Semiparametric Mixed-effects Ordinary Differential Equation Models with Heavy-tailed Distributions

Baisen Liu, Liangliang Wang, Yunlong Nie and Jiguo Cao*

Abstract

Ordinary differential equation (ODE) models are popularly used to describe complex dynamical systems. When estimating ODE parameters from noisy data, a common distribution assumption is using the Gaussian distribution. It is known that the Gaussian distribution is not robust when abnormal data exist. In this article, we develop a hierarchical semiparametric mixed-effects ODE model for longitudinal data under the Bayesian framework. For robust inference on ODE parameters, we consider a class of heavy-tailed distributions to model the random effects of ODE parameters and observations errors. An MCMC method is proposed to sample ODE parameters from the posterior distributions. Our proposed method is illustrated by studying a gene regulation experiment. Simulation studies show that our proposed method provides satisfactory results for the semiparametric mixed-effects ODE models with finite samples.

Keywords: Metropolis-Hastings; Abnormal data; Scale mixture of multivariate normal distributions.

Baisen Liu is Assistant Professor of School of Statistics, Dongbei University of Finance and Economics, China; Liangliang Wang is Assistant Professors of Department of Statistics and Actuarial Science, Simon Fraser University, Canada; Yunlong Nie is Ph.D. Candidate of Department of Statistics and Actuarial Science, Simon Fraser University, Canada; Jiguo Cao is Associate Professors of Department of Statistics and Actuarial Science, Simon Fraser University, Canada. Corresponding author: jiguo_cao@sfu.ca

1 Introduction

Ordinary differential equations are widely used to model complex dynamical systems in many areas of science and technology. For example, ODE models have been used in the study of reconstruction of transcription factor activity for gene regulatory activities (Khanin et al., 2007; Rogers, Khanin and Girolami, 2007). Although ODE models are often proposed based on expert knowledge of the dynamical process of interest, the solver of the ODE relies heavily on the ODE parameters and the values of the ODE parameters are rarely known. It is an important but challenging statistical problem to estimate the ODE parameters from observational (noisy) data since that most ODEs have no analytic solutions and it is often computationally intensive to solve ODE numerically.

For estimating the ODE parameters, many efficient methods have been developed. For instance, Liang and Wu (2008) developed a two-step method using local polynomial regression. Ramsay et al. (2007) and Cao, Fussmann and Ramsay (2008) proposed a generalized profiling approach to estimate the ODE parameters. Hall and Ma (2014) suggested a class of fast, easy-to-use, genuinely one-step procedures for estimating unknown parameters in dynamical system models. Brunel, Clairon and d'Alché-Buc (2014) developed a gradient matching approach for estimating ODE parameters. In the Bayesian framework, many researches have been made for estimating the ODE models, for example, Campbell and Steele (2012) proposed a Bayesian smooth functional tempering method for the ODE models; Bhaumik and Ghosal (2015) considered a Bayesian two-step estimation approach; Dass et al. (2017) suggested a Laplace approximation method for obtaining the posterior inference of ODE parameters, among others. However, all of the above methods assume that the data are collected for one individual dynamical process, and cannot address the problem when multiple individuals dynamical processes are observed.

When the dynamical process are measured for multiple subjects, mixed-effects ODE models are developed to include random-effects to account for the within-subjects correla-

tions in the models. For instance, Huang and Wu (2006) studied a parametric hierarchical HIV dynamic model and provided an MCMC algorithm to sample from the posterior distribution of ODE parameters. Guedj, Thiébaut and Commenges (2007) used the maximum likelihood approach directly to estimate unknown parameters in mixed-effects ODE models. Fang, Wu and Zhu (2011) proposed a fast two-stage estimating procedure for mixed-effects dynamical systems and applied it to analyze longitudinal HIV virus data. Wang et al. (2014) proposed a semiparametric method to estimate a mixed-effects ODE model for the HIV combination therapy study. A common fundamental assumption of aforementioned methods is that the observations for the dynamical process follow a normal distribution, but this assumption may lack robustness when abnormal data exist, which may lead to biased inference. On the other hand, the aforementioned papers only studied the parametric mixed-effects ODE models; while the semiparametric dynamic system which includes time-varying coefficients in ODE models can have many valuable applications, especially in gene transcription experiments in which the regulatory activity of the transcriptional factor is often modeled by a varying function over time (see the Eq. (4.1) in Section 4).

To introduce the motivation of this paper, we revisit the experiment that studies statistical reconstruction of transcription factor activity (TFA) from time-course gene expression data (Khanin et al., 2007; Rogers, Khanin and Girolami, 2007). It is a central but difficult problem to link transcription factors (TFs) to their target genes in post-genomic biology, since that the changes in the expression of a TF are subtle and its activity is often controlled at levels of other than expression, for example, via post-transcriptional and post-transcriptional modifications.

Under the log-normal noise assumptions, Khanin et al. (2007) made statistical inference on the ODE parameters using a conjugate gradient method, and Rogers, Khanin and Girolami (2007) developed a fully Bayesian approach to reconstruct the functional TFA from time-course gene expression data. However, this assumption may lack the robustness against

3

departures from normality and/or abnormal data. Figure 1(a) displays the histogram of the obtained residuals for the semiparametric mixed-effects gene regulation ODE model (4.1) by assuming that the observations errors and random-effects both follow normal distributions. It seems that the obtained residuals presents a heavy-tailed feature. This is also confirmed by the normal Q-Q plot of the residuals shown in Figure 1(b). In addition, the p-value from the Shapiro-Wilk test is 4.232×10^{-11}. Thus, a normal distribution assumption is quite doubtful and may be too restrictive to provide an accurate inference of ODE models.

To account for this departure from normality, we propose to apply the scale mixture of multivariate normal (SMN) distributions (Andrews and Mallows, 1974), to model the observations errors of the dynamical process and random effects of ODE parameters. The SMN distributions are flexible heavy-tailed distributions which include the multivariate normal distribution as a special case. In the literature, the SMN distributions have been extensively applied to regression models (Lange and Sinsheimer, 1993; Liu, 1996), linear mixed-effects models (Choy and Smith, 1997; Rosa, Gianola and Padovani, 2004), and nonlinear mixed-effects models (Meza, Osorio and De la Cruz, 2012; De la Cruza, 2014), to obtain robust estimates against outlying observations. In this article, we intend to pursue the robust inference of the ODEs model based on the SMN distributions.

This article has three main contributions. (i) We develop a hierarchical semiparametric mixed-effects ODE model which considers the within-subject and between-subject variations simultaneously and makes the statistical inference by borrowing more information from all subjects. (ii) Our model applies a class of heavy-tailed distributions for random-effect ODE parameters and observations errors for the dynamical process, which is robust against the atypical subjects and the abnormal observations within individual subjects. (iii) Our proposed method can detect the subjects which are atypical or have abnormal observations by estimating latent parameters in the model.

The remainder of this article is organized as follows. Section 2 briefly reviews the scale

4

mixture of multivariate normal distributions. Section 3 introduces our proposed hierarchical semiparametric mixed-effects ODE models and provides a Bayesian estimation method. Section 4 demonstrates our proposed method in comparison with the conventional method by studying the reconstruction of transcription factor activity from a set of real time-course gene expression measurements. Section 5 evaluates the finite sample performance of our proposed method using some simulation studies. We end this article with conclusions and some discussions of future research in Section 6. The Matlab code for our simulation studies can be downloaded at `https://github.com/caojiguo/SMODE`.

2 Scale Mixture of Multivariate Normal Distributions

The class of SMN distributions consists of a group of heavy-tailed distributions and has been applied to make robust inference in the statistic analysis. In this section, we provide some brief knowledge of the SMN distributions. More details refer to the book of Azzalini and Capitanio (2014).

