Supplement to “Nonlinear Panel Data Models with Distribution-Free Correlated Random Effects"

Yu-Chin Hsu* Ji-Liang Shiu†

September 7, 2020

This supplement complements Appendix A of the main paper, and we use a sequential numbering that follows Appendix A. In Appendix B, we provide primitive conditions for Assumption 2.6 and work through the binary choice model in Eq. (6) of the main text. In Appendix C, we provide results for the sieve maximum likelihood estimators proposed in the main paper. In Appendix D, we provide a consistent estimator for Ω_{di} that is needed for the Hausman-type test proposed in Section 4.1. We present the Monte Carlo simulation results in Appendix E.

B. Discussion on Assumption 2.6

B.1. Primitive Conditions for Assumption 2.6

The internally consistent semiparametric density of observable variables has the following form:

(B.1)
$$f(y|x,\overline{w};\alpha)=\int_{\mathscr{C}} f_{Y|X,C}(y|x,c;\theta)f_{C|\overline{W}}(c|\overline{w};\alpha)dc,$$

where $f_{C|\overline{W}}(c|\overline{w};\alpha)=\frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty}e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi$ with

(B.2)
$$\phi_{v;\alpha}(\xi)=\frac{-\lambda_1\int_{\mathscr{Y}\times\mathscr{X}\times\overline{\mathscr{W}}}e^{-i\xi\sum_{k=1}\lambda_k\overline{w}_k}f_{Y|X,\overline{W}}(y|x,\overline{w})\Omega(y,x,\overline{w}_{-1})dydxd\overline{w}}{\int_{\mathscr{Y}\times\mathscr{X}\times\mathscr{C}}e^{-i\xi c}f_{Y|X,C}(y|x,c;\theta)\Omega(y,x)dydxdc}.$$

*Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University. E-mail: ychsu@econ.sinica.edu.tw.

†Institute for Economic and Social Research, Jinan University. E-mail: jishiu.econ@gmail.com.

To illustrate, we consider $\alpha = (\theta_1, \lambda_1, \lambda_2)$, i.e., θ is a scalar and λ is a 2×1 vector. Similar lines of argument show that the same conclusion holds for general cases. The scale factor can be written as follows:

$$
\text{(B.3)} \qquad \begin{aligned} c_\alpha(\overline{w}) \equiv \int_{\mathscr{C}} f_{v;\alpha}(c - \overline{w}\lambda) dc &= \frac{1}{2\pi} \int_{\mathscr{C}} \int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)} \phi_{v;\alpha}(\xi) d\xi dc \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{\mathscr{C}} e^{-i\xi c} dc \right) e^{i\xi\overline{w}\lambda} \phi_{v;\alpha}(\xi) d\xi. \end{aligned}
$$

The gradient of the log likelihood at the true value α_0 in this simple case is

$$
\text{(B.4)} \qquad \begin{aligned} &\frac{\partial}{\partial \alpha} \log f(Y|X, \overline{W}; \alpha_0) \\ &= \frac{1}{f(Y|X, \overline{W}; \alpha)} \Big(\frac{\partial}{\partial \theta_1} f(Y|X, \overline{W}; \alpha_0), \frac{\partial}{\partial \lambda_1} f(Y|X, \overline{W}; \alpha_0), \frac{\partial}{\partial \lambda_2} f(Y|X, \overline{W}; \alpha_0) \Big)', \end{aligned}
$$

with

$$
\text{(B.5)} \qquad \begin{aligned} \frac{\partial}{\partial \theta_1} f(Y|X, \overline{W}; \alpha_0) = &\int_{\mathscr{C}} \frac{\partial}{\partial \theta_1} f_{Y|X,C}(y|x, c; \theta_0) f_{C|\overline{W}}(c|\overline{w}; \alpha_0) dc \\ &+ \int_{\mathscr{C}} f_{Y|X,C}(y|x, c; \theta_0) \underbrace{\frac{\partial}{\partial \theta_1} f_{C|\overline{W}}(c|\overline{w}; \alpha_0)}_{\substack{\text{related to } f(y|x,\overline{w}) \\ \text{and } f_{Y|X,C}(y|x,c;\theta)}} dc, \end{aligned}
$$

$$
\text{(B.6)} \qquad \frac{\partial}{\partial \lambda_1} f(Y|X, \overline{W}; \alpha_0) = \int_{\mathscr{C}} f_{Y|X,C}(y|x, c; \theta_0) \frac{\partial}{\partial \lambda_1} f_{C|\overline{W}}(c|\overline{w}; \alpha_0) dc,
$$

$$
\text{(B.7)} \qquad \frac{\partial}{\partial \lambda_2} f(Y|X, \overline{W}; \alpha) = \int_{\mathscr{C}} f_{Y|X,C}(y|x, c; \theta_0) \frac{\partial}{\partial \lambda_2} f_{C|\overline{W}}(c|\overline{w}; \alpha_0) dc.
$$

