Supplementary Materials of "Green's matching: an efficient approach to parameter estimation in complex dynamic systems"

Author: Jianbin Tan, Guoyu Zhang, Xueqin Wang, Hui Huang, and Fang Yao.

Contents

1 Technical details

In this section, we present the technical details of Green's matching. As a beginning, we recall that $Y_i(t_j)$s are the observed dynamic data, where the time point of the observed data $t_j \in [0, C]$. We assume that the observed data follows the measurement model

$$Y_i(t_j) = X_i(t_j) + \varepsilon_i(t_j),$$

for $i = 1, \cdots, p$ and $j = 1, \cdots, n$, where $\varepsilon_i(t_j)$s are the i.i.d. mean-zero noise variables, and $X_i(t_j)$ is the mean of $Y_i(t_j)$ conditioning on the time point t_j. Moreover, $X_i(t)$s are the smooth curves over $t \in [0, C]$ and follow the differential equations

$$\mathcal{D}_i^K X_i(t) = f_i(\boldsymbol{X}(t), t; \boldsymbol{\beta}), \ \forall t \in [0, C], \tag{1}$$

for $i = 1, \cdots, p$, where $\mathcal{D}_i^K$ is the differential operator of order K as defined in (3) in the main text, $\boldsymbol{X}(t) = (X_1(t), \cdots, X_p(t))^T$, and $f_i(\cdot; \boldsymbol{\beta})$ is a driving function indexed by the unknown parameters $\boldsymbol{\beta}$.

1.1 Local polynomial regression

In order to bypass equation solving, we impose a pre-smoothing step to facilitate the model fitting. Noting that $X_i(\cdot)$s are at least K-times differentiable, we use the local polynomial regression [Fan and Gijbels, 2018] to implement smooth approximations. Essentially, the local polynomial regression utilizes the Taylor expansion of $X_i(t_j)$ on t

$$X_i(t_j) = X_i(t) + DX_i(t)(t_j - t) + D^2 X_i(t) \frac{(t_j - t)^2}{2!} + \cdots .$$

Given the highest order of polynomials $m_k \geq k$, we estimate $X_i^{(k)}(t)$ by $\widehat{D^k X}_i(t) = a_k$, with a_k minimizing the weighted least squares

$$\sum_{j=1}^{n} W\left(\frac{t_j - t}{h_k}\right) \left\{ Y_i(t_j) - \sum_{v=0}^{m_k} a_v \frac{(t_j - t)^v}{v!} \right\}^2, \tag{2}$$

where $W(\cdot)$ is a function to assign a weight for each observation and introduce smoothness, h_k is a bandwidth to control the weighted smoothing, and a_v is a coefficient to approximate $D^v X_i(t)$. Simple calculation shows that $\widehat{D^k X}_i(t)$ can be represented as

$$\sum_{j=1}^{n} W_{k,n}\left(\frac{t_j - t}{h_k}\right) Y_i(t_j),$$

where $W_{k,n}(\cdot)$ is an estimated weight function determined by $W(\cdot)$, the differential order k, and the sample size n; see Section 3.2.2 in Fan and Gijbels [2018] for more details.

Here, we use a bandwidth h_0 selected by a data-driven cross-validation method [Fan and Gijbels, 2018]. Meanwhile, we also utilize a similar scheme to choose the bandwidths h_ks for the higher-order derivatives. In detail, we select the bandwidth h_k for $\hat{X}_i^{(k)}(t)$ by minimizing

$$\sum_{j=1}^{J} \left\{ Y_i(t_j) - \hat{X}_i^{-j,h_k}(t_j) \right\}^2 w_{\text{bw}}(t_j),$$

where $\hat{X}_i^{-j,h_k}(t)$ is taken as a_0 obtained from (2) with the data $\{(t_{j'}, Y_i(t_{j'})); j' \neq j\}$ and h_k, and $w_{\mathrm{bw}}(t_j)$ is a specified weight at t_j to alleviate the low convergence rate caused by boundary effects.

1.2 A unified representation of Green's functions

To implement Green's matching, we construct unified representations for $\boldsymbol{\psi}_i(t)$ and $G_i^K(t,s)$ based on $\mathcal{D}_i^K$ with form $D^K + \sum_{k=0}^{K-1} \omega_{ik} D^k$.

Let $[\ ]_{m,m'}$, $[\ ]_{m,\cdot}$ and $[\ ]_{\cdot,m}$ be the operations to extract the $(m,m')^{th}$ element, m^{th} row and m^{th} column of a matrix, respectively, and denote $[\cdot]_m$ as the m^{th} component of a vector. We first present the following lemma.

Lemma 1. Assume that $G_i^K(0,s) = \frac{\partial G_i^K(t,s)}{\partial t}\Big|_{t=0} = \cdots = \frac{\partial^{K-1} G_i^K(t,s)}{\partial t^{K-1}}\Big|_{t=0} = 0$, $\forall\ s \in [0, C]$, then $G_i^K(t,s)$ could be represented as

$$G_i^K(t,s) = \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{v}_i(s) \mathbb{I}(t \geq s),$$

where $\boldsymbol{\psi}_i(t) \in \mathbb{R}^K$ is an arbitrary collection of basis functions of $\mathrm{Ker}(\mathcal{D}_i^K)$, and $\boldsymbol{v}_i(s) \in \mathbb{R}^K$ is uniquely determined by the equations

$$\frac{\partial^k G_i^K(t,s)}{\partial t^k}\Bigg|_{t=s^+} - \frac{\partial^k G_i^K(t,s)}{\partial t^k}\Bigg|_{t=s^-} = 0, \quad k = 0, \cdots, K-2, \tag{3}$$

$$\frac{\partial^{K-1} G_i^K(t,s)}{\partial t^{K-1}}\Bigg|_{t=s^+} - \frac{\partial^{K-1} G_i^K(t,s)}{\partial t^{K-1}}\Bigg|_{t=s^-} = 1. \tag{4}$$

The proof of Lemma 1 is given in Section 1 in Duffy [2015]. Define

$$\boldsymbol{A}(\boldsymbol{\omega}_i) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\omega_{i0} & -\omega_{i1} & -\omega_{i2} & \cdots & -\omega_{i(K-1)} \end{bmatrix}$$

with $\boldsymbol{\omega}_i = (\omega_{i0}, \cdots, \omega_{i(K-1)})$. Notice that if $\boldsymbol{z} : [0, C] \to \mathbb{R}^K$ satisfies

$$\frac{\mathrm{d}\boldsymbol{z}(t)}{\mathrm{d}t} = \boldsymbol{A}(\boldsymbol{\omega}_i)\boldsymbol{z}(t),$$

$\boldsymbol{z}(t)$ could be represented as $\mathbf{e}^{t\boldsymbol{A}(\boldsymbol{\omega}_i)}\boldsymbol{c}$ with any $\boldsymbol{c} \in \mathbb{R}^K$, where $\mathbf{e}^{(\cdot)}$ is the matrix exponential of a matrix. Notice that

$$\mathrm{Ker}(\mathcal{D}_i^K) = \left\{ x(t) \;\middle|\; x(t) = [\boldsymbol{z}(t)]_1 \text{ with } \boldsymbol{z}(t) \text{ s.t. } \frac{\mathrm{d}\boldsymbol{z}(t)}{\mathrm{d}t} = \boldsymbol{A}(\boldsymbol{\omega}_i)\boldsymbol{z}(t) \right\}.$$

Accordingly, a collection of basis functions for $\mathrm{Ker}(\mathcal{D}_i^K)$ could be constructed as

$$\{\boldsymbol{\psi}_i(t)\}^T = \left[\mathbf{e}^{t\boldsymbol{A}(\boldsymbol{\omega}_i)}\right]_{1,\cdot}.$$

Providing $\boldsymbol{\psi}_i^T(t) = \left[\mathbf{e}^{t\boldsymbol{A}(\boldsymbol{\omega}_i)}\right]_{1,\cdot}$, $\boldsymbol{v}_i(s)$ can be uniquely solved under the constraints (3) and (4), which gives

$$G_i^K(t,s) = \left[\mathbf{e}^{(t-s)\boldsymbol{A}(\boldsymbol{\omega}_i)}\right]_{1,K}\mathbb{I}(t \geq s).$$

For the cases of non-zero $\boldsymbol{\omega}_i$, the matrix exponential above could be efficiently calculated, for example, by **R** package expm [Goulet et al., 2017]. Besides, when $\boldsymbol{\omega}_i = \mathbf{0}$, the above equations indicate that $\boldsymbol{\psi}_i(t) = \left(1, t, \cdots, \frac{t^{K-1}}{(K-1)!}\right)^T$ and $G_i^K(t,s) = \frac{(t-s)^{K-1}}{(K-1)!}\mathbb{I}(t \geq s)$.

2 Proofs of theorems

For simplicity, we assume that $C = 1$ in this section, i.e., we only focus on the differential equations on $[0,1]$. We first present the following lemma to evaluate the convergence rate of the local linear estimation $\hat{X}_i(\cdot)$s, which proof can be found in Stone [1982].

Lemma 2. Under Assumptions A.1−A.4, and assuming that the bandwidth h_0 for $\hat{X}_i(t)$s satisfies $h_0 = o(1)$ such that $n^{-1}\ln n/h_0 \to 0$ as $n \to \infty$, then

$$\sup_{t\in[0,1]} |\hat{X}_i(t) - X_i(t)| = O_P(b_n),\ \forall i = 1, \cdots, p, \tag{5}$$

where $b_n = h_0^2 + n^{-\frac{1}{2}}h_0^{-\frac{1}{2}}\ln^{\frac{1}{2}} n$.

This lemma provides a uniform bound for the estimation error of local linear regressions. Defining $\|\boldsymbol{z}(\cdot)\|_2^2 = \int_0^1 ||\boldsymbol{z}(t)||^2\, \mathrm{d}t$ for a function $\boldsymbol{z} : [0,1] \to \mathbb{R}^p$, we next present similar results for the convergence rates of the estimated k^{th} derivatives, which proof can also be found in Stone [1982]. To simplify notations, we use $X_i^{(k)}(t)$ instead of $D^k X_i(t)$ in the following.

Lemma 3. Under Assumptions A.1-A.3, A.6, B.4 and the bandwidth for $\hat{X}_i^{(k)}(t)$s satisfies $h_k = o(1)$ such that $n^{-1}\ln n/h_k^{2k+1} \to 0$ as $n \to \infty$ for all $k = 0, \cdots, K$. Then

$$\sup_{t\in[0,1]} |\hat{X}_i^{(k)}(t) - X_i^{(k)}(t)| = O_P(b_{nk}(h_k)),\ k = 0, 1, \cdots, K, \text{ and } \forall i = 1, \cdots, p, \tag{6}$$

$$\|\hat{\boldsymbol{X}}^{(k)}(\cdot) - \boldsymbol{X}^{(k)}(\cdot)\|_2 = O_P(\delta_{nk}(h_k)),\ k = 0, 1, \cdots, K, \tag{7}$$

where $b_{nk}(h) = h^2 + n^{-\frac{1}{2}}h^{-k-\frac{1}{2}}\ln^{\frac{1}{2}} n$ and $\delta_{nk}(h) = h^2 + n^{-\frac{1}{2}}h^{-k-\frac{1}{2}}$.

2.1 Proof of Theorem 1

In the following, we denote that $a_n \lesssim b_n$ for two sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ if there exists a constant C independent of n s.t. $a_n \leq Cb_n$. For the identifiability of $\boldsymbol{\alpha}$, we focus on the cases where $\boldsymbol{\alpha}$ is estimated based on the Green's functions constructed in the main text.

