Supplemental Materials for "Generalized Ordinary Differential Equation Models"

Hongyu Miao, Hulin Wu and Hongqi Xue

A Proof of Theorem

Theorem 1. Assume that there exists a $\lambda > 0$ such that $\delta = O(N^{-\lambda})$, then under Assumptions A1-A8, we have $\hat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}_0$.

Proof: Denote

$$
\begin{aligned}
&\tilde{M}_n(\boldsymbol{\alpha}) \\
= \ & \frac{1}{N} L(\boldsymbol{\alpha}; \boldsymbol{y}, \tilde{\boldsymbol{x}}, \boldsymbol{z}) = \frac{1}{N} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \ell(\boldsymbol{\alpha}; y_{iks}, \tilde{\boldsymbol{x}}_i, \boldsymbol{z}_{ik}) \\
= \ & \frac{1}{N} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \left[\frac{y_{iks} \cdot b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}_{ik}, \tilde{\boldsymbol{x}}(t_i, \boldsymbol{\theta}), \boldsymbol{\beta}) - b\{b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}_{ik}, \tilde{\boldsymbol{x}}(t_i, \boldsymbol{\theta}), \boldsymbol{\beta})\}}{a(\phi)} + c(y_{iks}, \phi) \right],
\end{aligned}
$$

$$
\begin{aligned}
&M_n(\boldsymbol{\alpha}) \\
= \ & \frac{1}{N} L(\boldsymbol{\alpha}; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) = \frac{1}{N} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \ell(\boldsymbol{\alpha}; y_{iks}, \boldsymbol{x}_i, \boldsymbol{z}_{ik}) \\
= \ & \frac{1}{N} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \left[\frac{y_{iks} \cdot b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}(t_i, \boldsymbol{\theta}), \boldsymbol{\beta}) - b\{b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}(t_i, \boldsymbol{\theta}), \boldsymbol{\beta})\}}{a(\phi)} + c(y_{iks}, \phi) \right],
\end{aligned}
$$

and

$$
M(\boldsymbol{\alpha}) = \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z}) = \frac{y \cdot b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}, \boldsymbol{x}(t, \boldsymbol{\theta}), \boldsymbol{\beta}) - b\{b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}, \boldsymbol{x}(t, \boldsymbol{\theta}), \boldsymbol{\beta})\}}{a(\phi)} + c(y, \phi).
$$

First, we claim that $E_0[M(\boldsymbol{\alpha})]$ reaches its unique maximum at $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$. In fact, by the Jensen's inequality and the convexity of function $-\log \boldsymbol{x}$, we have

$$
E_0[M(\boldsymbol{\alpha})] - E_0[M(\boldsymbol{\alpha}_0)] = E_0[\ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z}) - \ell(\boldsymbol{\alpha}_0; y, \boldsymbol{x}, \boldsymbol{z})] \leq 0,
$$

i.e., $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ is one maximum point of function $E_0[M(\boldsymbol{\alpha})]$. Moreover, from Assumption A6, it follows that it is the unique maximum point. Thus the above claim holds.

Next, with the Taylor expansion, for any $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2 \in \Theta \times \mathcal{B} \times \Phi$, we can easily obtain

$$
|M_n(\boldsymbol{\alpha}_1) - M_n(\boldsymbol{\alpha}_2)| \leq C(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|, \tag{A.1}
$$

where $C(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ is bounded. By the weak law of large numbers, it follows that for every fixed $\boldsymbol{\alpha}$, $M_n(\boldsymbol{\alpha}) - E_0[M(\boldsymbol{\alpha})] \xrightarrow{p} 0$. Then under Assumption A1, by Corollary 2 in Andrews (1987), we have that $\sup_{\boldsymbol{\alpha}} |M_n(\boldsymbol{\alpha}) - E_0[M(\boldsymbol{\alpha})]| \xrightarrow{p} 0$. Thus it follows that $M_n(\hat{\boldsymbol{\alpha}}) - E_0[M(\hat{\alpha})] \xrightarrow{p} 0$ and $M_n(\boldsymbol{\alpha}_0) - E_0[M(\boldsymbol{\alpha}_0)] \xrightarrow{p} 0$.