Let $\boldsymbol{\mu}$ be an m-dimensional vector, and $\boldsymbol{\Sigma}$ be an $m \times m$ positive definite symmetric matrix. An m-dimensional random vector $\mathbf{Y}$ is said to follow a *scale mixture of multivariate normal distribution*, if it has a probability density function

$$p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \int_0^\infty u^{m/2} \exp(-\frac{uD^2(\mathbf{y})}{2}) \mathrm{d}H(u|\boldsymbol{\nu}), \qquad (2.1)$$

where $D^2(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$, and $H(u|\boldsymbol{\nu})$ is a univariate probability distribution function which depends on a scalar or vector valued parameter $\boldsymbol{\nu}$ and satisfies $H(0|\boldsymbol{\nu}) = 0$. We use the notation $\mathbf{Y} \sim SMN_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, H)$ to indicate that $\mathbf{Y}$ has the density (2.1). This class of distributions is flexible and includes many symmetric distributions as special cases, for instance, the normal, the Student-t, etc. When the mixture distribution function H is degenerate, $SMN_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, H)$ reduces to the usual multivariate normal distribution

$N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Let $\mathbf{Z} \sim N_m(\mathbf{0}, \boldsymbol{\Sigma})$ be an m-dimensional random variable, and $U \sim H(u|\boldsymbol{\nu})$ be an independent mixture variable. The SMN distributed random variables $\mathbf{Y}$ can be written as (Azzalini and Capitanio, 2014)

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}. \tag{2.2}$$

From (2.2), applying the iterative law of expectations, the mean and covariance of $\mathbf{Y}$ are given, respectively, by

$$\mathrm{E}(\mathbf{Y}) = \mathrm{E}[\mathrm{E}(\mathbf{Y}|U)] = \boldsymbol{\mu},$$

and

$$\mathrm{Cov}(\mathbf{Y}) = \mathrm{E}[\mathrm{Cov}(\mathbf{Y}|U)] + \mathrm{Cov}[\mathrm{E}(\mathbf{Y}|U)] = \mathrm{E}(U^{-1})\boldsymbol{\Sigma}. \tag{2.3}$$

Trivially, when $\mathrm{E}(U^{-1}) < \infty$, $\mathbf{Y}$ has a finite positive definite covariance matrix.

In fact, a more convenient stochastic representation is to use the following hierarchical representation (De la Cruza, 2014)

$$\mathbf{Y}|U \sim N_m(\boldsymbol{\mu}, U^{-1}\boldsymbol{\Sigma}), \quad U \sim H(u|\boldsymbol{\nu}). \tag{2.4}$$

A special distribution of the SMN class is the multivariate t distribution (Lange, Little and Taylor, 1989) that has been extensively applied in robust regressions. In the rest of this paper, we use the notation $Ga(\alpha, \beta)$ to denote a Gamma distribution with shape parameter α and rate parameter β, which has the following density

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x), \quad x > 0.$$

Then the multivariate t distribution is obtained by assuming that $U \sim Ga(\nu/2, \nu/2)$ in (2.4), where the parameter ν corresponds to the degrees of freedom of the multivariate Student-t distribution. If further letting $\nu \to \infty$, the Gaussian distribution is recovered.

3 Estimating Semiparametric Mixed-effects ODE

3.1 Bayesian Framework

Consider the dynamical process $X_i(t)$ for the i-th subject , $i = 1, ..., n$, which is defined by

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = F(X_i(t)|\eta(t), \boldsymbol{\theta}_i), \tag{3.1}$$

where t is continuous in some interval $[0, T]$, F is a known function, $\eta(t)$ is the population time-varying function, and $\boldsymbol{\theta}_i$ is a q-dimensional vector of individual ODE parameters for the i-th subject. Without loss of generality, we assume that $X_i(t)$ is one-dimensional dynamical curve in this article. We also assume that the subject-specific ODE parameters are expressed as $\boldsymbol{\theta}_i = \boldsymbol{\xi} + \mathbf{b}_i$, where $\boldsymbol{\xi}$ is a q-dimensional fixed effect and $\mathbf{b}_i$ is a q-dimensional random effect. Let $\mathbf{Y}_i = (y_{i1}, ..., y_{in_i})^T$ denote the vector of observations or measurements for the i-th subject at the observation times $\mathbf{t}_i = (t_{i1}, ..., t_{in_i})^T$ and $\mathbf{X}_i = (X_i(t_{i1}), ..., X_i(t_{in_i}))^T$ with $X_i(t)$ being the solution of the ODE (3.1) conditional on $\{X_i(0), \eta(t), \boldsymbol{\theta}_i\}$. Generally we assume that $\mathbf{Y}_i = h(\mathbf{X}_i) + \boldsymbol{\epsilon}_i$, where $h(\cdot)$ is a known function (e.g., $h(\cdot) = \log(\cdot)$ in many statistical analysis) and $\boldsymbol{\epsilon}_i$ are measurement errors. In conventional methods, a common assumption is that the random effects $\mathbf{b}_i$ and the data errors $\boldsymbol{\epsilon}_i$ both follow the multivariate normal distributions. However, as discussed in Section 1, such normality assumptions are vulnerable in the presence of atypical observations, which can seriously affect the estimation accuracy of the mixed-effects ODE model. Thus, more flexible distributions are necessary to replace the normality assumption. To this aim, we propose to use the scale mixture of multivariate normal distributions for ODE random effects $\mathbf{b}_i$ and within-subject data errors $\boldsymbol{\epsilon}_i$. In other words, we assume $\mathbf{b}_i \sim SMN_q(\mathbf{0}, \boldsymbol{\Sigma}, H_1)$ and $\boldsymbol{\epsilon}_i \sim SMN_{n_i}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n_i}, H_2)$ where $\boldsymbol{\Sigma}$ is an unknown positive definite symmetric matrix.

Applying the stochastic representation (2.4), the ODE model of interest is presented as

the following hierarchical structure

$$
\begin{aligned}
\mathbf{Y}_i|\boldsymbol{\theta}_i, U_i &\overset{ind.}{\sim} N_{n_i}(h(\mathbf{X}_i|\boldsymbol{\theta}_i, \eta(\cdot)), U_i^{-1}\sigma_\epsilon^2 \mathbf{I}_{n_i}) \,, \\
\boldsymbol{\theta}_i|W_i &\overset{ind.}{\sim} N_q(\boldsymbol{\xi}, W_i^{-1}\boldsymbol{\Sigma}) \,, \\
U_i &\overset{ind.}{\sim} H_1(\cdot|\boldsymbol{\nu}) \,, \\
W_i &\overset{ind.}{\sim} H_2(\cdot|\boldsymbol{\kappa}) \,,
\end{aligned}
\tag{3.2}
$$

where U_i and W_i are two latent variables with distributions $H_1(\cdot|\boldsymbol{\nu})$ and $H_2(\cdot|\boldsymbol{\kappa})$, respectively. One popular choice for the latent variable distribution H_1 (and H_2) is to use the gamma distributions (Azzalini and Capitanio, 2014). For example, in this article, we assume that $U_i \sim Ga(\nu/2, \nu/2)$ with the shape parameter $\nu/2$ and the rate parameter $\nu/2$, which leads to a multivariate t distribution for $\mathbf{Y}_i$. Similarly, we assume that $W_i \sim Ga(\kappa/2, \kappa/2)$ with the shape parameter $\kappa/2$ and the rate parameter $\kappa/2$, which leads to a multivariate t distribution for $\boldsymbol{\theta}_i$. When U_i and W_i have degenerate distributions, model (3.2) reduces to the conventional model with the normal distribution assumption. However, when some U_i^{-1} has a large value, it indicates that the i-th subject may have outlying observations. When some W_i^{-1} has a large value, it indicates that the i-th subject may be an atypical subject with outlying ODE parameters. This outlier detection will be demonstrated in our applications at Section 4. Hence, our proposed model (3.2) is more flexible than the conventional model with the normal distribution assumption.

