Statistica Sinica **33** (2023), 1853-1878
doi:https://doi.org/10.5705/ss.202020.0429

MODEL CHECKING FOR PARAMETRIC ORDINARY DIFFERENTIAL EQUATIONS SYSTEMS

Ran Liu, Yun Fang and Lixing Zhu

Hong Kong Baptist University, Shanghai Normal University and Beijing Normal University at Zhuhai

Abstract: Model checking for parametric ordinary differential equations is a necessary step when checking whether the assumed models are plausible. In this paper, we first introduce a trajectory matching-based test for the whole model, which can also easily be applied to check partially observed systems. Then, we provide two tests to identify which component function is modeled incorrectly. The first is based on integral matching, and the second is based on gradient matching, with bias correction achieved using data splitting. We investigate their asymptotic properties under the null, global, and local alternative hypotheses. Because there are no results for relevant parameter estimations for alternative models in the literature, we also investigate the asymptotic properties of the nonlinear least squares estimation and the two-step estimation under both the null and the alternatives. To examine the performance of the tests, we conduct several numerical simulations and an analysis using a real-data example on immune cell kinetics and trafficking for influenza infection.

Key words and phrases: Local smoothing test, model checking, ordinary differential equations.

1. Introduction

To model how systems evolve over time, ordinary differential equations (ODEs) are widely applied in scientific fields such as physics, ecology (Goel, Maitra and Montroll (1971)), and neuroscience (FitzHugh (1961); Nagumo, Arimoto and Yoshizawa (1962)). A system of ODEs can be written as

$$X'(t) = \begin{bmatrix} \frac{dX_1(t)}{dt} \\ \vdots \\ \frac{dX_p(t)}{dt} \end{bmatrix} = \begin{bmatrix} f_1(t, X(t); \theta) \\ \vdots \\ f_p(t, X(t); \theta) \end{bmatrix} = f(t, X(t); \theta), \quad t \in [t_0, T], \tag{1.1}$$

Corresponding author: Lixing Zhu, Advanced Institute of Natural Sciences, Beijing Normal University at Zhuhai, Zhuhai 519085, China. E-mail: lzhu@bnu.edu.cn.

with an initial condition $X(t_0) = x_0$. Here, $X(t) = (X_1(t), \ldots, X_p(t))^\top$ is a p-dimensional state vector, and $f(t, X(t); \theta)$ is supposed to belong to a given parametric family of functions $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta \subset R^q\}$. Frequently, this system is measured on discrete time points with noise, say,

$$Y_i = X(t_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.2}$$

where the measurement error ϵ_i satisfying $E(\epsilon_i \mid t_i) = 0$ has a nonsingular variance-covariance matrix Σ_{ϵ_i}, and is independent of ϵ_j, for every $j \neq i$. The observation process may be more complicated in real-life applications. For example, the observable variable might be a combination or a functional of state variables. In this study, we focus on the simple observation process case represented by (1.2).

Here, we want to check

$$H_0 : X'(t) = f(t, X(t); \theta_0) \in \mathcal{F} \quad \text{versus} \quad H_1 : X'(t) \notin \mathcal{F}, \tag{1.3}$$

where θ_0 is an unknown parameter vector. If we reject H_0, we may further wish to identify the component(s) that are modeled incorrectly. In this situation, for the kth component function, the hypotheses are as follows:

$$H_{0k} : X_k'(t) = f_k(t, X(t); \theta_{0k}) \in \mathcal{F}_k \quad \text{versus} \quad H_{1k} : X_k'(t) \notin \mathcal{F}_k, \tag{1.4}$$

with $\mathcal{F}_k = \{f_k(\cdot, \theta_k) : \theta_k \in \Theta_k \subset R^{q_k}\}$. Here, we construct omnibus tests to check the whole parametric ODE system (1.3) and the individual components (1.4).

To this end, we first review relevant model checking methodologies for classical regressions in the literature, which motivate the new test constructions we need. For univariate response cases, there exist two broad classes of tests. First, the so-called local smoothing tests are constructed by using nonparametric estimations; see, for example, Härdle and Mammen (1993), Zheng (1996), Dette (1999), and Lavergne and Patilea (2012). Empirically, tests in this class are sensitive to alternative models that are oscillating/highly frequent, in general. Tests in the second class are based on residual-marked empirical processes, and take averages over an index set. Because averaging is a global smoothing step, these are called global smoothing tests; see, for example, Stute (1997), Stute, Thies and Zhu (1998), Zhu (2003), and Khmaladze and Koul (2004). Such tests have better asymptotic properties, but are less sensitive to oscillating alternative models; see González-Manteiga and Crujeiras (2013) for a comprehensive review. For multi-response regressions, Chen and van Keilegom (2009) constructed an empirical likelihood ratio test based on local smoothing.

However, the testing problems investigated here are rather different from those for classical parametric regression models. This is because the model structure is not directly assumed on the unknown function $X(\cdot)$, but on its derivative $X'(\cdot)$, and only part of the whole ODE system may be observed. Furthermore, any component of $X'(\cdot)$ is related to the whole original function $X(\cdot)$, rather than to any single corresponding component of $X(\cdot)$. This structure makes the testing problems complicated. We discuss these issues in the following sections. In the literature, to the best of our knowledge, there are only two relevant works on the ODE model checking problem. Hooker (2009) proposed a goodness-of-fit test based on estimated forcing functions for the whole ODE system. The ODE system is transferred to a multivariate linear model under the null hypothesis against this model, adding an empirical forcing function represented as a basis expansion under the alternative hypothesis. Then, a likelihood-ratio test is used to check the whole ODE system. Under normality, independent components, and homoscedasticity assumptions on the error terms when the transferred semiparametric model can be viewed as a mixed-effects model, its null distribution is tractable. The asymptotic properties under the global and local alternatives remain unknown. For nonlinear null ODE models, the author discusses an extension in the sense that the test is based on a linear approximation that distorts the null distribution of the test. This test cannot identify which individual components are incorrectly modeled. Without the regularity assumptions mentioned above, the asymptotic properties are not investigated. Hooker and Ellner (2015) discuss this further.

In this study, we construct three tests. The first checks the whole ODE system (1.3). It is based on a trajectory representation of the ODEs that solves (1.1), and represents the ODE system as a multi-response function of time:

$$X(t) = F(t;\theta), \tag{1.5}$$

where $F(t;\theta)$ denotes the trajectory solution of (1.1). This test is constructed by matching approximations of the two sides of (1.5), and thus is called a trajectory matching-based test (TM_n). It is also feasible for partially observed ODE systems (Dattner (2015)), of which some components are not measured. However, we show that the trajectory matching-based test fails to check each individual component (1.4), owing to mixed component and parameter effects. Thus, we consider two other tests for testing the components (1.4). We construct an integral matching-based test (IM_n) based on the integration representation of the ODEs: on the definite integral over the interval (t_0, t) for both sides of (1.1),

$$X(t) - X(t_0) = \int_{t_0}^{t} f\left(s, X\left(s\right); \theta\right) ds. \tag{1.6}$$

Although this integral matching-based test demonstrates consistency in theory, its empirical performance is not encouraging, owing to the cumulative error in the integral and the estimation error in the nonparametric estimation. Therefore, we also construct a gradient matching-based test (GM_n) that is based directly on a gradient representation of the ODEs (1.1), but with a data splitting technique to eliminate the bias. We refer to this test as the bias-corrected gradient matching-based test. The new tests are all local smoothing-type tests, and because some useful ODE models are highly oscillating, may better detect possible model departures from the null model.

To estimate parameters, we use the nonlinear least squares method based on the trajectory representation (1.5) for TM_n; see Xue, Miao and Wu (2010) and Ramsay and Hooker (2017) for details. However, when checking each component function, this estimation fails to work, because it involves all components of the system. Thus, two-step estimation methods are considered. They first replace the unknown functions in the integral representation (1.6) or the gradient representation (1.1) with their nonparametric estimators, and then construct pseudo least squares estimators; see Brunel (2008), Liang and Wu (2008), and Dattner and Klaassen (2015). We use a gradient matching two-step estimation method for IM_n and GM_n. Note that the estimation methods based on three model representations actually pair with the three proposed tests.

In the rest of this paper, we provide detailed constructions for the three tests, and study their asymptotic properties under the null, global, and local alternatives. We also investigate the properties of the corresponding estimators under both the null and alternative hypotheses, which is new, to the best of our knowledge. All technical proofs are provided in the online Supplementary Material. In this paper, the measurement time t_i is considered random, with the sampling probability density function $p(t)$, which is a convenient mathematical device used in some studies on ODE models, such as Liang and Wu (2008) and Ding and Wu (2014).

