

Research Article

Maximum Likelihood Inference for Univariate Delay Differential Equation Models with Multiple Delays

Ahmed A. Mahmoud,¹ Sarat C. Dass,¹ Mohana S. Muthuvalu,¹ and Vijanth S. Asirvadam²

¹Fundamental and Applied Sciences Department, Faculty of Science and Information Technology, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia

²Department of Electrical and Electronic Engineering, Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia

Correspondence should be addressed to Ahmed A. Mahmoud; adly@azhar.edu.eg

Received 11 May 2017; Revised 19 August 2017; Accepted 23 August 2017; Published 12 October 2017

Academic Editor: Fathalla A. Rihan

Copyright © 2017 Ahmed A. Mahmoud et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article presents statistical inference methodology based on maximum likelihoods for delay differential equation models in the univariate setting. Maximum likelihood inference is obtained for single and multiple unknown delay parameters as well as other parameters of interest that govern the trajectories of the delay differential equation models. The maximum likelihood estimator is obtained based on adaptive grid and Newton-Raphson algorithms. Our methodology estimates correctly the delay parameters as well as other unknown parameters (such as the initial starting values) of the dynamical system based on simulation data. We also develop methodology to compute the information matrix and confidence intervals for all unknown parameters based on the likelihood inferential framework. We present three illustrative examples related to biological systems. The computations have been carried out with help of mathematical software: MATLAB® 8.0 R2014b.

1. Introduction

Delay differential equations (DDEs) are widely used to model many real life phenomena, especially in science and engineering. Examples include the modeling of spread of infectious diseases, modeling of tumor growth and the growth of blood clots in the brain, population dynamics, traffic monitoring, and price fluctuations of commodities in economics; see [1–4]. A univariate delay differential equation model (DDEM) with multiple delays equates the real valued observations, y_i , as noisy realizations from an underlying DDE:

$$y_i = x(t_i) + \epsilon_i, \quad i = 0, 1, 2, \dots, n, \quad (1)$$

where ϵ_i 's are errors assumed to arise from a noise distribution with zero mean and unknown standard deviation $\sigma > 0$. In (1), $x(t_i)$ is the solution, $x(t)$, of the DDE

$$\dot{x}(t) = f(t, x(t), z_1(t), z_2(t), \dots, z_m(t), \theta) \quad (2)$$

evaluated at the n time points, t_i , $i = 0, 1, \dots, n$; in (2), $z_j(t) = x(t - \tau_j)$, $j = 1, 2, \dots, m$, is the j th delay term with

delay parameter $\tau_j > 0$, and $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ is a vector of other parameters of interest that govern the trajectories of the underlying DDE in (2). Equations (1) and (2) constitute a univariate DDEM in the most general form. In a DDEM, the parameters θ_r and τ_j are often unknown and have to be estimated based on observations y_i , $i = 0, 1, \dots, n$.

Not many methods appear in the statistical literature on parameter estimation and inference for DDEMs. Among the statistical approaches that have been suggested, many involve restrictions on the form of DDEMs that are being investigated. When such restrictions are relaxed, high computational costs and challenges arise. Typically, further inferential procedures such as obtaining standard errors and confidence intervals associated with parameter estimates involve further computational costs and challenges. We give a brief review of these works and approaches that have been reported in the literature in the following paragraph.

Ellner et al. [5] estimate the derivative of a univariate DDEM, which is assumed to be in an additive form, using nonparametric smoothing. Subsequently, they infer

the constant (single) delay parameter, τ , based on fitting a generalized additive model. Ellner's technique, although it unifies previous works, can thus be applied to DDEMs which satisfy the assumed additive form only. Wood [6] developed spline based model fitting techniques in the case when the DDEMs are partially specified. The spline based method involves high computational costs as cross-validation is used to select the smoothing coefficients associated with the penalty term as well as the unknown parameter estimates. A penalized semiparametric method is proposed by Wang and Cao [7] which involves maximizing an objective function consisting of two terms: a likelihood term and a penalty term which measures the discrepancy between an estimate of the derivative, $\dot{x}(t)$, and the right hand side of the DDEM in (1). The selection of smoothing coefficients is done, similar to [6], via cross-validation, whereas standard errors of parameter estimates are obtained by bootstrapping. It follows that the method of [7], like [6], involves high computational costs. Further, Wang and Cao consider only univariate DDEMs with a single delay parameter. An estimation method based on Least Squares Support Vector Machines (LS-SVMs) for approximating constant as well as time-varying parameters in deterministic parameter-affine DDEMs is presented by Mehrkanoon et al. [8]. We note that Mehrkanoon performs parameter estimation only; no standard errors of estimates or confidence intervals are reported. Further, only single delays (either constant or time varying) are considered in [8].

In this paper, we consider parameter estimation and inference for univariate DDEMs with multiple delays based on the maximum likelihood. The method of maximum likelihood, as advocated by Fisher in his important papers [9, 10], has become one of the most significant tools for estimation and inference available to statisticians. Maximum likelihood estimators (MLEs) are well defined once a distributional model is specified for the observations. MLEs have well-behaved and well-understood properties: Huber [11] presents general conditions whereby the MLE is consistent for the true value of the unknown parameters for large sample sizes. Wald [12] and Akaike [13] observed that the maximum likelihood estimator is a natural estimator for the parameters when the true distribution is unknown. The large sample theory and distributional properties of MLEs can be used to perform subsequent inference procedures such as obtaining standard errors and confidence intervals and performing tests of hypotheses at minimal additional computational costs. MLEs are also the basic estimators that are used in subsequent statistical inferential procedures such as model selection using Akaike Information Criteria (AIC), Bayes Information Criteria (BIC), and other model selection criteria. Model selection is an important issue in DDEMs, such as for partially specified DDEMs in [6], where several models can be elicited for an observed physical process, but one model needs to be selected among many which fits the observed data and is simple enough to understand (Occam's razor principle).

MLE can be developed for a large variety of estimation situations and is asymptotically efficient, which means that for large samples it produces the most precise estimates

compared to non-MLE based methods (such as [8]). These are the reasons why we preferred using MLE over all other estimators for DDEMs in this paper.

The remainder of this paper is organized as follows: we define univariate DDEMs in Section 2. In Section 3, the MLE approach for DDEMs is outlined and the MLE is obtained computationally using an adaptive grid procedure followed by a gradient descent algorithm. We also develop algorithms for obtaining the information matrix and construct standard errors and confidence intervals for the unknown parameters. Three examples of univariate DDEMs related to biological systems are presented, and the numerical solutions and results based on the proposed methodology are provided based on simulation in Section 4.

2. General Model Formulation

Recall the DDEM defined by (1) and (2). The observation $y_i \in R$ is obtained at the i th sampled time point, t_i , with $T_0 = t_0 < \dots < t_n = T_1$, where $y_i = x(t_i) + \epsilon_i$, $i = 0, 1, 2, \dots, n$. In the remainder of this paper, the errors are assumed to be independent and identically distributed according to a normal with mean zero and unknown standard deviation $\sigma > 0$, that is, $\epsilon_i \sim N(0, \sigma^2)$. The underlying dynamical system $x(t)$, $t \in [T_0, T_1] \subset R$, is expressed implicitly in terms of the DDE. The general form of DDE with multiple delays for $x(t) \in R$ is given by (2) as $\dot{x}(t) = f(t, x(t), z_1(t), z_2(t), \dots, z_m(t), \boldsymbol{\theta})$, where $z_j(t) = \{x(t - \tau_j) : t - \tau_j \geq 0, j = 1, 2, \dots, m\}$, $f(t, x(t), z_1(t), z_2(t), \dots, z_m(t), \boldsymbol{\theta})$ is 1-dimensional function, and $\dot{x}(t)$ denotes the first derivative of $x(t)$ with respect to time t . The quantities $\boldsymbol{\theta} \in R^p$ and $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_m)$ are unknown parameters of the DDEM, where $\boldsymbol{\theta}$ is a vector of unknown parameters of dimension p and $\boldsymbol{\tau}$ is a vector of time delays of dimension m . The complete trajectory of the function $x(t)$ on $[T_0, T_1]$ will be determined by (2), and initial condition function $\varphi : [t_0 - \max(\tau_1, \tau_2, \dots, \tau_m), t_0] \rightarrow R$, where $\varphi(t) = a$ for all $t \in [t_0 - \max(\tau_1, \tau_2, \dots, \tau_m), t_0]$, is also unknown in addition to the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. For given values of $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and a , we note that the solution $x(t)$, $x(t) = x(t, \boldsymbol{\theta}, \boldsymbol{\tau}, a)$, is a function of $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and a ; thus, we make it an explicit function of the unknown quantities $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and a based on the $x(t, \boldsymbol{\theta}, \boldsymbol{\tau}, a)$ notation. The observations y_0, y_1, \dots, y_n are collected at the $(n + 1)$ sampled time points $T_0 = t_0 < t_1 < \dots < t_n = T_1$, based on the observational model (1). Our goal is to estimate $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and a based on the observations y_0, y_1, \dots, y_n . Here, note that a is also unknown and thus appears as a nuisance parameter since properties of the dynamical system are governed by $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ and not a .

3. The Maximum Likelihood Estimation Approach for DDEMs

The likelihood of the DDEM for parameters $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and a , given observations y_0, y_1, \dots, y_n , is

$$L(\boldsymbol{\theta}, \boldsymbol{\tau}, a | \mathbf{y}) = \prod_{i=0}^n p(y_i | \boldsymbol{\theta}, \boldsymbol{\tau}, a), \quad (3)$$

where $\mathbf{y} = (y_0, y_1, \dots, y_n)$ is the collection of all $(n + 1)$ observations on \mathbf{y} with density

$$p(y_i | \boldsymbol{\theta}, \boldsymbol{\tau}, a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2\sigma^2)(y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a))^2}, \quad (4)$$

based on the normality assumption on ϵ_i 's in (1). The above expression for the likelihood can be simplified to

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\tau}, a | \mathbf{y}) &= \prod_{i=0}^n \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2\sigma^2)(y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a))^2} \right) \\ &= (2\pi\sigma^2)^{-(n+1)/2} e^{-(1/2\sigma^2) \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a))^2}. \end{aligned} \quad (5)$$

We assume for the moment that $\sigma > 0$ is fixed and known; the case when $\sigma > 0$ is unknown is dealt with later. Thus, the above likelihood is taken to be a function of $(\boldsymbol{\theta}, \boldsymbol{\tau}, a)$ for now. The usual practice for statistical inference is to use the natural logarithm of the likelihood function, namely, the log-likelihood function, which is given by

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\tau}, a | \mathbf{y}) &= \sum_{i=0}^n \ln p(y_i | \boldsymbol{\theta}, \boldsymbol{\tau}, a) \\ &= -\frac{(n+1)}{2} \ln(2\pi\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a))^2. \end{aligned} \quad (6)$$

Expressions of the log-likelihood l are often simpler than the likelihood function, L , since they are easier to differentiate and the results are more stable computationally.