We estimate the time-varying parameter $\eta(t)$ using a linear combination of spline basis functions:

$$
\eta(t) = \sum_{l=1}^{J} \phi_l(t)\zeta_l = \boldsymbol{\zeta}^T \boldsymbol{\phi}(t) \,.
\tag{3.3}
$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), ..., \phi_J(t))^T$ is a vector of basis functions with dimension J, and $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_J)^T$ is the corresponding vector of basis coefficients. In this article, we choose cubic

B-splines as basis functions for expanding $\eta(t)$ and the smoothness of $\eta(t)$ is controlled by penalizing the difference of adjacent B-splines coefficients. In the Bayesian framework, this penalization is conveniently implemented by assuming a kth-order random walk prior on the splines coefficients $\zeta_1, ..., \zeta_J$ (Lang and Brezger, 2004). The first-order random walk is defined as

$$\zeta_l = \zeta_{l-1} + e_l, \quad l \geq 2, \tag{3.4}$$

with $e_l \sim N(0, \lambda_\eta^{-1})$ and a diffuse prior $\zeta_1 \propto constant$. The second-order random walk is defined as

$$\zeta_l = 2\zeta_{l-1} - \zeta_{l-2} + e_l, \quad l \geq 3, \tag{3.5}$$

with $e_l \sim N(0, \lambda_\eta^{-1})$, and a diffuse prior $\zeta_1 \propto constant$ and $\zeta_2 \propto constant$. The smoothness of $\hat{\eta}(t)$ is controlled by the additional parameter $\lambda_\eta > 0$. In this article, we apply the second-order random walk prior (3.5). To avoid the possibility of improper posteriors of $\boldsymbol{\zeta}$, we assign weakly informative priors on (ζ_1, ζ_2): $\zeta_1 \sim N(0, 100)$ and $\zeta_2 \sim N(0, 100)$. Then, the prior of $\boldsymbol{\zeta}$ can be equivalently written in the following form

$$p(\boldsymbol{\zeta}|\lambda_\eta) \propto \lambda_\eta^{(J-2)/2} \exp\left(-\frac{1}{2}\boldsymbol{\zeta}^T \mathbf{M}(\lambda_\eta)\boldsymbol{\zeta}\right), \tag{3.6}$$

where $\mathbf{M}(\lambda_\eta) = \operatorname{diag}(0.01, 0.01, \lambda_\eta \mathbf{D}_2 \mathbf{D}_2^T)$ with

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{(J-2)\times J}. \tag{3.7}$$

The unknown parameter λ_η is assigned a hyper-prior $Ga(a_\lambda, b_\lambda)$ with the pre-specified values of the hyper-parameters (a_λ, b_λ).

In nonparametric regressions, the number of spline coefficients, J, may potentially impact

the fitted results. In the Bayesian scheme, there are two approaches to deal with the choice of J. One is fixing J, the other is treating J as an unknown variable to be estimated (for example, using the reversible jump MCMC algorithm, see Biller and Fahrmeir 2001). Lang and Brezger (2004) compared these two methods and found that the approach that fixes J performed better for functions with moderate curvature, while the approach that treats J as random performed better for highly oscillating functions. In this article, J is fixed and pre-determined in our approach [generally from 20 to 40, see Lang and Brezger (2004)]. The simulation results showed that our method is robust on the choice of J, see the Figure S5 in the Supplementary file.

Define $\mathbf{\Theta} = (\boldsymbol{\theta}_1^T, ..., \boldsymbol{\theta}_n^T)^T$ and let $\mathbf{U} = (U_1, ..., U_n)^T$ and $\mathbf{W} = (W_1, ..., W_n)^T$ be the latent variables. Then the joint likelihood is given by

$$L(\mathbf{Y}, \mathbf{\Theta}, \mathbf{U}, \mathbf{W} | \boldsymbol{\zeta}, \mathbf{\Sigma}, \boldsymbol{\xi}, \sigma_\epsilon^2, \nu, \kappa) = \prod_{i=1}^n L_i(\mathbf{Y}_i, \boldsymbol{\theta}_i, U_i, W_i | \boldsymbol{\zeta}, \mathbf{\Sigma}, \boldsymbol{\xi}, \sigma_\epsilon^2, \nu, \kappa),$$

where $L_i(\cdot|\cdot)$ is the likelihood function of the i-th subject, that is,

$$L_i(\mathbf{Y}_i, \boldsymbol{\theta}_i, U_i, W_i | \boldsymbol{\zeta}, \mathbf{\Sigma}, \boldsymbol{\xi}, \sigma_\epsilon^2, \nu, \kappa) = L_i(\mathbf{Y}_i, U_i | \boldsymbol{\theta}_i, \boldsymbol{\zeta}, \sigma_\epsilon^2, \nu) L_i(\boldsymbol{\theta}_i, W_i | \boldsymbol{\xi}, \mathbf{\Sigma}, \kappa),$$

with

$$L_i(\mathbf{Y}_i, U_i | \boldsymbol{\theta}_i, \boldsymbol{\zeta}, \sigma_\epsilon^2, \nu) = p(\mathbf{Y}_i | U_i, \boldsymbol{\theta}_i, \boldsymbol{\zeta}, \sigma_\epsilon^2) H_1(U_i | \nu),$$

and

$$L_i(\boldsymbol{\theta}_i, W_i | \boldsymbol{\xi}, \mathbf{\Sigma}, \kappa) = p(\boldsymbol{\theta}_i | \boldsymbol{\xi}, \mathbf{\Sigma}, W_i) H_2(W_i | \kappa).$$

To complete the Bayesian specification of the proposed model, the following prior distributions are assigned: $\sigma_\epsilon^{-2} \sim Ga(a_0, b_0)$, $\boldsymbol{\xi} \sim N(\boldsymbol{\xi}_0, \mathbf{\Psi}_0)$ and $\mathbf{\Sigma} \sim IW(\mathbf{S}_0, df)$, where the Gamma distribution $Ga(a_0, b_0)$ has the shape parameter a_0 and the rate parameter b_0, and the Inverse Wishart distribution $IW(\mathbf{S}_0, df)$ has the scale matrix $\mathbf{S}_0$ and degrees of freedom

df. The hyper-parameters a_0, b_0, $\boldsymbol{\xi}_0$, $\boldsymbol{\Psi}_0$, $\mathbf{S}_0$ and *df* are pre-specified. The priors for ν and κ depend on which distributions chosen for $H_1(\cdot|\nu)$ and $H_2(\cdot|\kappa)$. If both $H_1(\cdot|\nu)$ and $H_2(\cdot|\kappa)$ are chosen as gamma distributions, then the values of ν and κ must be chosen to ensure $\mathrm{E}(U^{-1}) < \infty$ and $\mathrm{E}(W^{-1}) < \infty$ which further lead to both $\mathbf{Y}_i$ and $\boldsymbol{\theta}_i$ have finite positive definite covariance matrices. For convenient conjugacy, in this article, we specify the priors of ν and κ as truncated exponential priors, $\lambda_\nu \exp(-\lambda_\nu \cdot \nu)I(\nu > 2.0)$ and $\lambda_\kappa \exp(-\lambda_\kappa \cdot \kappa)I(\kappa > 2.0)$, with pre-specified values λ_ν and λ_κ, respectively.

The joint posterior distribution of the parameters of the model conditional on the data is obtained by combining the joint likelihood and the prior distributions using the Bayes' theorem. The full conditional posterior distributions are sampled using the Monte Carlo methods which are presented in the Supplementary file.

3.2   Model Comparisons

To determine the best fitting model in a class of candidate models, in this article, we apply two common measures of model adequacy: the *deviance information criterion* (DIC; see, Spiegelhalter et al. 2002; Celeux et al. 2006) and *conditional predictive ordinate* (CPO; see, Chen, Shao and Ibrahim 2000). In this section, we briefly review the theory of DIC and CPO under a general Bayesian hierarchical framework.