Compared with the conventional CRE approach that $V \sim N(0,1)$, and $f_{C|\overline{W}}(c|\overline{w}; \lambda) = \phi(c - \overline{w}\lambda)$ where $\phi(\cdot)$ denotes the density function of the standard normal, the derivative term $\frac{\partial}{\partial \theta_1} f_{C|\overline{W}}(c|\overline{w}; \alpha)$ in Eq. (B.5) in our model is generally nonzero and related to the sample density and the proposed model. This implies that the term $\int_{\mathscr{C}} f_{Y|X,C}(y|x, c; \theta_0) \frac{\partial}{\partial \theta_1} f_{C|\overline{W}}(c|\overline{w}; \alpha_0)$ in our model is used to accommodate the modeling of the panel data structure $f_{Y|X,C}(y|x, c; \theta)$ and the distribution of the unobserved heterogeneity.

To provide a more primitive condition on the densities, we need to write out the derivatives of the composite distribution of the unobserved heterogeneity $\frac{\partial}{\partial \theta_1} f_{C|\overline{W};\alpha}$, $\frac{\partial}{\partial \lambda_1} f_{C|\overline{W};\alpha}$, and

$\frac{\partial}{\partial \lambda_2} f_{C|\overline{W};\alpha}$ at the true value α_0, where $f_{C|\overline{W};\alpha} \equiv f_{C|\overline{W}}(c|\overline{w};\alpha)$. We have

$$
\begin{aligned}
\frac{\partial \phi_{v;\alpha}(\xi)}{\partial \theta_1} &= \phi_{v;\alpha}(\xi) \times \frac{-\int_{\mathscr{Y}\times\mathscr{X}\times\mathscr{C}} e^{-i\xi c} \frac{\partial f_{Y|X,C}(y|x,c;\theta)}{\partial \theta_1} \Omega(y,x) dydxdc}{\int_{\mathscr{Y}\times\mathscr{X}\times\mathscr{C}} e^{-i\xi c} f_{Y|X,C}(y|x,c;\theta)\Omega(y,x)dydxdc} \equiv \phi_{v;\alpha}(\xi)\gamma_{\theta_1}(\xi;\theta), \\
\frac{\partial \phi_{v;\alpha}(\xi)}{\partial \lambda_1} &= \frac{1}{\lambda_1}\phi_{v;\alpha}(\xi) + \frac{i\lambda_1\xi \int_{\mathscr{Y}\times\mathscr{X}\times\overline{\mathscr{W}}} e^{-i\xi\sum_{k=1}\lambda_k\overline{w}_k}\overline{w}_1 f_{Y|X,\overline{W}}(y|x,\overline{w})\Omega(y,x,\overline{w}_{-1})dydxd\overline{w}}{\int_{\mathscr{Y}\times\mathscr{X}\times\mathscr{C}} e^{-i\xi c} f_{Y|X,C}(y|x,c;\theta)\Omega(y,x)dydxdc} \\
&\equiv \frac{1}{\lambda_1}\phi_{v;\alpha}(\xi) + i\lambda_1\xi\gamma_{\lambda_1}(\xi;\alpha), \\
\frac{\partial \phi_{v;\alpha}(\xi)}{\partial \lambda_2} &= \frac{i\lambda_1\xi \int_{\mathscr{Y}\times\mathscr{X}\times\overline{\mathscr{W}}} e^{-i\xi\sum_{k=1}\lambda_k\overline{w}_k}\overline{w}_2 f_{Y|X,\overline{W}}(y|x,\overline{w})\Omega(y,x,\overline{w}_{-1})dydxd\overline{w}}{\int_{\mathscr{Y}\times\mathscr{X}\times\mathscr{C}} e^{-i\xi c} f_{Y|X,C}(y|x,c;\theta)\Omega(y,x)dydxdc} \\
&\equiv i\lambda_1\xi\gamma_{\lambda_2}(\xi;\alpha).
\end{aligned}
$$

We note that these derivatives are mainly stated in terms of the proposed parametric nonlinear panel data model $f_{Y|X,C}(y|x,c;\theta)$ and the density of observables $f_{Y|X,\overline{W}}(y|x,\overline{w})$. These derivatives evaluated at the true parameter α_0 are