Since $\ln(x)$ is a monotonically increasing function of x , it follows that the maximization of (3) and (6) are equivalent in that the same optimized parameter is found. We denote the MLE of $(\boldsymbol{\theta}, \boldsymbol{\tau}, a)$ as $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}}, \hat{a})$. The MLE $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}}, \hat{a})$ is a point estimate such that

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}}, \hat{a}) = \operatorname{argmax}_{\boldsymbol{\theta}, \boldsymbol{\tau}, a} \{l(\boldsymbol{\theta}, \boldsymbol{\tau}, a)\} \quad (7)$$

and can be viewed as a random vector depending on the distribution of data, $\mathbf{y} = (y_0, y_1, \dots, y_n)$. We now consider the case when σ^2 is unknown. The MLE of σ^2 is denoted by $\hat{\sigma}^2$. After finding $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}}, \hat{a})$ as in (7), the log-likelihood equation in (6) is maximized as a function of σ^2 . The resulting estimate is available in closed form and is given by

$$\hat{\sigma}^2 = \frac{1}{n+1} \sum_{i=0}^n (y_i - x(t_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}}, \hat{a}))^2. \quad (8)$$

3.1. A Two-Stage Numerical Procedure for Finding the MLE. To find the MLE numerically, we develop a two-stage numerical procedure consisting of an adaptive grid procedure, then followed by a gradient descent algorithm. Two stages are needed as we wish to utilize the advantages of each algorithm

while avoiding the drawbacks of the other in each stage. Grid algorithms are able to find the global maximum of a function over a grid space. First, it evaluates values of the function on the grid space and then finds the grid value that corresponds to the maximum. Provided the grid space is refined enough, the grid value corresponding to this maximum will be close to the domain value that actually corresponds to the global maximum. So by gridding, we are able to ensure that we are close to the global maximum. The adaptive grid algorithm enhances the original gridding algorithm so that we will move closer and closer to the global maximum. However, the main drawback of any grid (and adaptive grid) algorithm is its slowness in convergence.

On the other hand, gradient descent algorithms can converge to maxima of a function sufficiently quickly. The main drawback of gradient descent algorithms is that it will find the nearest local maximum from the starting point. So, if the original starting point is not close to the global maximum, a gradient descent algorithm will not guarantee that the global maximum is found since it might get "stuck" at a local maximum only.

The function to maximize in our case is the log-likelihood in (6) and the domain value corresponding to this maximum is the MLE. Thus, our two-step method uses adaptive grid in the first stage to ensure that we are close to the MLE and then switches to the quasi-Newton algorithm to ensure rapid convergence to the MLE.

3.1.1. Grid Procedure. To find $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\tau}}$ as in (7), the value with largest (log) likelihood should be chosen. This can be done by an adaptive grid procedure. The gridding is carried out for $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$, and for each pair of $(\boldsymbol{\theta}, \boldsymbol{\tau})$ in the grid space, a Newton-Raphson numerical procedure is used to find the maximum value of a defined as

$$\hat{a} = \hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \operatorname{argmax}_a \{l(\boldsymbol{\theta}, \boldsymbol{\tau}, a)\}. \quad (9)$$

We use the grid space $\Theta = \{(\theta_r, \tau_s), r = 1, 2, \dots, R, s = 1, 2, \dots, S\}$, which covers RS values of (θ_r, τ_s) . For every fixed value of θ_r and τ_s in Θ , we find the MLE of a , $\hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau})$, given by maximizing the log-likelihood above. Since $\hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau})$ satisfies

$$\frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\tau}, \hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau}))}{\partial a} = 0, \quad (10)$$

the numerical problem is solved by using Newton-Raphson method:

$$a_{h+1}(\boldsymbol{\theta}, \boldsymbol{\tau}) = a_h(\boldsymbol{\theta}, \boldsymbol{\tau}) - \frac{(\partial l(\boldsymbol{\theta}, \boldsymbol{\tau}, a_h(\boldsymbol{\theta}, \boldsymbol{\tau})) / \partial a)}{(\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\tau}, a_h(\boldsymbol{\theta}, \boldsymbol{\tau})) / \partial a^2)}, \quad (11)$$

where

$$\frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial a} = \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)) \frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial a}, \quad (12)$$

$$\begin{aligned} & \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial a^2} \\ &= \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)) \frac{\partial^2 x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial a^2} \\ & \quad - \frac{1}{\sigma^2} \sum_{i=0}^n \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial a} \right)^2; \end{aligned} \quad (13)$$

in (12) and (13), $\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)/\partial a$ and $\partial^2 x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)/\partial a^2$ are, respectively, the first and second partial derivative of $x(t, \boldsymbol{\theta}, \boldsymbol{\tau}, a)$ with respect to a and then evaluated at $t = t_i$ for $i = 0, 1, 2, \dots, n$. As seen from (12) and (13), we need to calculate $\partial x/\partial a$ and $\partial^2 x/\partial a^2$ at each $t = t_i$, $i = 0, 1, 2, \dots, n$. This is done recursively as follows.

The first derivative process of $\partial x/\partial a$ is obtained by differentiating (2) with respect to a . Differentiating (2) with respect to a , where $\boldsymbol{\theta}$ and a are independent of each other, gives

$$\frac{\partial \dot{x}}{\partial a} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial a} + \sum_{j=1}^m \frac{\partial f}{\partial z_j} \frac{\partial z_j}{\partial a} \quad (14)$$

which implies that the first derivative process $\partial x/\partial a$ satisfies another DDE which is given by (14).

Similarly, the second derivative process $\partial^2 x/\partial a^2$ is obtained by differentiating (14) with respect to a , to obtain

$$\begin{aligned} \frac{\partial^2 \dot{x}}{\partial a^2} &= \frac{\partial^2 f}{\partial x^2} \left(\frac{\partial x}{\partial a} \right)^2 + \sum_{j=1, k=1}^m \frac{\partial^2 f}{\partial z_j \partial z_k} \left(\frac{\partial z_j}{\partial a} \right) \left(\frac{\partial z_k}{\partial a} \right) \\ & \quad + 2 \sum_{j=1}^m \frac{\partial^2 f}{\partial x \partial z_j} \frac{\partial x}{\partial a} \frac{\partial z_j}{\partial a} + \frac{\partial^2 x}{\partial a^2} \frac{\partial f}{\partial x} + \sum_{j=1}^m \frac{\partial^2 z_j}{\partial a^2} \frac{\partial f}{\partial z_j} \end{aligned} \quad (15)$$

which implies that $\partial^2 x/\partial a^2$ satisfies a DDEM depending on $\partial x/\partial a$. The above two DDEs can be solved numerically based on initial conditions that are specified below.

To obtain the initial conditions of the first and second derivative process, we note that

$$x(t) = a \quad \text{for } t \in (-\infty, t_0]. \quad (16)$$

Thus, $\partial x/\partial a = 1$ and $\partial^2 x/\partial a^2 = 0$ for $t \in (-\infty, t_0]$. Subsequently, we can get the value of $\partial l(\boldsymbol{\theta}, \boldsymbol{\tau}, a)/\partial a$ and $\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\tau}, a)/\partial a^2$ numerically at every value of t_i by numerically solving the DDEs using (12) and (13). We divide each $[t_{i-1}, t_i]$ into M equal segments. Here, M is a natural number.

The grid algorithm operates on the grid space of $(\boldsymbol{\theta}, \boldsymbol{\tau})$, and the Newton-Raphson procedure is nested within the adaptive grid algorithm. Thus, for every grid value pair (θ_r, τ_s) , the Newton-Raphson uses these values of $(\boldsymbol{\theta}, \boldsymbol{\tau})$ to find “ a ” via (11). On convergence of the Newton-Raphson method, we obtain the MLE of a , $\hat{a}(\theta_r, \tau_s)$, for each point grid (θ_r, τ_s) in Θ . The log-likelihood $l(\boldsymbol{\theta}, \boldsymbol{\tau}, a)$ is calculated based on (6) using $x_i = x(t_i, \theta_r, \tau_s, \hat{a}(\theta_r, \tau_s))$. Then the point based maximum is found by finding the maximum $l(\theta_r, \tau_s, \hat{a}(\theta_r, \tau_s))$

as a function of (θ_r, τ_s) . We define the MLE which is obtained from the gridding algorithm as

$$(\hat{\boldsymbol{\theta}}_G, \hat{\boldsymbol{\tau}}_G) = \underset{r,s}{\operatorname{argmax}} l(\theta_r, \tau_s). \quad (17)$$

3.1.2. The Adaptive Grid Procedure. The adaptive grid (AG) algorithm is a repeated application of the generic grid procedure over increasingly finer intervals for $(\boldsymbol{\theta}, \boldsymbol{\tau})$. The AG algorithm is as follows:

- (1) Choose an initial grid space $\Theta^{(0)}$ consisting of the grid points $(\theta_r^{(0)}, \tau_s^{(0)})$, $r = 1, 2, \dots, R$ and $s = 1, 2, \dots, S$.
- (2) Maximize $l(\theta_r, \tau_s, \hat{a}(\theta_r, \tau_s))$ with respect to r and s as described in the grid procedure above.
- (3) Obtain $(\hat{\boldsymbol{\theta}}_G^{(0)}, \hat{\boldsymbol{\tau}}_G^{(0)})$ as in (17).
- (4) Refine the grid: suppose $(\theta_{r_0}^{(0)}, \tau_{s_0}^{(0)}) \equiv (\hat{\boldsymbol{\theta}}_G^{(0)}, \hat{\boldsymbol{\tau}}_G^{(0)})$ as in #3). The new grid space $\Theta^{(1)}$ has lower and upper θ -grid points given by $(\theta_{r_0-1}^{(0)}, \theta_{r_0+1}^{(0)})$. The corresponding lower and upper τ -grid points are $(\tau_{s_0-1}^{(0)}, \tau_{s_0+1}^{(0)})$. If either the lower or upper bounds are not found, then the original grid space is enlarged so that the MLE occurs in the interior of $\Theta^{(0)}$.
- (5) Repeat steps #2)–#4) to obtain $(\hat{\boldsymbol{\theta}}_G^{(1)}, \hat{\boldsymbol{\tau}}_G^{(1)})$ based on the generic grid procedure. Repeat to generate the sequence $(\hat{\boldsymbol{\theta}}_G^{(k)}, \hat{\boldsymbol{\tau}}_G^{(k)})$, $k = 0, 1, 2, \dots$. Stop at k^* when $\|(\hat{\boldsymbol{\theta}}_G^{(k^*)}, \hat{\boldsymbol{\tau}}_G^{(k^*)}) - (\hat{\boldsymbol{\theta}}_G^{(k^*-1)}, \hat{\boldsymbol{\tau}}_G^{(k^*-1)})\| < \delta$, a prespecified threshold.
- (6) The final MLE based on the adaptive grid technique is

$$(\hat{\boldsymbol{\theta}}_{0,\text{MLE}}, \hat{\boldsymbol{\tau}}_{0,\text{MLE}}) = (\hat{\boldsymbol{\theta}}_G^{(k^*)}, \hat{\boldsymbol{\tau}}_G^{(k^*)}). \quad (18)$$

Remark 1. In step #1), the initial grid space $\Theta^{(0)}$ is chosen to be a large domain that is likely to contain the MLE. In our simulation experiments, since the true values of $(\boldsymbol{\theta}, \boldsymbol{\tau})$ are known, the domain is selected around these true values. In practice, we need to carry out an exhaustive search within the upper and lower bounds of $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. If the parameters are positive, say, as is usually the case, the lower bounds can be taken to be zero. Next, we can consider a large positive numbers, say B and C , and construct the grid in $[0, B]^p \times [0, C]^m$ consisting of H equidistant marginal grid points. The value of H need not be too large since we only aim to explore the log-likelihood profile. The log-likelihood can be evaluated at these grid points and plotted to visualize properties of the resulting surface. Depending on this plot, we can choose either to fix or increase B and C until we are certain that the MLE is within the selected domain.

After obtaining the first-step approximation to the MLE by the adaptive grid procedure above, we use the MATLAB function `fminunc` to obtain the final MLE by minimizing the negative log-likelihood function viewed as a function of the

unknown parameter vector $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_J) = (\boldsymbol{\theta}, \boldsymbol{\tau})$. We have $J = p + m$, and the final step MLE is defined as

$$\Gamma_{\text{MLE}} = \underset{\Gamma}{\operatorname{argmin}} [-l_0(\Gamma)], \quad (19)$$

where $l_0(\Gamma) = l_0((\boldsymbol{\theta}, \boldsymbol{\tau})) = l(\boldsymbol{\theta}, \boldsymbol{\tau}, \hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau}))$ with $\hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau})$ defined in (9). We require to input the gradient vector for this MATLAB function which is given by

$$\nabla l_0(\Gamma) = \begin{bmatrix} \frac{\partial l_0(\Gamma)}{\partial \Gamma_1} \\ \frac{\partial l_0(\Gamma)}{\partial \Gamma_2} \\ \vdots \\ \frac{\partial l_0(\Gamma)}{\partial \Gamma_j} \end{bmatrix}. \quad (20)$$

The explicit expression of each entry of $\nabla l_0(\Gamma)$ is provided in Appendix A. The MATLAB function `fminunc` uses, as an option, a quasi-Newton procedure that does not require the calculation of second derivatives and hence saves computational time.

3.2. Statistical Inference Based on MLEs

3.2.1. Information Matrix. Now we incorporate σ^2 into the estimation procedure as well. Once $(\hat{\boldsymbol{\theta}}_{\text{MLE}}, \hat{\boldsymbol{\tau}}_{\text{MLE}}, \hat{a}_{\text{MLE}})$ is obtained by the above two-stage procedure, the MLE of σ^2 is obtained analytically as

$$\hat{\sigma}_{\text{MLE}}^2 = \frac{1}{(n+1)} \sum_{i=0}^n (y_i - x(t_i, \hat{\boldsymbol{\theta}}_{\text{MLE}}, \hat{\boldsymbol{\tau}}_{\text{MLE}}, \hat{a}_{\text{MLE}}))^2. \quad (21)$$

Let $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_K) = (\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2)$ denote the $K \times 1$ vector of all unknown parameters (including σ^2) where $K = J + 2 = p + m + 2$. Subsequent inference based on the MLEs requires the computation of the Fisher information [14, 15]. The Fisher information matrix $I(\Gamma)$ is given by the $K \times K$ symmetric matrix whose (u, v) th element is the covariance between u th and v th first partial derivatives of the log-likelihood:

$$I(\Gamma)_{(u,v)} = \operatorname{Cov} \left[\frac{\partial l(\Gamma | \mathbf{y})}{\partial \Gamma_u}, \frac{\partial l(\Gamma | \mathbf{y})}{\partial \Gamma_v} \right]. \quad (22)$$

Based on the expected values of the second partial derivatives, the Fisher information matrix in (22) is equivalent to

$$I(\Gamma)_{(u,v)} = -E \left[\frac{\partial^2 l(\Gamma | \mathbf{y})}{\partial \Gamma_u \partial \Gamma_v} \right], \quad 1 \leq u, v \leq K. \quad (23)$$

The observed Fisher information matrix is simply $I(\hat{\Gamma}_{\text{MLE}})$, the information matrix evaluated at the maximum likelihood estimate, $\hat{\Gamma}_{\text{MLE}}$, of Γ . Further, its inverse evaluated at the MLE is an estimate of the asymptotic covariance matrix for $\hat{\Gamma}_{\text{MLE}}$ which is given by

$$\operatorname{COV}(\hat{\Gamma}_{\text{MLE}}) = [I(\hat{\Gamma}_{\text{MLE}})]^{-1}. \quad (24)$$

Since the log-likelihood function is given by

$$l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2 | \mathbf{y}) = -\frac{(n+1)}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a))^2, \quad (25)$$

the first-order partial derivative of $l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2 | \mathbf{y})$ with respect to each element of Γ is given by

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2 | \mathbf{y})}{\partial \Gamma_u} &= \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)) \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_u} \right) \end{aligned} \quad (26)$$

for $1 \leq u \leq K$. The second-order partial derivative is

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2 | \mathbf{y})}{\partial \Gamma_u \partial \Gamma_v} &= \frac{1}{\sigma^2} \left[\sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)) \left(\frac{\partial^2 x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_u \partial \Gamma_v} \right) \right. \\ &\quad \left. - \sum_{i=0}^n \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_u} \right) \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_v} \right) \right], \end{aligned} \quad (27)$$

for $1 \leq u, v \leq K$,

$$\begin{aligned} \frac{\partial^2 l}{\partial \Gamma_u \partial \sigma^2} &= \frac{-1}{(\sigma^2)^2} \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)) \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_u} \right) \end{aligned} \quad (28)$$

for $1 \leq u \leq K$, and

$$\begin{aligned} \frac{\partial^2 l}{\partial (\sigma^2)^2} &= \frac{1}{2\sigma^4} \left[(n+1) - \frac{2}{\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a))^2 \right]. \end{aligned} \quad (29)$$

Taking expectations on both sides of the above equations and from (23), we get

$$\begin{aligned} I(\Gamma)_{(u,v)} &= -E \left[\frac{\partial^2 l(\Gamma | \mathbf{y})}{\partial \Gamma_u \partial \Gamma_v} \right] \\ &= \frac{1}{\sigma^2} \sum_{i=0}^n \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_u} \right) \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_v} \right), \end{aligned} \quad (30)$$

for $1 \leq u, v \leq K$,

$$I(\Gamma)_{(u,\sigma^2)} = -E \left[\frac{\partial^2 l(\Gamma | \mathbf{y})}{\partial \Gamma_u \partial \sigma^2} \right] = 0 \quad (31)$$

for $1 \leq u \leq K$, and

$$I(\Gamma)_{(\sigma^2, \sigma^2)} = -E \left[\frac{\partial^2 l(\Gamma | \mathbf{y})}{\partial (\sigma^2)^2} \right] = \frac{n+1}{2\sigma^4}. \quad (32)$$

We compute each element (u, v) of the matrix in (23) for DDEMs with single and multiple delays; the explicit expressions are given in Appendices A and B.