Proof. For simplicity, we denote $(\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ and $(\boldsymbol{\beta}_0^T, \boldsymbol{\alpha}_0^T)^T$ as $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, respectively, and represent $\Omega_{\boldsymbol{\theta}} := \Omega_{\boldsymbol{\beta}} \times \Omega_{\boldsymbol{\alpha}}$. Define

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{i=1}^{p} \int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}) \text{ d}s \right]^2 w(t) \text{ d}t,$$

where $\boldsymbol{\psi}_i(t)$ is a basis of $\text{Ker}(\mathcal{D}_i^K)$ and $G_i^K(t,s)$ is the Green's function of $\mathcal{D}_i^K$. Accordingly, Let $\left((\hat{\boldsymbol{\beta}}^{\text{Gree}})^T, \hat{\boldsymbol{\alpha}}^T\right)^T$ be $\hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta} \in \Omega_{\boldsymbol{\theta}}} \mathcal{L}_n(\boldsymbol{\theta})$.

To obtain the consistency of $\hat{\boldsymbol{\beta}}^{\text{Gree}}$, we first divide the loss function $\mathcal{L}_n(\boldsymbol{\theta})$ into three parts

$$\begin{aligned}
\mathcal{L}_n(\boldsymbol{\theta}) =& \sum_{i=1}^{p} \int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s); \boldsymbol{\beta}) \text{d}s \right]^2 w(t) \text{ d}t \\
&+ \sum_{i=1}^{p} \int_0^1 \left[\int_0^1 G_i^K(t,s) \left\{ f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}) - f_i(\boldsymbol{X}(s); \boldsymbol{\beta}) \right\} \text{d}s \right]^2 w(t) \text{ d}t \\
&- 2 \sum_{i=1}^{p} \int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s); \boldsymbol{\beta}) \text{d}s \right] \\
&\cdot \left[\int_0^1 G_i^K(t,s) \left\{ f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}) - f_i(\boldsymbol{X}(s); \boldsymbol{\beta}) \right\} \text{d}s \right] w(t) \text{ d}t \\
=:& \, \mathcal{L}_n^*(\boldsymbol{\theta}) + I_1 + I_2.
\end{aligned}$$

Moreover, we define

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{p} \int_0^1 \left[X_i(t) - \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s); \boldsymbol{\beta}) \text{d}s \right]^2 w(t) \text{ d}t,$$

and we will prove that

$$\sup_{\boldsymbol{\theta} \in \Omega_{\boldsymbol{\theta}}} \left| \mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta}) \right| = o_P(1), \text{ as } n \to \infty. \tag{8}$$

Firstly, by noting that $w(\cdot)$ is bounded and $G_i^K(t,s)$ is a bounded function on $[0,1]^2$, we have

$$\begin{aligned}
I_1 =& \sum_{i=1}^{p} \int_0^1 \left[\int_0^1 G_i^K(t,s) \left\{ f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}) - f_i(\boldsymbol{X}(s); \boldsymbol{\beta}) \right\} \text{d}s \right]^2 w(t) \text{ d}t \\
\leq& \sum_{i=1}^{p} \sup_{t \in [0,1]} \left| f_i(\hat{\boldsymbol{X}}(t); \boldsymbol{\beta}) - f_i(\boldsymbol{X}(t); \boldsymbol{\beta}) \right|^2 \int_0^1 \int_0^1 \left\{ G_i^K(t,s) \right\}^2 w(t) \text{ d}s\text{d}t \\
\lesssim& \sum_{i=1}^{p} \sup_{t \in [0,1]} \left| f_i(\hat{\boldsymbol{X}}(t); \boldsymbol{\beta}) - f_i(\boldsymbol{X}(t); \boldsymbol{\beta}) \right|^2.
\end{aligned}$$

By using this inequality, we similarly achieve

$$\begin{aligned}
|I_2| =& 2\sum_{i=1}^{p}\int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T\boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\mathrm{d}s\right] \\
&\cdot \left[\int_0^1 G_i^K(t,s)\left\{f_i(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\right\}\mathrm{d}s\right] w(t)\ \mathrm{d}t \\
\lesssim& \sum_{i=1}^{p}\left[\int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T\boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\mathrm{d}s\right]^2 w(t)\ \mathrm{d}t\right]^{1/2} \\
&\cdot \left[\int_0^1 \left[\int_0^1 G_i^K(t,s)\left\{f_i(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\right\}\mathrm{d}s\right]^2 w(t)\ \mathrm{d}t\right]^{1/2} \\
\lesssim& \sum_{i=1}^{p}\left[\int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T\boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\mathrm{d}s\right]^2 w(t)\ \mathrm{d}t\right]^{1/2} \\
&\cdot \sup_{t\in[0,1]}\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right|.
\end{aligned}$$

Combing the above two inequalities, we obtain

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\left|\mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| \\
\lesssim& \sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\left|\mathcal{L}_n^*(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| \\
&+ \sum_{i=1}^{p}\sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}}\sup_{t\in[0,1]}\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right|^2 \\
&+ \sum_{i=1}^{p}\left[\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T\boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\mathrm{d}s\right]^2 w(t)\ \mathrm{d}t\right]^{1/2} \\
&\cdot \sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}}\sup_{t\in[0,1]}\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right|.
\end{aligned}$$

Based on the above inequality, it suffices to show that for any $i = 1\cdots,p$,

$$\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\left|\mathcal{L}_n^*(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| = o_P(1), \tag{9}$$

$$\sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}}\sup_{t\in[0,1]}\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right| = o_P(1), \tag{10}$$

$$\sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}}\int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T\boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\mathrm{d}s\right]^2 w(t)\ \mathrm{d}t = O_P(1), \tag{11}$$

as $n \to \infty$ for proving (8). Noting that (11) is simply implied by (9) and the fact that $\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\left|\mathcal{L}(\boldsymbol{\theta})\right| = O(1)$ (which is a consequence of Assumption B.2 in the main text), we just need to prove (9) and (10).

Proof for (9):

We begin with this equality

$$\mathcal{L}_n^*(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) + \sum_{i=1}^{p} \int_0^1 \left[\hat{X}_i(t) - \hat{X}_i(t)\right]^2 w(t)\ \mathrm{d}t$$
$$+2\sum_{i=1}^{p} \int_0^1 \left[X_i(t) - \hat{X}_i(t)\right] \cdot \left[X_i(t) - \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\mathrm{d}s\right] w(t)\ \mathrm{d}t.$$

Define

$$r_i(t) = X_i(t) - \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\mathrm{d}s,$$

then r_i is bounded among all $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}$ (by Assumption B.2 in the main text), and

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}} \left|\mathcal{L}_n^*(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| \\
\lesssim &\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}} \sum_{i=1}^{p} \int_0^1 r_i(t) \left[X_i(t) - \hat{X}_i(t)\right] w(t)\ \mathrm{d}t + \sum_{i=1}^{p} \int_0^1 \left[\hat{X}_i(t) - \hat{X}_i(t)\right]^2 w(t)\ \mathrm{d}t \\
\lesssim &\sum_{i=1}^{p} \sup_{t\in[0,1]} |\hat{X}_i(t) - X_i(t)| + \sum_{i=1}^{p} \sup_{t\in[0,1]} |\hat{X}_i(t) - X_i(t)|^2,
\end{aligned}$$

hence $\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}} \left|\mathcal{L}_n^*(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| = o_P(1)$ when $n \to \infty$ by Lemma 2.

Proof for (10):

We define the set

$$\Omega_r(\boldsymbol{X}) := \{\boldsymbol{\eta} \in \mathbb{R}^p \mid \exists s \in [0,1]\ s.t.\ \|\boldsymbol{\eta} - \boldsymbol{X}(s)\| \le r\},$$

where $\|\cdot\|$ is the Euclidean norm. By Lemma 2, we have $\sup_{t\in[0,1]} \|\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\| = O_P(b_n)$ as $n \to \infty$. Then $\forall \varepsilon > 0$, there exists a $M > 0$ and $N > 0$ such that

$$\mathbb{P}\left(\sup_{t\in[0,1]} |\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)| \le Mb_N\right) \ge 1 - \epsilon$$

for all $n \ge N$. Accordingly,

$$\mathbb{P}\left((\hat{\boldsymbol{X}}(t), \boldsymbol{\beta}) \in \Omega_{Mb_N}(\boldsymbol{X}) \times \Omega_\beta,\ \forall t \in [0,1]\right) \ge \mathbb{P}\left(\sup_{t\in[0,1]} |\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)| \le Mb_N\right) \ge 1 - \varepsilon.$$

Notice that $\Omega_{Mb_N}(\boldsymbol{X}) \times \Omega_\beta$ is compact, then by Assumption B.3 in the main text, f_i is uniformly continuous on $\Omega_{Mb_N}(\boldsymbol{X}) \times \Omega_{\boldsymbol{\beta}}$. It follows that, $\forall\ \varepsilon_1 > 0$, $\exists\ \delta > 0$ s.t.

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}}\sup_{t\in[0,1]}\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right| > \varepsilon_1\right) \\
\leq &\mathbb{P}\left(\sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}}\sup_{t\in[0,1]}\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right| > \varepsilon_1,\ (\hat{\boldsymbol{X}}(t),\boldsymbol{\beta}) \in \Omega_{Mb_N}(\boldsymbol{X}) \times \Omega_{\boldsymbol{\beta}},\ \forall t \in [0,1]\right) \\
&+ \varepsilon \\
\leq &\mathbb{P}\left(\sup_{t\in[0,1]}|\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)| > \delta, (\hat{\boldsymbol{X}}(t),\boldsymbol{\beta}) \in \Omega_{Mb_N}(\boldsymbol{X}) \times \Omega_{\beta},\ \forall t \in [0,1]\right) + \varepsilon.
\end{aligned}$$

Then by Lemma 2 again, we have

$$\sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}}\sup_{t\in[0,1]}\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right| = o_P(1),$$

as $n \to \infty$.

After proving (9) and (10), we complete the proof of

$$\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\left|\mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| = o_P(1), \tag{12}$$

as $n \to \infty$.

Moreover, by (10) in the main text, we have

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{p}\int_0^1\left[\{\boldsymbol{\psi}_i(t)\}^T(\boldsymbol{\alpha}_{0,i} - \boldsymbol{\alpha}_i) + \int_0^1 G_i^K(t,s)\left\{f_i(\boldsymbol{X}(s);\boldsymbol{\beta}_0) - f_i(\boldsymbol{X}(s);\boldsymbol{\beta})\right\}\mathrm{d}s\right]^2 w(t)\ \mathrm{d}t,$$

where $(\boldsymbol{\alpha}_{0,1}^T, \cdots, \boldsymbol{\alpha}_{0,p}^T)^T := \boldsymbol{\alpha}_0$. Under Assumption B.1 in the main text, we obtain a unique minimum of $\mathcal{L}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$. Since $\mathcal{L}(\boldsymbol{\theta})$ is continuous w.r.t. $\boldsymbol{\theta}$, we have

$$\inf_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\epsilon}\mathcal{L}(\hat{\boldsymbol{\theta}}) > \mathcal{L}(\boldsymbol{\theta}_0)$$

for any $\epsilon > 0$. Finially, by Theorem 5.7 in chapter 5 of Van der Vaart [2000], the conclusion in (12) implies that $\left((\hat{\boldsymbol{\beta}}^{\text{Gree}})^T, \hat{\boldsymbol{\alpha}}^T\right)^T$ is a consistent estimator of $\boldsymbol{\theta}_0$. $\square$

2.2 Proof of Theorem 2

For ease of notation, we denote h as h_0 in this subsection. Before proving theorem 2, we prove the following lemma for evaluating the convergence rate of local linear estimators.