2. Trajectory Matching-Based Test

2.1. Test statistic construction

Recall the hypotheses in (1.3) for the whole ODE system. According to the trajectory representation (1.5), the checking problem can be converted to testing whether the vector function $X(t) = F(t;\theta_0)$, for some $\theta_0 \in \Theta \subset R^q$.

Consider the $p = 1$ case to motivate our construction. Recall $p(t)$ denotes the sampling density of the measurement time t_i. Denote $\|\cdot\|$ as the Frobenius norm. Let $\varepsilon_i = Y_i - F(t_i;\theta^*_{NLS})$, with $\theta^*_{NLS} = \text{argmin}_\theta E\{\|Y_i - F(t_i;\theta)\|^2\}$. Under H_0, $\varepsilon_i = \epsilon_i$ and $E(\varepsilon_i \mid t_i) = 0$ leads to $E\left\{\varepsilon_i E(\varepsilon_i \mid t_i)p(t_i)\right\} = 0$, whereas under H_1, $E(\varepsilon_i \mid t_i) = X(t_i) - F(t_i;\theta^*_{NLS}) \neq 0$ and $E\{\varepsilon_i E(\varepsilon_i|t_i)p(t_i)\} = E[\{E(\varepsilon_i \mid t_i)\}^2 p(t_i)] > 0$. With the nonlinear least squares estimator $\hat{\theta}_{NLS} = \text{argmin}_\theta \sum_{i=1}^n \|Y_i - F(t_i;\theta)\|^2$, $e_i = Y_i - F(t_i;\hat{\theta}_{NLS})$ is the residual. Thus, we use the sample analogue of $E\{\varepsilon_i E(\varepsilon_i \mid t_i)p(t_i)\}$ to build the statistic

$$V_n^{Zh} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) e_i e_j,$$

where K is a kernel function and h is a bandwidth parameter. This is, in spirit, similar to the test suggested by Zheng (1996). A standardized test statistic T_n^{Zh} can be obtained easily by using V_n^{Zh} and its variance. In the multi-response case, we can obtain a vector version of V_n^{Zh} as $V_n^F = \left(V_{n1}^{Zh}, \ldots, V_{np}^{Zh}\right)^\top$. To summarize the information contained in V_n^F, we aggregate V_n^F to construct the test statistic TM_n:

$$TM_n = n^2 h V_n^{F\top} (\widehat{\Sigma}^F)^{-1} V_n^F.$$

Here, $\widehat{\Sigma}^F$ is a symmetric matrix used to normalize the test statistic:

$$\widehat{\Sigma}^F = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \frac{1}{h} K^2\left(\frac{t_i - t_j}{h}\right) (e_i \odot e_j)(e_i \odot e_j)^\top,$$

where $\odot$ denotes the element-wise product of two vectors, and e_i is the residual vector at the time point t_i. By using the quadratic form and taking $(\widehat{\Sigma}^F)^{-1}$ for the normalization, we eliminate the correlation between components, placing equal weight on each.

In practice, data are often only available for a subset of the components of the ODE system (Dattner (2015)). For instance, we cannot observe the susceptible population in the classical SIR model for infectious diseases. The proposed test TM_n can be extended easily to a modified test TM_n^o to handle this case by replacing the statistics of all components with their corresponding statistics of those observed components. See the Supplementary Material for details.

2.2. Asymptotic properties

To derive the asymptotic properties of TM_n, we suppose sets A and B of the assumptions in the Supplementary Material hold. The assumptions in set A give the basic setting of (t, Y) and the conditions on the kernel function and $X'(t)$. The assumptions in set B place restrictions on the primitive function $F(t;\theta)$, and have to do with the model identifiability in the nonlinear least squares estimation (see, e.g. White (1981)).

To investigate the power of a test, we consider two kinds of local alternatives for the ODE models. The first adds a local misspecification to the trajectory of the ODEs:

$$H_{1n}^F : X(t) = F\left(t;\theta_0\right) + \delta_n L\left(t\right), \tag{2.1}$$

where $L(t) = (L_1(t), \ldots, L_p(t))^\top$ is a bounded multiple response function such that $X(t) \neq F(t;\theta)$, for every $\theta \in \Theta$, and $\delta_n \to 0$ as $n \to \infty$. Here, H_{1n}^F in (2.1) is similar to that in the classical regression settings. However, things are more complicated for the ODE models, because another type of local disturbance directly affects the derivative function $X'(\cdot)$. That is, we consider the following sequence of local alternatives:

$$H_{1n}^f : X'(t) = f\left(t, X(t);\theta_0\right) + \delta_n l\left(t\right), \tag{2.2}$$

where $l(t) = (l_1(t), \ldots, l_p(t))^\top$ is a bounded multiple response function and $X'(t) \notin \mathcal{F}$. Hooker (2009) considered global alternatives in a similar manner using empirical forcing functions. We now state the relationship between H_{1n}^f and H_{1n}^F.

Proposition 1. *Given the set A of assumptions in the Supplementary Material, then, under H_{1n}^F in (2.1), the derivative has the form*

$$X'(t) = f\left(t, X(t);\theta_0\right) + \delta_n v_1\left(t\right) + o(\delta_n),$$

where $v_1\left(t\right) = L'\left(t\right) - (\partial f\left(t, X(t);\theta_0\right)/\partial X^\top)L\left(t\right)$. Under H_{1n}^f in (2.2), the original function can be expressed as

$$X(t) = F\left(t;\theta_0\right) + \delta_n v_2\left(t\right) + o(\delta_n),$$

where $v_2(t)$ is the solution of $v_2'\left(t\right) = (\partial f\left(t, X(t);\theta_0\right)/\partial X^\top)v_2\left(t\right) + l(t)$, with $v_2(0) = 0$.

Because any higher order small o term does not influence the asymptotic properties of the test under the local alternatives, H_{1n}^F is asymptotically equivalent to H_{1n}^f in this sense. However, in finite-sample cases, they may still affect the performance of the tests. In the following, we define $l(t) = v_1(t)$ under H_{1n}^F, and $L(t) = v_2(t)$ under H_{1n}^f. Then, we uniformly handle these two kinds of local alternatives.

Before exploring the limiting results of TM_n, we first study the asymptotic properties of the nonlinear least squares estimator. Recall $\theta_{NLS}^* = \operatorname{argmin}_\theta E\{\|Y_i - F(t_i;\theta)\|^2\}$. We give the following proposition.

Proposition 2. *Given sets A and B of the assumptions in the Supplementary Material, and supposing the numerical error of the numerical solution is negligible, then $\hat{\theta}_{NLS} - \theta_{NLS}^* = o_P(1)$. In addition, we have the following:*

1. *Under the null hypothesis $\theta_{NLS}^* = \theta_0$ and with*

$$H_{\dot{F}} = E\left\{\sum_{k=1}^{p} \frac{\partial F_k(t;\theta_0)}{\partial\theta}\frac{\partial F_k(t;\theta_0)}{\partial\theta^\top}\right\},$$

then

$$\sqrt{n}(\hat{\theta}_{NLS} - \theta_0) = H_{\dot{F}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{p}\left\{\epsilon_{ik}\frac{\partial F_k(t_i;\theta_0)}{\partial\theta}\right\} + o_P(1).$$

2. *Under the global alternative hypothesis H_1, writing $\theta_{NLS}^* = \theta_1^*$,*

$$\sqrt{n}(\hat{\theta}_{NLS} - \theta_1^*) = G_{\dot{F}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{p}\left[\{Y_{ik} - F_k(t_i;\theta_1^*)\}\frac{\partial F_k(t_i;\theta_1^*)}{\partial\theta}\right] + o_P(1),$$

where

$$G_{\dot{F}} = E\left\{\sum_{k=1}^{p} \frac{\partial F_k(t;\theta_1^*)}{\partial\theta}\frac{\partial F_k(t;\theta_1^*)}{\partial\theta^\top}\right\} - E\left[\sum_{k=1}^{p}\{X_k(t) - F_k(t;\theta_1^*)\}\frac{\partial^2 F_k(t;\theta_1^*)}{\partial\theta\partial\theta^\top}\right].$$

3. *Under the local alternative hypothesis H_{1n}^F in (2.1) or H_{1n}^f in (2.2), we have $\theta_{NLS}^* = \theta_0$ and*

$$\sqrt{n}(\hat{\theta}_{NLS} - \theta_0) = H_{\dot{F}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{p}\left\{\epsilon_{ik}\frac{\partial F_k(t_i;\theta_0)}{\partial\theta}\right\}$$

$$+\sqrt{n}\delta_n H_{\dot{F}}^{-1} E\left\{\sum_{k=1}^{p} L_k(t)\frac{\partial F_k\left(t;\theta_0\right)}{\partial\theta}\right\} + o_P(1).$$

This proposition is essentially a multivariate extension of the nonlinear least squares estimation in the literature (see, e.g., Jennrich (1969); White (1981); Li, Chiu and Zhu (2019)).