$$
\begin{aligned}
\frac{\partial \phi_{v;\alpha_0}(\xi)}{\partial \theta_1} &= \phi_v(\xi)\gamma_{\theta_1}(\xi;\theta_0), \\
\frac{\partial \phi_{v;\alpha_0}(\xi)}{\partial \lambda_1} &= \frac{1}{\lambda_1}\phi_v(\xi) + i\lambda_{01}\xi\gamma_{\lambda_1}(\xi;\alpha_0), \\
\frac{\partial \phi_{v;\alpha_0}(\xi)}{\partial \lambda_2} &= i\lambda_{01}\xi\gamma_{\lambda_2}(\xi;\alpha).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\text{(B.8)}\quad & \frac{\partial}{\partial \theta_1} f_{C|\overline{W};\alpha} \\
&= \frac{-1}{2\pi c_\alpha(\overline{w})^2}\frac{\partial c_\alpha(\overline{w})}{\partial \theta_1}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi + \frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\frac{\partial \phi_{v;\alpha}(\xi)}{\partial \theta_1}d\xi \\
&= \frac{-1}{2\pi c_\alpha(\overline{w})^2}\left(\int_{-\infty}^{\infty}(\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc)e^{i\xi\overline{w}\lambda}\frac{\partial \phi_{v;\alpha}(\xi)}{\partial \theta_1}d\xi\right)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi \\
&\quad + \frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\frac{\partial \phi_{v;\alpha}(\xi)}{\partial \theta_1}d\xi \\
&= \frac{-1}{2\pi c_\alpha(\overline{w})^2}\left(\int_{-\infty}^{\infty}(\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc)e^{i\xi\overline{w}\lambda}\phi_{v;\alpha}(\xi)\gamma_{\theta_1}(\xi;\theta)d\xi\right)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi \\
&\quad + \frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)\gamma_{\theta_1}(\xi;\theta)d\xi. \\
\text{(B.9)}\quad & \frac{\partial}{\partial \lambda_1} f_{C|\overline{W};\alpha} \\
&= \frac{-1}{2\pi c_\alpha(\overline{w})^2}\frac{\partial c_\alpha(\overline{w})}{\partial \lambda_1}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi + \frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} i\xi\overline{w}_1 e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi
\end{aligned}
$$

$$+\frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\frac{\partial\phi_{v;\alpha}(\xi)}{\partial\lambda_1}d\xi$$

$$=\frac{-1}{2\pi c_\alpha(\overline{w})^2}\left(\int_{-\infty}^{\infty}(\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc)e^{i\xi\overline{w}\lambda}(i\xi\overline{w}_1\phi_{v;\alpha}(\xi)+\frac{\partial\phi_{v;\alpha}(\xi)}{\partial\lambda_1})d\xi\right)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi.$$

$$+\frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} i\xi\overline{w}_1 e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi+\frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\frac{\partial\phi_{v;\alpha}(\xi)}{\partial\lambda_1}d\xi.$$

(B.10)

$$\frac{\partial}{\partial\lambda_2}f_{C|\overline{W};\alpha}$$

$$=\frac{-1}{2\pi c_\alpha(\overline{w})^2}\left(\int_{-\infty}^{\infty}(\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc)e^{i\xi\overline{w}\lambda}(i\xi\overline{w}_2\phi_{v;\alpha}(\xi)+\frac{\partial\phi_{v;\alpha}(\xi)}{\partial\lambda_2})d\xi\right)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi$$

$$+\frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} i\xi\overline{w}_2 e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi+\frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\frac{\partial\phi_{v;\alpha}(\xi)}{\partial\lambda_2}d\xi.$$

The derivatives at the true value α_0 can be stated as follows

$$\text{(B.11)}\quad \frac{\partial}{\partial\theta_1}f_{C|\overline{W};\alpha_0}=\frac{-1}{2\pi}\left(\int_{-\infty}^{\infty}(\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc)e^{i\xi\overline{w}\lambda_0}\phi_v(\xi)\gamma_{\theta_{01}}(\xi;\theta_0)d\xi\right)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi$$
$$+\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)\gamma_{\theta_1}(\xi;\theta_0)d\xi,$$

(B.12)

$$\frac{\partial}{\partial\lambda_1}f_{C|\overline{W};\alpha_0}$$

$$=\frac{-1}{2\pi}\left(\int_{-\infty}^{\infty}(\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc)e^{i\xi\overline{w}\lambda_0}(i\xi\overline{w}_{01}\phi_v(\xi)+\frac{\partial\phi_{v;\alpha_0}(\xi)}{\partial\lambda_1})d\xi\right)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi$$

$$+\frac{1}{2\pi}\int_{-\infty}^{\infty} i\xi\overline{w}_{01}e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi+\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\frac{\partial\phi_{v;\alpha_0}(\xi)}{\partial\lambda_1}d\xi,$$

(B.13)

$$\frac{\partial}{\partial\lambda_2}f_{C|\overline{W};\alpha_0}$$

$$=\frac{-1}{2\pi}\left(\int_{-\infty}^{\infty}(\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc)e^{i\xi\overline{w}\lambda_0}(i\xi\overline{w}_{02}\phi_v(\xi)+\frac{\partial\phi_{v;\alpha_0}(\xi)}{\partial\lambda_2})d\xi\right)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi$$

$$+\frac{1}{2\pi}\int_{-\infty}^{\infty} i\xi\overline{w}_{02}e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi+\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\frac{\partial\phi_{v;\alpha_0}(\xi)}{\partial\lambda_2}d\xi,$$

where $c_{\alpha_0}(\overline{w})=1$, and $\phi_{v;\alpha_0}(\xi)=\phi_v(\xi)$.

