |
|                | Logit link     | 0.063 (1.1E-3)   | 0.033 (6E-4)     | 924.0  | 924.1   | 44.89   | 0.055 (7E-4) |
|                | Linear monotone| 0.048 (9E-4)     | 0.031 (5E-4)     | 922.1  | 922.2   | 44.60   | 0.050 (6E-4) |
| Exp. decay     | 0.071 (1.2E-3) | 0.048 (4E-4)     | 932.4 (1.91)     | 932.4  | 46.45   | 0.061   | 0.061 (7E-4) |
| Log-logistic   | 0.035 (1.0E-3) | 0.018 (5E-4)     | 919.0 (1.86)     | 919.0  | 44.82   | 0.051   | 0.051 (6E-4) |
| S/U-shaped     | Linear link    | 0.040 (4E-4)     | 0.050 (8E-4)     | 1362.9 | 1363.0  | 245.6   | 0.423 (6E-4) |
|                | Logit link     | 0.039 (5E-4)     | 0.053 (8E-4)     | 1363.2 | 1363.3  | 245.6   | 0.423 (6E-4) |
|                | Linear monotone| 0.057 (4E-4)     | 0.129 (4E-4)     | 1396.9 | 1396.9  | 255.8   | 0.434 (5E-4) |
| Exp. decay     | 0.109 (3E-4)   | 0.130 (6E-4)     | 1425.0 (1.45)    | 1425.0 | 257.8   | 0.455   | 0.455 (4E-4) |
| Log-logistic   | 0.066 (5E-4)   | 0.153 (4E-4)     | 1407.2 (1.65)    | 1407.2 | 256.7   | 0.435   | 0.435 (5E-4) |
show comparable performance. On the other hand, when the true transition probability curves have non-concave shapes such as S-shaped or U-shaped, the nonparametric B-spline models significantly outperform the parametric models.

**A.2 Details of the Simulation of Knot Selection in 3.2**

Figure 10 illustrates the true transition probability curves to evaluate the performance of P-spline models in 3.2.

Table 4 summarizes the simulation results. Analogous to Fig. 3 and Fig. 4 in the main paper, P-spline models with higher orders of penalty perform better under a large number of internal knots. The internal knot of size 10 was found to be sufficient, and the P-spline model with penalty order 2 performs the best in general.
Table 4  Simulation results under the different number of internal knots; bold numbers imply the best model fit, i.e., the smaller the better

| Model        | RMISE $\theta_0$ | RMISE $\theta_1$ | WAIC  | LOOIC | NCLPL | Error rate | NCLPL$_4$ |
|--------------|------------------|------------------|-------|-------|-------|------------|-----------|
| **# internal knots = 5** | | | | | | | |
| No penalty   | 0.154 (2E-4)     | 0.196 (1E-4)     | 3406.7 (1.30) | 3406.7 (1.30) | 240.4 (0.36) | 0.410 (7E-4) | 399.8 (1.31) |
| Penalized 1  | 0.158 (2E-4)     | 0.200 (1E-4)     | 3418.0 (1.31) | 3418.0 (1.31) | 241.0 (0.35) | 0.414 (7E-4) | 415.2 (4.82) |
| Penalized 2  | 0.161 (3E-4)     | 0.203 (1E-4)     | 3437.6 (1.32) | 3437.7 (1.32) | **239.0** (0.36) | **0.408** (8E-4) | **348.6** (2.20) |
| Penalized 4  | 0.170 (2E-4)     | 0.212 (1E-4)     | 3459.6 (1.18) | 3459.6 (1.18) | 239.6 (0.34) | 0.422 (7E-4) | 368.0 (3.07) |
| **# internal knots = 10** | | | | | | | |
| No penalty   | 0.047 (5E-4)     | 0.062 (4E-4)     | 3138.3 (1.78) | 3138.3 (1.78) | **231.8** (0.38) | 0.389 (8E-4) | **356.2** (0.93) |
| Penalized 1  | 0.049 (4E-4)     | 0.066 (4E-4)     | 3137.5 (1.78) | 3137.6 (1.78) | 232.0 (0.38) | 0.388 (8E-4) | 389.4 (3.02) |
| Penalized 2  | **0.045** (4E-4) | **0.061** (3E-4) | **3137.2** (1.79) | **3137.2** (1.79) | 231.9 (0.38) | **0.385** (8E-4) | 433.2 (3.04) |
| Penalized 4  | 0.050 (5E-4)     | 0.119 (1.3E-3)   | 3202.9 (2.93) | 3203.0 (2.93) | 233.3 (0.40) | 0.397 (9E-4) | 602.2 (14.53) |
| **# internal knots = 15** | | | | | | | |
| No penalty   | 0.052 (4E-4)     | 0.055 (5E-4)     | 3135.0 (1.80) | 3135.1 (1.80) | **231.7** (0.38) | 0.393 (8E-4) | **307.7** (0.32) |
| Penalized 1  | **0.049** (4E-4) | 0.058 (5E-4)     | 3133.4 (1.80) | 3133.4 (1.80) | 231.8 (0.38) | 0.392 (8E-4) | 366.0 (2.04) |
| Penalized 2  | 0.045 (4E-4)     | **0.052** (4E-4) | **3131.5** (1.81) | **3131.5** (1.81) | **231.7** (0.38) | **0.389** (8E-4) | 364.5 (1.42) |
| Penalized 4  | 0.046 (4E-4)     | 0.079 (6E-4)     | 3153.7 (2.00) | 3153.7 (2.00) | 232.0 (0.38) | **0.389** (8E-4) | 871.9 (11.35) |
| **# internal knots = 20** | | | | | | | |
| No penalty   | 0.057 (4E-4)     | 0.060 (4E-4)     | 3139.1 (1.81) | 3139.2 (1.81) | **231.7** (0.38) | 0.393 (8E-4) | **307.9** (0.25) |
| Penalized 1  | 0.049 (4E-4)     | 0.058 (5E-4)     | 3135.1 (1.80) | 3135.1 (1.80) | **231.7** (0.38) | 0.393 (8E-4) | 351.1 (1.61) |
| Penalized 2  | **0.045** (4E-4) | **0.052** (4E-4) | **3132.5** (1.81) | **3132.5** (1.81) | 231.8 (0.38) | 0.390 (8E-4) | 356.3 (1.31) |
| Penalized 4  | **0.045** (4E-4) | 0.066 (4E-4)     | 3140.7 (1.81) | 3140.8 (1.81) | 231.9 (0.38) | **0.388** (8E-4) | 763.5 (7.89) |
References