The DIC has been extensively applied in Bayesian statistics, which measures the fit and the complexity of the candidate model. Suppose that $\mathbf{y} = (y_1, ..., y_n)^T$ is a sample with the probability distribution function $f(\mathbf{y}|\boldsymbol{\vartheta})$ which depends on the parameter vector $\boldsymbol{\vartheta}$. Then the deviance is given by

$$D(\boldsymbol{\vartheta}) = -2\log f(\mathbf{y}|\boldsymbol{\vartheta}) + 2\log g(\mathbf{y}),$$

where $g(\mathbf{y})$ is some fully specified normalized term. Define the effective number of parameters

$$p_D = \overline{D(\boldsymbol{\vartheta})} - D(\bar{\boldsymbol{\vartheta}}),$$

where $\overline{D(\boldsymbol{\vartheta})}$ is the posterior mean deviance which is defined as

$$\overline{D(\boldsymbol{\vartheta})} = E_{\boldsymbol{\vartheta}|\mathbf{y}}[D(\boldsymbol{\vartheta})] = E_{\boldsymbol{\vartheta}|\mathbf{y}}[-2\log f(\mathbf{y}|\boldsymbol{\vartheta})] + 2\log g(\mathbf{y}),$$

and $\bar{\boldsymbol{\vartheta}}$ is a posterior mean estimate of $\boldsymbol{\vartheta}$ depending on $\mathbf{y}$. Then the DIC statistic is defined as

$$\text{DIC} = \overline{D(\boldsymbol{\vartheta})} + p_D = 2\overline{D(\boldsymbol{\vartheta})} - D(\bar{\boldsymbol{\vartheta}}). \tag{3.8}$$

A smaller DIC value indicates a better model.

The CPO is a Bayesian diagnostic tool which is based on leave-one-out-cross-validation. The CPO can be used to detect the surprising observations. Let $\mathbf{y}_{-i} = (y_1, ..., y_{i-1}, y_{i+1}, ..., y_n)^T$ which is obtained from the sample $\mathbf{y}$ after omitting y_i. The CPO for y_i is defined as

$$\text{CPO}_i = f(y_i|\mathbf{y}_{-i}) = \frac{f(\mathbf{y})}{f(\mathbf{y}_{-i})} = \int f(y_i|\boldsymbol{\vartheta}, \mathbf{y}_{-i})p(\boldsymbol{\vartheta}|\mathbf{y}_{-i})\mathrm{d}\boldsymbol{\vartheta}, \tag{3.9}$$

which gives the predictive distribution of each data point conditional on the remainder of the data. Since that CPO_i involves a multiple integration, we estimate it based on the MCMC samples of $\boldsymbol{\vartheta}$ (Carlin and Louis, 2008). Let $\boldsymbol{\vartheta}_1, ..., \boldsymbol{\vartheta}_M$ be the posterior samples from the posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$ with the size M after the burn-in. A Monte Carlo estimate of CPO_i is given by

$$\widehat{\text{CPO}}_i = \left\{ \frac{1}{M} \sum_{\ell=1}^{M} \frac{1}{f(y_i|\boldsymbol{\vartheta}_\ell)} \right\}^{-1},$$

where $\{\boldsymbol{\vartheta}_\ell\}_{\ell=1}^{M}$ are the posterior samples of $\boldsymbol{\vartheta}$ (De la Cruza, 2014). More conveniently, a common summary statistic of CPO_i's is defined as $B = \sum_{i=1}^{n} \log(\widehat{\text{CPO}}_i)$, which is often

called the logarithm of the pseudo Bayes factor. A larger value of B indicates a better model. Furthermore,

4 Application: Gene Regulation Study

In this study, we consider the cell-cycle microarray data of S. *pombe* or fission yeast (Rustici et al., 2004) which studied three transcription factors that were involved in regulating three different groups of genes in the fission yeast cell-cycle. A single input motif (SIM) is a class of small regulatory subnetwork which consists of one transcription factor regulating a set of $i = 1, ..., n$ target genes. In this article, we only analyze one SIM, but our proposed model can be extended to multiple SIMs. Rustici et al. (2004) studied a SIM with one transcription factor, Sep1p, and 14 target genes. In addition, gene *ace2*, which is also cODEd as *Ace2p*, is known to be the target gene of *Sep1p* (Bahler, 2005). Hence, we include gene *ace2* in the SIM as another target gene. Thus, we have totally 15 target genes regulated by the same transcription factor, *Sep1p*.

The expressions of all genes are measured at 20 equally-spaced time points on $[0, 285]$. Let $\mathbf{X}_i = (X_i(t_{i1}), ..., X_i(t_{in_i}))^T$ be the expressions of the i-th gene at the common observation time points $\mathbf{t}_i = (t_{i1}, ..., t_{in_i})^T$. Let $\mathbf{Y}_i = (y_{i1}, ..., y_{in_i})^T$ be the logarithms of noisy measurements of gene expressions with $n_i = 20$. As discussed in section 1, it is doubtful that $\mathbf{Y}_i$ follows a multivariate normal distribution. Therefore, in this article, we assume that $\mathbf{Y}_i$ follows the scale mixture of multivariate normal distributions $SMN_{n_i}(\log \mathbf{X}_i, \sigma_\epsilon^2 \mathbf{I}_{n_i}, H_1)$, where H_1 is a gamma distribution with the shape parameter $\nu/2$ and rate parameter $\nu/2$. This is equivalent to assume a hierarchical structure $\mathbf{Y}_i | U_i \sim N_{n_i}(\log \mathbf{X}_i, U_i^{-1} \sigma_\epsilon^2 \mathbf{I}_{n_i})$ and $U_i \sim Ga(\nu/2, \nu/2)$.

One of often-used semiparametric gene regulation dynamical systems is given by (Khanin et al.,

2007; Rogers, Khanin and Girolami, 2007)

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = \alpha_i + \beta_i \frac{\eta(t)}{\eta(t) + \gamma_i} - \delta_i X_i(t) \, , \, i = 1, ..., n, \tag{4.1}$$

where $\eta(t)$ is the activity function of the transcriptional factor, $Sep1p$. To remove the positive constraints on the ODE parameters $(\alpha_i, \beta_i, \gamma_i, \delta_i)^T$, they are reparameterized in the logarithmic scales. Besides, the initial condition $X_i(0)$ is assumed unknown and also estimated together with the ODE parameters. Let $\boldsymbol{\theta}_i = (\log(X_i(0)), \log(\alpha_i), \log(\beta_i), \log(\gamma_i), \log(\delta_i))^T$. We assume that $\boldsymbol{\theta}_i$ follows the scale mixture of multivariate normal distribution $SMN_5(\boldsymbol{\xi}, \boldsymbol{\Sigma}, H_2)$, where $\boldsymbol{\xi} = (\log(X(0)), \log(\alpha), \log(\beta), \log(\gamma), \log(\delta))^T$ and H_2 is a gamma distribution with the shape parameter $\kappa/2$ and the rate parameter $\kappa/2$. This is equivalent to the hierarchical structure: $\boldsymbol{\theta}_i|W_i \sim N_5(\boldsymbol{\xi}, W_i^{-1}\boldsymbol{\Sigma})$ and $W_i \sim Ga(\kappa/2, \kappa/2)$.

For the identifiability problem, we rescale the transcriptional factor activity $\eta(t)$ to satisfy $\eta(0) = 1$. We use cubic B-spline basis functions with 20 equally-spaced knots on $[0, 285]$. Then $\eta(t)$ is approximated by a linear combination of B-spline basis functions:

$$\eta(t) = \exp\{\phi_0(t)\zeta_0 + \sum_{l=1}^{J} \phi_l(t)\zeta_l\} = \exp\{\phi_0(t)\zeta_0 + \boldsymbol{\zeta}^T\boldsymbol{\phi}(t)\}.$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), ..., \phi_J(t))^T$, ζ_0 and $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, ..., \zeta_J)^T$ are the corresponding basis coefficients. The property of B-spline basis functions allows that $\eta(0) = 1$ is equivalent to $\zeta_0 = 0$. We assign a prior for other basis coefficients $\zeta_1, ..., \zeta_J$: $p(\boldsymbol{\zeta}|\lambda_\eta) \propto \lambda_\eta^{(J-2)/2} \exp(-\frac{\lambda_\eta}{2}\boldsymbol{\zeta}^T\mathbf{D}_2^T\mathbf{D}_2\boldsymbol{\zeta})$ where $\mathbf{D}_2$ is given by (3.7).