When $\mathscr{C}=\mathbb{R}$, or the domain of C is a real line, using the relationship in Eq. (A.11) we obtain

$$\text{(B.14)}\qquad f_{C|\overline{W}}(c|\overline{w};\alpha)=\frac{1}{2\pi\phi_{v;\alpha}(0)}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi.$$

Applying the relations $\frac{1}{2\pi}\int_{\mathscr{C}} e^{-i\xi c}dc=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi c}dc=\delta(\xi)$ and $\int f(\xi)\delta(\xi)dt=f(0)$ into Eqs.

(B.11)-(B.13) yields

$$
\begin{aligned}
\text{(B.15)} \qquad \frac{\partial}{\partial\theta_1} f_{C|\overline{W};\alpha_0} &= \frac{-1}{2\pi}\phi_v(0)\gamma_{\theta_{01}}(0;\theta_0)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi \\
&\quad + \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)\gamma_{\theta_1}(\xi;\theta_0)d\xi, \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)\left(\gamma_{\theta_1}(\xi;\theta_0)-\gamma_{\theta_{01}}(0;\theta_0)\right)d\xi, \\
\frac{\partial}{\partial\lambda_1} f_{C|\overline{W};\alpha_0} &= \frac{-1}{2\pi}\frac{1}{\lambda_1}\phi_v(0)\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi + \frac{1}{2\pi}\int_{-\infty}^{\infty} i\xi\overline{w}_{01}e^{-i\xi(c-\overline{w}\lambda_0)}\phi_v(\xi)d\xi \\
&\quad + \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\left(\frac{1}{\lambda_1}\phi_v(\xi)+i\lambda_{01}\xi\gamma_{\lambda_1}(\xi;\alpha_0)\right)d\xi \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\left(i\xi\overline{w}_{01}+i\lambda_{01}\xi\gamma_{\lambda_1}(\xi;\alpha_0)\right)d\xi, \\
\frac{\partial}{\partial\lambda_2} f_{C|\overline{W};\alpha_0} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\left(i\xi\overline{w}_{02}\phi_v(\xi)+i\lambda_{01}\xi\gamma_{\lambda_2}(\xi;\alpha_0)\right)d\xi.
\end{aligned}
$$

The gradient of the log likelihood in Eq. (B.4) is therefore expressed in terms of the population panel data model $f_{Y|X,C}(y|x,c;\theta_0)$ and the density of observables $f_{Y|X,\overline{W}}(y|x,\overline{w})$. The negative definiteness of the information matrix in Assumption 2.6 is equivalent to the positive definiteness of the outer product of the gradient of the log likelihood. The property of the positive definiteness can be imposed by choosing the outer product $\mathrm{E}\left[\frac{\partial}{\partial\alpha}\log f(Y|X,\overline{W};\alpha_0)\cdot\frac{\partial}{\partial\alpha}\log f(Y|X,\overline{W};\alpha_0)'\Big|X=x,\overline{W}=\overline{w}\right]$ as a strictly diagonally dominant matrix. To provide a more transparent condition, we define the vector of the derivatives of the average likelihood as

$$
\text{(B.16)} \qquad Df(y|x,\overline{w};\alpha_0) = \begin{pmatrix} \int_{\mathscr{C}} \frac{\partial}{\partial\theta_1} f_{Y|X,C}(y|x,c;\theta_0) f_{C|\overline{W}}(c|\overline{w};\alpha_0)dc \\ \int_{\mathscr{C}} f_{Y|X,C}(y|x,c;\theta_0)\frac{\partial}{\partial\theta_1} f_{C|\overline{W}}(c|\overline{w};\alpha_0)dc \\ \int_{\mathscr{C}} f_{Y|X,C}(y|x,c;\theta_0)\frac{\partial}{\partial\lambda_1} f_{C|\overline{W}}(c|\overline{w};\alpha_0)dc \\ \int_{\mathscr{C}} f_{Y|X,C}(y|x,c;\theta_0)\frac{\partial}{\partial\lambda_2} f_{C|\overline{W}}(c|\overline{w};\alpha_0)dc \end{pmatrix}.
$$

The detailed formulae of $\frac{\partial}{\partial\theta_1} f_{C|\overline{W}}(c|\overline{w};\alpha_0)$, $\frac{\partial}{\partial\lambda_1} f_{C|\overline{W}}(c|\overline{w};\alpha_0)$, and $\frac{\partial}{\partial\lambda_2} f_{C|\overline{W}}(c|\overline{w};\alpha_0)$ are in Eqs. (B.11)-(B.13) respectively.