3.2.2. Confidence Intervals. A confidence interval for an unknown parameter gives the range of values most likely to cover the true value of the parameter with high probability. The standard form of a confidence interval is

$$\text{estimate} + / - \text{margin of error}. \quad (33)$$

To construct a level C confidence interval for any element of $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_K) = (\theta, \tau, a, \sigma^2)$, say, Γ_u , for $1 \leq u \leq K$, we need to find an estimate of the margin of error. First, the estimated standard error of the maximum likelihood estimate, $\hat{\Gamma}_{u, \text{MLE}}$, of Γ_u is given by

$$\text{SE}(\hat{\Gamma}_{u, \text{MLE}}) = \sqrt{\text{COV}(\hat{\Gamma}_{\text{MLE}})_{(u, u)}}, \quad (34)$$

where $\text{COV}(\hat{\Gamma}_{\text{MLE}})$ is the covariance matrix as given in (24). The explicit terms of the covariance matrix can be obtained by substituting (30) into (24). The confidence interval for Γ_u is

$$\hat{\Gamma}_{u, \text{MLE}} \pm z_{\alpha/2} \text{SE}(\hat{\Gamma}_{u, \text{MLE}}), \quad (35)$$

where $z_{\alpha/2} = \text{norminv}(1 - \alpha/2)$, $\alpha = 0.05$, is the desired significance level and $z_{\alpha/2} \text{SE}(\hat{\Gamma}_{u, \text{MLE}})$ is the margin of error. We can find confidence intervals for all components of $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_K) = (\theta, \tau, a, \sigma^2)$ in this way. In some cases, the estimated confidence interval in (35) may include some negative values which is unreasonable for a parameter that is known to be positive. In this case, we perform a logarithmic transformation of the parameter, construct the confidence interval for the log-transformed parameter, and then transform the confidence interval back to the original parameter space. The confidence interval for Γ_u based on this log-transformation procedure is

$$\exp \left(\log(\hat{\Gamma}_{u, \text{MLE}}) \pm \frac{z_{\alpha/2} \text{SE}(\hat{\Gamma}_{u, \text{MLE}})}{\hat{\Gamma}_{u, \text{MLE}}} \right). \quad (36)$$

4. Examples

We present three examples of DDEMs in the univariate case: we consider two models with a single delay and a third one with two delays.

4.1. Example 1. We consider the exponential delay differential equation model (EDDEM) with a single delay (i.e., $p = 1$, $m = 1$) which is the solution to the DDE

$$\dot{x}(t) = \theta x(t - \tau). \quad (37)$$

The EDDEM in (37) is a model for ideal population growth under infinite resources and no deaths, such as a protozoan or bacterial culture dividing under constant environmental conditions. The delay parameter τ can be taken to represent the gestation period or maturity period, that is, the time taken for individuals to be ready for division. The parameter θ represents the growth rate of the population. We numerically solve the DDE in (37) using the MATLAB function `dde23` with fixed parameters $(\theta, \tau, a, \sigma^2)$ values at $(0.5, 1, 5, 0.01)$. Sampled observations from the DDEM as in (1) were obtained at discrete time intervals of width $h = 0.1$ starting from $t_0 = 0$. The endpoint considered is $t_n = 10$ corresponding to $n+1 = 101$. The aim is to estimate θ , τ , a , and σ^2 based on y_0, y_1, \dots, y_n . Figure 1(a) illustrates the different behaviour of $x(t)$ based on different parameter specifications. Figure 1(b) shows the underlying trajectories of the solution $x(t)$ from the DDE model (37) and the $(n+1)$ sampled observations.

The initial grid space for the adaptive grid procedure was taken to be $\Theta^{(0)} = \{(\theta_r, \tau_s) : \theta_r = \theta_0 + rh, \tau_s = \tau_0 + sh, r = 1, 2, \dots, R; s = 1, 2, \dots, S\}$ with $(\theta_0, \tau_0) = (0.1, 0.6)$, $h = 0.1$, $R = S = 9$; see the remark that is given towards the end of this section regarding the selection of $\Theta^{(0)}$ for this example as well as for the rest. The stopping criteria threshold δ was chosen to be 0.0001. The adaptive grid and Newton-Raphson procedures were run; the results are given in Tables 1 and 2. Recall that M is the number of subdivisions of each interval $[t_{i-1}, t_i]$ needed for calculating the quantities $\partial x(t_i, \theta, \tau, a) / \partial a$ and $\partial^2 x(t_i, \theta, \tau, a) / \partial a^2$; see Section 3.1.1. As M increases, the MLE estimates become more accurate but at the cost of increased computational time. We note that, for $M = 50$ or $M = 100$, satisfactory results are already achieved in terms of closeness to the true parameter values of $(\theta, \tau, a, \sigma^2) = (0.5, 1, 5, 0.01)$. Subsequently, $M = 100$ is considered for finding the maximum of the log-likelihood function, for finding the information matrix and computing the confidence intervals of the parameters.

The Fisher information matrix as $h = 0.1$ and $M = 100$ for EDDEM is

$$I(\Gamma) = 10^9 \times \begin{bmatrix} 1.0887 & 0.0404 & 0.0459 & 0 \\ 0.0404 & 0.0015 & 0.0017 & 0 \\ 0.0459 & 0.0017 & 0.0020 & 0 \\ 0 & 0 & 0 & 0.0006 \end{bmatrix}, \quad (38)$$

and variance of $\hat{\Gamma}$ at the MLE is given by

$$\text{Var}(\hat{\Gamma}) = 10^{-4} \times \begin{bmatrix} 0.0004 & -0.0129 & 0.0015 & 0 \\ -0.0129 & 4.3309 & -3.5029 & 0 \\ 0.0015 & -3.5029 & 3.0445 & 0 \\ 0 & 0 & 0 & 0.0170 \end{bmatrix}. \quad (39)$$

The 95% confidence intervals for parameters $(\theta, \tau, a, \sigma^2)$ in EDDE with single delay are shown, respectively, in Table 3.

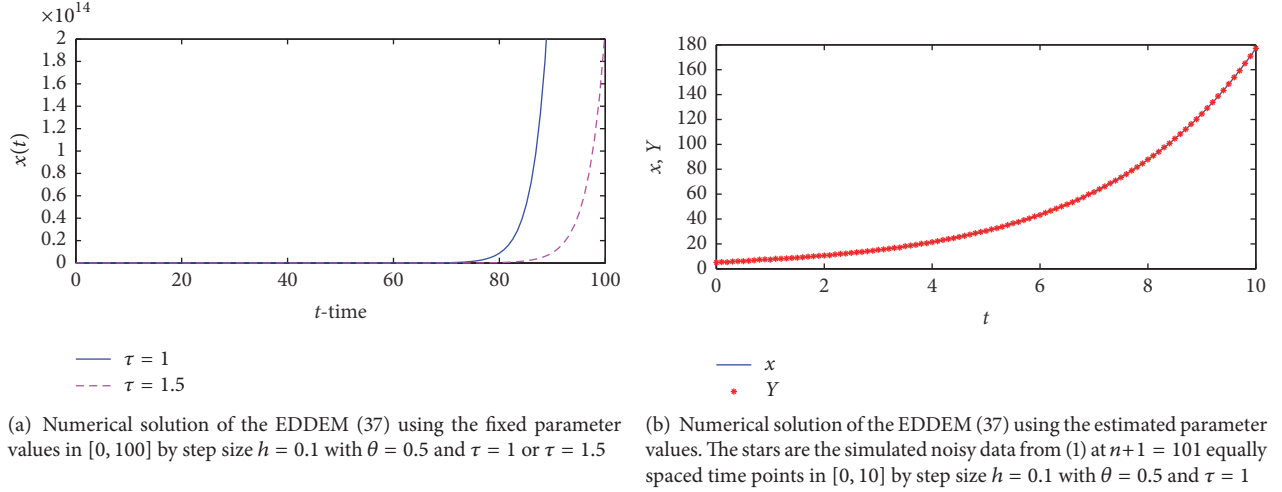


FIGURE 1

TABLE 1: Table showing values of $\text{Max.Val}(l)$, $\hat{a}(\hat{\theta}, \hat{\tau})$, $(\hat{\theta}, \hat{\tau})$, and $\hat{\sigma}^2$ for the EDDEM by using the adaptive grid procedure. Here, the number of equally spaced time points in $[0, 10]$ is $n + 1 = 101$ with $h = 0.1$.

M	$\text{Max.Val}(l)$	$\hat{a}(\hat{\theta}, \hat{\tau})$	$(\hat{\theta}, \hat{\tau})$	$\hat{\sigma}^2$
10	-0.5817	5.0181	(0.5000, 1.0000)	0.0114
20	-0.5037	5.0091	(0.5000, 1.0000)	0.0099
30	-0.4880	5.0061	(0.5000, 1.0000)	0.0096
40	-0.4821	5.0046	(0.5000, 1.0000)	0.0095
50	-0.4792	5.0037	(0.5000, 1.0000)	0.0094
100	-0.4748	5.0019	(0.5000, 1.0000)	0.0093
1000	-0.4724	5.0002	(0.5000, 1.0000)	0.0093

TABLE 2: Table showing value of $\text{Max.Val}(l)$, $\nabla l(\Gamma)$, and $(\hat{\theta}_{\text{MLE}}, \hat{\tau}_{\text{MLE}})$ for the EDDEM with $h = 0.1$.

M	$\text{Max.Val}(l)$	$\nabla l(\Gamma)$	$(\hat{\theta}_{0,\text{MLE}}, \hat{\tau}_{0,\text{MLE}})$	$(\hat{\theta}_{\text{MLE}}, \hat{\tau}_{\text{MLE}})$
100	0.4715	$10^{-2} \times \begin{bmatrix} 0.0006 \\ -0.1001 \end{bmatrix}$	(0.5000, 1.000)	(0.5001645, 1.0000017)

TABLE 3: Table showing the 95% confidence intervals for parameters θ , τ , a , and σ^2 in EDDEM.