Using the theory of U-statistics, we obtain the following asymptotic properties of TM_n under the null and global alternative hypotheses.

Theorem 1. *Given sets A and B of the assumptions in the Supplementary Material, if $h \to 0$ and $nh \to \infty$, then we have the following:*

1. *Under the null hypothesis,*

$$TM_n \to \chi_p^2, \text{ in distribution},$$

where χ_p^2 is the chi-square distribution with p degrees of freedom.

2. *Under the global alternative H_1,*

$$\frac{TM_n}{n^2 h} \to V'^{\top}\Sigma^{F\prime -1}V', \text{ in probability},$$

where $V' = E[\{X\left(t\right) - F\left(t;\theta_1^*\right)\}^2 \odot p\left(t\right)]$ and $\Sigma^{F\prime}$ is defined as follows: for any element (k_1, k_2), with $1 \leq k_1, k_2 \leq p$, $\sigma_{k1k2}(t) = E(\epsilon_{ik_1}\epsilon_{ik_2} \mid t)$,

$$\Sigma_{k_1k_2}^{F\prime} = 2\int K^2(u)du$$
$$\int \left[\sigma_{k1k2}(t) + \left\{X_{k_1}\left(t\right) - F_{k_1}\left(t,\theta_1^*\right)\right\}\left\{X_{k_2}\left(t\right) - F_{k_2}\left(t,\theta_1^*\right)\right\}\right]^2 p^2(t)dt.$$

We see that TM_n under H_1 diverges to infinity at a fast rate of order n^2h. We also study the asymptotic property of TM_n under the local alternatives.

Theorem 2. *Given sets A and B of the assumptions in the Supplementary Material, if $h \to 0$ and $nh \to \infty$, then under H_{1n}^F in (2.1) or H_{1n}^f in (2.2), with $n^{1/2}h^{1/4}\delta_n \to \infty$,*

$$\frac{TM_n}{n^2 h\delta_n^4} \to \mu^{\top}\Sigma^{F-1}\mu, \text{ in probability},$$

where μ is a p-dimensional vector with the ith component

$$\mu_i = E\left(\left[L_i\left(t\right) - \frac{\partial F_i\left(t;\theta_0\right)}{\partial\theta^{\top}} H_{\dot{F}}^{-1} E\left\{\sum_{k=1}^{p} L_k\left(t\right)\frac{\partial F_k\left(t;\theta_0\right)}{\partial\theta}\right\}\right]^2 p(t)\right).$$

In particular, if $\delta_n = n^{-1/2}h^{-1/4}$*, then*

$$TM_n \to \chi_p^2(\lambda), \textit{in distribution},$$

where $\chi_p^2(\lambda)$ *is noncentral chi-squared distribution, where the noncentrality parameter* $\lambda = \mu^\top \Sigma^{F-1}\mu$*, with* $\Sigma_{k_1k_2}^F = 2\int K^2(u)du \times \int \{\sigma_{k1k2}(t)\}^2 p^2(t)dt$*.*

This result shows that the test can detect the local alternatives distinct from the null at a rate of order $n^{-1/2}h^{-1/4}$, which is the typical rate local smoothing tests can achieve.

3. Checking the ODE Models

In the next three sections, we consider the hypotheses in (1.4) for each component. As mentioned before, we cannot use the trajectory matching-based test for each ODE component. We now give a detailed discussion with three aspects on checking the ODE models.

1. Mixed components effect. We use a toy example to explain why TM_n cannot identify the incorrectly modeled component(s). Let the hypothetical model be $(X_1'(t), X_2'(t)) = (X_1(t), X_1(t) + X_2(t))$. The true ODE system is $(X_1'(t), X_2'(t)) = (2X_1(t), X_1(t) + X_2(t))$, with the initial values $(X_1(0), X_2(0)) = (1, 1)$. Here, only the first component is modeled incorrectly. However, because the second component of $X'(t)$ involves both components of $X(t)$, the trajectory of the second component in the hypothetical model is $(1+t)\exp(t)$, whereas in the true system, it is $\exp(2t)$. Therefore, if we use a test based on the trajectory of the ODE for the second component, the decision of TM_n based on the second component is strongly disturbed by the model correctness of the first component, and the correctly modeled second component is rejected. Thus, to construct the tests, we should decouple the relationships between the components. We do so by applying the model-free nonparametric estimators $\hat{X}(t)$ and $\hat{X}'(t)$.

2. Mixed parameters effect. If different components share some of the same parameters, the incorrectly modeled component(s) may also make the estimators deviate from the underlying values, causing inconsistency. Thus, any test relying on these estimators is ineffective. To avoid this problem, we need only use the equation of the tested component to build an objective function for the estimation. This can be achieved using two-step methods. To estimate θ_{0k}, which consists of the parameters in the kth component, we

use the following gradient matching two-step method:

$$\hat{\theta}_{TS}^{k} = \underset{\theta_k}{\operatorname{argmin}} \sum_{j=1}^{m} \left\{ \hat{X}_k' \left(t_j^*\right) - f_k \left(t_j^*, \hat{X}\left(t_j^*\right); \theta_k\right)\right\}^2 \omega_k \left(t_j^*\right), \tag{3.1}$$

with $\omega_k(t)$ being a selected weight function, and t_j^* being the selected fitted time grid set by the user (Ding and Wu (2014)). Because only the modeled form of the kth component is used in the objective function, this method avoids the problem of parameters being shared by different components.

In (3.1), the number m can be larger than n. Here, $\hat{X}(t)$ is the local linear estimator for $X(t)$, and $\hat{X}'(t)$ is the local quadratic estimator for $X'(t)$ in the vector version, the kth components of which, $\hat{X}_k(t)$ and $\hat{X}_k'(t)$, respectively, are the corresponding local polynomial estimators for $X_k(t)$ and $X_k'(t)$ with the bandwidth h_e.

3. The choice of the smoothing method. Different smoothing procedures are employed in the following tests. Choosing a suitable smoothing method needs careful consideration. If the purpose is to construct a test by generating a sample analogue of some quantities at the population level, a simple method such as the Nadaraya–Watson method can be used to make the test simpler in form. However, if the purpose is to give plug-in estimators in the constructed test, we need to reduce the estimation error that would affect the limiting null distribution. In the following, we choose the local linear and quadratic methods to estimate $X(t)$ and $X'(t)$, because they have better estimation performance.

4. Integral Matching-Based Test

4.1. Test statistic construction

Hereafter, we omit the indicator k in θ_{0k}, θ_k, and $\hat{\theta}_{TS}^k$ to simplify the notation without confusion. Similarly, as the integral matching method for the estimation, we construct pseudo-residuals based on the integration representation (1.6):

$$\hat{e}_{ik} = Y_{ik} - X_k(t_0) - \int_{t_0}^{t_i} f_k\left(t, \hat{X}(t); \hat{\theta}_{TS}\right) dt.$$

Here, we use the local linear estimator $\hat{X}(t)$ with the bandwidth h_0 and the two-step estimator $\hat{\theta}_{TS}$. Because $\hat{F}_k(t_i; \hat{\theta}_{TS}) = X_k(t_0) + \int_{t_0}^{t_i} f_k(t, \hat{X}(t); \hat{\theta}_{TS}) dt$ is expected to converge to $F_k(t_i; \theta_0)$ under the null hypothesis, $\hat{e}_{ik}$ can be used as a surrogate to replace $e_{ik} = Y_{ik} - F_k(t_i; \hat{\theta}_{NLS})$ in the trajectory matching-based test.