Lemma 4. Under Assumptions A.1-A.4, B.4 and the bandwidth h for $\hat{X}_i(t)$s satisfies $h = o(1)$ such that $nh \to \infty$ and $n^{-1}h^{-1}\ln n \to 0$ as $n \to \infty$, we have for any $\forall\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$ $i = 1, \cdots, p$, and for all bounded function $g(t)$,

$$\int_0^1\left|f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta})\right|g(t)\ \mathrm{d}t = O_P(\delta_{n0}(h)), \tag{13}$$

$$\int_0^1 \left| f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta}) - \frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(t);\boldsymbol{\beta}) \left\{ \hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t) \right\} \right| g(t)\ \mathrm{d}t = O_P(\delta_{n0}^2(h)). \tag{14}$$

Proof of Lemma 4. First, notice that

$$\begin{aligned} &\int_0^1 \left| f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta}) \right| g(t)\ \mathrm{d}t \\ \le& \sup_{t,s\in[0,1]} \left\| \frac{\partial f_i}{\partial \boldsymbol{X}^T}(s\boldsymbol{X}(t) + (1-s)\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) \right\| \cdot \left\| \hat{\boldsymbol{X}}(\cdot) - \boldsymbol{X}(\cdot) \right\|_2 \cdot \sup_{t\in[0,1]} |g(t)|. \end{aligned}$$

Recall that $\Omega_r(\boldsymbol{X}) = \{\boldsymbol{\eta} \in \mathbb{R}^p \mid \exists s \in [0,1]\ s.t.\ \|\boldsymbol{\eta} - \boldsymbol{X}(s)\| \le r\}$, by a similar argument in proving (10), there exist some M and N such that $\hat{\boldsymbol{X}}(t)$ belongs to $\Omega_{Mb_N}(\boldsymbol{X})$, $\forall t$, with a high probability. By the compactness of $\Omega_{Mb_N}(\boldsymbol{X})$ and the continuity of $\frac{\partial f_i}{\partial \boldsymbol{X}}$ (Assumption B.4 in the main text), we have

$$\sup_{t\in[0,1]} \sup_{s\in[0,1]} \left\| \frac{\partial f_i}{\partial \boldsymbol{X}}(s\boldsymbol{X}(t) + (1-s)\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) \right\| = O_P(1).$$

Combining with (7) in Lemma 3, we have

$$\int_0^1 \left| f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta}) \right| g(t)\mathrm{d}t = O_P(\delta_{n0}(h)).$$

Similarly, (14) can be proved in the same way by noticing that

$$\begin{aligned} &\int_0^1 \left| f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta}) - \frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(t);\boldsymbol{\beta}) \left\{ \hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t) \right\} \right| g(t)\ \mathrm{d}t \\ \le& \sup_{t,s\in[0,1]} \left\| \frac{\partial^2 f_i}{\partial \boldsymbol{X} \partial \boldsymbol{X}^T}(s\boldsymbol{X}(t) + (1-s)\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) \right\|_{\mathrm{op}} \cdot \sup_{t\in[0,1]} |g(t)| \cdot \left\| \hat{\boldsymbol{X}}(\cdot) - \boldsymbol{X}(\cdot) \right\|_2^2, \end{aligned}$$

and

$$\sup_{t,s\in[0,1]} \left\| \frac{\partial^2 f_i}{\partial \boldsymbol{X} \partial \boldsymbol{X}^T}(s\boldsymbol{X}(t) + (1-s)\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}) \right\|_{\mathrm{op}} = O_P(1),$$

where $\|\cdot\|_{\mathrm{op}}$ is the largest singular value of a matrix. □

With Lemma 4, we now begin to prove Theorem 2.

Proof. Recall that

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{i=1}^p \int_0^1 \left[\hat{X}_i(t) - \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_i - \int_0^1 G_i^K(t,s) f_i(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta})\ \mathrm{d}s \right]^2 w(t)\ \mathrm{d}t \tag{15}$$

is the loss function of Green's matching, and $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Omega_\theta} \mathcal{L}_n(\boldsymbol{\theta})$ with $\boldsymbol{\theta}_0$ being the true value of $\hat{\boldsymbol{\theta}}$. Define $\theta_i = [\boldsymbol{\theta}]_i$, $\theta_{0,i} = [\boldsymbol{\theta}_0]_i$, and $\hat{\theta}_i = [\hat{\boldsymbol{\theta}}]_i$. Here, we apply the mean value theorem on $\frac{\partial \mathcal{L}_n}{\partial \theta_i}(\boldsymbol{\theta})$ w.r.t. θ_j between $\hat{\theta}_j$ and $\theta_{0,j}$, i.e.,

$$\mathbf{0} = \frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) + \boldsymbol{M}(\boldsymbol{\theta}^*)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),$$

where $\boldsymbol{M}(\boldsymbol{\theta}^*)$ is defined as $\left(\frac{\partial^2 \mathcal{L}_n}{\partial \theta_i \partial \theta_j}(\theta^*_{ij})\right)_{i,j=1,2,\cdots}$ with $\boldsymbol{\theta}^* = (\theta^*_{ij})_{i,j=1,2,\cdots}$ and $\theta^*_{ij} = v_i \theta_{0,j} + (1 - v_i)\hat{\theta}_j$ for some $v_i \in (0,1)$. We will show that $\boldsymbol{M}(\boldsymbol{\theta}^*)$ is asymptotically invertible, and

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left\{\boldsymbol{M}(\boldsymbol{\theta}^*)\right\}^{-1} \frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0). \tag{16}$$

Define

$$\begin{aligned}
l_i(t) :=& \{\boldsymbol{\psi}_i(t)\}^T \boldsymbol{\alpha}_{0,i} + \int_0^1 G_i^K(t,s) f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}_0)\ \mathrm{d}s - \hat{X}_i(t) \\
=& \int_0^1 G_i^K(t,s) \left\{ f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}_0) - f_i(\boldsymbol{X}(s); \boldsymbol{\beta}_0) \right\} \mathrm{d}s + X_i(t) - \hat{X}_i(t) \\
=& \int_0^1 G_i^K(t,s) \frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(s); \boldsymbol{\beta}_0) \left\{ \hat{\boldsymbol{X}}(s) - \boldsymbol{X}(s) \right\} \mathrm{d}s + X_i(t) - \hat{X}_i(t) \\
&+ \int_0^1 G_i^K(t,s) \cdot \Bigg[f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}_0) - f_i(\boldsymbol{X}(s); \boldsymbol{\beta}_0) \\
&\qquad - \frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(s); \boldsymbol{\beta}_0) \left\{ \hat{\boldsymbol{X}}(s) - \boldsymbol{X}(s) \right\} \Bigg] \mathrm{d}s.
\end{aligned}$$

Then

$$\begin{aligned}
\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{i=1}^{p} \int_0^1 \Bigg[& \{\boldsymbol{\psi}_i(t)\}^T (\boldsymbol{\alpha}_{0,i} - \boldsymbol{\alpha}_i) \\
& + \int_0^1 G_i^K(t,s) \left\{ f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}) - f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta}_0) \right\}\ \mathrm{d}s - l_i(t) \Bigg]^2 w(t)\ \mathrm{d}t.
\end{aligned}$$

Recall that $\boldsymbol{A} \otimes \boldsymbol{B}$ is the Kronecker product between matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and $\boldsymbol{e}_i$ is a p-dimensional vector with the i^{th} element being 1 and 0 elsewhere. Then $\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)$ in (16) could be represented as

$$
\begin{aligned}
&\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)\\
=&2\sum_{i=1}^{p}\int_0^1 l_i(t)\bigg(\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}_0)ds,\ \boldsymbol{e}_i^T\otimes\{\boldsymbol{\psi}_i(t)\}^T\bigg)^T w(t)\ \mathrm{d}t\\
=&2\sum_{i=1}^{p}\int_0^1 \left(\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)\left\{\hat{\boldsymbol{X}}(s)-\boldsymbol{X}(s)\right\}\mathrm{d}s + X_i(t)-\hat{X}_i(t)\right)\\
&\cdot\left(\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}_0)ds,\ \boldsymbol{e}_i^T\otimes\{\boldsymbol{\psi}_i(t)\}^T\right)^T w(t)\ \mathrm{d}t\\
&+2\sum_{i=1}^{p}\int_0^1 \bigg[\int_0^1 G_i^K(t,s)\cdot\bigg[f_i(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}_0)\\
&\qquad -f_i(\boldsymbol{X}(s);\boldsymbol{\beta}_0)-\frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)\left\{\hat{\boldsymbol{X}}(s)-\boldsymbol{X}(s)\right\}\bigg]\mathrm{d}s\bigg]\\
&\cdot\left(\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}_0)ds,\ \boldsymbol{e}_i^T\otimes\{\boldsymbol{\psi}_i(t)\}^T\right)^T w(t)\ \mathrm{d}t\\
=&2\sum_{i=1}^{p}\int_0^1 \left[\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)\left\{\hat{\boldsymbol{X}}(s)-\boldsymbol{X}(s)\right\}\mathrm{d}s + X_i(t)-\hat{X}_i(t)\right]\\
&\cdot\left(\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)ds,\ \boldsymbol{e}_i^T\otimes\{\boldsymbol{\psi}_i(t)\}^T\right)^T w(t)\ \mathrm{d}t\\
&+2\sum_{i=1}^{p}\int_0^1 \left[\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)\left\{\hat{\boldsymbol{X}}(s)-\boldsymbol{X}(s)\right\}\mathrm{d}s + X_i(t)-\hat{X}_i(t)\right]\\
&\cdot\left(\int_0^1 G_i^K(t,s)\left[\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}_0)-\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)\right]ds,\ \boldsymbol{0}\right)^T w(t)\ \mathrm{d}t\\
&+2\sum_{i=1}^{p}\int_0^1 \bigg[\int_0^1 G_i^K(t,s)\cdot\bigg[f_i(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}_0)-f_i(\boldsymbol{X}(s);\boldsymbol{\beta}_0)-\\
&\qquad \frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)\left\{\hat{\boldsymbol{X}}(s)-\boldsymbol{X}(s)\right\}\bigg]\mathrm{d}s\bigg]\\
&\cdot\left(\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}_0)ds,\ \boldsymbol{e}_i^T\otimes\{\boldsymbol{\psi}_i(t)\}^T\right)^T w(t)\ \mathrm{d}t\\
:=&\boldsymbol{I}_1+\boldsymbol{I}_2+\boldsymbol{I}_3. \qquad (17)
\end{aligned}
$$

Using Lemma 4, we obtain that

$$\|\boldsymbol{I}_2\| = O_P(\delta_{n0}^2(h)), \qquad (18)$$

$$\|\boldsymbol{I}_3\| = O_P(\delta_{n0}^2(h)), \qquad (19)$$

Using the Theorem 3.1 in Fan and Gijbels [2018], we have

$$\|\mathbb{E}\boldsymbol{I}_1\| = O(h^2). \qquad (20)$$

To calculate the variance of $\boldsymbol{I}_1$, we rewrite $\hat{X}_i(t)$ as

$$\hat{X}_i(t) = \sum_{j=1}^{n} W_0^n\left(\frac{t_j - t}{h}\right) Y_i(t_j), \tag{21}$$

where $W_0^n(\cdot)$ is a kernel function induced by the weight function $W(\cdot)$ used in local linear regression; see Section 3.2.2 in Chapter 3 in Fan and Gijbels [2018] for more details. Moreover, under Assumptions A.1-A.4 in the main text, Fan and Gijbels [2018] proved that the kernel $W_0^n(t)$ can be approximated by

$$W_0^n(\cdot) = \frac{1}{nhf_T(t)} W(\cdot)\left\{1 + O_P(h^2 + \frac{1}{\sqrt{nh}})\right\} \tag{22}$$

as $h \to 0$ such that $nh \to \infty$. For simplicity, we denote $\frac{\partial f_i(s)}{\partial \boldsymbol{\beta}} = \frac{\partial f_i}{\partial \boldsymbol{\beta}}(\boldsymbol{X}(s); \boldsymbol{\beta}_0)$ and $\frac{\partial f_i(s)}{\partial \boldsymbol{X}} = \frac{\partial f_i}{\partial \boldsymbol{X}}(\boldsymbol{X}(s); \boldsymbol{\beta}_0)$, etc., and define

$$\boldsymbol{d}_i(t) = \left(\int_0^1 G_i^K(t,s) \frac{\partial f_i(s)}{\partial \boldsymbol{\beta}^T} ds,\ \boldsymbol{e}_i^T \otimes \{\boldsymbol{\psi}_i(t)\}^T \right)^T. \tag{23}$$