Assumption B.1. *(Strictly Diagonally Dominant)*

Assume that every element of $Df(y|x,\overline{w};\alpha_0)$ is nonzero and the outer product of the derivatives *of the average likelihood*

$$E\Big[Df(y|x,\overline{w};\alpha_0)\cdot Df(y|x,\overline{w};\alpha_0)'\Big|X=x,\overline{W}=\overline{w}\Big] \tag{B.17}$$

is strictly diagonally dominant.[1]

Assumption B.1 implies that the expectation of the squared term of each derivative of the average likelihood is larger than the sum of the expectation of the magnitudes of the products of the derivative and other derivatives of the average likelihood. It is straightforward to verify that Assumption B.1 implies that every element of the gradient of the log likelihood $\frac{\partial}{\partial\alpha}\log f(Y|X,\overline{W};\alpha_0)$ is nonzero and its outer product is strictly diagonally dominant. Because every diagonal element of the matrix $\mathrm{E}\Big[\frac{\partial}{\partial\alpha}\log f(Y|X,\overline{W};\alpha_0)\cdot\frac{\partial}{\partial\alpha}\log f(Y|X,\overline{W};\alpha_0)'\Big|X=x,\overline{W}=\overline{w}\Big]$ is positive, symmetric, and strictly diagonally dominant, the matrix is positive definite.[2]

B.2. Primitive Conditions for the Binary Choice Model

To better elucidate these sufficient conditions, we work through the binary choice model in Eq. (6) by assuming that ε_t is symmetrically distributed, and $m(X_t,C;\theta)=\theta X_t+C$. It follows that

$$f_{Y_t|X_t,C}(y_t|x_t,c;\theta)=\big(1-F_{\varepsilon_t}(\theta X_t+C)\big)^{1-y_t}F_{\varepsilon_t}(\theta X_t+C)^{y_t}.$$

Denote $\vec{1}=(1_1,\dots,1_T)$, and suppose that the conditional density function $f(x,\overline{w}_{-1}|y=\vec{1})\neq 0$. We choose a weighting function concentrating at $y=\vec{1}$, i.e., $\Omega(y=\vec{1},x,\overline{w}_{-1})=f(x,\overline{w}_{-1}|y=\vec{1})$ for all $(x,\overline{w}_{-1})$ and zero elsewhere. It follows that $\Omega(y=\vec{1},x)=f(x|y=\vec{1})$ for all $\overline{w}_{-1}$ and zero elsewhere.

The denominator term of $\phi_{v;\alpha}(\xi)$ in Eq. (B.2) becomes

$$\begin{aligned}\int_{\mathscr{Y}\times\mathscr{X}\times\mathscr{C}}e^{-i\xi c}f_{Y|X,C}(y|x,c;\theta)\Omega(y,x)dydxdc&=\int_{\mathscr{X}\times\mathscr{C}}e^{-i\xi c}\prod_{t=1}^{T}F_{\varepsilon_t}(\theta x_t+c)f(x|y=\vec{1})dxdc\\&=\mathrm{E}\Big[\int_{\mathscr{C}}e^{-i\xi c}\prod_{t=1}^{T}F_{\varepsilon_t}(\theta X_t+c)dc\Big|Y=\vec{1}\Big].\end{aligned}$$

[1]A square matrix $A=[a_{ij}]$ is strictly diagonally dominant if $|a_{ii}|>\sum_{j\neq i}|a_{ij}|$ for all i.
[2]The result can be found as Corollary 6.1.10. in Horn and Johnson (1985).

Similarly, the numerator term of $\phi_{v;\alpha}(\xi)$ becomes

$$
\begin{aligned}
&-\lambda_1 \int_{\mathscr{Y}\times\mathscr{X}\times\overline{\mathscr{W}}} e^{-i\xi\sum_{k=1}\lambda_k\overline{w}_k} f_{Y|X,\overline{W}}(y|x,\overline{w})\Omega(y,x,\overline{w}_{-1})dydxd\overline{w} \\
=&-\lambda_1 \int_{\mathscr{X}\times\overline{\mathscr{W}}_{-1}} \left(\int_{\overline{\mathscr{W}}_1} e^{-i\xi\sum_{k=1}\lambda_k\overline{w}_k} f_{Y|X,\overline{W}}(y=\vec{1}|x,\overline{w}_1,\overline{w}_{-1})d\overline{w}_1\right) f(x,\overline{w}_{-1}|y=\vec{1})dxd\overline{w}_{-1} \\
=&-\lambda_1 \mathrm{E}\Big[\int_{\overline{\mathscr{W}}_1} e^{-i\xi\left(\lambda_1\overline{w}_1+\sum_{k=2}\lambda_k\overline{W}_k\right)} f_{Y|X,\overline{W}}(Y|X,\overline{w}_1,\overline{W}_{-1})d\overline{w}_1\Big|Y=\vec{1}\Big].
\end{aligned}
$$