Based on the formulation of ℓ in Section 3.1, for a variate x, $\frac{\partial}{\partial x}\ell(\boldsymbol{\alpha}; y_{iks}, x, \boldsymbol{z}_{ik})$ is continuous and bounded. In addition, the function $\boldsymbol{x}_i = \boldsymbol{x}(t_i, \boldsymbol{\theta})$ is continuous and bounded. Thus $\frac{\partial}{\partial x}\ell(\boldsymbol{\alpha}; y_{iks}, x, \boldsymbol{z}_{ik})|_{\boldsymbol{x}=\boldsymbol{x}(t_i,\boldsymbol{\theta})}$ is $O_p(1)$. It is similar for $\tilde{\boldsymbol{x}}_i$. Then by Lemma 1 in Xue et al. (2010) and the first-order Taylor expansion, it follows

$$\begin{aligned}
\tilde{M}_n(\boldsymbol{\alpha}) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \ell(\boldsymbol{\alpha}; y_{iks}, \tilde{\boldsymbol{x}}_i, \boldsymbol{z}_{ik}) \\
&= \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \ell(\boldsymbol{\alpha}; y_{iks}, \boldsymbol{x}_i + O(N^{-\lambda(p\wedge 4)}), \boldsymbol{z}_{ik}) \\
&= \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \ell(\boldsymbol{\alpha}; y_{iks}, \boldsymbol{x}_i, \boldsymbol{z}_{ik}) + O_p(N^{-\lambda(p\wedge 4)}) \\
&= M_n(\boldsymbol{\alpha}) + O_p(N^{-\lambda(p\wedge 4)}).
\end{aligned}$$

Then

$$\tilde{M}_n(\hat{\boldsymbol{\alpha}}) - E_0[M(\boldsymbol{\alpha}_0)] \le \tilde{M}_n(\hat{\boldsymbol{\alpha}}) - E_0[M(\hat{\boldsymbol{\alpha}})] = M_n(\hat{\boldsymbol{\alpha}}) + O_p(N^{-\lambda(p\wedge 4)}) - E_0[M(\hat{\boldsymbol{\alpha}})].$$

and

$$\tilde{M}_n(\hat{\boldsymbol{\alpha}}) - E_0[M(\boldsymbol{\alpha}_0)] \ge \tilde{M}_n(\boldsymbol{\alpha}_0) - E_0[M(\boldsymbol{\alpha}_0)] = M_n(\boldsymbol{\alpha}_0) + O_p(N^{-\lambda(p\wedge 4)}) - E_0 M(\boldsymbol{\alpha}_0).$$

Hence $\tilde{M}_n(\hat{\boldsymbol{\alpha}}) - E_0[M(\boldsymbol{\alpha}_0)] \xrightarrow{p} 0$, when $N \to \infty$ and $\lambda > 0$. Thus

$$|E_0[M(\hat{\boldsymbol{\alpha}})] - E_0[M(\boldsymbol{\alpha}_0)]| \le |\tilde{M}_n(\hat{\boldsymbol{\alpha}}) - E_0[M(\hat{\boldsymbol{\alpha}})]| + |\tilde{M}_n(\hat{\boldsymbol{\alpha}}) - E_0[M(\boldsymbol{\alpha}_0)]| \xrightarrow{p} 0. \tag{A.2}$$

Now, we claim that $\hat{\boldsymbol{\alpha}}$ weakly converges to $\boldsymbol{\alpha}_0$. Otherwise, for sequence $\{\hat{\boldsymbol{\alpha}}\}$ in the compact subset $\Theta \times \mathcal{B} \times \Phi$, there exists a convergent subsequence, $\{\hat{\boldsymbol{\alpha}}_{n_k}\}$, such that $\hat{\boldsymbol{\alpha}}_{n_k} \xrightarrow{p} \boldsymbol{\alpha}_*$ and $\boldsymbol{\alpha}_* \neq \boldsymbol{\alpha}_0$. From the continuous mapping theorem, it follows that $E_0[M(\hat{\boldsymbol{\alpha}}_{n_k})] \xrightarrow{p} E_0[M(\boldsymbol{\alpha}_*)]$. From (A.2), we have $E_0[M(\hat{\boldsymbol{\alpha}}_{n_k})] \xrightarrow{p} E_0[M(\boldsymbol{\alpha}_0)]$. Then $E_0[M(\boldsymbol{\alpha}_*)] = E_0[M(\boldsymbol{\alpha}_0)]$. Since $\boldsymbol{\alpha}_0$ is the unique maximum point of $E_0[M(\boldsymbol{\alpha})]$, we have $\boldsymbol{\alpha}_* = \boldsymbol{\alpha}_0$, which is contradictory to $\boldsymbol{\alpha}_* \neq \boldsymbol{\alpha}_0$. Thus $\hat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}_0$. □