We can use (21) to represent $\boldsymbol{I}_1$ as below

$$\begin{aligned}
\boldsymbol{I}_1 =& 2\sum_{i,l=1}^{p} \int_0^1 \boldsymbol{d}_i(t) \Big\{ \int_0^1 G_i^K(t,s) \sum_{l=1}^{p} \frac{\partial f_i(s)}{\partial X_l} \cdot \Big(\sum_{j=1}^{n} W_0^n(\frac{t_j - s}{h}) X_l(t_j) - X_l(s) \Big) \mathrm{d}s \\
&\qquad + X_i(t) - \sum_{j=1}^{n} W_0^n(\frac{t_j - t}{h}) X_i(t_j) \Big\} w(t)\ \mathrm{d}t \\
&+ 2\sum_{i=1}^{p}\sum_{j=1}^{n} \epsilon_i(t_j) \int_0^1 \Big\{ \sum_{l=1}^{p} \boldsymbol{d}_l(t) \int_0^1 G_l^K(t,s) \frac{\partial f_l(s)}{\partial X_i} W_0^n(\frac{t_j - s}{h}) \mathrm{d}s \\
&\qquad - \boldsymbol{d}_i(t) W_0^n(\frac{t_j - t}{h}) \Big\} w(t)\ \mathrm{d}t.
\end{aligned}$$

The variance of $\boldsymbol{I}_1$ can be calculated with (22):

$$\begin{aligned}
n \cdot \mathrm{Var}\left[\boldsymbol{I}_1\right] =& 4\sum_{i,l,u=1}^{p} \iiint_{[0,1]^3} \frac{\sigma_i^2(\tau)}{f_T(\tau)} G_l^K(t,\tau) G_u^K(s,\tau) \frac{\partial f_l(\tau)}{\partial X_i} \frac{\partial f_u(\tau)}{\partial X_i} \\
& \boldsymbol{d}_l(t) \cdot \boldsymbol{d}_u(s)^T w(t) w(s)\ \mathrm{d}t\ \mathrm{d}s\ \mathrm{d}\tau \\
& -8\sum_{i,l=1}^{p} \iint_{[0,1]^2} \frac{\sigma_i^2(\tau)}{f_T(\tau)} G_l^K(t,\tau) \frac{\partial f_l(\tau)}{\partial X_i} \boldsymbol{d}_l(t) \cdot \boldsymbol{d}_i(\tau)^T w(t) w(\tau)\ \mathrm{d}t\ \mathrm{d}\tau \\
& +4\sum_{i=1}^{p} \int_{[0,1]} \frac{\sigma_i^2(\tau)}{f_T(\tau)} \boldsymbol{d}_i(\tau) \cdot \boldsymbol{d}_i(\tau)^T\ w^2(\tau)\ \mathrm{d}\tau + O(\frac{1}{\sqrt{nh}} + h) \qquad (24) \\
=& :\boldsymbol{\Sigma}_2 + O(\frac{1}{\sqrt{nh}} + h).
\end{aligned} \tag{25}$$

With (17)-(20) and (25), we have

$$\sqrt{n}\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) = \sqrt{n}\left(\boldsymbol{I}_1 - \mathbb{E}[\boldsymbol{I}_1]\right) + \sqrt{n}O_P\left(\delta_{n0}^2(h) + h^2\right) = \sqrt{n}\cdot\left(\boldsymbol{I}_1 - \mathbb{E}[\boldsymbol{I}_1]\right) + o_P(1) \quad (26)$$

by conditions on bandwidths. Applying the central limit theorem, we have

$$\sqrt{n}\left(\boldsymbol{I}_1 - \mathbb{E}[\boldsymbol{I}_1]\right) \to_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_2),$$

by (26), which is equivalent to

$$\sqrt{n}\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) \to_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_2).$$

Finally, we show that $\boldsymbol{M}(\boldsymbol{\theta}^*)$ converges to a constant invertible matrix, hence we complete the proof by Slusky's theorem due to (16). Define $\boldsymbol{A}^{\otimes 2} = \boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{\theta}^* = \left((\boldsymbol{\beta}^*)^T, (\boldsymbol{\alpha}^*)^T\right)^T$, and let

$$l_i^*(t;\boldsymbol{\theta},\hat{\boldsymbol{X}}) = \{\boldsymbol{\psi}_i(t)\}^T\boldsymbol{\alpha}_i + \int_0^1 G_i^K(t,s) f_i(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta})\ \mathrm{d}s - \hat{X}_i(t).$$

We calculate that

$$\begin{aligned}\boldsymbol{M}(\boldsymbol{\theta}^*) =& 2\sum_{i=1}^p \int_0^1 \left(\int_0^1 G_i^K(t,s)\frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s);\boldsymbol{\beta}^*)ds, \boldsymbol{e}_i^T\otimes\{\boldsymbol{\psi}_i(t)\}^T\right)^{\otimes 2} w(t)\ \mathrm{d}t\\ &+2\sum_{i=1}^p \int_0^1 l_i^*(t;\boldsymbol{\theta}^*,\hat{\boldsymbol{X}})\frac{\partial^2 l_i^*}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(t;\boldsymbol{\theta}^*,\hat{\boldsymbol{X}})w(t)\ \mathrm{d}t.\end{aligned}$$

By the same proof as (10), we can show that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\sup_{t\in[0,1]}\left|l_i^*(t;\boldsymbol{\theta},\hat{\boldsymbol{X}}) - l_i^*(t;\boldsymbol{\theta},\boldsymbol{X})\right| &= o_P(1),\\ \sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\sup_{t\in[0,1]}\left\|\frac{\partial l_i^*}{\partial\boldsymbol{\theta}}(t;\boldsymbol{\theta},\hat{\boldsymbol{X}}) - \frac{\partial l_i^*}{\partial\boldsymbol{\theta}}(t;\boldsymbol{\theta},\boldsymbol{X})\right\| &= o_P(1),\\ \sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\sup_{t\in[0,1]}\left\|\frac{\partial^2 l_i^*}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(t;\boldsymbol{\theta},\hat{\boldsymbol{X}}) - \frac{\partial^2 l_i^*}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(t;\boldsymbol{\theta},\boldsymbol{X})\right\| &= o_P(1)\end{aligned}$$

as $n\to\infty$. This gives

$$\sup_{\boldsymbol{\theta}\in\Omega_{\boldsymbol{\theta}}}\left\|\boldsymbol{M}(\boldsymbol{\theta}) - \frac{\partial^2\mathcal{L}_n}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(\boldsymbol{\theta})\right\|_{\mathrm{op}} = o_P(1).$$

Together with the consistency of $\hat{\boldsymbol{\theta}}$, we have

$$\begin{aligned}\boldsymbol{M}(\boldsymbol{\theta}^*) =& \frac{\partial^2 \mathcal{L}_n}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(\boldsymbol{\theta}_0) + o_P(1) \\ =& 2\sum_{i=1}^{p}\int_0^1 \{\boldsymbol{d}_i(t)\}^{\otimes 2} w(t)\ \mathrm{d}t + 2\sum_{i=1}^{p}\int_0^1 l_i(t)\frac{\partial^2 l_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(t)w(t)\ \mathrm{d}t + o_P(1).\end{aligned}$$

Define

$$\boldsymbol{\Sigma}_1 := 2\sum_{i=1}^{p}\int_0^1 \{\boldsymbol{d}_i(t)\}^{\otimes 2} w(t)\ \mathrm{d}t. \tag{27}$$

By Assumption B.5 in the main text, $\boldsymbol{\Sigma}_1$ is invertible. Since $\sum_{i=1}^{p}\int_0^1 l_i(t)\frac{\partial^2 l_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(t)w(t)\ \mathrm{d}t = o_P(1)$ as $n \to \infty$, we have

$$\boldsymbol{M}(\boldsymbol{\theta}^*) \to_d \boldsymbol{\Sigma}_1, \tag{28}$$

and this gives that $\boldsymbol{M}(\boldsymbol{\theta}^*)$ is asymptotically invertible. □

Remark: In practice, we may incorporate a discrete approximation for the integration in Green's matching. Specifically, we can choose an equally spaced finite time grid $\{s_l \in [0, C]; h = 1, \cdots, L\}$ with L being sufficiently large, and then approximate the integration terms in (15), which are

$$\int_0^1 G_i^K(t,s) f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta})\ \mathrm{d}s$$

by $\frac{1}{L}\sum_{l=1}^{L} G_i^K(t,s_l) f_i(\hat{\boldsymbol{X}}(s_l); \boldsymbol{\beta})$. It can be seen that for each $i = 1, \cdots, p$,

$$\sup_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}} \sup_{t,s\in[0,C]} \left|\frac{\partial^2}{\partial s^2} G_i^K(t,s) f_i(\hat{\boldsymbol{X}}(s); \boldsymbol{\beta})\right| < \infty,$$

with the Assumptions A.4, B.3, and B.4 in the main text. Thus, **?** shows that if we take the number of grids L grows faster than $\sqrt{n}$, the error

$$\int_0^1 G_i^K(t,s) f_i(\hat{\boldsymbol{X}}(s), s; \boldsymbol{\beta})\ \mathrm{d}s - \frac{1}{L}\sum_{l=1}^{L} G_i^K(t,s_l) f_i(\hat{\boldsymbol{X}}(s_l), s_l; \boldsymbol{\beta}) = o_P(n^{-1}),$$

for each $i = 1, \cdots, p$. As a result, if L grows faster than $\sqrt{n}$ as $n \to \infty$, the asymptotic properties of Green's matching would remain the same as presented in Theorems 1 and 2 in the main text, according to Theorem 5.7 and 5.23 in Van der Vaart [2000].

2.3 Proof of Theorem 3

(a) To derive the convergence rate of Green's matching, we utilize (16):

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left\{\boldsymbol{M}(\boldsymbol{\theta}^*)\right\}^{-1} \frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0).$$

First, we derive the convergence rate of $\left\|\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)\right\|$. In (17), we decompose $\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)$ into $\boldsymbol{I}_1 + \boldsymbol{I}_2 + \boldsymbol{I}_3$, and we separately bound these three terms. Based on (20) and (25), we have

$$\|\boldsymbol{I}_1\| = O_P(\|\mathbb{E}\boldsymbol{I}_1\| + \sqrt{\mathrm{Var}\left[\boldsymbol{I}_1\right]}) = O_P(h_0^2 + n^{-\frac{1}{2}}).$$

Combing the above rate with (18) and (19), we have

$$\left\|\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)\right\| = O_P(h_0^2 + \delta_{n0}^2(h) + n^{-\frac{1}{2}}).$$

Since $\left\{\boldsymbol{M}(\boldsymbol{\theta}^*)\right\}^{-1}$ is $O_P(1)$ by (28), we get the convergence rate of Green's matching.