We set a gamma prior $Ga(0.01, 0.01)$ for σ_ϵ^{-2}, a gamma prior $Ga(1, 0.01)$ for λ_η, an Inverse Wishart prior $IW(\boldsymbol{\Omega}_0, f_0)$ for $\boldsymbol{\Sigma}$, and a multivariate normal prior $N_5(\boldsymbol{\xi}_0, \boldsymbol{\Psi}_0)$ for $\boldsymbol{\xi}$. Moreover, we chose the following values for the hyper-parameters: $\boldsymbol{\xi}_0 = (0, 0, 0, 0, 0)^T$, $\boldsymbol{\Omega}_0 = \mathrm{diag}(10, 10, 10, 10, 10)$, $\boldsymbol{\Psi}_0 = \mathrm{diag}(1000, 1000, 1000, 1000, 1000)$, and $f_0 = 6$. For the hyper priors of ν and κ, we choose $\frac{1}{3}\exp(-\frac{\nu}{3})I(\nu > 2.0)$ and $\frac{1}{3}\exp(-\frac{\kappa}{3})I(\kappa > 2.0)$.

14

We compare our proposed SMN model which assumes both the individual ODE parameters and the observations errors to follow the scale mixture of multivariate normal distribution with the conventional Normal model which assumes both the individual ODE parameters and the observations errors to follow the normal distributions. To improve mixing convergence of the posterior samples, the adaptive proposals are applied in the Metropolis-Hastings algorithm. The details are provided in Section S1 of the supplementary document. For each of SMN model and Normal model, the proposed MCMC algorithm is run for 150,000 iterations. With the 'burn-in' of the first 50,000 iterations, the samples of $(\alpha, \beta, \gamma, \delta)^T$ are chosen with a lag of size 20 from the rest of Markov chains under each of the models. The convergence of the MCMC chains was monitored using trace plots and Gelman–Rubin convergence diagnostic ($\hat{R}$). We report the results of the convergence diagnostic results in Table S1 of the Supplementary document. Gelman and Rubin (1992) suggested that diagnostic $\hat{R}$ values greater than 1.2 for any of the model parameters should indicate non-convergence. Here, we use a rule of $\hat{R} < 1.1$ to declare convergence of the Markov chains. All these diagnostics indicate a good convergence of the Markov chains. We calculate the values of the DIC using the formula (3.8) and the pseudo Bayes factor $B = \sum_{i=1}^n \log(\widehat{\text{CPO}}_i)$ based on the formula (3.9). The results are: for the SMN model, DIC $= -2186.90$ and $B = 174.16$, whereas for the Normal model, DIC $= -1502.30$ and $B = 146.04$. Obviously, our proposed model using the SMN model has a smaller value of DIC and a large value of the logarithm of the pseudo Bayes factor $B = \sum_{i=1}^n \log(\widehat{\text{CPO}}_i)$ than the conventional model using the normal distribution, which indicates that our proposed model is better. The values of CPO_i for each individual are calculated and reported in Table S2 of the Supplementary document. Obviously, the overall CPO is driven by a few overly influential values under the multivariate normal model, e.g., the 12th gene. In contrast, the influence of those points is greatly attenuated by assuming a scaled mixture of normal distributions. As the requirement of a referee, we also calculate CPO_i using the bridge sampling method (Meng and Wang, 1996) and the

15

results are: for the SMN model, $B = 702.24$, whereas for the Normal model, $B = 110.57$, which further demonstrated that the SMN model is better than the Normal model.

Table 1 shows the posterior means, the standard errors and the corresponding 95% equal-tail credible intervals for the fixed-effects $(\alpha, \beta, \gamma, \delta)^T$ in our proposed model. Table 2 displays the estimated latent parameters in the gene regulation mixed-effects ODE model (4.1) under the assumption that the individual ODE parameters and observations errors follow the SMN distributions. All of $\widehat{W}_i^{-1}$ and most of $\widehat{U}_i^{-1}$ are around 1.0, but Gene 12 have very large values of $\widehat{U}_i^{-1}$. This implies that Gene 12 may have abnormal observations. The left panel of Figure 2 displays the measured gene expressions and the estimated $X_i(t)$ for Gene 12. As expected, the estimate of gene expression under the Normal model is much affected by the abnormal observations. In contrast, the estimate of gene expression under the SMN model is robust against the abnormal observations. Furthermore, from Figure 1(a), it seems that the residuals present a feature of skewness. We performed a mixed-effects ODE model by assuming that the ODE parameters and observations errors follow the scale mixture of multivariate skew-normal distributions; however, it is found that the skewness is not significant. Finally, the estimated transcriptional factor activity is displayed in the right panel of Figure 2. It shows that the transcriptional factor activity has two periodic cycles, which coincided with the results of Rogers, Khanin and Girolami (2007).

5   Simulation Studies

In this section, we implement four simulation studies to evaluate the finite sample performance of our proposed semiparametric mixed-effects ODE models.

5.1   Scenario I

As stated in the Section 1, the semiparametric ordinary differential equations has valuable applications in the gene transcription experiments. To mimic the transcription activity in gene

transcription experiments, we consider the following semiparametric ODE (Barenco et al. 2006):

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = \alpha_i + \beta_i\eta(t) - \delta_iX_i(t), \quad t \in [0, 8], \tag{5.1}$$

where the changing level of a gene i's expression, $X_i(t)$, is given by a combination of a basal transcription rate, α_i, a sensitivity, β_i, to its governing TF's activity, $\eta(t)$, and the decay rate of the mRNA, δ_i. We generate the logarithm of the individual ODE parameters $\boldsymbol{\theta}_i = \log\{(\alpha_i, \beta_i, \delta_i)^T\}$ from a multivariate normal distribution $N(\boldsymbol{\xi}, \boldsymbol{\Sigma})$ with the mean vector $\boldsymbol{\xi}$ and the covariance matrix $\boldsymbol{\Sigma} = 0.04\cdot\mathbf{R}$, where the true values of $\boldsymbol{\xi}$ is chosen as $\boldsymbol{\xi} = \log\{(\alpha, \beta, \delta)^T\}$ with $\alpha = 1.2$, $\beta = 3.5$ and $\delta = 1.0$, and the correlation matrix $\mathbf{R} = (r_{ij})_{3\times3}$ with $r_{12} = 0.45$, $r_{13} = -0.25$ and $r_{23} = -0.45$. The activity function of the transcriptional factor is given by a mixture of three Gaussian-like bumps, $\eta(t) = \exp[-(t - 2)^2/1.44] + \exp[-(t - 5)^2] + 0.5\exp[-(t-6)^2/0.64] - 0.062$, which was considered in Chapter 9 of Lawrence et al. (2010) with a similar formula.

The true initial condition $X_i(0)$ is generated as $X_i(0) \sim N(1.0, 0.04)$. Then, our simulated data are generated as $Y_i(t_{ij}) = X_i(t_{ij}) + \epsilon_{ij}$, where $X_i(t_{ij})$ is the numerical solution of ODE (5.1) via the fourth-order Runge-Kutta algorithm evaluated at 201 equally-spaced time points on $[0, 8]$, and ϵ_{ij} are the measurement errors which are independently generated from a normal distribution $N(0, 0.25)$. We then estimate the mixed-effects ODE (5.1) by assuming the ODE parameter $\boldsymbol{\theta}_i$ and the data error ϵ_{ij} follow the SMN distributions. This proposed model is also compared with the conventional model which assumes $\boldsymbol{\theta}_i$ and ϵ_{ij} follow the normal distribution. With the 'burn-in' of the first 10,000 iterations, we obtain 1,000 posterior samples from the rest of the Markov chains with a lag of size 10. The above procedure is repeated for 100 simulation replicates.