Theorem 2. (i) For $\delta = O(N^{-\lambda})$ with $\lambda > 1/(p \wedge 4)$ where p is the order of the numerical method, under Assumptions A1-A10, we have that $\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{H}^{-1})$;
(ii) For $\delta = O(N^{-\lambda})$ with $0 < \lambda \leq 1/(p \wedge 4)$, under Assumptions A1-A10, we have that $\sqrt{N}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}) \xrightarrow{d} \mathcal{N}(0, \tilde{\mathbf{H}}^{-1})$ with $\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\| = O_p(\delta^{(p\wedge 4)/2}) = O_p(N^{-\lambda(p\wedge 4)/2})$ and $\|\tilde{\mathbf{H}} - \mathbf{H}\| = O_p(\delta^{(p\wedge 4)/2}) = O_p(N^{-\lambda(p\wedge 4)/2})$.

Proof: For the proof of Part (i), it suffices to verify conditions of Theorem 3.1 in Newey and McFadden (1994) on asymptotic normality for a M-estimator. Denote

$$\tilde{G}_n(\boldsymbol{\alpha}) = \begin{pmatrix} \frac{\partial \tilde{M}_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \\ \frac{\partial \tilde{M}_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \\ \frac{\partial \tilde{M}_n(\boldsymbol{\alpha})}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}; y_{iks}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \tilde{\boldsymbol{x}}, \boldsymbol{\beta})}{\partial \tilde{\boldsymbol{x}}}\big|_{\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta})} \frac{\partial \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}; y_{iks}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}; y_{iks}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}), \boldsymbol{z}_{ik})}{\partial \phi} \end{pmatrix},$$

$$G_n(\boldsymbol{\alpha}) = \begin{pmatrix} \frac{\partial M_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \\ \frac{\partial M_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \\ \frac{\partial M_n(\boldsymbol{\alpha})}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}, \boldsymbol{\beta})}{\partial \boldsymbol{x}}\big|_{\boldsymbol{x}=\boldsymbol{x}(t_i,\boldsymbol{\theta})} \frac{\partial \boldsymbol{x}(t_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}(t_i,\boldsymbol{\theta}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}), \boldsymbol{z}_{ik})}{\partial \phi} \end{pmatrix},$$

$$G(\boldsymbol{\alpha}) = \begin{pmatrix} \frac{\partial E_0[M(\boldsymbol{\alpha})]}{\partial \boldsymbol{\theta}} \\ \frac{\partial E_0[M(\boldsymbol{\alpha})]}{\partial \boldsymbol{\beta}} \\ \frac{\partial E_0[M(\boldsymbol{\alpha})]}{\partial \phi} \end{pmatrix} = \begin{pmatrix} E_0\big[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\beta})}{\partial \boldsymbol{x}}\big|_{\boldsymbol{x}=\boldsymbol{x}(t,\boldsymbol{\theta})} \frac{\partial \boldsymbol{x}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big] \\ E_0\big[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\big] \\ E_0\big[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial \phi}\big] \end{pmatrix},$$

where $\frac{\partial \ell}{\partial g^*} = \frac{\partial}{\partial g^*}\ell(., g^*, .)|_{g^*=g^*(\boldsymbol{z},\boldsymbol{x},\boldsymbol{\beta})}$ with $g^*(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\beta})$ defined in (2.5). Since $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$ are maximum points of $\tilde{M}_n(\boldsymbol{\alpha})$ and $E_0[M(\boldsymbol{\alpha})]$, respectively, $\tilde{G}_n(\hat{\boldsymbol{\alpha}}) = 0$ and $G(\boldsymbol{\alpha}_0) = 0$.

