Web-based Supplementary Materials for "Semiparametric Mixed-effects Ordinary Differential Equation Models with Heavy-tailed Distributions"

S1 The MCMC algorithm

We use the Markov chain Monte Carlo (MCMC) method which consists of the Metropolis-Hastings algorithm and the Gibbs sampling method to sample the parameters $\boldsymbol{\theta}_i$, $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$, $\boldsymbol{\Sigma}$, σ_ϵ^{-2}, U_i, W_i, λ_η, ν and κ. In this appendix, the symbol $\|\mathbf{a}\|^2_{\mathbf{A}}$ denotes $\mathbf{a}^T\mathbf{A}\mathbf{a}$ for the vector $\mathbf{a}$ and the matrix $\mathbf{A}$. When $\mathbf{A} = \mathbf{I}$, a symbol $\|\mathbf{a}\|^2$ is used instead. Define $\mathbf{X}_i = (X_i(t_{i1}), ..., X_i(t_{in_i}))^T, i = 1, ..., n$. The full conditional distributions for $\boldsymbol{\theta}_i$, $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$, $\boldsymbol{\Sigma}$, σ_ϵ^{-2}, U_i, W_i, λ_η, ν and κ are displayed as follows (where $\sim$ denotes all variables except the one to be sampled):
(a) Full conditional distributions of $\boldsymbol{\theta}_i$ for $i = 1, ..., n$.

$$p(\boldsymbol{\theta}_i| \sim) \propto \exp\left\{-\frac{U_i}{2\sigma_\epsilon^2}\|\mathbf{Y}_i - \mathbf{X}_i\|^2\right\} \exp\left\{-\frac{W_i}{2}\|\boldsymbol{\theta}_i - \boldsymbol{\xi}\|^2_{\boldsymbol{\Sigma}^{-1}}\right\}.$$

Since that the solution $X_i(t)$ generally does not have an explicit expression, the full conditional distribution for $\boldsymbol{\theta}_i$ does not have closed form. We apply the Metropolis-Hastings method to sample $\boldsymbol{\theta}_i$. Generally, the initial conditions $X_i(0)$'s are rarely known. We incorporate them into $\boldsymbol{\theta}_i$ to be estimated from the noisy data. It is well known that the convergence efficiency of the traditional Metropolis-Hastings algorithm can be improved if the proposals are adapted properly. Let $\boldsymbol{\theta}_i^{(k)}$ be the sample of $\boldsymbol{\theta}_i$ at the kth iteration. Then at iteration $(k+1)$, the proposal distribution used in the MH algorithm is $N\left(\boldsymbol{\theta}_i^{(k)}, s_d[\mathbf{C}_i^{(k)} + \varrho\mathbf{I}_{p+1}]\right)$, where $\mathbf{C}_i^{(k)}$ is updated by (Liang, Liu and Carroll, 2010)

$$\mathbf{C}_i^{(k+1)} = \mathbf{C}_i^{(k)} + \tilde{\gamma}_{k+1}[(\boldsymbol{\theta}_i^{(k+1)} - \tilde{\boldsymbol{\mu}}_i^{(k)})(\boldsymbol{\theta}_i^{(k+1)} - \tilde{\boldsymbol{\mu}}_i^{(k)})^T - \mathbf{C}_i^{(k)}],$$

and $\tilde{\boldsymbol{\mu}}_i^{(k)}$ is updated by

$$\tilde{\boldsymbol{\mu}}_i^{(k+1)} = \tilde{\boldsymbol{\mu}}_i^{(k)} + \tilde{\gamma}_{k+1}(\boldsymbol{\theta}_i^{(k+1)} - \tilde{\boldsymbol{\mu}}_i^{(k)}).$$

The gain factor sequence $\{\tilde{\gamma}_k\}$ satisfy the conditions of $\sum_{k=1}^{\infty} \tilde{\gamma}_k = \infty$ and $\sum_{k=1}^{\infty} \tilde{\gamma}_k^{1+\delta_0} < \infty$ for some $\delta_0 \in (0, 1]$. s_d is used for tuning the acceptance rate and a very small ϱ is chosen to avoid the singularity of the covariance matrix.
(b) Full conditional distribution of $\boldsymbol{\zeta}$.