Combining the results yields

$$
\phi_{v;\alpha}(\xi) = \frac{-\lambda_1 \mathrm{E}\Big[\int_{\overline{\mathscr{W}}_1} e^{-i\xi\left(\lambda_1\overline{w}_1+\sum_{k=2}\lambda_k\overline{W}_k\right)} f_{Y|X,\overline{W}}(Y|X,\overline{w}_1,\overline{W}_{-1})d\overline{w}_1\Big|Y=\vec{1}\Big]}{\mathrm{E}\Big[\int_{\mathscr{C}} e^{-i\xi c}\prod_{t=1}^T F_{\varepsilon_t}(\theta X_t+c)dc\Big|Y=\vec{1}\Big]} \equiv \frac{-\lambda_1 D(\xi;\lambda)}{M(\xi;\theta)},
$$

where $D(\xi;\lambda)$ is related to the population density and $M(\xi;\theta)$ is related to the panel data model. Assume that the support $\mathscr{C}$ is compact and the domain of C is a real line. Thus, we have

$$
c_\alpha(\overline{w}) = \phi_{v;\alpha}(0) = \frac{-\lambda_1 \mathrm{E}\Big[\int_{\overline{\mathscr{W}}_1} f_{Y|X,\overline{W}}(Y|X,\overline{w}_1,\overline{W}_{-1})d\overline{w}_1\Big|Y=\vec{1}\Big]}{\mathrm{E}\Big[\int_{\mathscr{C}} \prod_{t=1}^T F_{\varepsilon_t}(\theta X_t+c)dc\Big|Y=\vec{1}\Big]} \equiv \frac{-\lambda_1 D(0)}{M(0;\theta)}.
$$

Therefore,

$$
f_{C|\overline{W}}(c|\overline{w};\alpha) = \frac{1}{2\pi c_\alpha(\overline{w})}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\phi_{v;\alpha}(\xi)d\xi = \frac{M(0;\theta)}{2\pi D(0)}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda)}\frac{D(\xi;\lambda)}{M(\xi;\theta)}d\xi.
$$

We consider $\alpha=(\theta_1,\lambda_1,\lambda_2)$ and denote the derivatives of $D(\xi;\lambda)$ and $M(\xi;\theta)$ with respect to λ and θ, respectively, as

$$
\begin{aligned}
\frac{\partial M(\xi;\theta)}{\partial\theta_1} &= \mathrm{E}\Big[\int_{\mathscr{C}} e^{-i\xi c}\sum_{s=1}^T\prod_{t=1,t\neq s}^T F_{\varepsilon_t}(\theta X_t+c)f_{\varepsilon_s}(\theta X_s+c)X_s dc\Big|Y=\vec{1}\Big], \\
\frac{\partial D(\xi;\lambda)}{\partial\lambda_1} &= -\mathrm{E}\Big[\int_{\overline{\mathscr{W}}_1} i\xi\overline{w}_1 e^{-i\xi\left(\lambda_1\overline{w}_1+\sum_{k=2}\lambda_k\overline{W}_k\right)} f_{Y|X,\overline{W}}(Y|X,\overline{w}_1,\overline{W}_{-1})d\overline{w}_1\Big|Y=\vec{1}\Big], \\
\frac{\partial D(\xi;\lambda)}{\partial\lambda_2} &= -\mathrm{E}\Big[\int_{\overline{\mathscr{W}}_1} i\xi\overline{W}_2 e^{-i\xi\left(\lambda_1\overline{w}_1+\sum_{k=2}\lambda_k\overline{W}_k\right)} f_{Y|X,\overline{W}}(Y|X,\overline{w}_1,\overline{W}_{-1})d\overline{w}_1\Big|Y=\vec{1}\Big].
\end{aligned}
$$

The derivatives of $f_{C|\overline{W}}(c|\overline{w};\alpha)$ with respect to θ_1, λ_1, and λ_2 at the true value α_0 can then be stated as

$$
\frac{\partial}{\partial\theta_1}f_{C|\overline{W};\alpha_0} = \frac{\frac{\partial M(0;\theta_0)}{\partial\theta_1}}{2\pi D(0)}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\frac{D(\xi;\lambda_0)}{M(\xi;\theta_0)}d\xi - \frac{M(0;\theta_0)}{2\pi D(0)}\int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)}\frac{D(\xi;\lambda_0)}{M(\xi;\theta_0)^2}\frac{\partial M(\xi;\theta_0)}{\partial\theta_1}d\xi,
$$