First, we verify the result that $\sqrt{N}[\tilde{G}_n(\boldsymbol{\alpha}_0) - G(\boldsymbol{\alpha}_0)] \xrightarrow{d} \mathcal{N}(0, \mathbf{H})$ with $\mathbf{H}$ defined by (4.18). For fixed t, by the Landau-Kolmogorov inequality between different derivatives of a function, we have $\|\frac{\partial \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \boldsymbol{x}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\|_\infty \leq C\|\frac{\partial^2 \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} - \frac{\partial^2 \boldsymbol{x}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\|_\infty^{1/2}\|\tilde{\boldsymbol{x}}(t,\boldsymbol{\theta}) - \boldsymbol{x}(t,\boldsymbol{\theta})\|_\infty^{1/2} \leq C'\|\tilde{\boldsymbol{x}}(t,\boldsymbol{\theta}) - \boldsymbol{x}(t,\boldsymbol{\theta})\|_\infty^{1/2}$ for two constants C and C', where the second inequality holds because of the uniform boundedness of both $\frac{\partial^2 \boldsymbol{x}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}$ and $\frac{\partial^2 \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}$ under Assumptions A7-A8, and $\|\Upsilon(\boldsymbol{\theta})\|_\infty = \sup_{\boldsymbol{\theta}}|\Upsilon(\boldsymbol{\theta})|$ is the supremum norm of a function Υ. Based on $\sup_{t\in[t_0,T]}\|\tilde{\boldsymbol{x}}(t,\boldsymbol{\theta}) - \boldsymbol{x}(t,\boldsymbol{\theta})\|_\infty = O(N^{-\lambda(p\wedge 4)})$ from Lemma 1 in Xue et al. (2010), it follows that $\|\frac{\partial \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \boldsymbol{x}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\| = O(N^{-\lambda(p\wedge 4)/2})$.

Then we have

$$
\begin{aligned}
&\sqrt{N}[\tilde{G}_n(\boldsymbol{\alpha}_0) - G(\boldsymbol{\alpha}_0)] \\
=& \begin{pmatrix}
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \tilde{\boldsymbol{x}}, \boldsymbol{\beta}_0)}{\partial \tilde{\boldsymbol{x}}} |_{\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}_0)} \frac{\partial \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} \\
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}_0), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0} \\
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \tilde{\boldsymbol{x}}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial \phi_0}
\end{pmatrix} \\
=& \begin{pmatrix}
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} [\frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial g^*} + O_p(N^{-\lambda(p\wedge 4)})][\frac{\partial g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{x}} |_{\boldsymbol{x}=\boldsymbol{x}(t_i,\boldsymbol{\theta}_0)} + O_p(N^{-\lambda(p\wedge 4)})] \\
\times [\frac{\partial \boldsymbol{x}(t_i,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} + O_p(N^{-\lambda(p\wedge 4)/2})] \\
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} [\frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial g^*} + O_p(N^{-\lambda(p\wedge 4)})][\frac{\partial g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0} + O_p(N^{-\lambda(p\wedge 4)})] \\
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} [\frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial \phi_0} + O_p(N^{-\lambda(p\wedge 4)})]
\end{pmatrix} \\
=& \begin{pmatrix}
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{x}} |_{\boldsymbol{x}=\boldsymbol{x}(t_i,\boldsymbol{\theta}_0)} \frac{\partial \boldsymbol{x}(t_i,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} \\
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}_{ik}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0} \\
\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{s=1}^{S_{ik}} \frac{\partial \ell(\boldsymbol{\alpha}_0; y_{iks}, \boldsymbol{x}(t_i,\boldsymbol{\theta}_0), \boldsymbol{z}_{ik})}{\partial \phi_0}
\end{pmatrix} + O_p(N^{-\lambda(p\wedge 4)/2+1/2}) \\
=& \sqrt{N}[G_n(\boldsymbol{\alpha}_0) - G(\boldsymbol{\alpha}_0)] + O_p(N^{-\lambda(p\wedge 4)/2+1/2}).
\end{aligned}
$$