$$p(\boldsymbol{\zeta}|\sim) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^n U_i\|\mathbf{Y}_i - \mathbf{X}_i\|^2\right\}\exp\left\{-\frac{\lambda_\eta}{2}\|\mathbf{D}_2\boldsymbol{\zeta}\|^2\right\},$$

which is not the standard distribution. It is sampled by the Metropolis-Hastings algorithm.
(c) Full conditional distributions of $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$.

$$\begin{aligned} p(\boldsymbol{\xi}|\sim) &\propto \prod_{i=1}^n \exp\left\{-\frac{W_i}{2}\|\boldsymbol{\theta}_i - \boldsymbol{\xi}\|^2_{\boldsymbol{\Sigma}^{-1}}\right\}\exp\left\{-\frac{1}{2}\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2_{\boldsymbol{\Omega}_0}\right\}, \\ p(\boldsymbol{\Sigma}|\sim) &\propto |\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^n W_i\|\boldsymbol{\theta}_i - \boldsymbol{\xi}\|^2_{\boldsymbol{\Sigma}^{-1}}\right\}|\boldsymbol{\Sigma}|^{-(df+q+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{S}_0\boldsymbol{\Sigma}^{-1})\right\}. \end{aligned}$$

Then the full conditional posterior distribution of $\boldsymbol{\xi}$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{\boldsymbol{\xi}} = \mathbf{B}(\sum_{i=1}^n W_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_i + \boldsymbol{\Omega}_0\boldsymbol{\xi}_0)$ and covariance matrix $\mathbf{B} = (\sum_{i=1}^n W_i\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}_0)^{-1}$. The full conditional posterior distribution of $\boldsymbol{\Sigma}$ is an Inverse Wishart distribution with the scale matrix $\mathbf{S}_0 + \sum_{i=1}^n W_i\|\boldsymbol{\theta}_i - \boldsymbol{\xi}\|^2$ and degrees of freedom $n + q + 2$.
(d) Full conditional distributions of U_i and W_i.

$$\begin{aligned} p(U_i|\sim) &\propto H_1(U_i|\nu)U_i^{n_i/2}\exp\left\{-\frac{U_i}{2\sigma_\epsilon^2}\|\mathbf{Y}_i - \mathbf{X}_i\|^2\right\}, \\ p(W_i|\sim) &\propto H_2(W_i|\kappa)W_i^{q/2}\exp\left\{-\frac{W_i}{2}\|\boldsymbol{\theta}_i - \boldsymbol{\xi}\|^2_{\boldsymbol{\Sigma}^{-1}}\right\}. \end{aligned}$$

Assuming that $U_i \sim Ga(\nu/2, \nu/2)$, then the full conditional posterior distribution of U_i is still a Gamma distribution with shape parameter $n_i/2 + \nu/2$ and rate parameter $\nu/2 + \frac{1}{2\sigma_\epsilon^2}\|\mathbf{Y}_i - \mathbf{X}_i\|^2$. Similarly, the full conditional posterior distribution of W_i is a Gamma distribution with shape parameter $\kappa/2 + q/2$ and rate parameter $\kappa/2 + \frac{1}{2}\|\boldsymbol{\theta}_i - \boldsymbol{\xi}\|^2_{\boldsymbol{\Sigma}^{-1}}$.

(e) Full conditional distributions of ν and κ.

$$
\begin{aligned}
p(\nu| \sim) &\propto p(\nu)\prod_{i=1}^{n} H_1(U_i|\nu), \\
p(\kappa| \sim) &\propto p(\kappa)\prod_{i=1}^{n} H_2(W_i|\kappa).
\end{aligned}
$$

Assuming that $U_i \sim Ga(\nu/2, \nu/2)$ and a truncated exponential prior $\lambda_\nu \exp(-\lambda_\nu \cdot \nu)I(\nu > 2.0)$ is assigned on ν, then the full conditional posterior distribution of ν is proportional to $\left\{(\nu/2)^{\nu/2}/\Gamma(\nu/2)\right\}^n \prod_{i=1}^n U_i^{\nu/2-1} \exp(-\nu U_i/2) \exp(-\lambda_\nu \cdot \nu)I(\nu > 2.0)$. This is not a standard distribution; however, we can apply the Metropolis-Hastings algorithm to sample it. In the same way, under the assumption of $W_i \sim Ga(\kappa/2, \kappa/2)$ and the prior $p(\kappa) \propto \exp(-\lambda_\kappa \cdot \kappa)I(\kappa > 2.0)$, the full conditional posterior distribution of κ is given by

$$
p(\kappa| \sim) \propto \left\{(\kappa/2)^{\kappa/2}/\Gamma(\kappa/2)\right\}^n \prod_{i=1}^{n} W_i^{\kappa/2-1} \exp(-\kappa W_i/2) \exp(-\lambda_\kappa \cdot \kappa) I(\kappa > 2.0),
$$

which is also sampled by the Metropolis-Hastings algorithm.
(f) Sample σ_ϵ^{-2}.