M	(θ_L, θ_U)	(τ_L, τ_U)	(a_L, a_U)	(σ_L^2, σ_U^2)
100	(0.4996, 0.5004)	(0.9592, 1.0408)	(4.9677, 5.0361)	(0.0068, 0.0119)

4.2. *Example 2.* A delay differential equation in population ecology given by

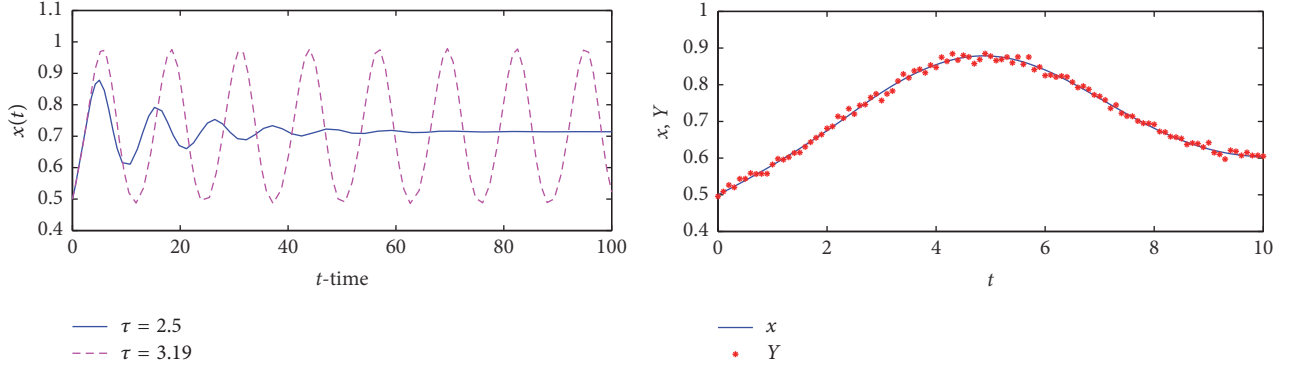
$$\dot{x}(t) = \theta x(t) \left(1 - \frac{x(t-\tau)}{K}\right) \quad (40)$$

is known as Hutchinson's equation [16], where x is the population at that instant, θ is the intrinsic growth rate, and K is the carrying capacity of the population. Both θ and K are positive constants and τ is a positive constant delay parameter.

Define $\theta_1 = \theta$ and $\theta_2 = \theta/K$; then (40) can be rewritten as

$$\dot{x}(t) = \theta_1 x(t) - \theta_2 x(t) x(t-\tau). \quad (41)$$

The observations y_0, y_1, \dots, y_n are collected at the $(n + 1)$ sampled time points $T_0 = t_0 < t_1 < \dots < t_n = T_1$, based on the observational model (1), and the aim is to estimate $\theta_1, \theta_2, \tau, a$, and σ^2 based on the observations y_0, y_1, \dots, y_n . The DDE in (41) is solved numerically by using the MATLAB function `dde23` with fixed parameters $(\theta_1, \theta_2, \tau, a, \sigma^2)$ values. Figure 2(a) shows the underlying trajectories (mean function) of the solution $x(t)$ from the DDE model (41). Figure 2(a) illustrates the different behaviour of $x(t)$ based on different parameter specifications which reflect both stable and unstable solutions. Subsequently, we fix the parameter specification at $(\theta_1, \theta_2, \tau, a, \sigma^2) = (0.5, 0.7, 2.5, 0.5, 0.0001)$. Sampled observations from the DDE as in (1) were obtained



(a) Numerical solution of the DLDEM with two delays (41) using the fixed parameter values in $[0, 100]$ by step size $h = 0.1$ with $\theta_1 = 0.5$ and $\theta_2 = 0.7$. The steady state is stable when $\tau = 2.5$ and becomes unstable when $\tau = 3.19$

(b) Numerical solution of the DLDEM with two delays (41) using the estimated parameter values. The stars are the simulated noisy data by adding noises to the DLDE solutions at $n + 1 = 101$ equally spaced time points in $[0, 10]$ by step size $h = 0.1$ with $\theta_1 = 0.5$, $\theta_2 = 0.7$, and $\tau = 2.5$

FIGURE 2

TABLE 4: Table showing values of $\text{Max.Val}(l)$, $\hat{a}(\hat{\theta}_1, \hat{\theta}_2, \hat{\tau})$, $(\hat{\theta}_1, \hat{\theta}_2, \hat{\tau})$, and $\hat{\sigma}^2$ for the DLDEM with single delay by using an adaptive grid. Here, the number of equally spaced time points in $[0, 10]$ is $n + 1 = 101$ with $h = 0.1$.

M	$\text{Max.Val}(l)$	$\hat{a}(\theta_1, \theta_2, \tau)$	$(\hat{\theta}_1, \hat{\theta}_2, \hat{\tau})$	$\hat{\sigma}^2$
10	-0.0049	0.5008	(0.5, 0.7, 2.5)	0.000095
20	-0.0048	0.5005	(0.5, 0.7, 2.5)	0.000095
30	-0.0048	0.5005	(0.5, 0.7, 2.5)	0.000095
40	-0.0048	0.5004	(0.5, 0.7, 2.5)	0.000095
50	-0.0048	0.5004	(0.5, 0.7, 2.5)	0.000095
100	-0.0048	0.5003	(0.5, 0.7, 2.5)	0.000095
1000	-0.0048	0.5003	(0.5, 0.7, 2.5)	0.000095

TABLE 5: Table showing value of $\text{Max.Val}(l)$, $\nabla l(\Gamma)$, and $(\hat{\theta}_{\text{MLE}}, \hat{\tau}_{\text{MLE}})$ for the DLDEM with single delay with $h = 0.1$.

M	$\text{Max.Val}(l)$	$\nabla l(\Gamma)$	$(\hat{\theta}_{10, \text{MLE}}, \hat{\theta}_{20, \text{MLE}}, \hat{\tau}_{0, \text{MLE}})$	$(\hat{\theta}_{1, \text{MLE}}, \hat{\theta}_{2, \text{MLE}}, \hat{\tau}_{\text{MLE}})$
100	0.0048	$10^{-5} \times \begin{bmatrix} -0.4019 \\ -0.2653 \\ 0.9364 \end{bmatrix}$	(0.5, 0.7, 2.5)	(0.5007149, 0.7006367, 2.5004988)

at discrete time intervals of width $h = 0.1$ starting from $t_0 = 0$. The endpoint considered is $t_n = 10$ corresponding to $n = 100$ where the number of sampled time points are $(n + 1)$. Figure 2(b) shows the underlying trajectory of the solution $x(t)$ from the DDE model (41) and the $(n+1)$ sampled observations for the time range selected.

The initial grid space for the adaptive grid procedure was taken to be $\Theta^{(0)} = \{(\theta_{1u}, \theta_{2v}, \tau_r) : \theta_{1u} = \theta_{10} + uh, \theta_{2v} = \theta_{20} + vh, \tau_r = \tau_0 + rh\}$ with $(\theta_{10}, \theta_{20}, \tau_0) = (0.4, 0.6, 2.4)$, $h = 0.1$ and $U = V = R = 3$. The stopping criteria threshold δ was chosen to be 0.0001. The adaptive grid and Newton-Raphson procedures were run; the results are given in Tables 4 and 5. As in the previous example, as M increases, the $\hat{a}(\hat{\theta}_1, \hat{\theta}_2, \hat{\tau})$ becomes more accurate and very close to be the true value $a(\theta_1, \theta_2, \tau) = 0.5$ but at the cost of increased computational time.

The Fisher information matrix as $h = 0.1$ and $M = 100$ for DLDEM with single delay is

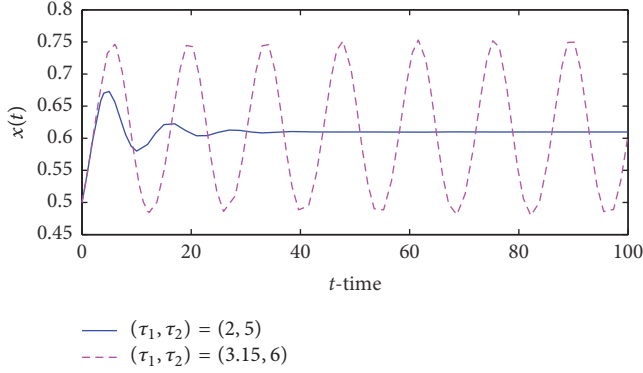
$$I(\Gamma) = 10^6 \times \begin{bmatrix} 3.7030 & -2.2042 & -0.1819 & -0.5757 & 0 \\ -2.2042 & 1.4665 & 0.2069 & 0.3285 & 0 \\ -0.1819 & 0.2069 & 0.0774 & 0.0059 & 0 \\ -0.5757 & 0.3285 & 0.0059 & 0.3164 & 0 \\ 0 & 0 & 0 & 0 & 5.6613 \times 10^3 \end{bmatrix}, \quad (42)$$

and variance of $\hat{\Gamma}$ by using MLE is given by

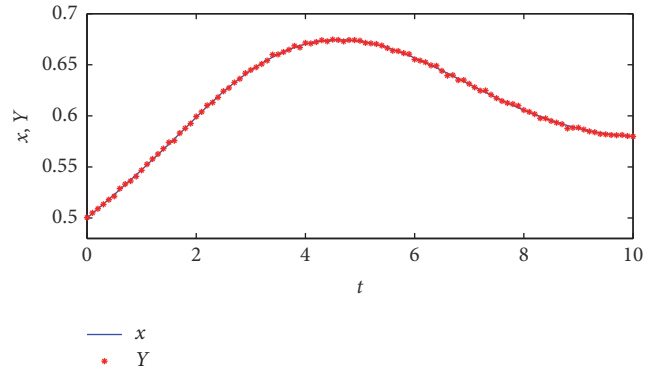
$$\text{Var}(\hat{\Gamma}) = 10^{-3} \times \begin{bmatrix} 0.0241 & 0.0463 & -0.0669 & -0.0030 & 0 \\ 0.0463 & 0.0907 & -0.1330 & -0.0075 & 0 \\ -0.0669 & -0.1330 & 0.2102 & 0.0125 & 0 \\ -0.0030 & -0.0075 & 0.0125 & 0.0052 & 0 \\ 0 & 0 & 0 & 0 & 1.7664 \times 10^{-7} \end{bmatrix}. \quad (43)$$

TABLE 6: Table showing the 95% confidence intervals for parameters $\theta_1, \theta_2, \tau, a$, and σ^2 in DLDE with single delay.