When $\lambda > 1/(p\wedge 4)$, $O_p(N^{-\lambda(p\wedge 4)/2+1/2}) = o_p(1)$. So for the above expression, we have $\sqrt{N}[\tilde{G}_n(\boldsymbol{\alpha}_0) - G(\boldsymbol{\alpha}_0)] = \sqrt{N}[G_n(\boldsymbol{\alpha}_0) - G(\boldsymbol{\alpha}_0)] + o_p(1)$. Based on the special structure of ξ_{ik} and ϕ in (2.4), it is easy to follow that $E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z})}{\partial(\boldsymbol{\theta}^T, \boldsymbol{\beta}^T)} \frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z})}{\partial \phi}]|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = \mathbf{0}$, that is, the dispersion parameter ϕ is orthogonal to parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ in the sense of Cox and Reid (1987). From the standard central limit theorem, we have $\sqrt{N}[G_n(\boldsymbol{\alpha}_0) - G(\boldsymbol{\alpha}_0)] \xrightarrow{d} \mathcal{N}(0, \mathbf{J})$, where

$$
\mathbf{J} = \begin{pmatrix}
E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z})}{\partial(\boldsymbol{\theta}^T, \boldsymbol{\beta}^T)^T} \frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z})}{\partial(\boldsymbol{\theta}^T, \boldsymbol{\beta}^T)}]|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} & \mathbf{0} \\
\mathbf{0}^T & E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z})}{\partial \phi}]^2|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}
\end{pmatrix}
$$

Thus it follows that $\sqrt{N}[\tilde{G}_n(\boldsymbol{\alpha}_0) - G(\boldsymbol{\alpha}_0)] \xrightarrow{d} \mathcal{N}(0, \mathbf{J})$.

Next, by similar arguments to those in the proof of Theorem 1, it follows that $\sup_{\boldsymbol{\alpha}\in\Delta} \| \frac{\partial^2 M_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T} + \mathbf{H}(\boldsymbol{\alpha}) \| \xrightarrow{P} 0$ with a neighborhood Δ of $\boldsymbol{\alpha}_0$. By Lemma 1 in Xue et al. (2010), it follows that $\sup_{\boldsymbol{\alpha}\in\Delta} \| \frac{\partial^2 \tilde{M}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T} - \frac{\partial^2 M_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T} \| \xrightarrow{p} 0$. Then we have $\sup_{\boldsymbol{\alpha}\in\Delta} \| \frac{\partial^2 \tilde{M}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T} + \mathbf{H}(\boldsymbol{\alpha}) \| \xrightarrow{p} 0$. By Theorem 1, Assumptions A9 and A10, and $\mathbf{H} = \mathbf{J}$, all conditions of Theorem 3.1 in Newey and McFadden (1994) are satisfied, thus Theorem 2(i) holds with a variance-covariance matrix $(-\mathbf{H})^{-1}\mathbf{J}(-\mathbf{H})^{-1} = \mathbf{J}^{-1} = \mathbf{H}^{-1}$.

Now, we consider the proof of Theorem 2(ii). Define

$$
\begin{aligned}
\tilde{M}(\boldsymbol{\alpha}) &= \ell(\boldsymbol{\alpha}; y, \tilde{\boldsymbol{x}}, \boldsymbol{z}) \\
&= \frac{y \cdot b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}, \tilde{\boldsymbol{x}}(t, \boldsymbol{\theta}), \boldsymbol{\beta}) - b\{b'^{-1} \circ g^{-1} \circ g^*(\boldsymbol{z}, \tilde{\boldsymbol{x}}(t, \boldsymbol{\theta}), \boldsymbol{\beta})\}}{a(\phi)} + c(y, \phi),
\end{aligned}
$$

and

$$\tilde{G}(\boldsymbol{\alpha}) = \begin{pmatrix} \frac{\partial \tilde{M}(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \\ \frac{\partial \tilde{M}(\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \\ \frac{\partial \tilde{M}(\boldsymbol{\alpha})}{\partial \phi} \end{pmatrix} = \begin{pmatrix} E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}, \tilde{\boldsymbol{x}}, \boldsymbol{\beta})}{\partial \tilde{\boldsymbol{x}}}|_{\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{x}}(t,\boldsymbol{\theta})} \frac{\partial \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}] \\ E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}, \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}] \\ E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \tilde{\boldsymbol{x}}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial \phi}] \end{pmatrix}.$$