(b) For the gradient matching with order K, we define

$$\mathcal{L}_n^{\mathrm{Grad}}(\boldsymbol{\beta}) = \sum_{i=1}^{p} \int_0^1 \left\{\hat{X}_i^{(K)}(t) + \sum_{k=0}^{K-1} \omega_{ik} \hat{X}_i^{(k)}(t) - f_i(\hat{\boldsymbol{X}}(t); \boldsymbol{\beta})\right\}^2 w(t)\ \mathrm{d}t,$$

and $\hat{\boldsymbol{\beta}}_K^{\mathrm{Grad}} = \arg\min_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}}} \mathcal{L}_n^{Grad}(\boldsymbol{\beta})$. For ease of notation, we omit the superscript "Grad" in the following. The strategy for proving (b) when $K_0 = K$ is the same as Green's matching in (a), and the key step is to calculate the convergence order of $\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}_0)$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$. Similar to (a), we will decompose $\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}_0)$ into $\boldsymbol{I}_1 + \boldsymbol{I}_2 + \boldsymbol{I}_3$ below and separately bound these terms.

Define

$$\begin{aligned}
l_i(t) &= \hat{X}_i^{(K)}(t) + \sum_{k=0}^{K-1} \omega_{ik} \hat{X}_i^{(k)}(t) - f_i(\hat{\boldsymbol{X}}(t); \boldsymbol{\beta}_0) \\
&= \hat{X}_i^{(K)}(t) - X^{(K)}(t) + \sum_{k=0}^{K-1} \omega_{ik} \hat{X}_i^{(k)}(t) - \sum_{k=0}^{K-1} \omega_{ik} X_i^{(k)}(t) \\
&\quad + f_i(\hat{\boldsymbol{X}}(t); \boldsymbol{\beta}_0) - f_i(\boldsymbol{X}(t); \boldsymbol{\beta}_0) \\
&= \hat{X}_i^{(K)}(t) - X^{(K)}(t) + \sum_{k=0}^{K-1} \omega_{ik} \left(\hat{X}_i^{(k)}(t) - X_i^{(k)}(t)\right) \\
&\quad + \frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(t); \boldsymbol{\beta}_0) \left(\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\right) \\
&\quad + f_i(\hat{\boldsymbol{X}}(t); \boldsymbol{\beta}_0) - f_i(\boldsymbol{X}(t); \boldsymbol{\beta}_0) - \frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(t); \boldsymbol{\beta}_0) \left(\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\right),
\end{aligned}$$

and let $\frac{\partial f_i(s)}{\partial \boldsymbol{\beta}} = \frac{\partial f_i}{\partial \boldsymbol{\beta}}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)$, $\frac{\partial f_i(s)}{\partial \boldsymbol{X}} = \frac{\partial f_i}{\partial \boldsymbol{X}}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)$, etc.. We can decompose $\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}_0)$ as follows

$$
\begin{aligned}
&\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}_0) \\
=& -2\sum_{i=1}^{p}\int_0^1 l_i(t)\frac{\partial f_i}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}_0)w(t)\ \mathrm{d}t \\
=& -2\sum_{i=1}^{p}\int_0^1 \Bigg(\hat{X}_i^{(K)}(t) - X^{(K)}(t) + \sum_{k=0}^{K-1}\omega_{ik}\left(\hat{X}_i^{(k)}(t) - X_i^{(k)}(t)\right) \\
&\qquad\qquad + \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\right)\Bigg)\frac{\partial f_i}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}_0)w(t)\ \mathrm{d}t \\
& -2\sum_{i=1}^{p}\int_0^1 \Bigg[f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}_0) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta}_0) \\
&\qquad\qquad - \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\right)\Bigg]\frac{\partial f_i}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}_0)w(t)\ \mathrm{d}t \\
=& -2\sum_{i=1}^{p}\int_0^1 \Bigg(\hat{X}_i^{(K)}(t) - X^{(K)}(t) + \sum_{k=0}^{K-1}\omega_{ik}\left(\hat{X}_i^{(k)}(t) - X_i^{(k)}(t)\right) \\
&\qquad\qquad + \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\right)\Bigg)\frac{\partial f_i(t)}{\partial \boldsymbol{\beta}}w(t)\ \mathrm{d}t \\
& -2\sum_{i=1}^{p}\int_0^1 \Bigg(\hat{X}_i^{(K)}(t) - X^{(K)}(t) + \sum_{k=0}^{K-1}\omega_{ik}\left(\hat{X}_i^{(k)}(t) - X_i^{(k)}(t)\right) \\
&\qquad\qquad + \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\right)\Bigg)\left(\frac{\partial f_i}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}_0) - \frac{\partial f_i}{\partial \boldsymbol{\beta}}(t)\right)w(t)\ \mathrm{d}t \\
& -2\sum_{i=1}^{p}\int_0^1 \Bigg[f_i(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}_0) - f_i(\boldsymbol{X}(t);\boldsymbol{\beta}_0) \\
&\qquad\qquad - \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\right)\Bigg]\frac{\partial f_i}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{X}}(t);\boldsymbol{\beta}_0)w(t)\ \mathrm{d}t \\
=:& \boldsymbol{I}_1 + \boldsymbol{I}_2 + \boldsymbol{I}_3.
\end{aligned}
$$

Using (7) and the same proof of Lemma 4, we see that

$$\|\boldsymbol{I}_2\| = O_P\left(\sum_{k=0}^{K}\delta_{nk}(h_k)\delta_{n0}(h_0)\right), \tag{29}$$

$$\|\boldsymbol{I}_3\| = O_P\left(\delta_{n0}^2(h_0)\right), \tag{30}$$

and by Theorem 3.1 in Fan and Gijbels [2018]

$$\mathbb{E}[\boldsymbol{I}_1] = O\left(\sum_{k=0}^{K}h_k^2\right). \tag{31}$$

The rest of the calculation is similar to the proof of Theorem 2. To calculate the variance of $\boldsymbol{I}_1$, we need the expression of the equivalence kernel for the local polynomial regression like (21):

$$\hat{X}_i^{(k)}(t) = \frac{1}{nh^{1+k} f_T(t)} \sum_{j=1}^{n} W_k^{*,n}(\frac{t_j - t}{h_k}) Y_i(t_j), \tag{32}$$

where $W_k^{*,n}(\cdot)$ is the equivalent kernel satisfies

$$\int u^q W_k^{*,n}(u) \mathrm{d}u = \delta_{q,k}.$$

See Section 3.2.2 in Chapter 3 in Fan and Gijbels [2018] for more details. With this expression, we can show that

$$\sqrt{n}(\boldsymbol{I}_1 - \mathbb{E}[\boldsymbol{I}_1]) \to_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{2,K}),$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_{2,K} = 4 \sum_{i=1}^{p} \sum_{k,h=0}^{K} \omega_{ik}\omega_{ih} \int_0^1 \frac{\partial^k}{\partial t^k} \left(\frac{w(t)}{f_T(t)} \frac{\partial f_i}{\partial \boldsymbol{\beta}} \right) \cdot \frac{\partial^h}{\partial t^h} \left(\frac{w(t)}{f_T(t)} \frac{\partial f_i}{\partial \boldsymbol{\beta}^T} \right) \cdot f_T(t) \mathrm{d}t \\
+ 4 \sum_{i,u=1}^{p} \sum_{k=0}^{K} \omega_{ik} \int_0^1 \frac{\partial^k}{\partial t^k} \left(\frac{w(t)}{f_T(t)} \frac{\partial f_i}{\partial \boldsymbol{\beta}} \right) \cdot w(t) \frac{\partial f_u}{\partial X_i} \frac{\partial f_u}{\partial \boldsymbol{\beta}^T} \mathrm{d}t \\
+ 4 \sum_{i,u,l=1}^{p} \int_0^1 \frac{w^2(t)}{f_T(t)} \frac{\partial f_u}{\partial X_i} \frac{\partial f_l}{\partial X_i} \frac{\partial f_u}{\partial \boldsymbol{\beta}} \frac{\partial f_l}{\partial \boldsymbol{\beta}^T} \mathrm{d}t.
\end{aligned}$$

Here we denote $\omega_{iK} = 1$ for simplicity.

Additionally, we can show that

$$\frac{\partial^2 \mathcal{L}_n}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \to_d 2 \sum_{i=1}^{p} \int_0^1 \frac{\partial f_i}{\partial \boldsymbol{\beta}} \frac{\partial f_i}{\partial \boldsymbol{\beta}^T} w(t) \mathrm{d}t := \boldsymbol{\Sigma}_{1,K}.$$

With this, we define

$$\boldsymbol{Z}_{n,K} = \boldsymbol{\Sigma}_{1,K}^{-1} \cdot \sqrt{n}(\boldsymbol{I}_1 - \mathbb{E}[\boldsymbol{I}_1]) \text{ and } \boldsymbol{W}_{n,K} = \sqrt{n}(\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_0) - \boldsymbol{Z}_{n,K}.$$

It follows that

$$\boldsymbol{Z}_{n,K} \to_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{1,K}^{-1} \boldsymbol{\Sigma}_{2,K} \boldsymbol{\Sigma}_{1,K}^{-1}), \tag{33}$$

and

$$\boldsymbol{W}_{n,K} = \sqrt{n} \cdot O_P(\mathbb{E}[\boldsymbol{I}_1] + \boldsymbol{I}_2 + \boldsymbol{I}_3) = O_P \left(\sum_{k=0}^{K} (n^{\frac{1}{2}} h_k^2 + n^{-\frac{1}{2}} h_0^{-\frac{1}{2}} h_k^{-k-\frac{1}{2}}) \right). \tag{34}$$

Combining (29)–(31) and (33)–(34), we have the the bias of order K gradient matching is of order

$$O_P\left(\sum_{k=0}^{K}(h_k^2+\delta_{nk}(h_k)\delta_{n0}(h_0))\right)=O_P\left(\sum_{k=0}^{K}(h_k^2+n^{-1}h_0^{-\frac{1}{2}}h_k^{-k-\frac{1}{2}})\right).$$

Moreover, for proving (b) when $K_0=K-1$, we abuse the notation $\mathcal{L}_n^{\text{Grad}}(\boldsymbol{\theta})$ to denote

$$\mathcal{L}_n^{\text{Grad}}(\boldsymbol{\theta})=\sum_{i=1}^{p}\int_0^1\left[\hat{X}_i^{(K-1)}(t)-\alpha_i-\int_0^t\left(f_i(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta})-\sum_{k=0}^{K-1}\omega_{ik}\hat{X}_i^{(k)}(s)\right)\,\mathrm{d}s\right]^2 w(t)\,\mathrm{d}t,$$

where $(\boldsymbol{\beta}^T,\boldsymbol{\alpha}^T)^T$ and $(\boldsymbol{\beta}_0^T,\boldsymbol{\alpha}_0^T)^T$ are denoted as $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, respectively. As such,

$$\hat{\boldsymbol{\theta}}_{K-1}^{\text{Grad}}=\arg\min_{\boldsymbol{\beta}\in\Omega_{\boldsymbol{\beta}},\boldsymbol{\alpha}\in\Omega_\alpha}\mathcal{L}_n^{\text{Grad}}(\boldsymbol{\theta}).$$

Again, we omit the superscript "Grad" below. Define

$$\begin{aligned}
l_i(t)=&\hat{X}_i^{(K-1)}(t)-\alpha_{0,i}-\int_0^t\left(f_i(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)-\sum_{k=0}^{K-1}\omega_{ik}\hat{X}_i^{(k)}(s)\right)\,\mathrm{d}s\\
=&\hat{X}_i^{(K-1)}(t)-X_i^{(K-1)}(t)+\sum_{k=0}^{K-1}\omega_{ik}\int_0^t\left(\hat{X}_i^{(k)}(s)-X_i^{(k)}(s)\right)\,\mathrm{d}s\\
&-\int_0^t\left(f_i(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)-f_i(\boldsymbol{X}(s),\boldsymbol{\beta}_0)\right)\,\mathrm{d}s\\
=&\hat{X}_i^{(K-1)}(t)-X_i^{(K-1)}(t)+\sum_{k=0}^{K-1}\omega_{ik}\int_0^t\left(\hat{X}_i^{(k)}(s)-X_i^{(k)}(s)\right)\,\mathrm{d}s\\
&-\int_0^t\frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(t);\boldsymbol{\beta}_0)\left(\hat{\boldsymbol{X}}(t)-\boldsymbol{X}(t)\right)\,\mathrm{d}s\\
&-\int_0^t\left(f_i(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)-f_i(\boldsymbol{X}(s),\boldsymbol{\beta}_0)-\frac{\partial f_i}{\partial \boldsymbol{X}^T}(\boldsymbol{X}(t);\boldsymbol{\beta}_0)\left(\hat{\boldsymbol{X}}(t)-\boldsymbol{X}(t)\right)\right)\,\mathrm{d}s,
\end{aligned}$$

and let $\frac{\partial f_i(s)}{\partial\boldsymbol{\beta}}=\frac{\partial f_i}{\partial\boldsymbol{\beta}}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)$, $\frac{\partial f_i(s)}{\partial\boldsymbol{X}}=\frac{\partial f_i}{\partial\boldsymbol{X}}(\boldsymbol{X}(s);\boldsymbol{\beta}_0)$, etc.