$$
p(\sigma_\epsilon^{-2}| \sim) \propto p(\sigma_\epsilon^{-2})(\sigma_\epsilon^{-2})^{N/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n} U_i \|\mathbf{Y}_i - \mathbf{X}_i\|^2\right\}.
$$

Assuming that σ_ϵ^{-2} has a Gamma prior $Ga(a_0, b_0)$, then the full conditional posterior distribution of σ_ϵ^{-2} is a Gamma distribution with shape parameter $a_0 + N/2$ and rate parameter $b_0 + \frac{1}{2}\sum_{i=1}^n U_i \|\mathbf{Y}_i - \mathbf{X}_i\|^2$ where $N = \sum_{i=1}^n n_i$.
(g) Sample λ_η.

$$
p(\lambda_\eta|\cdot) \propto p(\lambda_\eta)\lambda_\eta^{(J-2)/2} \exp\{-\frac{\lambda_\eta}{2}\|\mathbf{D}_2\boldsymbol{\zeta}\|^2\}.
$$

Assuming that λ_η has a Gamma prior $Ga(a_\lambda, b_\lambda)$, then the full conditional posterior distribution of λ_η is a Gamma distribution with shape parameter $a_\lambda + (J-2)/2$ and rate parameter $b_\lambda + \frac{1}{2}\|\mathbf{D}_2\boldsymbol{\zeta}\|^2$.

S2 Additional results of the Gene Regulation Study

Table S1: The effective sample sizes and Gelman–Rubin convergence diagnostics for the gene regulation mixed-effects ODE models. The values of $\hat{R}$ is less than 1.1 indicate the convergence of Markov chains.

Method	Parameters	Effective sample size			Gelman–Rubin
		Chain 1	Chain 2	Chain 3	$\hat{R}$
SMN	$\ln(\alpha)$	5297.52	4846.53	4842.77	1.02
	$\ln(\beta)$	1189.24	1282.37	1667.38	1.07
	$\ln(\gamma)$	678.04	633.69	867.72	1.03
	$\ln(\delta)$	14637.24	19793.41	18442.82	1.03
Normal	$\ln(\alpha)$	1370.34	2257.52	2765.37	1.08
	$\ln(\beta)$	1472.13	966.08	774.09	1.08
	$\ln(\gamma)$	815.65	588.15	524.95	1.02
	$\ln(\delta)$	9132.64	12569.69	9687.08	1.04