Since $E_0[\tilde{M}(\boldsymbol{\alpha})]$ reaches its maximum at $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$, then the first-order derivative of $E_0[\tilde{M}(\boldsymbol{\alpha})]$ at $\tilde{\boldsymbol{\alpha}}$ equals to 0, i.e., $\tilde{G}(\tilde{\boldsymbol{\alpha}}) = 0$. Then similar to the proof of Case (i) above, we have

$$\begin{aligned} \tilde{G}(\boldsymbol{\alpha}) &= \begin{pmatrix} E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} + O_p(N^{-\lambda(p\wedge 4)})][\frac{\partial g^*(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\beta})}{\partial \boldsymbol{x}}|_{\boldsymbol{x}=\boldsymbol{x}(t,\boldsymbol{\theta})} + O_p(N^{-\lambda(p\wedge 4)})] \\ \times[\frac{\partial \boldsymbol{x}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + O_p(N^{-\lambda(p\wedge 4)/2})] \\ E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} + O_p(N^{-\lambda(p\wedge 4)})][\frac{\partial g^*(\boldsymbol{z}, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + O_p(N^{-\lambda(p\wedge 4)})] \\ E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial \phi} + O_p(N^{-\lambda(p\wedge 4)})] \end{pmatrix} \\ &= \begin{pmatrix} E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{\beta})}{\partial \boldsymbol{x}}|_{\boldsymbol{x}=\boldsymbol{x}(t,\boldsymbol{\theta})} \frac{\partial \boldsymbol{x}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}] \\ E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial g^*} \frac{\partial g^*(\boldsymbol{z}, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}] \\ E_0[\frac{\partial \ell(\boldsymbol{\alpha}; y, \boldsymbol{x}(t,\boldsymbol{\theta}), \boldsymbol{z})}{\partial \phi}] \end{pmatrix} + O_p(N^{-\lambda(p\wedge 4)/2}) \\ &= G(\boldsymbol{\alpha}) + O_p(N^{-\lambda(p\wedge 4)/2}). \end{aligned}$$

It follows that $\tilde{G}(\tilde{\boldsymbol{\alpha}}) = G(\tilde{\boldsymbol{\alpha}}) + O_p(N^{-\lambda(p\wedge 4)/2})$, then $G(\tilde{\boldsymbol{\alpha}}) = O_p(N^{-\lambda(p\wedge 4)/2})$ from $\tilde{G}(\tilde{\boldsymbol{\alpha}}) = 0$. The Taylor series expansion yields that there exist constants $0 < c_1, c_2 < \infty$ such that

$$c_1\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\| \leq |G(\tilde{\boldsymbol{\alpha}}) - G(\boldsymbol{\alpha}_0)| \leq c_2\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|.$$

Thus $\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\| = O_p(N^{-\lambda(p\wedge 4)/2})$ from $G(\boldsymbol{\alpha}_0) = 0$. It follows that $\tilde{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}_0$. By Assumption A9, we have that $\tilde{\boldsymbol{\alpha}}$ is also an interior point of $\Theta \times \mathcal{B} \times \Phi$ for sufficiently large N. Similarly we can show that $\|\tilde{\mathbf{H}} - \mathbf{H}\| = O_p(N^{-\lambda(p\wedge 4)/2})$. By Assumption A10, we have that $\tilde{\mathbf{H}}$ is also nonsingular for sufficiently large N. Then it is straightforward to derive the asymptotic normality of Theorem 2(ii) by verifying the conditions of Theorem 3.3 in Newey and McFadden (1994) on asymptotic normality for MLE. □

Theorem 3. For the weighted MLE $\hat{\boldsymbol{\alpha}}^*$ in (3.14), under the same assumptions as those in Theorem 2 as well as Assumption A11, we have that,
(i) for $\lambda > 1/(p \wedge 4)$, $\sqrt{N/v_0}(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})$ has the same conditional limiting distribution as $\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ has unconditionally, i.e., $\left(\sqrt{N/v_0}(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})\big|\{t_i, \boldsymbol{z}_{ik}, y_{iks}\}\right) \xrightarrow{d} \sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$. Thus $\sqrt{N}(\hat{\boldsymbol{\alpha}}^* - \boldsymbol{\alpha}_0) \xrightarrow{d} N(0, (1 + v_0)\mathbf{H}^{-1})$.
(ii) for $0 < \lambda \le 1/(p \wedge 4)$, $\left(\sqrt{N/v_0}(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})\big|\{t_i, \boldsymbol{z}_{ik}, y_{iks}\}\right) \xrightarrow{d} \sqrt{N}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}})$. Thus $\sqrt{N}(\hat{\boldsymbol{\alpha}}^* - \tilde{\boldsymbol{\alpha}}) \xrightarrow{d} N(0, (1 + v_0)\tilde{\mathbf{H}}^{-1})$.