Recall

$$\frac{\partial\mathcal{L}_n}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_0)=2\sum_{i=1}^{p}\int_0^1 l_i(t)\left(\int_0^t\frac{\partial f_i}{\partial\boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)\,\mathrm{d}s,\ -\boldsymbol{e}_i^T\right)^T w(t)\,\mathrm{d}t,$$

we thus have

$$
\begin{aligned}
&\frac{\partial \mathcal{L}_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)\\
=&2\sum_{i=1}^{p}\int_0^1 \Bigg(\hat{X}_i^{(K-1)}(t)-X_i^{(K-1)}(t)+\sum_{k=0}^{K-1}\omega_{ik}\int_0^t\left(\hat{X}_i^{(k)}(s)-X_i^{(k)}(s)\right)\,\mathrm{d}s\\
&\quad -\int_0^t \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t)-\boldsymbol{X}(t)\right)\,\mathrm{d}s\Bigg)\cdot\left(\int_0^t \frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta})\,\mathrm{d}s,\ -\boldsymbol{e}_i^T\right)^T w(t)\,\mathrm{d}t\\
&+2\sum_{i=1}^{p}\int_0^1 \Bigg(\int_0^t (f_i(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)-f_i(\boldsymbol{X}(s),\boldsymbol{\beta}_0)\\
&\quad -\frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t)-\boldsymbol{X}(t)\right))\,\mathrm{d}s\Bigg)\cdot\left(\int_0^t \frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta})\,\mathrm{d}s,\ -\boldsymbol{e}_i^T\right)^T w(t)\,\mathrm{d}t\\
=&2\sum_{i=1}^{p}\int_0^1 \Bigg(\hat{X}_i^{(K-1)}(t)-X_i^{(K-1)}(t)+\sum_{k=0}^{K-1}\omega_{ik}\int_0^t\left(\hat{X}_i^{(k)}(s)-X_i^{(k)}(s)\right)\,\mathrm{d}s\\
&\quad -\int_0^t \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t)-\boldsymbol{X}(t)\right)\,\mathrm{d}s\Bigg)\cdot\left(\int_0^t \frac{\partial f_i(s)}{\partial \boldsymbol{\beta}^T}\,\mathrm{d}s,\ -\mathbf{0}^T\right)^T w(t)\,\mathrm{d}t\\
&+2\sum_{i=1}^{p}\int_0^1 \Bigg(\hat{X}_i^{(K-1)}(t)-X_i^{(K-1)}(t)+\sum_{k=0}^{K-1}\omega_{ik}\int_0^t\left(\hat{X}_i^{(k)}(s)-X_i^{(k)}(s)\right)\,\mathrm{d}s\\
&\quad -\int_0^t \frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t)-\boldsymbol{X}(t)\right)\,\mathrm{d}s\Bigg)\\
&\quad \cdot\left(\int_0^t \frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)-\frac{\partial f_i(s)}{\partial \boldsymbol{\beta}^T}\,\mathrm{d}s,\ -\mathbf{0}^T\right)^T w(t)\,\mathrm{d}t\\
&+2\sum_{i=1}^{p}\int_0^1 \Bigg(\int_0^t (f_i(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)-f_i(\boldsymbol{X}(s),\boldsymbol{\beta}_0)\\
&\quad -\frac{\partial f_i(t)}{\partial \boldsymbol{X}^T}\left(\hat{\boldsymbol{X}}(t)-\boldsymbol{X}(t)\right))\,\mathrm{d}s\Bigg)\cdot\left(\int_0^t \frac{\partial f_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{X}}(s),\boldsymbol{\beta}_0)\,\mathrm{d}s,\ -\boldsymbol{e}_i^T\right)^T w(t)\,\mathrm{d}t\\
=:&\,\boldsymbol{I}_1+\boldsymbol{I}_2+\boldsymbol{I}_3.
\end{aligned}
$$

By the same argument as in Lemma 4, we get

$$
\|\boldsymbol{I}_1\| = O_P\left(h^2+n^{-\frac{1}{2}}\right),\ \mathbb{E}[\boldsymbol{I}_1]=O_P\left(\sum_{k=0}^{K-1}h_k^2+n^{-\frac{1}{2}}\right),
$$

$$
\|\boldsymbol{I}_2\| = O_P\left(\sum_{k=0}^{K-1}\delta_{nk}(h_k)\delta_{n0}(h_0)\right),
$$

$$
\|\boldsymbol{I}_3\| = O_P\left(\delta_{n0}^2(h_0)\right).
$$

The rest of the calculation is similar to the proof of Theorem 2 and we can prove the asymptotic normality of $(\boldsymbol{I}_1-\mathbb{E}[\boldsymbol{I}_1])$ and define $\boldsymbol{Z}_{n,K-1}$ and $\boldsymbol{W}_{n,K-1}$ as the case when

$K_0 = K$ above. Combining these bias of $\boldsymbol{I}_1$, $\boldsymbol{I}_2$ and $\boldsymbol{I}_3$, we obtain the bias of order $(K-1)$ gradient matching is of order

$$O_P\left(\sum_{k=0}^{K-1}(h_k^2 + \delta_{nk}(h_k)\delta_{n0}(h_0))\right) = O_P\left(\sum_{k=0}^{K-1}(h_k^2 + n^{-1}h_0^{-\frac{1}{2}}h_k^{-k-\frac{1}{2}})\right),$$

and therefore, we finish the proof.

3 Additional supporting results

3.1 Illustration of test models

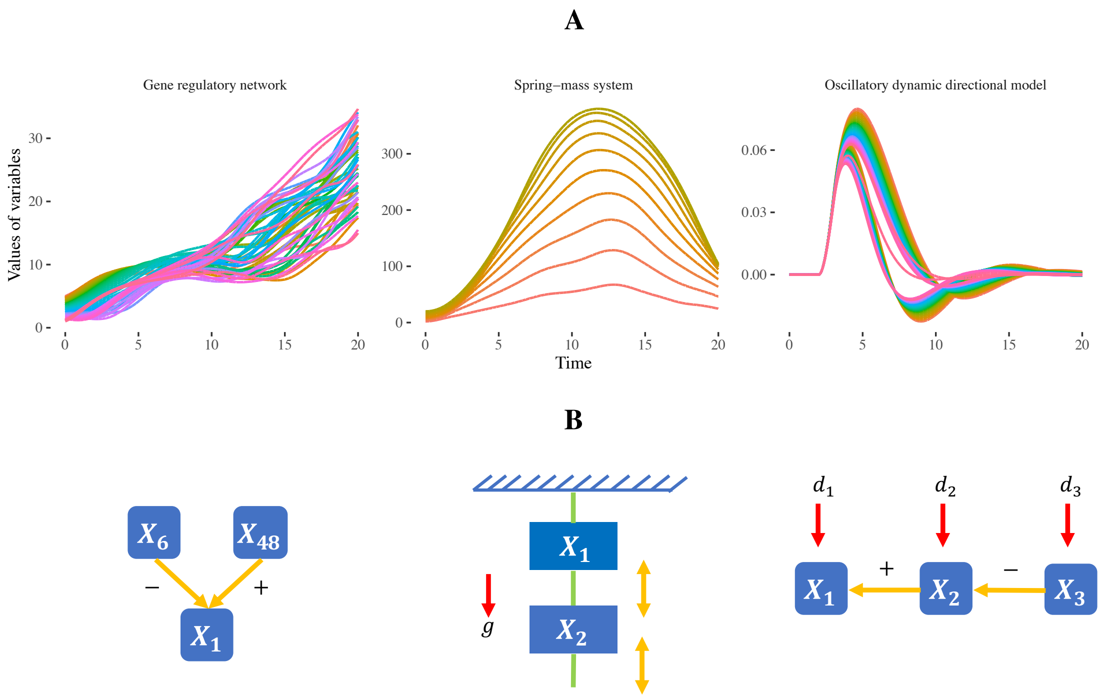

Figure 1: Dynamic curves of three test models (**A**) and their corresponding states' interaction diagrams (**B**), where the yellow arrows in panel **B** represent the directed interactions, with the symbols "+" and "−" to respectively denote the positive and negative interaction. Besides, we use red arrows to denote the introducing force for the spring-mass system (gravity) and the oscillatory dynamic directional model (activation signal).

3.2 Green's matching with a different weight function

In this subsection, we implement Green's matching with pre-smoothing using a different weight function. The results are given in Table 1, which shows that the RRMSEs of Green's matching are nearly the same as those in the results of Table 1 in the main text. This indicates that Green's matching is not sensitive to the choice of weight functions.

Table 1: RRMSEs of Green's matching with the weight function $W^2(t)$.

RRMSE(%)	$\gamma = 3\%$	$\gamma = 5\%$	$\gamma = 7\%$
Gene regulatory network			
$n = 50$	9.61	15.07	20.22
$n = 150$	5.88	9.29	12.48
$n = 250$	4.73	7.46	10.06
Spring-mass system			
$n = 50$	13.30	20.98	29.49
$n = 150$	6.13	11.90	16.61
$n = 250$	4.51	8.58	12.94
Oscillatory dynamic directional model			
$n = 50$	9.70	16.07	22.60
$n = 150$	5.46	9.05	12.89
$n = 250$	4.32	7.05	9.97

3.3 Reconstructed curves

In this subsection, we describe the procedures to reconstruct curves. For the reconstruction of curves, we first obtain an estimation of $\boldsymbol{\beta}$, denoted as $\hat{\boldsymbol{\beta}}$, and the pre-smoothed initial conditions, which are $\{\widehat{D^k X_i}(0); i = 1, \cdots, p, k = 0, \cdots, K-1\}$. After that, we employ the Runge–Kutta method [Soetaert et al., 2010] to simulate the curves $D^k X_i(\cdot)$ given the pre-smoothed initial conditions and $\hat{\boldsymbol{\beta}}$.

For the cases of $n = 250$ and $\gamma = 7\%$, the reconstructed trajectories $X_1(\cdot)$, $DX_1(\cdot)$, and $D^2X_1(\cdot)$ in the oscillatory dynamic directional model have been given in Fig. 3 **A** in the main text. Here, we also present the reconstructed results of $n = 250$ and $\gamma = 3\%$ or 5% in Fig. 2 or 3, respectively. By these results, we can see that the order 2 gradient matching falls short of fitting the true curves, especially for the discontinuous points of $D^2X_1(\cdot)$.