This replacement is critical, because $\hat{X}(t)$ always captures the true form $X(t)$, which eliminates the influence of latent incorrectly modeled components. Note that we require data on all components. Consequently, we obtain the following integral matching-based test, IM_n:

$$IM_n = \sqrt{\frac{n-1}{n}} \frac{nh^{1/2}V_n^{\hat{F}}}{\sqrt{\widehat{\Sigma}^{\hat{F}}}} = \frac{\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n} K\left((t_i-t_j)/h\right)\hat{e}_{ik}\hat{e}_{jk}}{\left\{\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n} 2K^2\left((t_i-t_j)/h\right)\hat{e}_{ik}^2\hat{e}_{jk}^2\right\}^{1/2}}$$

where

$$V_n^{\hat{F}} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\frac{1}{h}K\left(\frac{t_i-t_j}{h}\right)\hat{e}_{ik}\hat{e}_{jk},$$

$$\widehat{\Sigma}^{\hat{F}} = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\frac{1}{h}K^2\left(\frac{t_i-t_j}{h}\right)\hat{e}_{ik}^2\hat{e}_{jk}^2.$$

4.2. Asymptotic properties

As in Section 2.2, we suppose that sets A and B of the assumptions in the Supplementary Material hold. Note that assumptions (3) and (4) in set A ensure the uniform convergence rate for the nonparametric estimation (Hansen (2008)). We also suppose that the assumptions in set C hold. These assumptions are similar to those in Liang and Wu (2008), and contain the conditions on $f(t, X(t);\theta)$ necessary for the two-step method.

In the following, we also consider two local alternatives corresponding to (2.1) and (2.2) for any component function:

$$H_{1kn}^{F} : X_k(t) = F_k\left(t;\theta_0\right) + \delta_n L_k\left(t\right), \tag{4.1}$$

with $X_k(t) \neq F_k\left(t;\theta\right)$, for every $\theta \in \Theta$, and

$$H_{1kn}^{f} : X_k'(t) = f_k\left(t, X(t);\theta_0\right) + \delta_n l_k\left(t\right), \tag{4.2}$$

with $X_k'(t) \neq f_k\left(t, X(t);\theta\right)$, for every $\theta \in \Theta$. Here, the subscript k represents the kth component. Using Proposition 1, we can define the counterpart functions $l_k(t)$ under H_{1kn}^{F} and $L_k(t)$ under H_{1kn}^{f}, and then deal with these two kinds of local alternatives uniformly.

We first give the asymptotic properties of $\hat{\theta}_{TS}$ defined in (3.1) under different hypotheses, which are not available in the literature. Assumption (3) in Set C ensures that there exists a unique minimizer $\theta_{TS}^{*} = \text{argmin}_\theta E_{p^*}[\{X_k'(t) -$

$f_k(t, X(t), \theta)\}^2 w_k(t)]$, where $p^*(t)$ is the probability density function of the selected fitted time point. Denote $\Lambda(t) = \hat{X}(t) - X(t)$, $\Delta(t) = \hat{X}'(t) - X'(t)$. Then, we have the following.

Proposition 3. *Given sets A and C of the assumptions in the Supplementary Material,* $\log n/(nh_e^3) = o(1)$*, the two-step estimator* $\hat{\theta}_{TS}$ *is consistent with* θ^*_{TS} *and has the following asymptotic representations:*

1. *Under the null hypothesis, we have* $\theta^*_{TS} = \theta_0$*, and letting*

$$H_{\dot{f}} = E_{p^*}\left\{\omega_k(t)\frac{\partial f_k(t, X(t); \theta_0)}{\partial \theta}\frac{\partial f_k(t, X(t); \theta_0)}{\partial \theta^\top}\right\},$$

we have

$$\begin{aligned}\hat{\theta}_{TS} - \theta_0 = H_{\dot{f}}^{-1}\frac{1}{m}\sum_{j=1}^{m}\Bigg\{&\Delta_k(t_j^*)\omega_k(t_j^*)\frac{\partial f_k(t_j^*, X(t_j^*); \theta_0)}{\partial \theta}\\ &- \omega_k(t_j^*)\frac{\partial f_k(t_j^*, X(t_j^*); \theta_0)}{\partial \theta}\frac{\partial f_k(t_j^*, X(t_j^*); \theta_0)}{\partial X^\top}\Lambda\left(t_j^*\right)\Bigg\}\\ &+ o_P(n^{-1/2}),\end{aligned}$$

which is a term of order $O_P(n^{-1/2})$.

2. *Under the global alternative hypothesis* H_1*, we have* $\theta^*_{TS} = \theta_1$ *and*

$$\begin{aligned}&\sqrt{n}(\hat{\theta}_{TS} - \theta_1)\\ &= G^{-1}\frac{\sqrt{n}}{m}\sum_{j=1}^{m}\Bigg\{\Delta_k(t_j^*)\omega_k(t_j^*)\frac{\partial f_k(t_j^*, X(t_j^*); \theta_1)}{\partial \theta}\\ &\quad -\omega_k(t_j^*)\frac{\partial f_k(t_j^*, X(t_j^*); \theta_1)}{\partial \theta}\frac{\partial f_k(t_j^*, X(t_j^*); \theta_1)}{\partial X^\top}\Lambda\left(t_j^*\right)\Bigg\} + o_P(1),\end{aligned}$$

where

$$\begin{aligned}G =& E_{p^*}\left\{\omega_k(t)\frac{\partial f_k(t, X(t); \theta_1)}{\partial \theta}\frac{\partial f_k(t, X(t); \theta_1)}{\partial \theta^\top}\right\}\\ &- E_{p^*}\left[\{X_k(t) - f_k(t, X(t); \theta_1)\}\,\omega_k(t)\frac{\partial^2 f_k(t, X(t); \theta_1)}{\partial \theta \partial \theta^\top}\right].\end{aligned}$$

3. *Under the local alternative hypothesis* H^F_{1kn} *in (4.1) or* H^f_{1kn} *in (4.2), with*

$\delta_n \to 0$, *we have* $\theta^*_{TS} = \theta_0$ *and*

$$\sqrt{n}(\hat{\theta}_{TS} - \theta_0) = H_{\dot{f}}^{-1} \frac{\sqrt{n}}{m} \sum_{j=1}^{m} \left\{ \Delta_k(t_j^*)\omega_k(t_j^*) \frac{\partial f_k(t_j^*, X(t_j^*); \theta_0)}{\partial \theta} \right.$$
$$\left. - \omega_k(t_j^*) \frac{\partial f_k(t_j^*, X(t_j^*); \theta_0)}{\partial \theta} \frac{\partial f_k(t_j^*, X(t_j^*); \theta_0)}{\partial X^\top} \Lambda\left(t_j^*\right) \right\}$$
$$+ \sqrt{n}\delta_n H_{\dot{f}}^{-1} E_{p^*} \left\{ l(t)\omega(t) \frac{\partial f_k\left(t, X\left(t\right); \theta_0\right)}{\partial \theta} \right\} + o_P(1).$$

Given Proposition 3, we next study the asymptotic properties of IM_n. Let $\varepsilon_{ik} = Y_{ik} - X_k(t_0) - \int_{t_0}^{t_i} f_k\left(t, X\left(t\right); \theta^*_{TS}\right) dt$ be the residual. By denoting $\hat{e}_{ik} = \hat{e}_{ik} + \varepsilon_{ik} - \varepsilon_{ik}$, $V_n^{\hat{F}}$ is an asymptotic U-statistic. To handle this U-statistic, we give the following proposition about a non-degenerate U-statistic of order m^* with a kernel varying with n.

Proposition 4. *Suppose* U_n *is a non-degenerate U-statistic with the kernel* $h_n(z_1, \ldots, z_{m^*})$ *of order* m^**. If* $E[\|h_n\left(z_1, \ldots, z_{m^*}\right)\|^2] = o(n)$*, then*

$$\sqrt{n}\left(U_n - \hat{U}_n\right) = o_P(1),$$

where

$$\hat{U}_n = E\left\{h_n\left(z_1, \ldots, z_{m^*}\right)\right\} + \frac{m^*}{n} \sum_{i=1}^{n} \left\{E\left[h_n\left(z_1, \ldots, z_{m^*}\right) \mid z_i\right] - E\left[h_n\left(z_1, \ldots, z_{m^*}\right)\right]\right\}$$

is the projection of U_n*.*

Applying Proposition 4, we can prove that replacing $\hat{e}_{ik}$ with e_{ik} has no effect on the limiting null distribution of IM_n. Using this property, we state the asymptotic properties of IM_n in the following theorem. Recall that $\hat{X}(t)$ is the local linear estimator of $X(t)$ with the bandwidth h_0. Define $a_n(h_0) = h_0^2 + n^{-1/2} h_0^{-1/2} \log n^{1/2}$, which is the uniform convergence rate of the local linear estimator (Hansen (2008)).