Proof: Replace all likelihood functions $\ell(.)$ in the proofs of Theorems 1-2 with the weight likelihood functions $w\ell(.)$ and denote

$$
\begin{aligned}
\tilde{M}_n^*(\boldsymbol{\alpha}) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} w_{iks}\ell(\boldsymbol{\alpha}; y_{iks}, \tilde{\boldsymbol{x}}_i, \boldsymbol{z}_{ik}),\\
M_n^*(\boldsymbol{\alpha}) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{s=1}^{S_{ik}} w_{iks}\ell(\boldsymbol{\alpha}; y_{iks}, \boldsymbol{x}_i, \boldsymbol{z}_{ik}),\\
M^*(\boldsymbol{\alpha}) &= w\ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z}),
\end{aligned}
$$

and $\tilde{G}_n^*(\boldsymbol{\alpha}) = \frac{\partial \tilde{M}_n^*(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$, $G_n^*(\boldsymbol{\alpha}) = \frac{\partial M_n^*(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$, $G^*(\boldsymbol{\alpha}) = \frac{\partial E_0[M^*(\boldsymbol{\alpha})]}{\partial \boldsymbol{\alpha}}$.

First, we claim that under Assumptions A1-A9 and A11, we have $\hat{\boldsymbol{\alpha}}^* - \boldsymbol{\alpha}_0 \to 0$, almost surely under $P_{\boldsymbol{\alpha}_0}$, i.e., Theorem 1 still holds for the weighted MLE $\hat{\boldsymbol{\alpha}}^*$. In fact, by the Jensen's inequality, the convexity of function $-\log x$, $E(w) = 1$ and the independence between w and $(t, y, \boldsymbol{z})$, we have

$$
\begin{aligned}
& E_0[M^*(\boldsymbol{\alpha})] - E_0[M^*(\boldsymbol{\alpha}_0)]\\
= \;& E_0[w\ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z}) - w\ell(\boldsymbol{\alpha}_0; y, \boldsymbol{x}, \boldsymbol{z})]\\
= \;& E(w)E_0[\ell(\boldsymbol{\alpha}; y, \boldsymbol{x}, \boldsymbol{z}) - \ell(\boldsymbol{\alpha}_0; y, \boldsymbol{x}, \boldsymbol{z})]\\
\le \;& 0,
\end{aligned}
$$

i.e., $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ is still the maximum point of function $E_0[M^*(\boldsymbol{\alpha})]$. In addition, from Assumption A6, it follows that it is the unique maximum point. Under Assumption A11, similar to (A.1), for any $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \Theta \times \mathcal{B} \times \Phi$, we have

$$
|M_n^*(\boldsymbol{\alpha}_1) - M_n^*(\boldsymbol{\alpha}_2)| \le Q|M_n(\boldsymbol{\alpha}_1) - M_n(\boldsymbol{\alpha}_2)| \le C(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})Q\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|.
$$

The remaining steps to prove this claim are similar to those in the proof of Theorem 1.

Next, we claim that $\sqrt{N}(\hat{\boldsymbol{\alpha}}^* - \boldsymbol{\alpha}_0) \xrightarrow{d} \mathcal{N}(0, (1 + v_0)\mathbf{H}^{-1})$ or $\sqrt{N}(\tilde{\boldsymbol{\alpha}}^* - \tilde{\boldsymbol{\alpha}}) \xrightarrow{d} \mathcal{N}(0, (1 + v_0)\tilde{\mathbf{H}}^{-1})$ for $\lambda > 1/(p \wedge 4)$ and $0 < \lambda \le 1/(p \wedge 4)$, respectively. We verify it as follows.

Since $\hat{\boldsymbol{\alpha}}^*$ and $\boldsymbol{\alpha}_0$ are the maximum points of $\tilde{M}_n^*(\boldsymbol{\alpha})$ and $E_0[M^*(\boldsymbol{\alpha})]$, respectively, $\tilde{G}_n^*(\hat{\boldsymbol{\alpha}}^*) = 0$ and $G^*(\boldsymbol{\alpha}_0) = 0$. By similar arguments to those in the proof of Theorem 2, we have $\sqrt{N}[\tilde{G}_n^*(\boldsymbol{\alpha}_0) - G^*(\boldsymbol{\alpha}_0)] = \sqrt{N}[G_n^*(\boldsymbol{\alpha}_0) - G^*(\boldsymbol{\alpha}_0)] + o_p(1)$. From the standard central limit theorem and $E(w^2) = 1 + v_0$, we have $\sqrt{N}[G_n^*(\boldsymbol{\alpha}_0) - G^*(\boldsymbol{\alpha}_0)] \xrightarrow{d} \mathcal{N}(0, (1 + v_0)\mathbf{J})$. Thus it follows that $\sqrt{N}[\tilde{G}_n^*(\boldsymbol{\alpha}_0) - G^*(\boldsymbol{\alpha}_0)] \xrightarrow{d} \mathcal{N}(0, (1 + v_0)\mathbf{J})$. In addition, by the weak law of large number and $E(w) = 1$, it follows that