M	$(\theta_{1L}, \theta_{1U})$	$(\theta_{2L}, \theta_{2U})$	(τ_L, τ_U)	(a_L, a_U)	(σ_L^2, σ_U^2)
100	(0.4904, 0.5096)	(0.6813, 0.7187)	(2.4716, 2.5284)	(0.4959, 0.5048)	(0.00007, 0.00012)



(a) Numerical solution of the DLDEM with two delays (45) using the fixed parameter values in $[0, 100]$ by step size $h = 0.1$ with $\theta_1 = 0.5, \theta_2 = 0.7$, and $\theta_3 = 0.12$. The steady state is stable when $\tau_1 = 2$ and $\tau_2 = 5$ and becomes unstable when $\tau_1 = 3.15$ and $\tau_2 = 6$



(b) Numerical solution of the DLDEM with two delays (45) using the estimated parameter values. The stars are the simulated noisy data by adding noises to the DLDE solutions at $n + 1 = 101$ equally spaced time points in $[0, 10]$ by step size $h = 0.1$ with $\theta_1 = 0.5, \theta_2 = 0.7, \theta_3 = 0.12, \tau_1 = 2$, and $\tau_2 = 5$

FIGURE 3

The 95% confidence intervals for parameters $(\theta_1, \theta_2, \tau, a, \sigma^2)$ in DLDE with single delay are shown in Table 6.

4.3. *Example 3.* The delayed logistic differential equation model (DLDEM) with two delays proposed by Braddock and van den Driessche [17] is the solution to the DDE

$$\dot{x}(t) = \theta x(t) \left(1 - \frac{x(t - \tau_1)}{k_1} - \frac{x(t - \tau_2)}{k_2} \right), \quad (44)$$

where θ, k_1, k_2, τ_1 , and τ_2 are positive constants. DDEs with two delays appear in many applications such as epidemiological models [18], physiological models [19], neurological models [20], and medical models [21]. In such equations [22, 23], very rich dynamics have been observed. Denoting $\theta_1 = \theta, \theta_2 = \theta/k_1, \theta_3 = \theta/k_2, z_1(t) = x(t - \tau_1)$, and $z_2(t) = x(t - \tau_2)$, we obtain

$$\dot{x}(t) = \theta_1 x(t) - \theta_2 z_1(t) - \theta_3 z_2(t). \quad (45)$$

By using the MATLAB function `dde23` with fixed parameters $(\theta_1, \theta_2, \theta_3, \tau_1, \tau_2, a, \sigma^2)$ values, we obtain the trajectories of

the solution $x(t)$. As in Example 2, we note the different characteristics of the solution depending on the parameter specifications as shown in Figure 3(a). Subsequently, the parameters are fixed at $(0.5, 0.7, 0.12, 2, 5, 0.5, (0.001)^2)$ and observations y_0, y_1, \dots, y_n are collected at the $(n + 1)$ sampled time points $T_0 = t_0 < t_1 < \dots < t_n = T_1$, based on the observational model (1) at discrete time intervals of width $h = 0.1$ starting from $t_0 = 0$. The endpoint considered is $t_n = 10$ corresponding to $n + 1 = 101$. Figure 3(b) shows the underlying trajectories of the solution $x(t)$ from the DDE model (41) and the $(n + 1)$ sampled observations.

The initial grid space for the adaptive grid procedure is taken to be $\Theta^{(0)} = \{(\theta_{1u}, \theta_{2v}, \theta_{3w}, \tau_{1r}, \tau_{2s}) : \theta_{1u} = \theta_{10} + uh, \theta_{2v} = \theta_{20} + vh, \theta_{3w} = \theta_{30} + wh, \tau_{1r} = \tau_{10} + rh, \tau_{2s} = \tau_{20} + sh\}$ with $(\theta_{10}, \theta_{20}, \theta_{30}, \tau_{10}, \tau_{20}) = (0.4, 0.6, 0.02, 1.9, 4.9)$, $h = 0.1$ and $U = V = W = R = S = 3$. The stopping criteria threshold δ was chosen to be 0.0001. The adaptive grid and Newton-Raphson procedures were run; the results are given in Tables 7 and 8.

The Fisher information matrix as $h = 0.1$ for DLDEM with two delays at $M = 100$ is

$$I(\Gamma) = 10^8 \times \begin{bmatrix} 1.5952 & -0.9359 & -0.8343 & -0.2912 & -0.0098 & -0.2324 & 0 \\ -0.9359 & 0.5630 & 0.4973 & 0.1809 & 0.0074 & 0.1328 & 0 \\ -0.8343 & 0.4973 & 0.4432 & 0.1619 & 0.0063 & 0.1121 & 0 \\ -0.2912 & 0.1809 & 0.1619 & 0.0711 & 0.0033 & 0.0408 & 0 \\ -0.0098 & 0.0074 & 0.0063 & 0.0033 & 0.0003 & 0.0001 & 0 \\ -0.2324 & 0.1328 & 0.1121 & 0.0408 & 0.0001 & 0.1519 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 305361 \end{bmatrix}, \quad (46)$$

TABLE 7: Table showing value of $\text{Max.Val}(l)$, $\hat{a}(\theta_1, \theta_2, \theta_3, \tau_1, \tau_2)$, $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\tau}_1, \hat{\tau}_2)$, and $\hat{\sigma}^2$ in DLDEM with two delays by using an adaptive grid for $(\theta_1, \theta_2, \theta_3, \tau_1, \tau_2) = (0.5, 0.7, 0.12, 2, 5)$ and $n + 1 = 101$ equally spaced time points in $[0, 10]$ at $h = 0.1$.

M	$\text{Max.Val}(l)$	$\hat{a}(\theta_1, \theta_2, \theta_3, \tau_1, \tau_2)$	$(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\tau}_1, \hat{\tau}_2)$	$\hat{\sigma}^2$
10	-6.4697×10^{-5}	0.5007	(0.5, 0.7, 0.12, 2, 5)	1.2686×10^{-6}
20	-6.5220×10^{-5}	0.5006	(0.5, 0.7, 0.12, 2, 5)	1.2788×10^{-6}
30	-6.5478×10^{-5}	0.5005	(0.5, 0.7, 0.12, 2, 5)	1.2839×10^{-6}
40	-6.5622×10^{-5}	0.5005	(0.5, 0.7, 0.12, 2, 5)	1.2867×10^{-6}
50	-6.5714×10^{-5}	0.5005	(0.5, 0.7, 0.12, 2, 5)	1.2885×10^{-6}
100	-6.5910×10^{-5}	0.5005	(0.5, 0.7, 0.12, 2, 5)	1.2923×10^{-6}
1000	-6.6098×10^{-5}	0.5004	(0.5, 0.7, 0.12, 2, 5)	1.2960×10^{-6}

TABLE 8: Table showing value of $\text{Max.Val}(l)$, $\nabla l(\Gamma)$, and $(\hat{\theta}_{1\text{MLE}}, \hat{\theta}_{2\text{MLE}}, \hat{\theta}_{3\text{MLE}}, \hat{\tau}_{1\text{MLE}}, \hat{\tau}_{2\text{MLE}})$ for DLDEM with two delays with $h = 0.1$.

M	$\text{Max.Val}(l)$	$\nabla l(\Gamma)$	$(\hat{\theta}_{10\text{MLE}}, \hat{\theta}_{20\text{MLE}}, \hat{\theta}_{30\text{MLE}}, \hat{\tau}_{10\text{MLE}}, \hat{\tau}_{20\text{MLE}})$	$(\hat{\theta}_{1\text{MLE}}, \hat{\theta}_{2\text{MLE}}, \hat{\theta}_{3\text{MLE}}, \hat{\tau}_{1\text{MLE}}, \hat{\tau}_{2\text{MLE}})$
100	6.4697×10^{-5}	$10^{-4} \times \begin{bmatrix} 0.1600 \\ -0.1495 \\ -0.1815 \\ -0.2596 \\ -0.0146 \end{bmatrix}$	(0.5, 0.7, 0.12, 2, 5)	(0.5001029, 0.6999608, 0.1199824, 2.0000706, 5.0000044)

and variance of $\hat{\Gamma}$ by using MLE is given by

$$\text{Var}(\hat{\Gamma}) = 10^{-3} \times \begin{bmatrix} 0.0642 & 0.0331 & 0.0983 & -0.0161 & -0.6515 & 0.0014 & 0 \\ 0.0331 & 0.0215 & 0.0457 & -0.0063 & -0.3586 & 0.0002 & 0 \\ 0.0983 & 0.0457 & 0.1569 & -0.0275 & -0.9709 & 0.0027 & 0 \\ -0.0161 & -0.0063 & -0.0275 & 0.0066 & 0.1444 & -0.0007 & 0 \\ -0.6515 & -0.3586 & -0.9709 & 0.1444 & 6.8315 & -0.0102 & 0 \\ 0.0014 & 0.0002 & 0.0027 & -0.0007 & -0.0102 & 0.0002 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3.3000 \times 10^{-11} \end{bmatrix}. \quad (47)$$

The 95% confidence intervals for parameters $(\theta_1, \theta_2, \theta_3, \tau_1, \tau_2, a, \sigma^2)$ in DLDE with two delays are shown in Table 9.

Remark 2. As mentioned earlier, the adaptive grid procedure in the first stage of our two-step procedure needs to select a sufficiently large domain that is likely to contain the MLE. The MLE should be close to the true parameter values that generated the data as standard MLE theory [11–13] dictates. This has also been established in the three examples considered. Hence, since the true value of (θ, τ) is known in our simulation examples, we selected the initial domain of the grid procedure to contain these true values in its interior. Thus, the notation (θ_0, τ_0) denotes the lower bound of the parameters which is used for the grid procedure. The grid $\Theta^{(0)} = \{(\theta_r, \tau_s) : \theta_r = \theta_0 + rh, \tau_s = \tau_0 + sh, r = 1, 2, \dots, R; s = 1, 2, \dots, S\}$ is ensured to contain the true parameter values based on selection of R, S , and h in all the

examples. Other than this consideration, the true values that were used in the simulation were selected rather arbitrarily, only chosen so as to be representative parameter values that exhibit the typical nature of trajectories of the underlying DDEs as shown in the figures.

5. Conclusion

In this paper, we presented the method of maximum likelihood for estimating parameters in delayed differential equations. As examples we considered the exponential differential equation model, delayed logistic differential equation model with single delay, and delayed logistic differential equation model with two delays; then we estimated the unknown parameters in these models. Two-step approach using an adaptive grid followed by a gradient descent procedure is proposed. Our methodology estimates the delay parameter

TABLE 9: Table showing the 95% confidence intervals for parameters $\theta_1, \theta_2, \theta_3, \tau_1, \tau_2, a$, and σ^2 in DLDE with two delays.