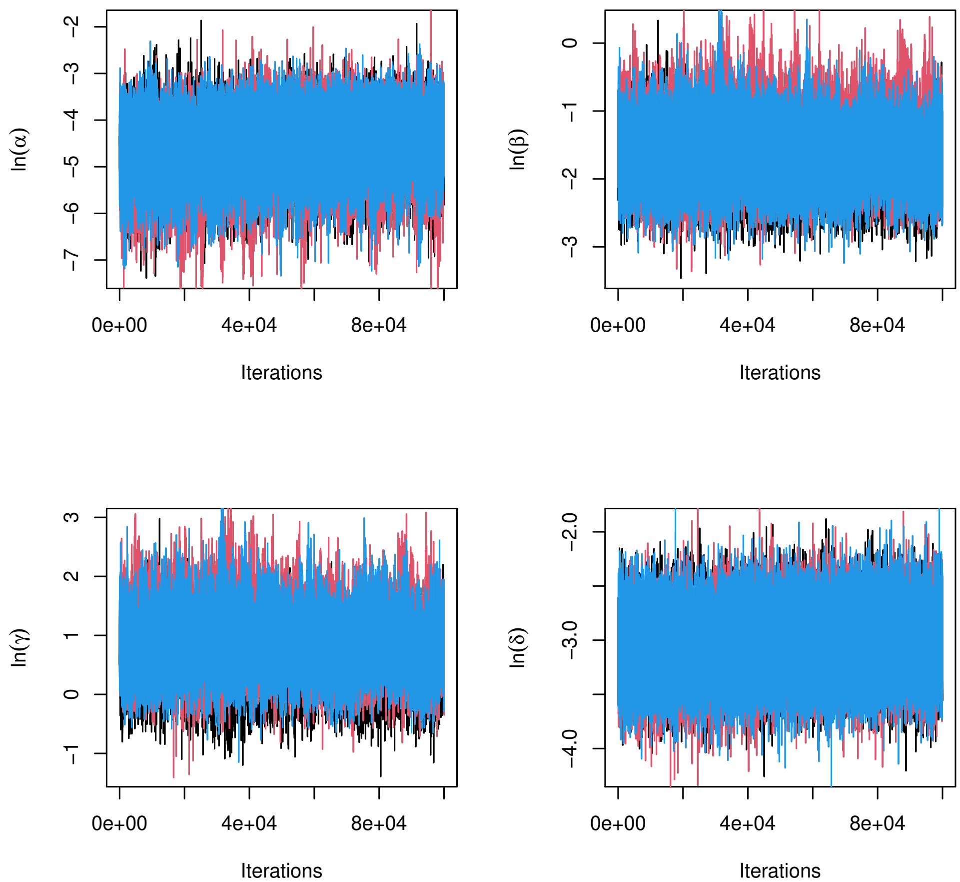


Figure S1: Trace plot to informally check convergence of MCMC samples based on the rest of Markov chains for the fixed-effects $\{\ln(\alpha), \ln(\beta), \ln(\gamma), \ln(\delta)\}$ in the gene regulation mixed-effects ODE model (4.1) under the SMN model.

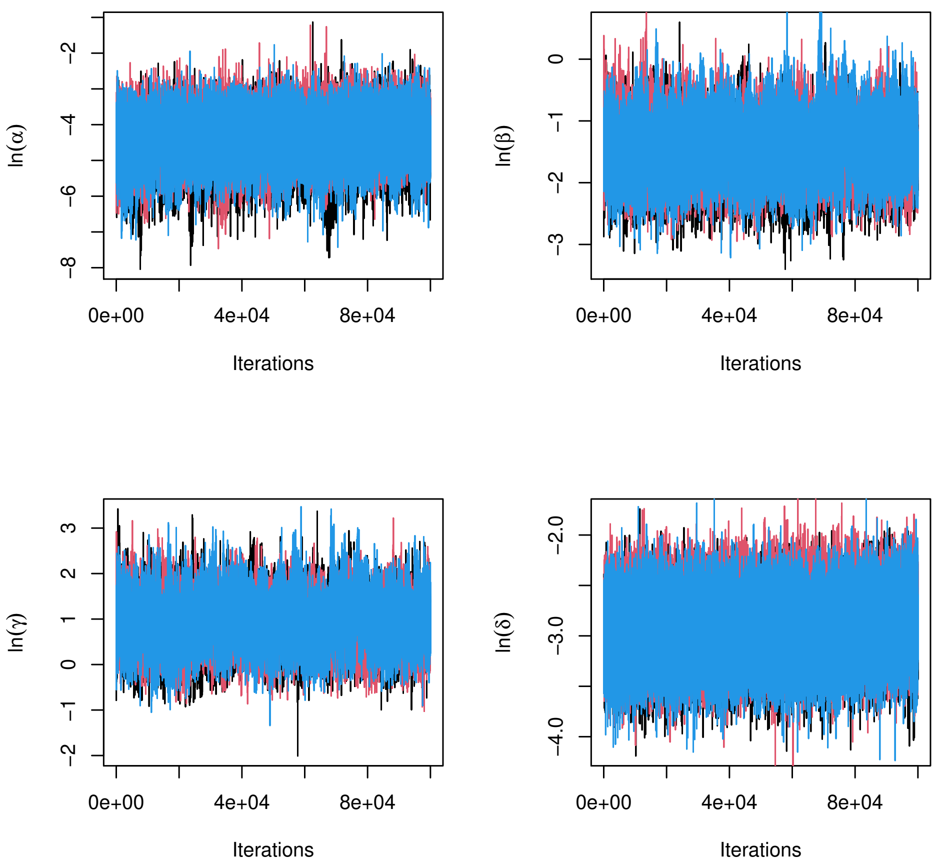


Figure S2: Trace plot to informally check convergence of MCMC samples based on the rest of Markov chains for the fixed-effects $\{\ln(\alpha), \ln(\beta), \ln(\gamma), \ln(\delta)\}$ in the gene regulation mixed-effects ODE model (4.1) under the Normal model.