Theorem 3. *Given sets A–C of the assumptions in the Supplementary Material, if* $h \to 0$, $nh \to \infty$, $n^{-1/2}h^{-1/2} = o(h_0)$*, and* $a_n^2(h_0) = o(n^{-1}h^{-1/2})$*, then we have the following:*

1. *Under the null hypothesis,*

$$IM_n \to N(0, 1), \textit{in distribution.}$$

2. *Under the global alternative,* $IM_n/(nh^{1/2}) \to IM_{H1}$ *in probability, where*

$$IM_{H1} = \frac{E\left[\{X_k(t) - F_k^*(t;\theta_1)\}^2 p(t)\right]}{\left(2\int K^2(u)du \int \left[\sigma_k^2(t) + \left\{X_k(t) - F_k^*(t;\theta_1)\right\}^2\right]^2 p^2(t)dt\right)^{1/2}},$$

with $F_k^*(t;\theta_1) = X_k(t_0) + \int_{t_0}^t f_k(t, X(s);\theta_1)\, ds$ *and* $\sigma_k(t) = E(\epsilon_{ik}^2 \mid t)$.

Theorem 3 shows that this test is consistent and diverges to infinity at a rate of order $nh^{1/2}$ under H_{1k}. The following theorem states the asymptotic property of IM_n under the local alternatives.

Theorem 4. *Assume the conditions in Theorem* 3, *with* $n^{1/2}h^{1/4}\delta_n \to \infty$. *Then, under* H_{1kn}^F *in* (4.1) *or* H_{1kn}^f *in* (4.2), *we have*

$$\frac{IM_n}{nh^{1/2}\delta_n^2} \to \frac{\mu_I}{\sigma_k^*}, \text{ in probability,}$$

where

$$\begin{aligned}\mu_I = {} & E\left(\left[L_k(t) - \frac{\partial F_k(t;\theta_0)}{\partial\theta^\top} H_{\dot{f}}^{-1} E_{p^*}\left\{l_k(t)\,\omega_k(t)\frac{\partial f_k(t, X(t);\theta_0)}{\partial\theta}\right\}\right]^2 p(t)\right) \\ & + 2E\left(\left[\frac{\partial F_k(t;\theta_0)}{\partial\theta^\top} H_{\dot{f}}^{-1} E_{p^*}\left\{l_k(t)\,\omega_k(t)\frac{\partial f_k(t, X(t);\theta_0)}{\partial\theta}\right\} - L_k(t)\right] p(t) \right. \\ & \left. \times \int_{t_0}^t \frac{\partial f_k(s, X(s);\theta_0)}{\partial X^T} L(s)\, ds\right) + E\left[\left\{\int_{t_0}^t \frac{\partial f_k(s, X(s);\theta_0)}{\partial X^T} L(s) ds\right\}^2 p(t)\right], \\ \sigma_k^* = {} & \left[2\int K^2(u)du \int \left\{\sigma_k^2(t)\right\}^2 p^2(t)dt\right]^{1/2}.\end{aligned}$$

In particular, when $\delta_n = n^{-1/2}h^{-1/4}$,

$$IM_n \to N\left(\frac{\mu_I}{\sigma_k^*}, 1\right), \text{ in distribution.}$$

5. Gradient Matching-Based Test

5.1. Test statistic construction

As noted for the integral matching-based test, the surrogate $\hat{F}_k(t_i;\hat{\theta}_{TS}) = X_k(t_0) + \int_{t_0}^{t_i} f_k(t, \hat{X}(t);\hat{\theta}_{TS})dt$ involves both the estimation of $X(t)$ and the integral over the function $f_k(\cdot)$. In other words, the cumulative estimation error could be large in finite-sample scenarios. Our numerical studies confirm this problem. Thus, in this section, we consider a test that is based directly on the gradient representation (1.1). With $n^* \leq n$ observations, we give the Nadaraya–Watson kernel estimator of $X'(t)$ as

$$\tilde{X}'(t) = \frac{\hat{\psi}'(t)\hat{p}(t) - \hat{\psi}(t)\hat{p}'(t)}{\hat{p}^2\,(t)},$$

where

$$\hat{\psi}(t) = \frac{1}{n^*}\sum_{i=1}^{n^*} \frac{1}{h} K\left(\frac{t-t_i}{h}\right) Y_i, \quad \hat{\psi}'(t) = \frac{1}{n^*}\sum_{i=1}^{n^*} \frac{1}{h^2} K'\left(\frac{t-t_i}{h}\right) Y_i,$$
$$\hat{p}(t) = \frac{1}{n^*}\sum_{i=1}^{n^*} \frac{1}{h} K\left(\frac{t-t_i}{h}\right), \quad \hat{p}'(t) = \frac{1}{n^*}\sum_{i=1}^{n^*} \frac{1}{h^2} K'\left(\frac{t-t_i}{h}\right).$$

Note that $e_f(t) = X_k'(t) - f_k(t, X(t); \theta_0) = 0$ corresponds to the null hypothesis, otherwise to the alternative hypothesis. Thus, if we replace $X_k'(t)$ with $\tilde{X}_k'(t)$ and $f_k(t, X(t); \theta_0)$ by $f_k(t, \hat{X}(t); \hat{\theta}_{TS})$ with $\hat{\theta}_{TS}$ defined in (3.1), the pseudo-residual $\hat{e}_f(t) = \tilde{X}_k'(t) - f_k(t, \hat{X}(t); \hat{\theta}_{TS})$ is expected to converge to zero in probability. Then, $E\{\hat{e}_f^2(t_i)\hat{p}^4\,(t_i)\}$ is expected to converge to zero under the null, and to a positive constant under the alternatives, where $\hat{p}^4(t_i)$ is used to eliminate the denominator in the nonparametric estimation. Therefore, its empirical version seems reasonably to be a test statistic:

$$\begin{aligned}
V_{n^*}^f =& \frac{1}{n^* h^2} \sum_{d=1}^{n^*} \left\{ \tilde{X}_k'(t_d) - f_k(t, \hat{X}(t_d); \hat{\theta}_{TS}) \right\}^2 \hat{p}^4\,(t_d) \\
=& \frac{1}{n^* h^2} \sum_{d=1}^{n^*} \left[\frac{1}{n^2} \sum_{i=1}^{n^*}\sum_{j=1}^{n^*} \left\{ \frac{1}{h^3} K'\left(\frac{t_d - t_i}{h}\right) K\left(\frac{t_d - t_j}{h}\right) (Y_{ik} - Y_{jk}) \right.\right. \\
& \left.\left. - \frac{1}{h^2} K\left(\frac{t_d - t_i}{h}\right) K\left(\frac{t_d - t_j}{h}\right) f_k(t, \hat{X}(t_d); \hat{\theta}_{TS}) \right\} \right]^2.
\end{aligned}$$

Here, $1/h^2$ is added to obtain a non-degenerate limit, $\hat{e}_f(t)$ converges to zero rather than a zero mean random variable in probability under the null.

However, because the nonparametric kernel estimation is biased, when we choose $n^* = n$, V_n^f has a non-negligible bias, even under H_0. This bias is difficult to estimate, which, in turn, makes it difficult to analyze the limiting null distribution. Thus, we suggest using a data-splitting method to build a new test.

To this end, we randomly partition the original sample into two subsamples, $n^* = \tilde{n}$ and $n^* = n - \tilde{n}$, where $\tilde{n} = \lfloor n/2 \rfloor$. Using these two subsamples, we construct two statistics $V^f_{\tilde{n}1}$ and $V^f_{(n-\tilde{n})2}$. Because $n - 2\tilde{n} \le 1$, the asymptotic properties of $V^f_{(n-\tilde{n})2}$ should be the same as those of $V^f_{\tilde{n}2}$. Thus, we assume that, without loss of generality, $n = 2\tilde{n}$ is even. Therefore, the difference $V^f_{\tilde{n}1} - V^f_{\tilde{n}2}$ can serve as a statistic that is symmetric about zero to determine the limiting null distribution. To increase the power of the test, we use another statistic $\hat{S} = 1/h^2 \int \{f_k(t, \hat{X}(t); \hat{\theta}_{TS}) - \hat{X}'_k(t)\}^2 dt$, which is an estimator of $S = 1/h^2 \int \{f_k(t, X(t); \theta) - X'_k(t)\}^2 dt$ that is equal to zero under the null, and is larger than zero under the alternatives. The new test statistic is their convex combination, $V^f_{\tilde{n}1} - V^f_{\tilde{n}2} + c\hat{S}$. The key point is that, using a proper choice of the bandwidth parameters, $\hat{S}$ has a faster rate of convergence to zero than does $V^f_{\tilde{n}1} - V^f_{\tilde{n}2}$ under the null hypothesis. Thus, it does not change the limiting distribution under the null hypothesis, but provides power under the alternatives. The constant c is a tuning parameter set by user.