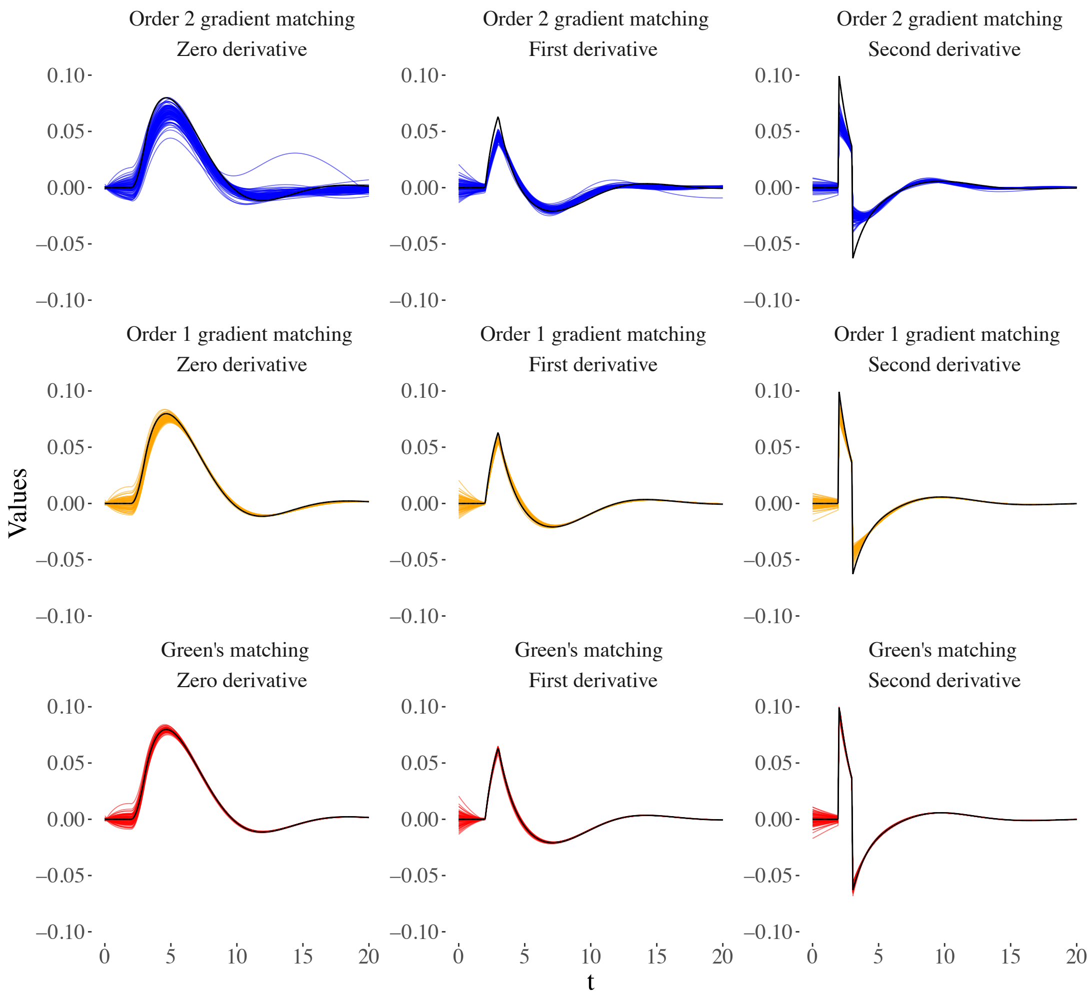

Figure 2: Reconstructed curves for $X_1(\cdot)$, $DX_1(\cdot)$, and $D^2X_1(\cdot)$ (in the oscillatory dynamic directional model with $n = 250$ and $\gamma = 3\%$) by three methods from 100 simulations, where the solid black lines indicate the corresponding true curves.

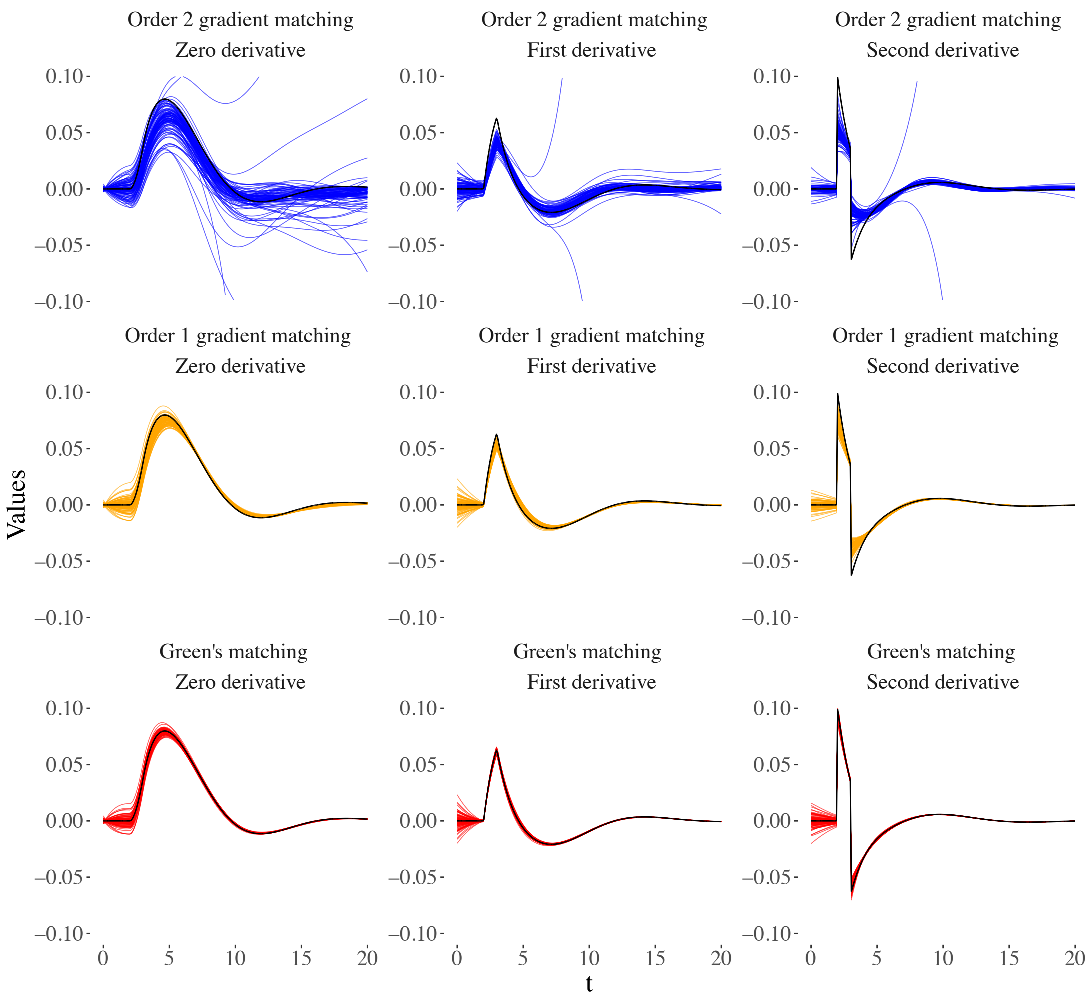

Figure 3: Reconstructing curves for $X_1(\cdot)$, $DX_1(\cdot)$, and $D^2X_1(\cdot)$ (in the oscillatory dynamic directional model with $n = 250$ and $\gamma = 5\%$) by three methods from 100 simulations, where the solid black lines indicate the corresponding true curves.

3.4 Confidence intervals

In this subsection, we calculate the confidence intervals of $\boldsymbol{\beta}$ by a bootstrap procedure for a two-step estimator $\hat{\boldsymbol{\beta}}$. In detail, we assume that the time points $\{t_j; j = 1, \cdots, n\}$ are fixed. We first obtain the pre-smoothed data $\{\hat{X}_i(t_j); j = 1, \cdots, n\}$ by the local polynomial fitting. After that, we estimate the variance of Gaussian noises σ_i^2 by the consistent estimator

$$\hat{\sigma}_i^2 := \frac{\sum_{j=1}^n \left\{Y_i(t_j) - \hat{X}_i(t_j)\right\}^2}{n - \sum_{j=1}^n w_{jj} + \sum_{j=1}^n \sum_{j'}^n w_{jj'}^2},$$

where $w_{jj'}$ is taken as $W_{n,0}(\frac{t_j - t_{j'}}{h_0})$ as defined in Part 1.1; see Hall and Marron [1990] for the derivation details. Accordingly, we regenerate the observed data $\tilde{Y}_i(t_j)$ as $\hat{X}_i(t_j)$ by adding a mean-zero Gaussian noise with variance $\hat{\sigma}_i^2$, and obtain a bootstrap sample of $\hat{\boldsymbol{\beta}}$ based on the parameter estimation on the newly regenerated data.

Since the two-step estimator $\hat{\boldsymbol{\beta}}$ generally possesses estimation biases, we apply the bias-corrected percentiles [Diciccio and Romano, 1988] to the bootstrap samples for constructing the confidence intervals of $\boldsymbol{\beta}$. Let z_a and $\Phi(\cdot)$ be the a percentile and the distribution function of a standard normal random variable, and $\hat{F}(\cdot)$ be the empirical distribution of the bootstrap samples of $\hat{\boldsymbol{\beta}}$. When a transformation of $\hat{\boldsymbol{\beta}}$ is asymptotic normal distributed, we can use the $\Phi(z_{a/2} + 2\Phi^{-1}(\hat{F}(\hat{\boldsymbol{\beta}})))$ and $\Phi(z_{1-a/2} + 2\Phi^{-1}(\hat{F}(\hat{\boldsymbol{\beta}})))$ percentiles of the bootstrap samples for $\hat{\boldsymbol{\beta}}$ to construct the $(1-a)\%$ confidence intervals. Specially, when $\Phi^{-1}(\hat{F}(\hat{\boldsymbol{\beta}})) = 0$, i.e., $\hat{\boldsymbol{\beta}}$ is the median of $\hat{F}(\cdot)$, the bias-corrected percentiles degenerate to the usual percentiles.

Table 2: Proportions of the 95% confidence intervals that cover the true values of parameters in the oscillatory dynamic directional model ($n = 250$).

Proportions (%)		a_1	b_1	c_1	d_1
$\gamma = 3\%$	Order 2 gradient matching	7	7	14	0
	Order 1 gradient matching	0	0	0	0
	Green's matching	94	95	95	92
$\gamma = 5\%$	Order 2 gradient matching	21	21	31	3
	Order 1 gradient matching	0	0	0	0
	Green's matching	91	92	96	90
$\gamma = 7\%$	Order 2 gradient matching	39	42	47	7
	Order 1 gradient matching	0	0	1	0
	Green's matching	94	92	96	88

In Theorem 2 in the main text, we have proven that the estimator of Green's matching is asymptotic normal distributed. Hence we can use the above bias-corrected percentiles of the bootstrap samples to construct the confidence intervals for $\boldsymbol{\beta}$. Besides, we adopt similar strategies to obtain the confidence intervals of $\boldsymbol{\beta}$ by gradient matching with different orders. The corresponding results for the parameters in oscillatory dynamic directional models are given in Fig. 4, Fig. 5, and Fig. 3 **B** in the main text, for the cases of $n = 250$ and $\gamma = 3\%$, 5%, or 7%, respectively. Based on these results, we find that the confidence intervals