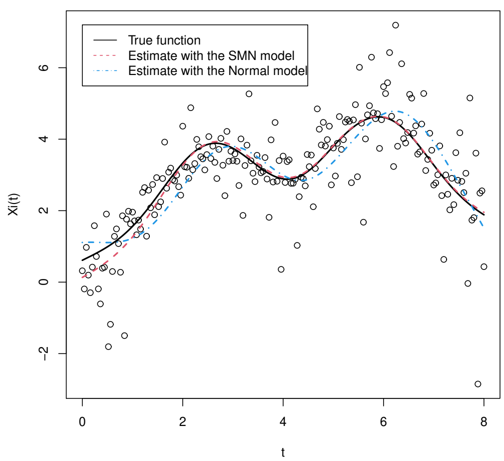
Table S2: The individual CPO_i for fifteen genes using the formula (3.9). The bold values indicate the influential genes.

Criterion	Distribution assumptions	
	SMN distributions	Normal distributions
$\widehat{\mathrm{CPO}}_1$	7.01×10^5	2.11×10^6
$\widehat{\mathrm{CPO}}_2$	1.03×10^7	2.21×10^6
$\widehat{\mathrm{CPO}}_3$	38.33	612.11
$\widehat{\mathrm{CPO}}_4$	1.45×10^8	5.54×10^6
$\widehat{\mathrm{CPO}}_5$	1.86×10^8	8.80×10^6
$\widehat{\mathrm{CPO}}_6$	6.46×10^7	8.65×10^6
$\widehat{\mathrm{CPO}}_7$	3.11×10^6	1.71×10^5
$\widehat{\mathrm{CPO}}_8$	25.07	175.68
$\widehat{\mathrm{CPO}}_9$	104.07	2.10×10^4
$\widehat{\mathrm{CPO}}_{10}$	4.73×10^7	3.39×10^6
$\widehat{\mathrm{CPO}}_{11}$	6.07×10^9	2.51×10^7
$\widehat{\mathrm{CPO}}_{12}$	$\mathbf{6.71 \times 10^{-8}}$	$\mathbf{7.03 \times 10^{-16}}$
$\widehat{\mathrm{CPO}}_{13}$	2.62×10^7	7.77×10^6
$\widehat{\mathrm{CPO}}_{14}$	1.29×10^4	2.56×10^4
$\widehat{\mathrm{CPO}}_{15}$	1.67×10^5	2.93×10^5
$B = \sum_{i=1}^{15} \log(\widehat{\mathrm{CPO}}_i)$	**174.16**	146.04
DIC	**-2186.90**	-1502.30

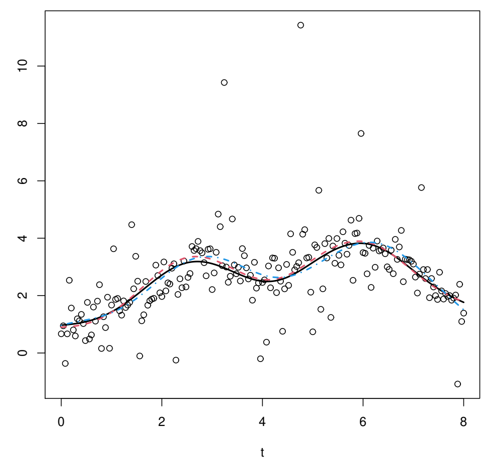
S3 Additional results of simulation studies

Table S3: 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 I based on 100 simulation replicates. The true values of fixed effects $\alpha = 1.2$, $\beta = 3.5$ and $\delta = 1.0$.