To evaluate the performance of the estimate for the time-varying ODE parameter $\eta(t)$,

we define the following mean absolute deviation error (MADE)

$$\text{MADE}(\hat{\eta}) = \frac{1}{MN} \sum_{\ell=1}^{M} \sum_{j=1}^{N} |\hat{\eta}^{(\ell)}(t_j) - \eta(t_j)|,$$

where $t_j, j = 1, ..., N$, are $N = 201$ equally-spaced time points in $[0, 8]$, and $\hat{\eta}^{(\ell)}(t)$ is the estimate of $\eta(t)$ in the ℓ-th simulation replicate for $\ell = 1, ..., M$. Table S3 of the Supplement file shows the posterior means, standard deviations as well as the mean absolute deviation errors for the fixed effect $(\alpha, \beta, \delta)^T$. In Table S3, for α and $\eta(\cdot)$, our proposed model based on the SMN distribution has a little larger MADE values but has smaller MADE values for β and δ in comparison with the conventional model using the normal distribution for $n = 50$ and 100. We think that our proposed model using the SMN distribution is compatible with the conventional model using the normal distribution.

5.2 Scenario II

To study the robustness of the proposed method, we generate the logarithm of the individual ODE parameters $\boldsymbol{\theta}_i = \log\{(\alpha_i, \beta_i, \delta_i)^T\}$ from a multivariate Student-t distribution $t_\kappa(\boldsymbol{\xi}, \boldsymbol{\Sigma})$ with the location parameter $\boldsymbol{\xi}$, shape matrix $\boldsymbol{\Sigma} = 0.04 \cdot \mathbf{R}$ which are the same as in Scenario I and degrees of freedom $\kappa = 3$. The time-varying fixed-effect function $\eta(\cdot)$ is also the same as in Scenario I.

The true initial condition $X_i(0)$ is generated from the $X_i(0) = 1.0 + 0.1b_{i0}$ where b_{i0} follows a Student-t distribution, $t(3)$, with degree of freedom 3. Then, our simulated data are generated as $Y_i(t_{ij}) = X_i(t_{ij}) + \epsilon_{ij}$, where $X_i(t_{ij})$ is the numerical solution of ODE (5.1) via the fourth-order Runge-Kutta algorithm evaluated at 201 equally-spaced time points on $[0, 8]$, and ϵ_{ij} are the measurement errors. We consider two cases for generating the measurement errors: (i) $\epsilon_{ij} \sim 0.3 \times t(3)$; (ii) $\epsilon_{ij} \sim 0.6 \times t(3)$. With the 'burn-in' of the first 10,000 samples, we obtain 1,000 posterior samples from the rest of the Markov chain

with a lag of size 10. Table 3 displays the simulation results for the estimates of fixed effect $(\alpha, \beta, \delta)^T$ and time-varying function $\eta(t)$ based on 100 simulation replicates. It shows that our proposed model using the SMN distribution is preferred than the conventional model using the normal distribution.

5.3 Scenario III

To study the robustness of the proposed method for mis-specified ODEs model, we consider the following two experiment designs:

III-(a): Suppose that the true ODEs are given by

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = -\delta_i X_i(t) + \frac{D_i \delta_i K a_i}{Cl_i(\delta_i - Ka_i)} \exp(-Ka_i t) + \beta_i \frac{(t-5)^2 + 119}{144}, \quad t \in [0, 12], \quad (5.2)$$

where $D_i = 400, 1 \leq i \leq 50$, the logarithms of the individual ODE parameters $\{Ka_i, \beta_i, \delta_i, Cl_i\}$ follow a multivariate normal distribution $N(\boldsymbol{\xi}, \boldsymbol{\Sigma})$ with $\boldsymbol{\xi} = \log\{(1.0, 0.5, 0.3, 21.0)^T\}$ and $\boldsymbol{\Sigma} = \mathrm{diag}\{0.1, 0.1, 0.1, 0.25\}$. The true initial condition $X_i(0)$ is generated from a normal distribution $N(1.0, 0.01)$. Then, our simulated data are generated as $Y_i(t_{ij}) = X_i(t_{ij}) + \epsilon_{ij}$, where $X_i(t_{ij})$ is the numerical solution of ODE (5.2) via the fourth-order Runge-Kutta algorithm evaluated at 201 equally-spaced time points on $[0, 12]$, and the measurement errors $\epsilon_{ij} \sim N(0, 0.25)$.

III-(b): Suppose that the true ODEs are given by

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = -\delta_i X_i(t) + \frac{D_i \delta_i K a_i}{Cl_i(\delta_i - Ka_i)} \exp(-Ka_i t), \quad t \in [0, 12], \quad (5.3)$$

where the values of individual parameters are taken the same as in Scenario III-(a).

To implement the traditional method which is based on the normal distributions assump-

19

tions, the working ODEs model is specified by

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = -\delta_i X_i(t) + \frac{D_i \delta_i K a_i}{Cl_i(\delta_i - K a_i)} \exp(-K a_i t), \tag{5.4}$$

which is a mis-specified model in Scenario of III-(a) and is a correct model in Scenario of III-(b). For our proposed method, the working ODEs model is specified by

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = -\delta_i X_i(t) + \frac{D_i \delta_i K a_i}{Cl_i(\delta_i - K a_i)} \exp(-K a_i t) + \beta_i \eta(t), \tag{5.5}$$

We use the following mean integrated square error (MISE) criterion:

$$\text{MISE} = \frac{1}{50} \sum_{i=1}^{50} \int_0^{12} [\hat{X}_i(t) - X_i(t)]^2 \mathrm{d}t,$$

to evaluate the performances of our proposed method and the traditional method. The results are displayed in Figure 3, which show that our proposed method is robust on the mis-specified ODEs model.

6   Conclusions and discussions

Ordinary differential equation models are elegant and popular models for describing the mechanism of complex dynamical systems. In this paper, we propose a semiparametric mix-effects ODE model that contains both constant and time-varying parameters and considers the within-subject and between-subject variations simultaneously. We propose to use a class of scale mixture of multivariate normal distributions to model the random effects of ODE parameters and measurement errors in the data to obtain a robust estimation for the ODE parameters when the abnormal subjects and the abnormal measurement errors exist in the data.

Our proposed model can be framed in a Bayesian hierarchical model in which two latent variables are introduced to identify abnormal individuals and abnormal measurement errors. As one reviewer points out, one interesting feature of the proposed hierarchical model is the potential interpretability of the latent parameters as indicators of abnormal individuals. Another benefit of the proposed hierarchical model is feasibly performing an MCMC method to estimate the ODE parameters and latent variables. To improve the mixing convergence of the algorithm, an adaptive proposal was applied in the MH step. Finally, our proposed method is demonstrated by studying a genetic regulation experiment in which the regulatory activity of the transcriptional factor is modeled by a varying function over time. We show that our proposed model using the scale mixture of multivariate normal distribution is preferred in comparison with the conventional model using the normal distribution. Our simulation studies also show that our proposed model can obtain a more robust estimation for ODE parameters when using the scale mixture of multivariate normal distributions. As an example of applications, a class of often-used gene regulatory dynamic systems was studied in real data analysis, however, our proposed approach is very flexible and can be applied to provide robust inference results for other semiparametric mixed-effects dynamic systems.

Furthermore, the ODEs are generally designed based on the experts knowledge or the prior research. Our simulation results showed that there will have serious bias when applying a mis-specified ODEs model. A conservative approach is to consider a flexible semiparametric ODEs model:

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = F(X_i(t); \boldsymbol{\theta}_i) + r_i(t), \tag{6.1}$$

where the parameters $\boldsymbol{\theta}_i$ belong to a parameter space $\boldsymbol{\Theta}$, the function F is known based on the experts knowledge, and $r_i(t)$ describes the uncertainty of the ODEs. In the simulation studies, we have investigated the case of $r_i(t) = \alpha_i + \beta_i \eta(t)$ and demonstrated the efficiency of our proposed method. We will go on to study this issue in the future. It is possible that the

heavy tailed residuals arise from stochastic transcriptional factor activity rather than how it is currently modelled as a smooth deterministic trend. We will study how to distinguish them in our future study.

Acknowledgements

References

Andrews, D. F., and Mallows, C. L. (1974), 'Scale Mixtures of Normal Distributions', *Journal of the Royal Statistics Society*, Ser. B, 36, 99-102.