1. Shoko C, Chikobvu D (2018) Time-homogeneous markov process for hiv/aids progression under a combination treatment therapy: cohort study, South Africa. Theoret Biol Med Model 15(1):1–14
2. Ma J, Chan W, Tilley BC (2018) Continuous time markov chain approaches for analyzing transtheoretical models of health behavioral change: a case study and comparison of model estimations. Stat Methods Med Res 27(2):593–607
3. Jackson CH et al (2011) Multi-state models for panel data: the MSM package for r. J Stat Softw 38(8):1–29
4. Hubbard RA, Inoue L, Fann J (2008) Modeling nonhomogeneous markov processes via time transformation. Biometrics 64(3):843–850
5. Jones RH, Xu S, Grunwald GK (2006) Continuous time Markov models for binary longitudinal data. Biom J 48(3):411–419
6. Zeng L, Cook RJ (2007) Transition models for multivariate longitudinal binary data. J Am Stat Assoc 102(477):211–223
7. Andersen PD, Iversen A, Madsen H, Rode C (2014) Dynamic modeling of presence of occupants using inhomogeneous Markov chains. Energy Build 69:213–223
8. Iversen EB, Møller JK, Morales JM, Madsen H (2016) Inhomogeneous Markov models for describing driving patterns. IEEE Trans Smart Grid 8(2):581–588
9. de Boor C (2002) Spline basics. Handbook of Computer Aided Geometric Design 141–164
10. Meyer MC (2012) Constrained penalized splines. Canad J Stat 40(1):190–206
11. Shin SJ, Ghosh SK (2017) A comparative study of the dose-response analysis with application to the target dose estimation. J Stat Theory Pract 11(1):145–162
12. McKay Curtis S, Ghosh SK (2011) A variable selection approach to monotonic regression with Bernstein polynomials. J Appl Stat 38(5):961–976
13. Lorentz GG (2013) Bernstein polynomials (American Mathematical Society)
14. Eilers PH, Marx BD (1996) Flexible smoothing with b-splines and penalties. Stat Sci 11:89–102
15. Lang S, Brezger A (2004) Bayesian p-splines. J Comput Graph Stat 13(1):183–212
16. Currie ID, Durban M (2002) Flexible smoothing with p-splines: a unified approach. Stat Model 2(4):333–349
17. Kauermann G, Krivobokova T, Fahrmeir L (2009) Some asymptotic results on generalized penalized spline smoothing. J Royal Stat Soc Series B (Stat Methodol) 71(2):487–503
18. Yoshida T, Naito K (2014) Asymptotics for penalised splines in generalised additive models. J Nonparam Stat 26(2):269–289
19. Diggle P et al (2002) Analysis of longitudinal data. Oxford University Press, Oxford
20. Fitzmaurice GM, Laird NM, Ware JH (2012) Applied longitudinal analysis, vol 998. John Wiley & Sons, Hoboken
21. Watanabe S, Opper M (2010) Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 11(12)
22. Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and waic. Stat Comput 27(5):1413–1432
23. Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for bayesian models. Stat Comput 24(6):997–1016
24. Gelman A et al. (2013) Bayesian data analysis
25. Geweke J, Amisano G (2010) Comparing and evaluating Bayesian predictive distributions of asset returns. Int J Forecast 26(2):216–230
26. De Backer M, De Vroey C, Lesaffre E, Scheys I, De Keyser P (1998) Twelve weeks of continuous oral therapy for toenail onychomycosis caused by dermatophytes: a double-blind comparative trial of terbinafine 250 mg/day versus itraconazole 200 mg/day. J Am Acad Dermatol 38(5):S57–S63
27. Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econom J Econ Soc 357–384
28. Hamilton JD, Susmel R (1994) Autoregressive conditional heteroskedasticity and changes in regime. J Econom 64(1–2):307–333
29. De Angelis L, Paas LJ (2013) A dynamic analysis of stock markets using a hidden Markov model. J Appl Stat 40(8):1682–1700
30. Sasikumar R, Abdullah AS (2016) Forecasting the stock market values using hidden Markov model. Int J Bus Anal Intell 4:17–21
31. Cheng T, Gao J, Yan Y (2018) A new regime switching model with state-varying endogeneity. J Manag Sci Eng 3(4):214–231
32. Probasco Jim (2022) The Dow Jones Industrial Average is one of the world’s most influential stock indexes. Here’s how it works. https://www.businessinsider.com/personal-finance/what-is-dow-jones . Accessed: 2022-10-01
33. Grewal JK, Krzywinski M, Altman N (2019) Markov models-hidden Markov models. Nat Methods 16(9):795–796
34. Lenk PJ, Choi T (2017) Bayesian analysis of shape-restricted functions using gaussian process priors. Stat Sin 43–69

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.