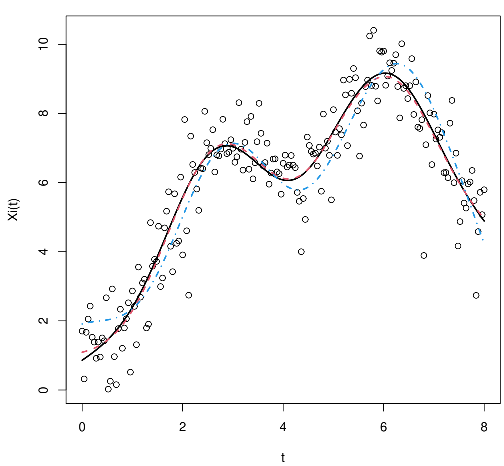
n	Fixed-effects	Distribution assumptions					
		SMN distributions			Normal distributions		
		Bias	SD	MADE	Bias	SD	MADE
50	α	0.403	0.141	0.426	0.311	0.128	0.336
	β	-0.255	0.190	0.317	-0.366	0.215	0.424
	δ	-0.298	0.098	0.313	-0.3388	0.109	0.354
	$\eta(t)$	N/A	N/A	0.458	N/A	N/A	0.401
100	α	0.430	0.099	0.441	0.343	0.109	0.360
	β	-0.281	0.145	0.316	-0.384	0.162	0.417
	δ	-0.324	0.071	0.331	-0.361	0.078	0.368
	$\eta(t)$	N/A	N/A	0.490	N/A	N/A	0.427



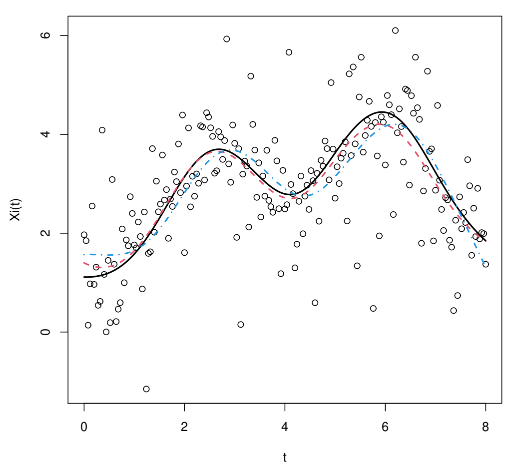
(a)



(b)



(c)



(d)

Figure S3: The fitted ODEs in the simulation Scenario II where the observations errors were generated from the distribution $0.6 \times t(3)$.