$$\frac{\partial}{\partial\lambda_1} f_{C|\overline{W};\alpha_0} = \frac{M(0;\theta_0)}{2\pi D(0)} \int_{-\infty}^{\infty} i\xi\overline{w}_1 e^{-i\xi(c-\overline{w}\lambda_0)} \frac{D(\xi;\lambda_0)}{M(\xi;\theta_0)} d\xi + \frac{M(0;\theta_0)}{2\pi D(0)} \int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)} \frac{\frac{\partial D(\xi;\lambda_0)}{\partial\lambda_1}}{M(\xi;\theta_0)} d\xi,$$
$$\frac{\partial}{\partial\lambda_2} f_{C|\overline{W};\alpha_0} = \frac{M(0;\theta_0)}{2\pi D(0)} \int_{-\infty}^{\infty} i\xi\overline{w}_2 e^{-i\xi(c-\overline{w}\lambda_0)} \frac{D(\xi;\lambda_0)}{M(\xi;\theta_0)} d\xi + \frac{M(0;\theta_0)}{2\pi D(0)} \int_{-\infty}^{\infty} e^{-i\xi(c-\overline{w}\lambda_0)} \frac{\frac{\partial D(\xi;\lambda_0)}{\partial\lambda_2}}{M(\xi;\theta_0)} d\xi.$$

Let f_{ε_t} be the PDF of ε_t. In the binary choice model, we have

$$f_{Y|X,C}(y|x,c;\theta_0) = \prod_{t=1}^{T} \left(1 - F_{\varepsilon_t}(\theta_0 X_t + C)\right)^{1-y_t} F_{\varepsilon_t}(\theta_0 X_t + C)^{y_t},$$
$$\frac{\partial}{\partial\theta_1} f_{Y|X,C}(y|x,c;\theta_0) = \sum_{s=1}^{T} \prod_{t=1, t\neq s}^{T} \frac{\partial}{\partial\theta_1} f_{Y_t|X_t,C}(y_t|x_t,c;\theta_0)$$
$$= \sum_{s=1}^{T} \prod_{t=1, t\neq s}^{T} \frac{f_{\varepsilon_t}(\theta_0 X_t + C) X_t \left(y_t - F_{\varepsilon_t}(\theta_0 X_t + C)\right)}{\left(1 - F_{\varepsilon_t}(\theta_0 X_t + C)\right)^{y_t} F_{\varepsilon_t}(\theta_0 X_t + C)^{1-y_t}}.$$

With the detailed formulae, we can construct the vector of the derivatives of the average likelihood $Df(y|x,\overline{w};\alpha_0)$ defined in Eq. (B.16) for the binary choice model. Assumption B.1 requires that the outer product of $Df(y|x,\overline{w};\alpha_0)$ is strictly diagonally dominant. Thus, Assumption B.1 is intuitive and provides a sufficient condition for Assumption 2.6.

C. Sieve Maximum Likelihood Estimators

C.1. Sufficient Conditions and Proof for Theorem 3.1

We provide sufficient conditions and proof for Theorem 3.1. We first present conditions such that the estimator is consistent. Let $\widehat{g} \equiv (\widehat{\alpha}, \widehat{f}_1)$.

Assumption C.1. *For any $\epsilon > 0$, there exists a nonincreasing positive sequence $c_N(\epsilon)$ such that for all $N \geq 1$ we have*

$$E[\psi(Z,g_o)] - \sup_{g\in\mathscr{G}_N:\|g-g_o\|_{\mathscr{G}}\geq\epsilon} E[\psi(Z,g)] \geq c_N(\epsilon), \tag{C.1}$$

and $\liminf c_N(\epsilon) > 0$.

Define $\mu_N(f) = N^{-1}\sum_{i=1}^{N}[f(Z_i) - E[f(Z_i)]]$, which denotes the empirical process indexed by f.

Assumption C.2. *Assume that (i) for all N, there exists some $g_N \in \mathcal{G}_N$ such that*

$$|E[\psi(Z, g_N)] - E[\psi(Z, g_o)]| = o(1); \tag{C.2}$$

(ii) $\sup_{g \in \mathcal{G}_N} |\mu_N(\psi(Z, g))| = o_P(1)$.

Theorem C.1. *If Assumptions C.1 and C.2 hold, then* $\|\widehat{g} - g_o\| = o_p(1)$.

Let C denote a positive finite number. Define $\mathcal{N}_{g,C} = \{g \in \mathcal{G} : \|g - g_o\| \leq C\}$ and $\mathcal{N}_{g,N,C} = \{g \in \mathcal{G}_N : \|g - g_o\| \leq C\}$. By Theorem C.1, we have $\widehat{g} \in \mathcal{N}_{g,N,C}$ with probability approaching 1. Next, we provide conditions to derive the convergence rate of $\widehat{g}$.