Azzalini, A., and Capitanio, A. (2014), *The Skew-Normal and Related Families*, Chapman and Hall, London.

Bahler, J. (2005), 'Cell-Cycle Control of Gene Expression in Budding and Fission Yeast', *Annual Review of Genetics*, 39, 69-94.

Barenco, M., Tomescu, D., Brewer, D., Callard, J., Stark, R., and Hubank, M. (2006), 'Ranked prediction of p53 targets using hidden variable dynamic modeling', *Genome Biology*, 7, pg. R25.

Bhaumik, P., and Ghosal, S. (2015), 'Bayesian Two-step Estimation in Differential Equation Models', *Electronic Journal of Statistics*, 9, 3124-3154.

Biller, C. and Fahrmeir, L. (2001), 'Bayesian varying-coefficient models using adaptive regression splines', *Statistical Modelling*, 1, 195-211.

Brunel, N. J., Clairon, Q., and d'Alché-Buc, F. (2014), 'Parametric Estimation of Ordinary Differential Equations With Orthogonality Conditions', *Journal of the American Statistical Association*, 109, 173-185.

Campbell, D., and Steele, R. J. (2012), 'Smooth Functional Tempering for Nonlinear Differential Equation Models', *Statistics and Computing*, 22, 429-443.

Cao, J., Fussmann, G. F., and Ramsay, J. O. (2008), 'Estimating a Predator-Prey Dynamical Model With the Parameter Cascades Method', *Biometrics*, 64, 959-967.

Carlin, B. P., and Louis, T. A. (2008), *Bayesian Methods for Data Analysis* (3rd ed.), Chapman/Hall, London.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D.M. (2006), 'Deviance Information Criteria for Missing Data Models', *Bayesian Analysis*, 1, 651-674.

Chen, M-H., Shao, Q-M., and Ibrahim, J. G. (2000), *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag Inc., New York.

Choy, S. T. B., and Smith, A. F. M. (1997), 'Hierarchical Models With Scale Mixtures of Normal Distributions', *Test*, 6, 205-221.

De la Cruza, R. (2014), 'Bayesian Analysis for Nonlinear mixed-effects Models Under Heavy-Tailed Distributions', *Pharmaceutical Statistics*, 13, 81-93.

Dass, S. C., Lee, J., Lee, K., and Park, J.(2017), 'Laplace Based Approximate Posterior Inference for Differential Equation Models', *Statistics and Computing*, 27, 679-698.

Fang, Y., Wu, H., and Zhu, L. X. (2011), 'A Two-Stage Estimation Method for Random-Coefficient Differential Equation Models With Application to Longitudinal HIV Dynamic Data', *Statistica Sinica*, 21, 1145-1170.

Gelman, A. and Rubin, D. B. (1992), 'Inference from Iterative Simulation using Multiple Sequences', *Statistical Science*, 7, 457-511.

Guedj, J., Thiébaut, R., and Commenges, D. (2007), 'Maximum Likelihood Estimation in Dynamical Models of HIV', *Biometrics*, 63, 1198-1206.

Hall, P., and Ma, Y. (2014), 'Quick and Easy Kernel Based One-Step Estimation of Parameters in Differential Equations', *Journal of the Royal Statistical Society*, Ser. B, 76, 735-748.

Huang, Y., and Wu, H. (2006), 'A Bayesian Approach for Estimating Antiviral Efficacy in Hiv Dynamic Models', *Journal of Applied Statistics*, 33, 155-174.

Khanin, R., Vinciotti, V., Mersinias, M., Smith, C., and Wit, E. (2007), 'Statistical Reconstruction of Transcription Factor Activity Using Michaelis-Menten Kinetics', *Biometrics*, 63, 816-823.

Lang, S., and Brezger, A. (2004), 'Bayesian P-splines', *Journal of Computation and Graphical Statistics*, 131, 183-212.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989), 'Robust Statistical Modeling Using the t Distribution', *Journal of the American Statistical Association*, 84, 881-896.

Lange, K., and Sinsheimer, J. (1993), 'Normal/Independent Distributions and Their Applications in Robust Regression', *Journal of Computational and Graphical Statistics*, 2, 175-198.

Lawrence, N. D., Girolami, M., Rattray, M., and Sanguinetti, G. (2000), *Learning and Inference in Computational Systems Biology*, Cambridge, MA, The MIT Press.

Liang, H., and Wu, H. (2008), 'Parameter Estimation for Differential Equation Models Using A Framework of Measurement Error in Regression', *Journal of the American Statistical Association*, 103, 1570-1583.

Liu, C. (1996), 'Bayesian Robust Multivariate Linear Regression With Incomplete Data', *Journal of the American Statistical Association*, 91, 1219-1227.

Meng, X.-L., and Wong, W. H. (1996), 'Simulating ratios of normalizing constants via a simple identity: A theoretical exploration', *Statistica Sinica*, 6, 831-886.

Meza, C., Osorio, F., and De la Cruz, R. (2012), 'Estimation in Nonlinear mixed-effects Models Using Heavy-Tailed Distributions', *Statistics and Computing*, 22, 121-139.

Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007), 'Parameter Estimation for Differential Equations: A Generalized Smoothing Approach (with discussion)', *Journal of the Royal Statistical Society*, Ser. B, 69, 741-796.

Rogers, S., Khanin, R., and Girolami, M. (2007), 'Bayesian Model-Based Inference of Transcription Factor Activity', *BMC Bioinformatics*, 8, S2.

Rosa, G. J. M., Gianola, D., and Padovani, C. R. (2004), 'Bayesian Longitudinal Data Analysis With Mixed Models and Thick-Tailed Distributions Using MCMC', *Journal of Applied Statistics*, 31, 855-873.

Rustici, G., Mata, J., Kivinen, K., Lio, P., Penkett, C. J., Burns, G., Hayles, J., Brazma, A., Nurse, P., and Bahler, J. (2004), 'Periodic Gene Expression Program of The Fission Yeast Cell Cycle', *Nature Genetics*, 36, 809-817.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), 'Bayesian Measures of Model Complexity and Fit', *Journal of the Royal Statistics Society*, Ser. B, 64, 583-639.

Wang, L., Cao, J., Ramsay, J. O., Burger, D., Laporte, C., and Rockstrohk, J. (2014), 'Estimating mixed-effects Differential Equation Models', *Statistics and Computing*, 24, 111-121.

Table 1: The posterior means, standard deviations (SD), and 95% credible intervals of the population ODE parameters $(\alpha, \beta, \gamma, \delta)^T$ in the gene regulation mixed-effects ODE model (4.1) under the SMN model and Normal model, respectively.

Model	Parameters	Mean	SD	95% C.I.
SMN model	α	0.012	0.006	(0.004, 0.026)
	β	0.207	0.080	(0.108, 0.441)
	γ	2.656	1.222	(1.123, 5.791)
	δ	0.050	0.011	(0.031, 0.075)
Normal model	α	0.012	0.007	(0.003, 0.028)
	β	0.221	0.081	(0.108, 0.418)
	γ	2.702	1.265	(1.095, 5.828)
	δ	0.053	0.013	(0.032, 0.082)

Table 2: The estimated latent parameters in the gene regulation mixed-effects ODE model (4.1) under the SMN model which assumes that the ODE parameters and observations errors follow the scale mixture of multivariate normal distributions.