By dividing by the estimator of its variance, the final test statistic is

$$GM_n = \frac{\sqrt{\tilde{n}}(V^f_{\tilde{n}1} - V^f_{\tilde{n}2} + c\hat{S})}{\sqrt{2\widehat{\Sigma}^f}},$$

where

$$\begin{aligned}
\widehat{\Sigma}^f &= \frac{1}{n-1} \sum_{s=1}^{n} \left\{ \hat{w}_n(z_s) - \frac{1}{n} \sum_{i=1}^{n} \hat{w}_n(z_i) \right\}^2, \\
\hat{w}_n(z_s) &= \frac{1}{\lfloor (n-1)/4 \rfloor} \sum_{i=1}^{\lfloor (n-1)/4 \rfloor} W_n(z_{i_1}, z_{i_2}, z_{i_3}, z_{i_4}, z_s), \\
W_n(z_a, z_b, z_c, z_d, z_s) &= \frac{1}{5!} \sum_P W'_n(z_{i_1}, z_{i_2}, z_{i_3}, z_{i_4}, z_{i_5}), \\
W'_n(z_a, z_b, z_c, z_d, z_s) &= \frac{1}{h^2} K\left(\frac{t_s - t_a}{h}\right) K\left(\frac{t_s - t_b}{h}\right) \\
&\times \left\{ \frac{1}{h^3} K'\left(\frac{t_s - t_c}{h}\right)(Y_{ck} - Y_{ak}) - \frac{1}{h^2} K\left(\frac{t_s - t_c}{h}\right) f_k(t_s, \hat{X}(t_s); \hat{\theta}_{TS}) \right\} \\
&\times \left\{ \frac{1}{h^3} K'\left(\frac{t_s - t_d}{h}\right)(Y_{dk} - Y_{bk}) - \frac{1}{h^2} K\left(\frac{t_s - t_d}{h}\right) f_k(t_s, \hat{X}(t_s); \hat{\theta}_{TS}) \right\}.
\end{aligned}$$

Here, $W_n(\cdot)$ is the symmetric version of $W'_n(\cdot)$, and $\sum_P$ means that the sum is taken over all permutations $(i_1, i_2, i_3, i_4, i_5)$ of $\{a, b, c, d, s\}$. Local linear and quadratic smoothers are used to obtain $\hat{X}(t)$ and $\hat{X}'(t)$, respectively, with the corresponding bandwidths h_0 and h_1.

5.2. Asymptotic properties

To derive the properties, we first state sets A and C of the assumptions in the Supplementary Material. Let $b_n(h) = h^2 + n^{-1/2}h^{-3/2}\log n$, which is the uniform convergence rate of $X'(t)$ (Liang and Wu (2008)). Next, we state the asymptotic properties of GM_n under the null and global alternative hypotheses.

Theorem 5. *Given sets A and C of the assumptions in the Supplementary Material, if* $h^{-12} = o(n)$, $a_n^2(h_0)h^{-2} = o(n^{-1/2})$ *and* $b_n^2(h_1)h^{-2} = o(n^{-1/2})$, *recalling that* $\tilde{n} = \lfloor n/2 \rfloor$, *we have the following:*

1. *Under the null hypothesis,*

$$GM_n \to N(0,1), \quad \textit{in distribution.}$$

2. *Under the global alternative,*

$$\frac{GM_n}{\sqrt{\tilde{n}}} \to \frac{c\int \left[f_k(t,X(t);\theta_1) - X_k'(t)\right]^2 dt}{\sqrt{2\Sigma^{f\prime}}} > 0, \quad \textit{in probability,}$$

where

$$\begin{aligned}\Sigma^{f\prime} = \int \Bigg[&25\left\{f_k(t,X(t);\theta_1) - X_k'(t)\right\}^4 p^8(t) \\ &+ 4\left\{f_k'(t,X(t);\theta_1) - X_k^{(2)}(t)\right\}^2 \sigma_k^2(t)p^8(t)\Bigg] dt \\ &- 25\left[\int \left\{f_k(t,X(t);\theta_1) - X_k'(t)\right\}^2 p^4(t)dt\right]^2.\end{aligned}$$

Theorem 5 shows that the test is consistent and diverges to infinity at a rate of $\sqrt{n}$ under the global alternatives, although, as in the local smoothing tests, we use a nonparametric technique. The following theorem states the asymptotic power of GM_n under H_{1kn}^F and H_{1kn}^f.

Theorem 6. *Assume the conditions in Theorem* 5 *hold. Then, under* H_{1kn}^F *in* (4.1) *or* H_{1kn}^f *in* (4.2), *with* $\tilde{n}^{1/4}h^{-1}\delta_n \to \infty$ *and* $\delta_n h^{-1} = o(1)$,

$$\frac{GM_n}{\tilde{n}^{1/2}h^{-2}\delta_n^2} \to \frac{c\mu_G}{\sqrt{2\Sigma^f}}, \quad \textit{in probability,}$$

where

$$\mu_G = \left[H_{\dot{f}}^{-1} E_{p^*}\left\{l_k(t)\,\omega_k(t)\,\frac{\partial f_k(t, X(t);\theta_0)}{\partial\theta}\right\}\right]^\top \times \left\{\int \frac{\partial f_k(t,X(t);\theta_0)}{\partial\theta}\frac{\partial f_k(t,X(t);\theta_0)}{\partial\theta^\top}dt\right\} \times \left[H_{\dot{f}}^{-1} E_{p^*}\left\{l_k(t)\,\omega_k(t)\,\frac{\partial f_k(t, X(t);\theta_0)}{\partial\theta}\right\}\right] + \int l_k^2(t)dt - 2\left\{\int l_k(t)\frac{\partial f_k(t,X(t);\theta_0)}{\partial\theta^\top}dt\right\} H_{\dot{f}}^{-1} E_{p^*}\left\{l_k(t)\,\omega_k(t)\,\frac{\partial f_k(t, X(t);\theta_0)}{\partial\theta}\right\},$$

$$\Sigma^f = \frac{1}{9}\left\{\int u^3 K'(u)du\right\}^2 \int \left\{X_k^{(4)}(t)\right\}^2 \sigma_k^2(t)p^8(t)dt.$$

In particular, if $\delta_n = \tilde{n}^{-1/4}h$,

$$GM_n \to N\left(\frac{c\mu_G}{\sqrt{2\Sigma^f}}, 1\right), \text{ in distribution.}$$

6. Numerical Studies

6.1. Simulations

We conduct four simulation studies to show the performance of the proposed tests in finite-sample scenarios. In Examples 1–3, we use TM_n to check the whole ODE system, and use IM_n and GM_n to check each component. The subscript denotes which component the tests check. For example, GM_{n1} means the GM_n test for the first component. In Example 4, we reconsider the models in Examples 1–3 as partially observed ODE systems. Suppose that only data of the second component are measured, and that we use TM_n^o to check these models. As a competitor of TM_n, we also apply Hooker (2009) test T^H to check the whole ODE system.

In particular, the simulation results show that the empirical size of IM_n is often very large in the complex ODE model settings. This may be because the nonparametric estimations $\hat{X}(t)$ for all time points t have estimation errors, and the integral over the surrogate $\hat{e}_{ik}$ of e_{ik} in finite-sample scenarios could cause a very large cumulative error of IM_n. Thus, to control the empirical size of IM_n, we use an adjusted version in the simulation.

First, we restrict the integral to be in the shorter interval $(0.1, 0.9)$, rather than use the whole time interval $(0, 1)$, to avoid the boundary effect. Second, to reduce the error caused by the integration, we split the interval into $n_l = 8$

equidistant parts, $\mathcal{T}_l = (l/10, (l+1)/10)$, for $l = 1, 2, \ldots, 8$, and define

$$\hat{e}^l_{ik} = \sum_{l=1}^{8} \left\{ Y_{ik} - \hat{X}_k\left(\frac{l}{10}\right) - \int_{min((l/10),t_i)}^{t_i} f_k\left(t, \hat{X}(t); \hat{\theta}_{TS}\right) dt \right\} I(t_i \in \mathcal{T}_l),$$

$$IM^l_n = \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} K\left((t_i - t_j)/h\right) \hat{e}^l_{ik} \hat{e}^l_{jk}}{\left\{ \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} 2K^2\left((t_i - t_j)/h\right) \hat{e}^{l2}_{ik} \hat{e}^{l2}_{jk} \right\}^{1/2}},$$

where $I(\cdot)$ is the characteristic function. Then, we define the test statistic as

$$IM^*_n = \frac{\sum_{l=1}^{8} IM^l_n}{2\sqrt{2}\nu_n},$$

where $\nu_n = 1 + 2n^{-1/2}$ is used to further reduce the magnitude. It can be shown this test has the same asymptotic normality as the original test under the null by using the Cramér–Wald device and the continuous mapping theorem. See the Supplementary Material for additional details on selecting the smoothing parameters and other simulation settings.