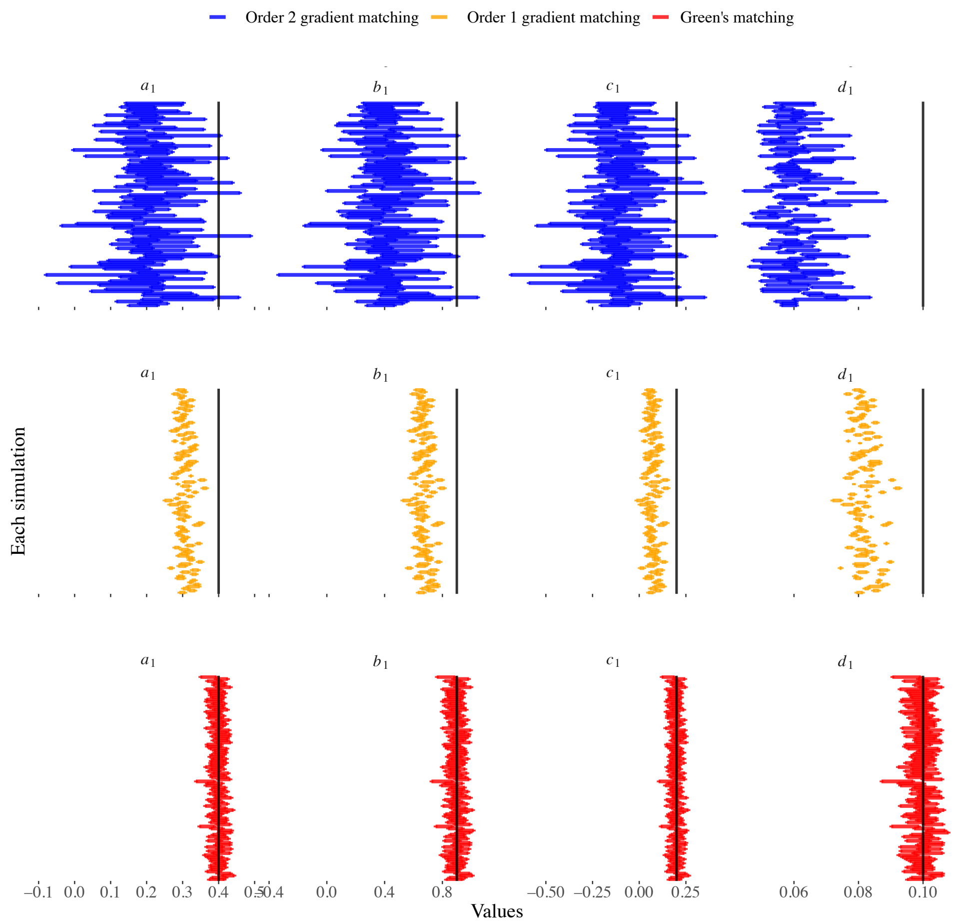

Figure 4: Confidence intervals of the parameters a_1, b_1, c_1, and d_1 (in the oscillatory dynamic directional model with $n = 250$ and $\gamma = 3\%$) by three methods from 100 simulations, where the solid black lines represent the true values of parameters.

of the gradient matching approaches are significantly biased with incorrect covering proportions, while nearly 95% of the 95% confidence intervals constructed by Green's matching cover the true values of the model parameters; see Table 2 for the corresponding covering proportions.

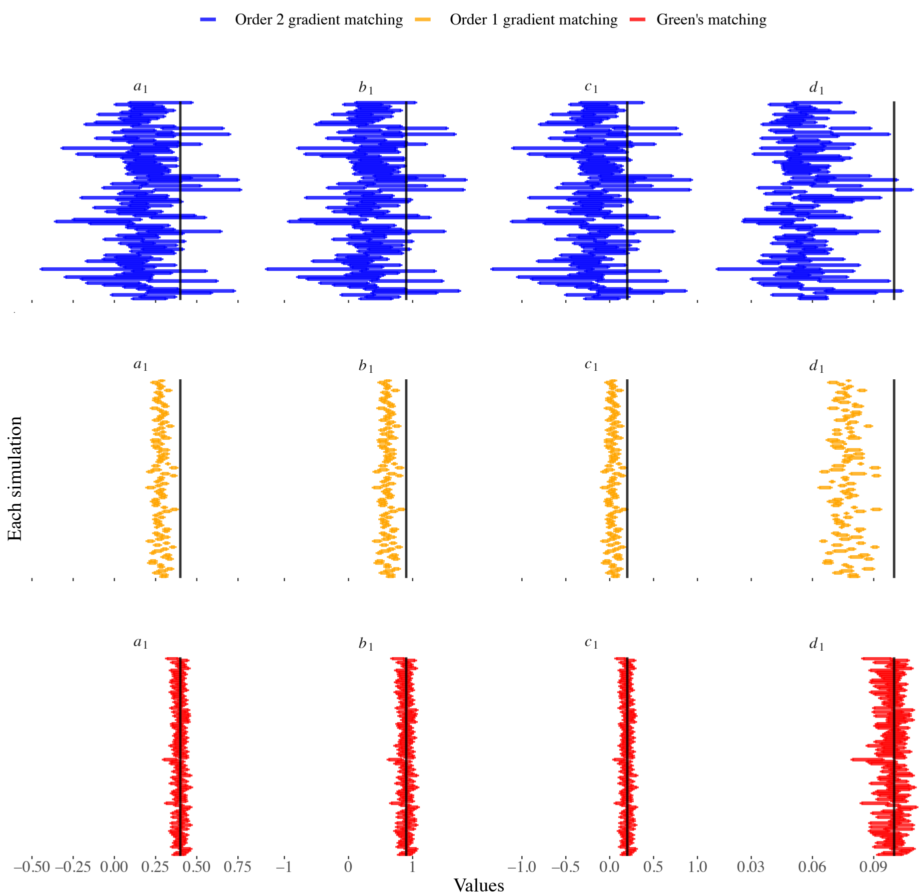

Figure 5: Confidence intervals of the parameters a_1, b_1, c_1, and d_1 (in the oscillatory dynamic directional model with $n = 250$ and $\gamma = 5\%$) by three methods from 100 simulations, where the solid black lines represent the true values of parameters.

4 Data illustration

We utilize the Chinese handwriting data from **?** to further illustrate our methods. This dataset contains replications of a Chinese script, specifically the three characters for "statistics" in Chinese. These characters are recorded by cameras mounted on the pen at approximately 400 times per second, with each replication taking roughly 6 seconds to produce. The production of the Chinese script involves moving the pen onto the writing surface, and external forces are applied to outline the Chinese characters.

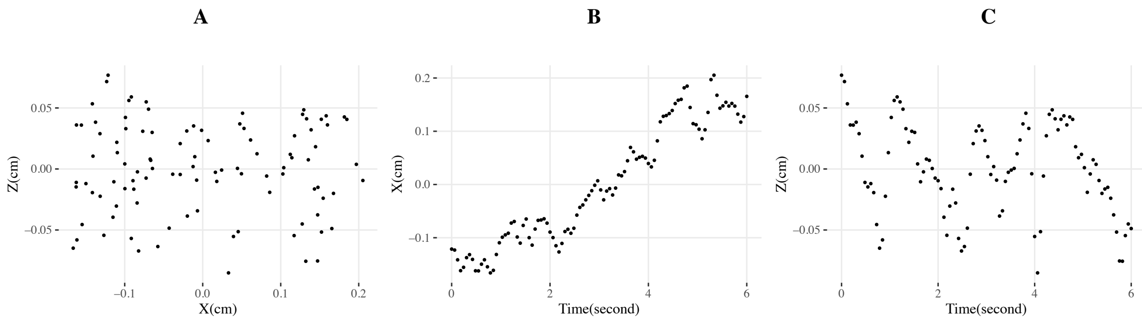

Figure 6: Scatter plot of the selected Chinese script on a writing surface (**A**), and the dynamic curves of the corresponding recorded locations on horizontal (**B**) and vertical (**C**) directions, respectively.

We select one of the replications of the Chinese script as our fitting data, presented in Fig. 6. In this data, X and Z represent the horizontal and vertical directions of the writing surface, respectively. The corresponding observed trajectories in the horizontal and vertical directions are shown in Fig. 6 **B** and **C**, respectively.

To characterize the writing dynamics of the Chinese script, **?** employed the harmonic equations in Section 4.2 in the main text with $Q = 46$. In this section, we apply three two-step methods and GSA to fit the data using this model. To evaluate the estimation behaviors of the different methods, we use the data at the time points $\{t_j; j = 1, \cdots, n\}$, uniformly spaced in $[0, 6]$ with $n = 100$, as fitting data. We then view the data outside the fitting time points as a test set and compare the reconstructed trajectories (see Section 3.3 for more detials) based on the estimated parameters from the four methods. The reconstruction results are presented in Figure 7. It is observed that the parameters estimated by gradient matching and GSA fail to accurately capture the data patterns of the writing dynamics. On the other hand, Green's matching offers a more accurate reconstruction of the Chinese script compared to the other three methods. These results suggest that the parameters estimated by Green's matching are more suitable in capturing the evolution pattern from the limited data samples.

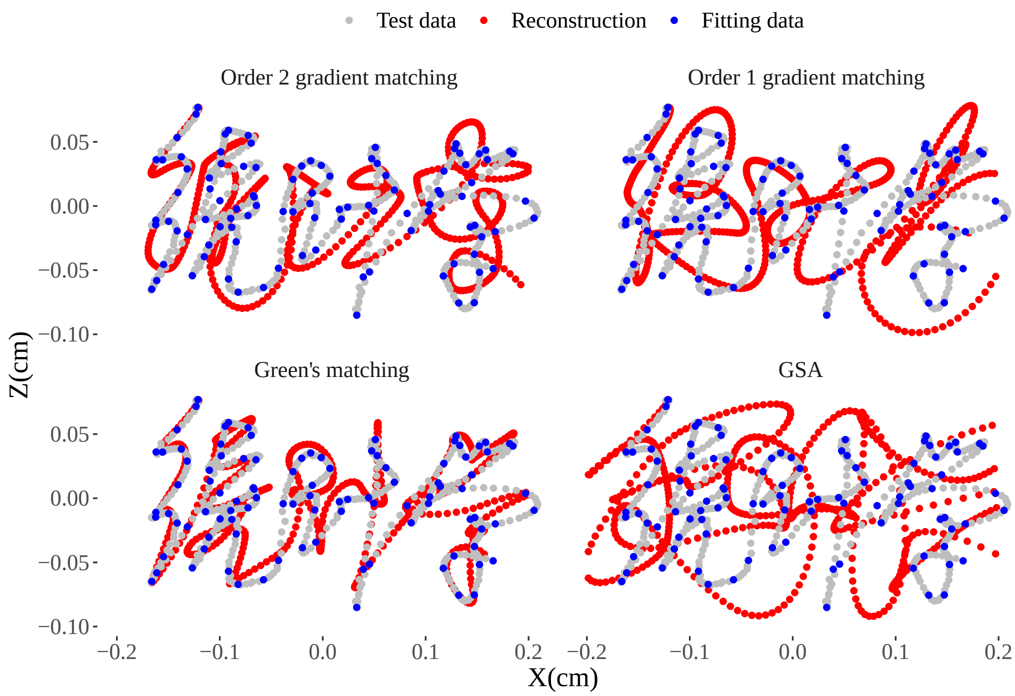

Figure 7: Scatter plots of $Z(t)$ plotted against $X(t)$, where the time grids t are selected as the 600 time points that are uniformly spaced in $[0, 6]$.

References

P. J. Davis and P. Rabinowitz. *Methods of numerical integration*. Courier Corporation, 2007.

T. J. Diciccio and J. P. Romano. A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):338–354, 1988.

D. G. Duffy. *Green's functions with applications*. Chapman and Hall/CRC, 2015.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*. Routledge, 2018.

V. Goulet, C. Dutang, M. Maechler, D. Firth, M. Shapira, M. Stadelmann, et al. expm: Matrix exponential. *R package version 0.99-0*, 2013.

P. Hall and J. Marron. On variance estimation in nonparametric regression. *Biometrika*, 77 (2):415–419, 1990.

J. Ramsay and G. Hooker. *Dynamic data analysis*. Springer, 2017.

K. Soetaert, T. Petzoldt, and R. W. Setzer. Solving differential equations in R: package desolve. *Journal of statistical software*, 33:1–25, 2010.

C. J. Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, 10:1040–1053, 1982.

A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.