$$
\frac{\partial^2 M_n^*(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \xrightarrow{p} E_0\left[\frac{\partial^2 M^*(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T}\right] = E(w)E_0\left[\frac{\partial^2 M(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T}\right] = -\mathbf{H}.
$$

The remaining steps to prove this claim are similar to those in the proof of Theorem 2. Combining the second claim and Theorem 2, we have that

$$\left(\sqrt{N/v_0}(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})\big|\{t_i, \boldsymbol{z}_{ik}, y_{iks}\}\right) \xrightarrow{d} \sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$$

for $\lambda > 1/(p \wedge 4)$ and

$$\left(\sqrt{N/v_0}(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})\big|\{t_i, \boldsymbol{z}_{ik}, y_{iks}\}\right) \xrightarrow{d} \sqrt{N}(\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}})$$

for $0 < \lambda \leq 1/(p \wedge 4)$, respectively. □

References

Andrews, D. W. K. (1987), "Consistency in nonlinear econometric models: a general uniform law of large numbers," *Econometrica*, 55, 1465–1471.

Cox, D. R., and Reid, N. (1987), "Parameter orthogonality and approximate conditional inference," *Journal of Royal Statistical Society, Series B*, 49(1), 1–39.

Newey, W., and McFadden, D. (1994), "Large sample estimation and hypothesis testing," in *Handbook of Econometrics, Vol. 4*, eds. R. Engle, and D. McFadden, Amsterdam, The Netherlands: Elsevier, chapter 36, pp. 2111–2245.

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(4), 2351–2387.

B Additional Simulation Result

Table B.1: Evaluation of the GODE method by calculating average relative errors (ARE) in parameter estimates based on 500 simulation runs with ρ_E fixed at zero.

distribution	# of obs. per dilu. factor	variance	ARE(%) of GODE					ARE(%) of ESB				
			ρ_E	β_E	δ_{E^*}	c_V	β	ρ_E	β_E	δ_{E^*}	c_V	β
binomial	5	-	fixed	8.53%	32.5%	27.7%	3.97%	fixed	91.9%	72.8%	36.5%	3.85%
	10	-	fixed	6.49%	18.9%	18.9%	2.85%	fixed	96.3%	72.1%	30.1%	2.83%
	20	-	fixed	4.27%	12.4%	12.5%	2.02%	fixed	87.5%	38.6%	19.2%	2.33%
Poisson	5	-	fixed	13.0%	45.6%	42.1%	7.05%	fixed	68.1%	85.3%	38.1%	6.95%
	10	-	fixed	8.67%	31.5%	28.8%	4.91%	fixed	72.8%	66.7%	29.4%	4.94%
	20	-	fixed	6.38%	21.2%	19.8%	3.45%	fixed	69.2%	51.2%	19.5%	3.49%
Gamma	5	3.0	fixed	10.3%	48.3%	58.1%	84.3%	fixed	86.0%	86.5%	58.6%	84.6%
	10	1.5	fixed	6.46%	48.8%	44.3%	76.9%	fixed	107%	83.8%	29.8%	77.1%
	20	0.75	fixed	11.8%	35.9%	25.0%	69.5%	fixed	113%	89.4%	21.2%	69.7%
normal	5	3.0	fixed	47.7%	72.4%	116%	54.8%	fixed	66.3%	74.2%	46.1%	71.8%
	10	1.5	fixed	20.1%	53.9%	60.2%	13.9%	fixed	66.7%	61.8%	31.8%	15.4%
	20	0.75	fixed	9.11%	30.5%	27.8%	6.41%	fixed	62.5%	52.0%	22.2%	6.33%