M	100
$(\theta_{1L}, \theta_{1U})$	(0.4843, 0.5157)
$(\theta_{2L}, \theta_{2U})$	(0.6909, 0.7091)
$(\theta_{3L}, \theta_{3U})$	(0.0955, 0.1445)
(τ_{1L}, τ_{1U})	(1.9950, 2.0050)
(τ_{2L}, τ_{2U})	(4.8380, 5.1620)
(a_L, a_U)	(0.4996, 0.5013)
(σ_L^2, σ_U^2)	(0.0000009, 0.0000016)

as well as the initial starting value of the dynamical system correctly based on simulation data. Confidence intervals and information matrix by using maximum likelihood are obtained and are found to contain the true values of the parameter based on simulation data.

In this paper, we took the initial value function of the DDE as an unknown constant “ a ”. However, it is possible to extend the constant initial value assumption to a more general linear or nonlinear function, say $\varphi(x)$, for $x \in [-\tau, 0]$. Two complications arise here. First, we need additional unknown parameters to represent $\varphi(x)$; for example, if $\varphi(x) = a + bx$ is chosen to be linear, we have to estimate parameter b in addition to a . Higher order functions offer greater flexibility in modeling the initial function but at the expense of estimating extra parameters and slowing down the computational procedure. A second issue that follows the first is the selection of a “best” initial value function—either constant, linear, quadratic, or others. Thus, further research is required to address this concern and we hope to report some results in this direction in the future.

Appendix

A. Quasi-Newton Procedure

As input to the quasi-Newton procedure, we require to compute the gradient vector, $\nabla l_0(\Gamma)$, as given in (20), which consists of the partial derivatives of the log-likelihood function with respect to the entries of Γ . Recall that the DDEM has multiple delays parameters given by $\tau = (\tau_1, \tau_2, \dots, \tau_m)$, evolution parameters given by $\theta = (\theta_1, \theta_2, \dots, \theta_p)$, and unknown initial condition a . Based on the log-likelihood

$$l(\theta, \tau, a) = -\frac{1}{2\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \theta, \tau, a))^2, \quad (\text{A.1})$$

recall that we define $l(\theta, \tau, \hat{a}(\theta, \tau))$ as

$$l(\theta, \tau, \hat{a}(\theta, \tau)) = -\frac{1}{2\sigma^2} \sum_{i=0}^n (y_i - x(t_i, \theta, \tau, \hat{a}(\theta, \tau)))^2. \quad (\text{A.2})$$

Denoting $l_0(\theta, \tau) \equiv l(\theta, \tau, \hat{a}(\theta, \tau))$ and $x_i = x(t_i, \theta, \tau, \hat{a}(\theta, \tau))$, the first-order partial derivatives are as follows:

$$\frac{\partial l_0}{\partial \theta_u} = \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x_i) \left[\frac{\partial x_i}{\partial \theta_u} + \frac{\partial x_i}{\partial a} \frac{\partial \hat{a}}{\partial \theta_u} \right], \quad (\text{A.3})$$

$$\frac{\partial l_0}{\partial \tau_v} = \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x_i) \left[\frac{\partial x_i}{\partial \tau_v} + \frac{\partial x_i}{\partial a} \frac{\partial \hat{a}}{\partial \tau_v} \right].$$

The above equations involve derivatives of x_i with respect to the parameters. Each derivative expression of $\partial x_i / \partial \theta_u$, $\partial x_i / \partial \tau_v$, and $\partial x_i / \partial a$ for $i = 0, 1, 2, \dots, n$ can be numerically obtained from respective DDEs which are derived from the initial model in (2) by differentiating it with respect to the quantity of interest. In the case of $\partial x_i / \partial \theta_u$, differentiating (2) with respect to θ_u gives a new DDE for $\partial x / \partial \theta_u$:

$$\left(\frac{\partial x}{\partial \theta_u} \right) = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \theta_u} + \sum_{j=1}^m \frac{\partial f}{\partial z_j} \frac{\partial z_j}{\partial \theta_u} + \frac{\partial f}{\partial \theta_u}, \quad (\text{A.4})$$

where $(\partial x / \partial \theta_u)$ is the derivative of $\partial x / \partial \theta_u$ with respect to t and $\partial z_j / \partial \theta_u$ is the delayed version of $\partial x / \partial \theta_u$; that is,

$$\frac{\partial z_j}{\partial \theta_u}(t) = \frac{\partial x}{\partial \theta_u}(t - \tau_j). \quad (\text{A.5})$$

The initial condition for the DDE in (A.4) is $\partial x / \partial \theta_u = 0$ since the derivative of the initial value a with respect to θ_u is 0. Based on this initial condition, the above DDE model can be numerically solved and the values of $\partial x_i / \partial \theta_u$ can be obtained from $\partial x(t) / \partial \theta_u$ for each $t = t_i$, $i = 0, 1, 2, \dots, n$.

In a similar way, each value of $\partial x_i / \partial \tau_v$ can be determined by differentiating (2) with respect to τ_v . The new DDE for $\partial x / \partial \tau_v$ is

$$\left(\frac{\partial x}{\partial \tau_v} \right) = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \tau_v} + \sum_{j=1}^m \frac{\partial f}{\partial z_j} \left[\frac{\partial x(t - \tau_j)}{\partial \tau_v} - f(t - \tau_v), \right. \\ \left. x(t - \tau_v), z_1(t - \tau_v), z_2(t - \tau_v), \dots, z_m(t - \tau_v), \right. \\ \left. \theta \right], \quad (\text{A.6})$$

where $(\partial x / \partial \tau_v)$ is the derivative of $\partial x / \partial \tau_v$ with respect to t and $\partial z_j / \partial \tau_v$ is the delayed version of $\partial x / \partial \tau_v$; that is,

$$\frac{\partial z_j}{\partial \tau_v}(t) = \frac{\partial x(t - \tau_j)}{\partial \tau_v} - f(t - \tau_v, x(t - \tau_v), \\ z_1(t - \tau_v), z_2(t - \tau_v), \dots, z_m(t - \tau_v), \theta). \quad (\text{A.7})$$

The initial condition for the DDE in (A.6) is $\partial x / \partial \tau_v = 0$ since, again, the derivative of the initial value a with respect to τ_v is 0. Based on this initial condition, the above DDE model can be numerically solved and the values of $\partial x_i / \partial \tau_v$ can be obtained from $\partial x(t) / \partial \tau_v$ for each $t = t_i$, $i = 0, 1, 2, \dots, n$. The case of $\partial x_i / \partial a$ is similar and has been discussed in the main text when presenting the Newton-Raphson procedure.

The expressions in (A.3) also involve $\partial\hat{a}/\partial\theta_u$ and $\partial\hat{a}/\partial\tau_v$. They can be obtained from differentiating the equation satisfied by $\hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau})$ in (10) for every pair $(\boldsymbol{\theta}, \boldsymbol{\tau})$:

$$\frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\tau}, \hat{a}(\boldsymbol{\theta}, \boldsymbol{\tau}))}{\partial a} = 0 \quad \forall \boldsymbol{\theta}, \boldsymbol{\tau}. \quad (\text{A.8})$$

Differentiating with respect to θ_u , we get

$$\frac{\partial^2 l}{\partial \theta_u \partial a} + \frac{\partial^2 l}{\partial a^2} \frac{\partial \hat{a}}{\partial \theta_u} = 0; \quad (\text{A.9})$$

thus

$$\frac{\partial \hat{a}}{\partial \theta_u} = -\frac{\partial^2 l / \partial \theta_u \partial a}{\partial^2 l / \partial a^2} \quad (\text{A.10})$$

for $1 \leq u \leq p$. Similarly, from differentiating (10) with respect to τ_v , we get

$$\frac{\partial^2 l}{\partial \tau_v \partial a} + \frac{\partial^2 l}{\partial a^2} \frac{\partial \hat{a}}{\partial \tau_v} = 0, \quad (\text{A.11})$$

and hence,

$$\frac{\partial \hat{a}}{\partial \tau_v} = -\frac{\partial^2 l / \partial \tau_v \partial a}{\partial^2 l / \partial a^2}, \quad (\text{A.12})$$

for $1 \leq v \leq m$. We give the explicit expressions for the second-order derivatives of l with respect to its arguments in Appendix B (as given by (B.6), (B.7), and (B.11) in Appendix B).

B. Information Matrix

For the DDE in (2) with multiples delays $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_m)$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$, recall that $l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2) = -((n+1)/2) \ln(2\pi\sigma^2) - (1/2\sigma^2) \sum_{i=0}^n (y_i - x_i)^2$, where $x_i = x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)$. The first-order partial derivatives of $l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2)$ are as follows:

$$\begin{aligned} \frac{\partial l}{\partial \theta_u} &= \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x_i) \left(\frac{\partial x_i}{\partial \theta_u} \right), \\ \frac{\partial l}{\partial \tau_v} &= \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x_i) \left(\frac{\partial x_i}{\partial \tau_v} \right), \\ \frac{\partial l}{\partial a} &= \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x_i) \left(\frac{\partial x_i}{\partial a} \right) \end{aligned} \quad (\text{B.1})$$

for $1 \leq u \leq p$, $1 \leq v \leq m$. As mentioned in Appendix A, these partial derivatives of $l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2)$ can be evaluated numerically from the derivative expression of $\partial x_i / \partial \theta_u$, $\partial x_i / \partial \tau_v$, and $\partial x_i / \partial a$ for $i = 0, 1, 2, \dots, n$, since each of them forms an additional DDE derived from (2) by differentiating it with respect to the quantity of interest. Further we have

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2\sigma^2} \left[(n+1) - \frac{1}{\sigma^2} \sum_{i=0}^n (y_i - x_i)^2 \right]. \quad (\text{B.2})$$