Subject	1	2	3	4	5
Residual errors $(\widehat{U}_i^{-1})$	1.02	0.92	2.11	0.75	0.72
Random effects $(\widehat{W}_i^{-1})$	0.95	1.09	0.94	0.93	0.95
Subject	6	7	8	9	10
Residual errors $(\widehat{U}_i^{-1})$	0.75	0.99	2.22	1.53	0.84
Random effects $(\widehat{W}_i^{-1})$	0.94	0.95	0.97	1.28	0.99
Subject	11	12	13	14	15
Residual errors $(\widehat{U}_i^{-1})$	0.60	**8.05**	0.74	1.54	1.14
Random effects $(\widehat{W}_i^{-1})$	0.96	1.16	0.96	1.05	1.18
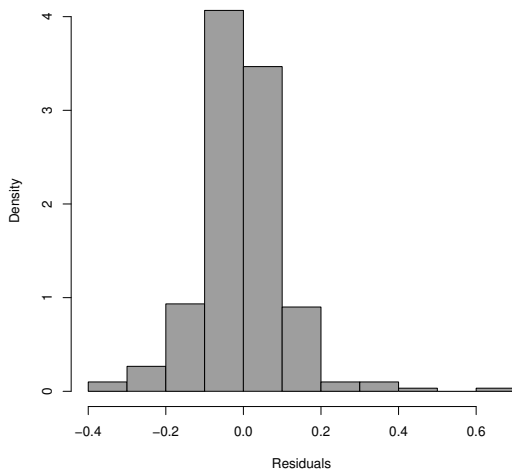
(a) Histgram
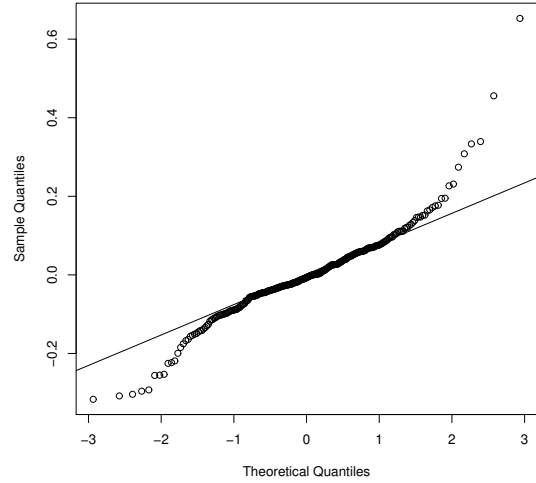
(b) Normal Q-Q plot

Figure 1: The histogram and the normal Q-Q plot of the obtained residuals assuming normal distributions for observations and random-effects in the gene regulation ODE model (4.1) with the fisson yeast cells.

Table 3: The bias, standard deviation(SD) and mean absolute deviation error (MADE) of estimates for the fixed effects of the mixed-effects ODE model (5.1) under Scenario II based on 100 simulation replicates. The true values of fixed effects $\alpha = 1.2$, $\beta = 3.5$ and $\delta = 1.0$.

| Case | n | Fixed-effects | Distribution assumptions | | | | | |
| | | | SMN distributions | | | Normal distributions | | |
			Bias	SD	MADE	Bias	SD	MADE
(i)	50	α	0.155	0.150	0.214	0.222	0.156	0.272
		β	-0.261	0.269	0.374	-0.420	0.261	0.493
		δ	-0.244	0.182	0.303	-0.370	0.146	0.397
		$\eta(t)$	N/A	N/A	0.384	N/A	N/A	0.497
	100	α	0.139	0.149	0.202	0.263	0.143	0.311
		β	-0.207	0.235	0.311	-0.434	0.197	0.476
		δ	-0.217	0.174	0.277	-0.388	0.090	0.397
		$\eta(t)$	N/A	N/A	0.364	N/A	N/A	0.481
(ii)	50	α	0.274	0.188	0.331	0.314	0.229	0.388
		β	-0.206	0.268	0.336	-0.216	0.305	0.372
		δ	-0.257	0.123	0.282	-0.250	0.150	0.289
		$\eta(t)$	N/A	N/A	0.492	N/A	N/A	0.517
	100	α	0.207	0.148	0.252	0.307	0.162	0.346
		β	-0.275	0.199	0.339	-0.367	0.215	0.424
		δ	-0.28	0.120	0.303	-0.359	0.097	0.371
		$\eta(t)$	N/A	N/A	0.421	N/A	N/A	0.491

(a)

(b)
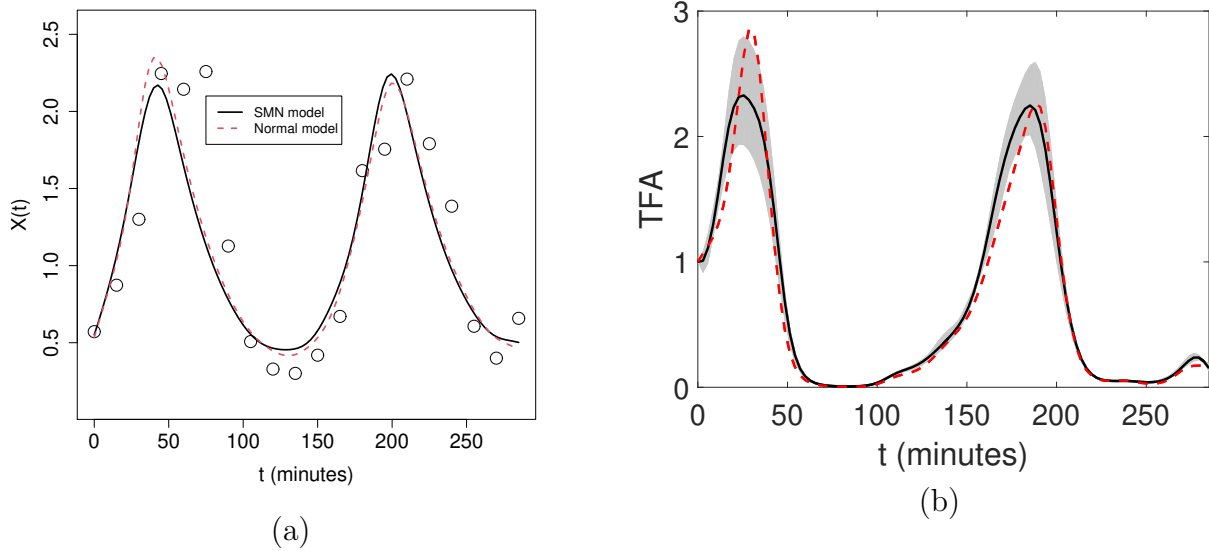
Figure 2: Example of inference applying the SMN model (solid lines) and Normal model (slashed lines) with real microarray data from fission yeast. (a) The observations of gene expressions (circles) and the numerical solutions of the gene regulation mixed-effects ODE model (4.1) for Gene 12; (b) The posterior mean estimates of $\eta(t)$; the shaded area denotes the 95% credible intervals of the TFA $\eta(t)$ under the SMN model.
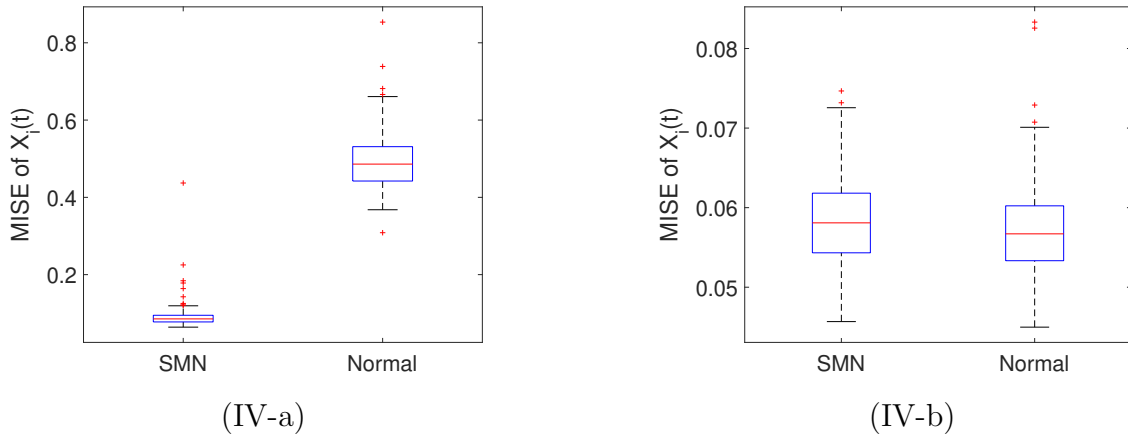

(IV-a)

(IV-b)

Figure 3: The MISEs for $X_i(t)$ in the Scenario III by performing the SMN model and Normal model,respectively.