Example 1. Data sets are generated from the following ODE models:

$$H_{11}: X'(t) = \begin{bmatrix} \frac{dX_1}{dt} \\ \frac{dX_2}{dt} \end{bmatrix} = \tau \begin{bmatrix} aX_1 + 0.4\alpha cos(aX_1) \\ aX_1 + bX_2 + 0.4\beta cos(aX_1 + bX_2) \end{bmatrix},$$

$$H_{12}: X'(t) = \begin{bmatrix} \frac{dX_1}{dt} \\ \frac{dX_2}{dt} \end{bmatrix} = \tau \begin{bmatrix} aX_1 + 0.1\alpha(aX_1)^3 \\ aX_1 + bX_2 + 0.1\beta(aX_1 + bX_2)^3 \end{bmatrix},$$

$$H_{13}: X'(t) = \begin{bmatrix} \frac{dX_1}{dt} \\ \frac{dX_2}{dt} \end{bmatrix} = \tau \begin{bmatrix} aX_1 + 2\alpha exp(aX_1) \\ aX_1 + bX_2 + 5\beta exp(aX_1 + bX_2) \end{bmatrix}.$$

We consider three cases in which the linear null ODE models are added with different disturbance terms to form alternative ODE models. The alternatives are oscillating functions of X in H_{11}, and are low-frequent functions of X in both H_{12} and H_{13}. In each case, $\alpha = 0$ and $\beta = 0$ correspond to the null, otherwise to the alternative hypothesis. When only one of α and β is nonzero, only one element ODE function is different under the alternative hypothesis. When α and β are both nonzero, both components are then changed. In addition, τ is a timescale parameter that transforms the arbitrary length of the time interval to one. We set the true parameter $(a, b) = (-0.06, -0.24)$, $\tau = 10$, and $\sigma_\epsilon = 0.05$, and the sample size is 300. The empirical size and power of the tests with a 0.05 significance level are presented in Table 1.

The results show that TM_n maintains the significance level. It also has very good power under all of the alternative models, which are significantly larger than IM_n and GM_n. This is not surprising, because TM summarizes the deviation of all the components from the trajectory of the null model. Furthermore, T^H maintains the significance level well, with good power.

Recall that the subscript represents the component to be checked. In general, IM_{n1} and IM_{n2} maintain the significance level, although in some cases, the empirical size of IM_{n1} is slightly lower than the significance level. Furthermore, IM_{n1} and IM_{n2} have high power in most cases, and GM_{n1} and GM_{n2} tend to maintain the significance level. In addition, GM_{n1} has good power in all three cases, whereas that of GM_{n2} varies in different cases. In the last two cases, when $(\alpha, \beta) = (1, 1)$, GM_{n2} has low power $(0.600, 0.095)$, but when $(\alpha, \beta) = (0, 1)$, it has higher power $(0.734, 0.120)$. This phenomenon worth nothing, and is quite different to the classical testing for regressions. A possible explanation is that the extra α suppresses the influence of the β term, making the disturbance term in $\hat{X}'(t)$ less important.

In the third case, GM_{n1} shows greater power than that of IM_{n1}. However, for GM_{n2} and IM_{n2}, the situation is the opposite. Because IM_n and GM_n measure different deviation indices and have diverse normalizing factors, they are distinctly superior in terms of sensitivity in different settings.

Example 2. Data sets are generated from the following ODE models:

$$H_2 : X'(t) = \begin{bmatrix} \frac{dX_1}{dt} \\ \frac{dX_2}{dt} \end{bmatrix} = \tau \begin{bmatrix} a(X_1 + X_2 - \frac{X_1^3}{3}) + \alpha X_1 X_2 \\ -\frac{X_1 + bX_2 - c}{a} + 0.4\beta X_1 X_2 \end{bmatrix}.$$

This is the famous FitzHugh–Nagumo ODE system, which describes the behavior of spike potentials in the giant axon of squid neurons (FitzHugh (1961); Nagumo, Arimoto and Yoshizawa (1962)). Following Ding and Wu (2014), we set the true parameter $(a, b, c) = (3, 0.2, 0.34)$, $\tau = 10$, $\sigma_\epsilon = 0.05$, and the initial values $(X_1(0), X_2(0)) = (1, -1)$. The sample size is 300. The empirical size and power of the tests with a 0.05 significance level are reported in Table 2.

Here, TM_n and T^H still perform very well for checking this complex nonlinear ODE model. Furthermore, GM_{n1} and GM_{n2} also work well in most cases. Owing to the complex interaction between the components of the ODE system, GM_{n1} and GM_{n2} when $(\alpha, \beta) = (0.5, 0.5)$ seem to perform differently to the $(\alpha, \beta) = (1, 1)$ setting. These simulation results again show the complexity of the ODE testing problem.

Obviously, the size of IM_{n1} is out of control. Unreported results show that

Table 1. Empirical sizes and powers in Example 1.

Hypothesis	α	β	TM_n	IM_{n1}	IM_{n2}	GM_{n1}	GM_{n2}	T^H
H_{11}	0	0	0.045	0.036	0.058	0.038	0.050	0.047
	0.5	0	1.000	0.510	0.059	0.279	0.034	1.000
	0	0.5	1.000	0.030	0.999	0.042	0.191	1.000
	0.5	0.5	1.000	0.499	0.998	0.270	0.193	1.000
	1	0	1.000	1.000	0.059	1.000	0.044	1.000
	0	1	1.000	0.021	1.000	0.035	0.994	1.000
	1	1	1.000	0.997	1.000	1.000	0.994	1.000
H_{12}	0	0	0.043	0.023	0.047	0.035	0.048	0.049
	0.5	0	1.000	1.000	0.052	0.915	0.036	1.000
	0	0.5	1.000	0.026	1.000	0.039	0.661	1.000
	0.5	0.5	1.000	1.000	1.000	0.919	0.606	1.000
	1	0	1.000	1.000	0.051	0.998	0.045	1.000
	0	1	1.000	0.026	0.998	0.031	0.734	1.000
	1	1	1.000	1.000	0.993	0.997	0.600	1.000
H_{13}	0	0	0.046	0.026	0.049	0.037	0.050	0.059
	0.5	0	1.000	0.126	0.037	0.294	0.034	1.000
	0	0.5	1.000	0.033	0.849	0.040	0.078	1.000
	0.5	0.5	1.000	0.134	0.721	0.296	0.062	1.000
	1	0	1.000	0.171	0.044	0.906	0.048	1.000
	0	1	1.000	0.025	0.982	0.031	0.120	1.000
	1	1	1.000	0.161	0.875	0.893	0.095	1.000

Table 2. Empirical sizes and powers in Example 2.

Hypothesis	α	β	TM_n	IM_{n1}	IM_{n2}	GM_{n1}	GM_{n2}	T^H
H_2	0	0	0.048	0.816	0.085	0.071	0.048	0.045
	0.5	0	1.000	1.000	0.070	0.131	0.048	1.000
	0	0.5	1.000	0.846	1.000	0.055	0.159	1.000
	0.5	0.5	1.000	1.000	1.000	0.129	0.203	1.000
	1	0	1.000	1.000	0.089	0.514	0.059	1.000
	0	1	1.000	0.846	1.000	0.069	0.993	1.000
	1	1	1.000	1.000	0.206	0.990	0.053	1.000

even when the sample size is increased to 10,000, the empirical size can be greatly reduced, which suggests consistency, but is still too large to make sense. This reminds that we must be careful to use IM_n to check complex nonlinear ODE models. Here, IM_{n2} performs acceptably, because the hypothetical model is now linear with $(\alpha, \beta) = (0, 0)$.

Example 3. The null ODE system is the standard Lotka–Volterra model, which is well known for modeling the evolution of prey–predator populations (Goel, Maitra and Montroll (1971)). Because the tests perform similarly to those in the last example, we present the results in the Supplementary Material.

Example 4. The hypotheses are the same as those in the previous three examples. Here, we only collect the data of the second component only: $Y_{i2} = X_2(t) + \epsilon_{i2}$. The results show that TM_n^o can be applied for partially observed ODE systems. See the Supplementary Material for the detailed results.