Differentiating the above again with respect to the arguments of $l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2)$, the second-order partial derivatives have the general forms

$$\frac{\partial^2 l}{\partial \theta_u \partial \theta_{u^*}} = \frac{1}{\sigma^2} \left[\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial \theta_u \partial \theta_{u^*}} \right) \right. \quad (\text{B.3})$$

$$\left. - \sum_{i=0}^n \left(\frac{\partial x_i}{\partial \theta_u} \right) \left(\frac{\partial x_i}{\partial \theta_{u^*}} \right) \right],$$

$$\frac{\partial^2 l}{\partial \theta_u \partial \tau_{v^*}} = \frac{1}{\sigma^2} \left[\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial \theta_u \partial \tau_{v^*}} \right) \right. \quad (\text{B.4})$$

$$\left. - \sum_{i=0}^n \left(\frac{\partial x_i}{\partial \theta_u} \right) \left(\frac{\partial x_i}{\partial \tau_{v^*}} \right) \right],$$

$$\frac{\partial^2 l}{\partial \tau_v \partial \tau_{v^*}} = \frac{1}{\sigma^2} \left[\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial \tau_v \partial \tau_{v^*}} \right) \right. \quad (\text{B.5})$$

$$\left. - \sum_{i=0}^n \left(\frac{\partial x_i}{\partial \tau_v} \right) \left(\frac{\partial x_i}{\partial \tau_{v^*}} \right) \right],$$

$$\frac{\partial^2 l}{\partial \theta_u \partial a} = \frac{1}{\sigma^2} \left[\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial \theta_u \partial a} \right) \right. \quad (\text{B.6})$$

$$\left. - \sum_{i=0}^n \left(\frac{\partial x_i}{\partial \theta_u} \right) \left(\frac{\partial x_i}{\partial a} \right) \right],$$

$$\frac{\partial^2 l}{\partial \tau_v \partial a} = \frac{1}{\sigma^2} \left[\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial \tau_v \partial a} \right) \right. \quad (\text{B.7})$$

$$\left. - \sum_{i=0}^n \left(\frac{\partial x_i}{\partial \tau_v} \right) \left(\frac{\partial x_i}{\partial a} \right) \right],$$

$$\frac{\partial^2 l}{\partial \theta_u \partial \sigma^2} = \frac{-1}{(\sigma^2)^2} \sum_{i=0}^n (y_i - x_i) \left(\frac{\partial x_i}{\partial \theta_u} \right), \quad (\text{B.8})$$

$$\frac{\partial^2 l}{\partial \tau_v \partial \sigma^2} = \frac{-1}{(\sigma^2)^2} \sum_{i=0}^n (y_i - x_i) \left(\frac{\partial x_i}{\partial \tau_v} \right), \quad (\text{B.9})$$

$$\frac{\partial^2 l}{\partial a \partial \sigma^2} = \frac{-1}{(\sigma^2)^2} \sum_{i=0}^n (y_i - x_i) \quad (\text{B.10})$$

$$\frac{\partial^2 l}{\partial a^2} = \frac{1}{\sigma^2} \left[\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial a^2} \right) - \sum_{i=0}^n \left(\frac{\partial x_i}{\partial a} \right)^2 \right], \quad (\text{B.11})$$

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{1}{2(\sigma^2)^2} \left[(n+1) - \frac{2}{\sigma^2} \sum_{i=0}^n (y_i - x_i)^2 \right]. \quad (\text{B.12})$$

Equations (B.6), (B.7), and (B.10) are required for the quasi-Newton procedure in Appendix A.

To obtain the information matrix, we need to take the expectation of the negative of the second-order derivatives of $l(\boldsymbol{\theta}, \boldsymbol{\tau}, a, \sigma^2) = l(\boldsymbol{\Gamma} | \mathbf{y})$ with respect to its arguments:

$$I(\boldsymbol{\Gamma})_{(u,v)} = -E \left[\frac{\partial^2 l(\boldsymbol{\Gamma} | \mathbf{y})}{\partial \Gamma_u \partial \Gamma_v} \right]. \quad (\text{B.13})$$

Taking negative on the LHS of (B.3)–(B.12), followed by expectation under the sampling distribution of each y_i , we note that the general terms of the form

$$\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial \Gamma_u \partial \Gamma_v} \right) \quad (\text{B.14})$$

or

$$\sum_{i=0}^n (y_i - x_i) \left(\frac{\partial^2 x_i}{\partial \Gamma_u} \right) \quad (\text{B.15})$$

on the RHS become zero since the expectation of y_i equals x_i . Hence we get

$$\begin{aligned} I(\boldsymbol{\Gamma})_{(u,v)} &= -E \left[\frac{\partial^2 l(\boldsymbol{\Gamma} | \mathbf{y})}{\partial \Gamma_u \partial \Gamma_v} \right] \\ &= \frac{1}{\sigma^2} \sum_{i=0}^n \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_u} \right) \left(\frac{\partial x(t_i, \boldsymbol{\theta}, \boldsymbol{\tau}, a)}{\partial \Gamma_v} \right) \end{aligned} \quad (\text{B.16})$$

as in (30) as well as

$$I(\boldsymbol{\Gamma})_{(u,\sigma^2)} = -E \left[\frac{\partial^2 l(\boldsymbol{\Gamma} | \mathbf{y})}{\partial \Gamma_u \partial \sigma^2} \right] = 0. \quad (\text{B.17})$$

Finally, taking expectation in (B.12),

$$\begin{aligned} E \left(-\frac{\partial^2 l}{\partial (\sigma^2)^2} \right) &= -\frac{1}{2(\sigma^2)^2} \left[(n+1) - \frac{2}{\sigma^2} \sum_{i=0}^n E(y_i - x_i)^2 \right] \\ &= -\frac{1}{2(\sigma^2)^2} \left[(n+1) - \frac{2}{\sigma^2} (n+1)\sigma^2 \right] = \frac{(n+1)}{2(\sigma^2)^2}. \end{aligned} \quad (\text{B.18})$$

Conflicts of Interest

The authors declare that there are no conflicts of interest with respect to the research, authorship, and/or publication of this paper.

Acknowledgments

The authors would like to thank Universiti Teknologi PETRONAS for the financial assistance under ERGS and HiCOE grants from the Ministry of Education Malaysia, as well as resource facilities provided by Center for Intelligent Signal and Imaging Research (CISIR).

References

- [1] Y. Kuang, *Delay Differential Equations with Applications in Population Dynamics*, Academic Press, New York, NY, USA, 1993.
- [2] J. J. Batzel and H. T. Tran, “Stability of the human respiratory control system. I. Analysis of a two-dimensional delay state-space model,” *Journal of Mathematical Biology*, vol. 41, no. 1, pp. 45–79, 2000.
- [3] T. Kalmár-Nagy, G. Stépán, and F. C. Moon, “Subcritical Hopf bifurcation in the delay equation model for machine tool vibrations,” *Nonlinear Dynamics*, vol. 26, no. 2, pp. 121–142, 2001.
- [4] A. Bellen and M. Zennaro, *Numerical Methods for Delay Differential Equations*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, UK, 2013.
- [5] S. P. Ellner, B. E. Kendall, S. N. Wood, E. McCauley, and C. J. Briggs, “Inferring mechanism from time-series data: Delay-differential equations,” *Physica D: Nonlinear Phenomena*, vol. 110, no. 3–4, pp. 182–194, 1997.
- [6] S. N. Wood, “Partially specified ecological models,” *Ecological Monographs*, vol. 71, no. 1, pp. 1–25, 2001.
- [7] L. Wang and J. Cao, “Estimating parameters in delay differential equation models,” *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 17, no. 1, pp. 68–83, 2012.
- [8] S. Mehrkanoon, S. Mehrkanoon, and J. A. Suykens, “Parameter estimation of delay differential equations: an integration-free LS-SVM approach,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 4, pp. 830–841, 2014.
- [9] R. A. Fisher, “On the mathematical foundations of theoretical statistics,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 222, no. 594–604, pp. 309–368, 1922.
- [10] R. A. Fisher, “Theory of statistical estimation,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 22, no. 5, pp. 700–725, 1925.
- [11] P. J. Huber, “The behavior of maximum likelihood estimates under nonstandard conditions,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
- [12] A. Wald, “Note on the consistency of the maximum likelihood estimate,” *Annals of Mathematical Statistics*, vol. 20, no. 4, pp. 595–601, 1949.
- [13] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in *Selected papers of Hirotugu Akaike*, pp. 199–213, Springer, 1998.
- [14] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Science & Business Media, 1998.
- [15] A. Spanos, *Probability Theory and Statistical Inference*, Cambridge University Press, Cambridge, UK, 1999.
- [16] G. E. Hutchinson, “Circular causal systems in ecology,” *Annals of the New York Academy of Sciences*, vol. 50, no. 4, pp. 221–246, 1948.
- [17] R. D. Braddock and P. van den Driessche, “On a two-lag differential delay equation,” *Australian Mathematical Society. Journal. Series B. Applied Mathematics*, vol. 24, no. 3, pp. 292–317, 1982/83.
- [18] K. L. Cooke and J. A. Yorke, “Some equations modelling growth processes and gonorrhoea epidemics,” *Mathematical Biosciences*, vol. 16, pp. 75–101, 1973.
- [19] A. Beuter, J. Bélair, C. Labrie, and J. Bélair, “Feedback and delays in neurological diseases: A modeling study using dynamical

- systems,” *Bulletin of Mathematical Biology*, vol. 55, no. 3, pp. 525–541, 1993.
- [20] J. Bélair and S. A. Campbell, “Stability and bifurcations of equilibria in a multiple-delayed differential equation,” *SIAM Journal on Applied Mathematics*, vol. 54, no. 5, pp. 1402–1424, 1994.
- [21] J. Bélair, M. C. Mackey, and J. M. Mahaffy, “Age-structured and two-delay models for erythropoiesis,” *Mathematical Biosciences*, vol. 128, no. 1-2, pp. 317–346, 1995.
- [22] J. K. Hale and W. Z. Huang, “Global geometry of the stable regions for two delay differential equations,” *Journal of Mathematical Analysis and Applications*, vol. 178, no. 2, pp. 344–362, 1993.
- [23] J. M. Mahaffy, K. M. Joiner, and P. J. Zak, “A geometric analysis of stability regions for a linear differential equation with two delays,” *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 5, no. 3, pp. 779–796, 1995.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