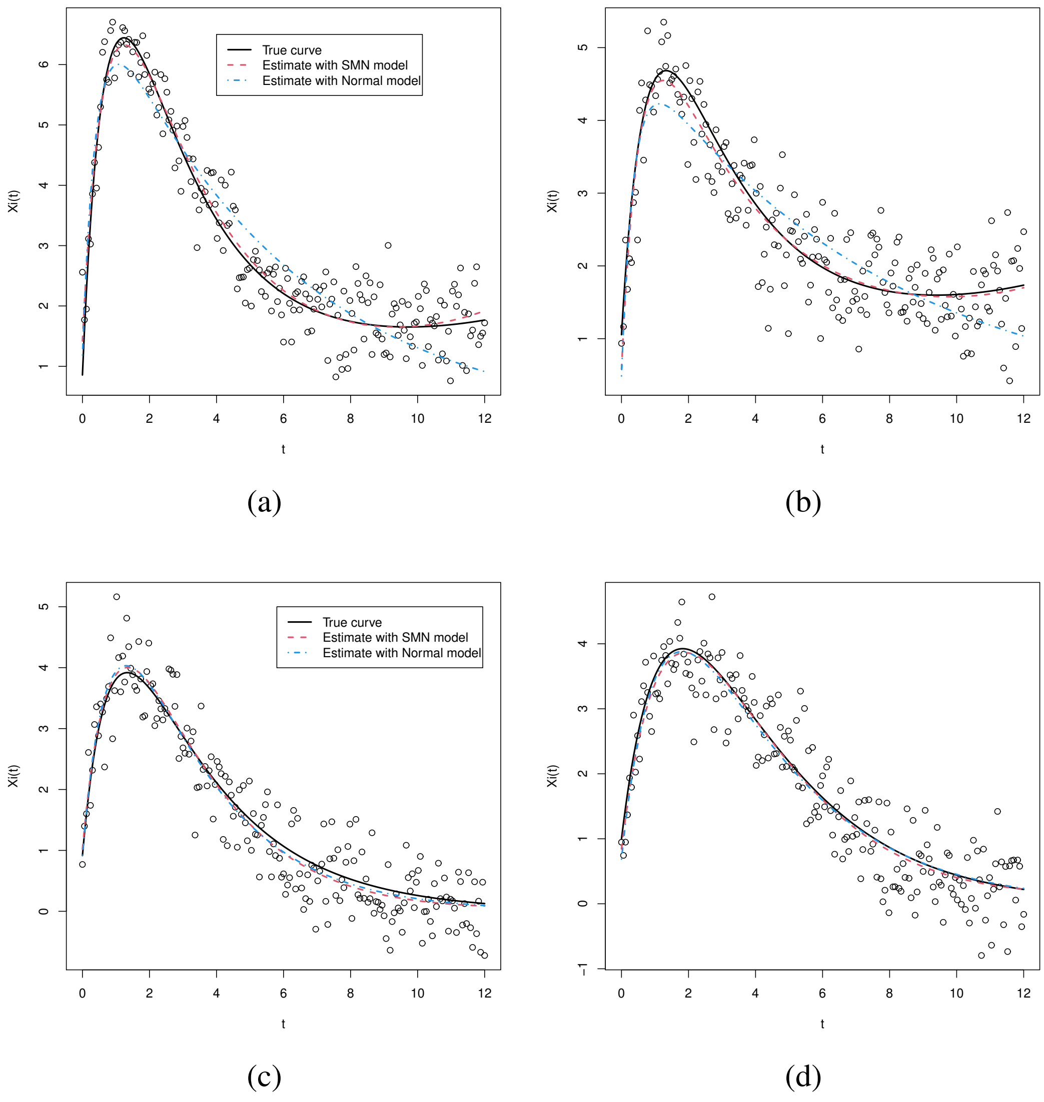
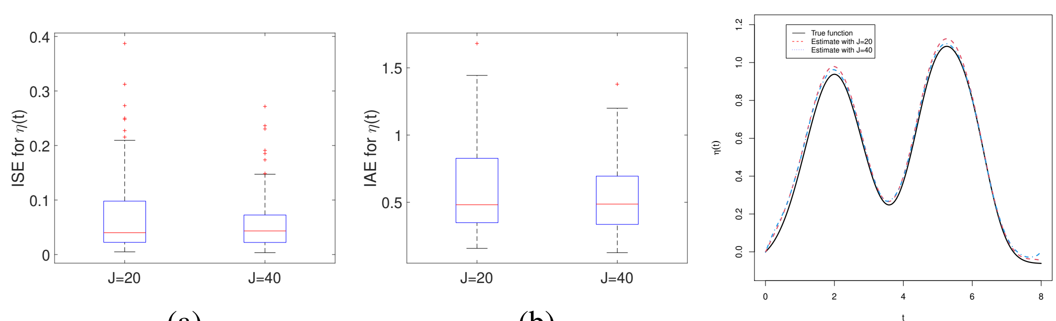


Figure S4: The upper panel shows the fitted ODEs in the simulation Scenario III-(a); the lower panel shows the fitted ODEs in the simulation Scenario III-(b).

To study the sensitivity of the proposed method to the choice of J, we conducted a sensitivity analysis regarding the different choice of J under the same simulation setting of Scenario II. The same datasets are reanalyzed with different $J = 20$ and 40. To evaluate the performance of the proposed method, we use the integrated squared error (ISE) and the integrated absolute error (IAE):

$$\mathrm{ISE} = \int_0^8 [\hat{\eta}(t) - \eta(t)]^2 \mathrm{d}t, \quad \text{and} \quad \mathrm{IAE} = \int_0^8 |\hat{\eta}(t) - \eta(t)| \mathrm{d}t.$$

Figure S5 displays the boxplots of ISE and IAE as well as the estimates of $\hat{\eta}(t)$ with different J based on 100 simulation replicates, which showed that the performance of our proposed method is robust to the choice of J.



(a) (b) (c)

Figure S5: The left panel and the middle panel show the ISEs and IAEs for $\hat{\eta}(t)$ in the mixed-effects ODE model (5.1) under the SMN model when fitting $\eta(t)$ using penalized splines with different number of splines basis functions, respectively. The right panel shows the true function as well as the estimates for $\eta(t)$ in the mixed-effects ODE model (5.1) under the SMN model when fitting $\eta(t)$ using penalized splines with different number of splines basis functions ('-': the true function; '$--$': the estimate with $J = 20$; '$-\cdot$': the estimate with $J = 40$).

Reference

1. Liang, F., Liu, C. and Carroll, R.J. (2010), Advanced Markov chain Monte Carlo: Learning from Past Samples, *Wiley*.