We conclude that TM_n and GM_n can maintain the significance level with good power in the conducted simulations. However, IM_n is only usable for checking linear ODEs, and TM_n^o is feasible for partially observed ODEs, and Compared with TM_n, T^H also performs well in these examples. Additional results provided in the Supplementary Material show that it may fail to maintain the significance level in cases with dependent error components or heteroscedasticity, where TM_n still works. Thus, these two tests are complementary, but our test could be more robust in all these scenarios.

6.2. A real-data example

Here, we apply our tests to a real data set downloaded from Hulin Wu Lab (`https://sph.uth.edu/dotAsset/3ac61148-e59e-493c-bbda-0a38ffe111e5.zip`). The data set has been analyzed to show the benefits of using a differential equation-constrained local polynomial regression for estimating the parameters in an ODE model for influenza virus-specific effector CD8+ T cells (Ding and Wu (2014)). Here, we employ the proposed tests to check the adequacy of this model. See the Supplementary Material for details of the model form and the data set.

The value of TM_n is 84.10 and the corresponding p-value is about zero. This suggests that the whole ODE model under the null is not plausible. Next, we use IM_n and GM_n to check each component function. The values of IM_n for the three component functions are $(3.17, 3.26, 4.46)$ and the p-values are $(0.00077, 0.00056, 0)$. However, as noted previously, this ODE model is not linear and, thus, we need to be careful when making a decision based on the result of IM_n only. The values of GM_n for the three component functions are $(13.44, 2.68, 25.96)$ and the p-values are $(0, 0.0037, 0)$. These results again suggest that none of the three component functions under the null are tenable. Therefore, we consider that the models may not fit the data well. On the other hand, we also realize that make sense statistically are only references for investigating whether they have biological meaning. However, this is not discussed further because it is beyond the scope of this study.

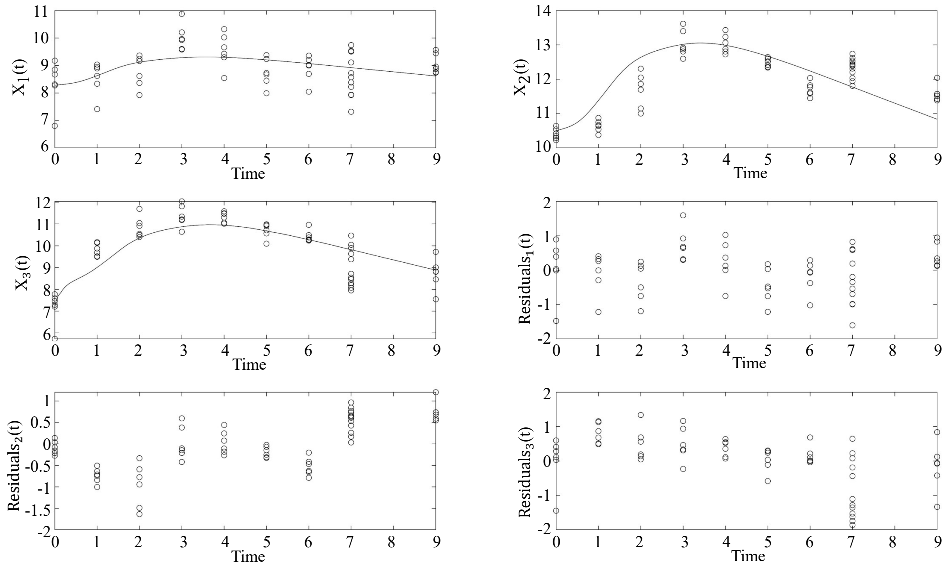

Figure 1. Time course of response and residuals.

7. Conclusion

In this paper, we have investigated model checking for parametric ODE systems and have proposed three tests.

Unlike TM_n, IM_n and GM_n cannot deal with partially observed ODE systems. If some components are unmeasured, we cannot obtain the kernel estimators to decouple the relationships between the components. Furthermore, the two-step method used to build the tests does not work, for the same reason. In the case, existing estimation methods usually need to draw support from the model structure (see, e.g., Dattner (2015)). However, it is difficult to use information about the model structure while eliminating the effects of the mixed components and mixed parameters in hypothesis testing. Thus we may need to use other methods, such as the semiparametric approach used by Hooker (2009) or the profiling method (Ramsay et al. (2007)). How to identify incorrectly modeled components in partially observed systems deserves further study.

We have discussed two kinds of alternatives in which the disturbances are represented as functions of time in the mathematical analysis. However, we have only tried the alternatives that have disturbances on $X'(t)$ depending on $X(t)$ our simulations. In finite-sample scenarios, the power may be quite different for alternatives with other disturbances, such as disturbances on $X(t)$ or momentary disturbances not depending on $X(t)$. Trying different kinds of alternatives is left to further research.

Several other issues are worth investigating in future studies. First, we find that GM_n outperforms IM_n, but is still not satisfactory in some cases. Solving this problem is important . Second, as seen in the limited simulations, IM_n finds it difficult to control the significance level, owing to its sensitivity to the non-parametric estimator. Modifing it is a nontrivial task. Third, for ODE models, when the ODE system is large, p is large, which is a challenging problem.

Supplementary Material

The online Supplementary Material includes additional conditions, remarks on the notation, lemmas, technical proofs, and other specific details.

Acknowledgments

This study was supported by grants from the University Grants Council of Hong Kong (HKBU12303921), NSFC (12131006), and Science and Technology Innovation Plan of Shanghai (20JC1414200). The authors thank the editor, associate editor, and referee for their helpful comments and suggestions.

References

Brunel, N. J.-B. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electronic Journal of Statistics* **2**, 1242–1267.

Chen, S. X. and van Keilegom, I. (2009). A goodness-of-fit test for parametric and semi-parametric models in multiresponse regression. *Bernoulli* **15**, 955–976.

Dattner, I. (2015). A model-based initial guess for estimating parameters in systems of ordinary differential equations. *Biometrics* **71**, 1176–1184.

Dattner, I. and Klaassen, C. A. J. (2015). Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *Electronic Journal of Statistics* **9**, 1939–1973.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *The Annals of Statistics* **27**, 1012–1040.

Ding, A. A. and Wu, H. (2014). Estimation of ordinary differential equation parameters using constrained local polynomial regression. *Statistica Sinica* **24**, 1613–1631.

FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal* **1**, 445–466.

Goel, N. S., Maitra, S. C. and Montroll, E. W. (1971). On the Volterra and other nonlinear models of interacting populations. *Reviews of Modern Physics* **43**, 231–276.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test* **22**, 361–411.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics* **21**, 1926–1947.

Hooker, G. (2009). Forcing function diagnostics for nonlinear dynamics. *Biometrics* **65**, 928–936.

Hooker, G. and Ellner, S. P. (2015). Goodness of fit in nonlinear dynamics: Misspecified rates or misspecified states? *The Annals of Applied Statistics* **9**, 754–776.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics* **40**, 633–643.

Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *The Annals of Statistics* **32**, 995–1034.

Lavergne, P. and Patilea, V. (2012). One for all and all for one: Regression checks with many regressors. *Journal of Business & Economic Statistics* **30**, 41–52.

Li, L., Chiu, S. N. and Zhu, L.-X. (2019). Model checking for regressions: An approach bridging between local smoothing and global smoothing methods. *Computational Statistics & Data Analysis* **138**, 64–82.

Liang, H. and Wu, H. (2008). Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association* **103**, 1570–1583.

Nagumo, J., Arimoto, S. and Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proceedings of the IRE* **50**, 2061–2070.

Ramsay, J. and Hooker, G. (2017). *Dynamic Data Analysis*. Springer, New York.

Ramsay, J. O., Hooker, G., Campbell, D. and Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 741–796.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics* **25**, 613–641.

Stute, W., Thies, S. and Zhu, L.-X. (1998). Model checks for regression: An innovation process approach. *The Annals of Statistics* **26**, 1916–1934.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* **76**, 419.

Xue, H., Miao, H. and Wu, H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *The Annals of Statistics* **38**, 2351–2387.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* **75**, 263–289.

Zhu, L.-X. (2003). Model checking of dimension-reduction type for regression. *Statistica Sinica* **13**, 283–296.

Ran Liu

Department of Mathematics, Hong Kong Baptist University, Hong Kong 111111, China.

E-mail: sdqdliuran@life.hkbu.edu.hk

Yun Fang

Department of Mathematics, Shanghai Normal University, Shanghai 200234, China.

E-mail: fangyun0919@shnu.edu.cn

Lixing Zhu
Advanced Institute of Natural Sciences, Beijing Normal University at Zhuhai, Zhuhai 519085, China.
E-mail: lzhu@bnu.edu.cn

(Received October 2020; accepted October 2021)