Assumption C.3. *There exist some finite, positive, and nonincreasing sequences δ_{1N}, δ_{2N}, and δ_{3N} that are all $o(1)$ such that (i)*

$$\sup_{g \in \mathcal{N}_{g,N,C}} \left| \mu_N(\psi(Z, g) - \psi(Z, g_o)) \right| = O_p(\delta_{1N}^2); \tag{C.3}$$

(ii) for all N large enough and for any $\delta > 0$ small enough,

$$E\Big[\sup_{g \in \mathcal{N}_{g,N,C} : \|g - g_o\|_{\mathcal{G}} \leq \delta} \left| \mu_N(\psi(Z, g) - \psi(Z, g_o)) \right| \Big] \leq \frac{c_1 \phi_N(\delta)}{\sqrt{N}}, \tag{C.4}$$

where c_1 is some positive number and $\phi_N(\cdot)$ is some function such that $\delta^{\nu} \psi_N(\delta)$ is a decreasing function for some $\nu \in (0, 2)$;

(iii) $\delta_{2N}^{-2} \phi(\delta_{2N}) \leq c_2 \sqrt{N}$ for some finite positive $c_2 > 0$; (iv) $\|g_N - g_o\|_{\mathcal{G}} = O(\delta_{3N})$.

Theorem C.2. *Suppose that Assumptions C.1, C.2 and C.3 hold. Then, we have $\|\widehat{g} - g_o\| = O_p(\epsilon_N^*)$, where $\epsilon_N^* = \max\{\delta_{1N}, \delta_{2N}, \delta_{3N}\} = o(1)$.*

Assumption C.4. *Assume that there exists $c_\psi < \infty$ such that $\|v\|_\psi \leq c_\psi \|v\|_{\mathcal{G}}$.*

Let $\epsilon_N = \epsilon_N^* \cdot \delta_N$ with $\delta_N \to \infty$ such that ϵ_N remains $o_p(1)$. Let $\mathcal{G}_N \equiv \mathcal{A} \times \mathcal{F}_{1N}$. Let $\mathcal{N}_g = \{(g \in \mathcal{G} : \|g - g_o\|_{\mathcal{G}} \leq \epsilon_N)\}$ and $\mathcal{N}_{g,N} = \mathcal{N}_g \cap \mathcal{G}_N$. It is true that $\widehat{g} \in \mathcal{N}_{g,N}$ with probability approaching one. Let Π_N denote the projection of g on $\mathcal{G}_N$ under the norm $\|\cdot\|_\psi$ and let $g_{o,N} = \Pi_N g_o$. Assumption C.4 implies that $g_{o,N} \in \mathcal{N}_{g,N}$.

Π_N is also the projection of v on $\mathcal{V}_N$, which is the closed linear span of $\mathcal{G}_N - g_{o,N}$. Let

$v_N^* \equiv (\Pi_N \cdot v_1^*, \dots, \Pi_N \cdot v_{d_\alpha}^*) \in \mathcal{V}_N^{d_\alpha}$. By construction, we have for all $v \in \mathcal{V}_N$

$$d\rho_\gamma(g_0)[v] = \langle \gamma' v_N^*, v \rangle_\psi.$$

Let κ_N denote a sequence of positive numbers and $\kappa_N = o(N^{-1/2})$. For any g, let $g_\lambda^* = g \pm \kappa_N \cdot \gamma' v_N^*$.

Assumption C.5. *Assume that (i)*

$$\sup_{g \in \mathcal{N}_{g,N}} \left| \mu_N \left(\psi(Z, g^*) - \psi(Z, g) - \Delta_\psi(Z, g)[\pm \kappa_N \cdot \gamma' v_N^*] \right) \right| = O_p(\kappa_N^2), \tag{C.5}$$

$$\sup_{g \in \mathcal{N}_{g,N}} \left| \mu_N \left(\Delta_\psi(Z, g)[\gamma' v_N^*] - \Delta_\psi(Z, g_o)[\gamma' v_N^*] \right) \right| = O_p(\kappa_N); \tag{C.6}$$

(ii) let $K_\psi(g) \equiv E[\psi(Z, g) - \psi(Z, g_o)]$*; then,*

$$\sup_{g \in \mathcal{N}_{g,N}} \left| K_\psi(g) - K_\psi(g_o) - \frac{\|g^* - g_o\|_\psi^2 - \|g - g_o\|_\psi^2}{2} \right| = O_p(\kappa_N^2); \tag{C.7}$$

(iii) $\kappa_N / \epsilon_N^* = o(1)$ *and (iv)* $\|v_j^*\|_\psi < \infty$ *for all* $j = 1, \dots, d_\alpha$.

Theorem C.3. *Suppose that Assumptions C.1, C.2, C.3, C.4, and C.5 hold. Then*

$$\lambda \sqrt{N} (\widehat{\alpha} - \alpha_0) \xrightarrow{d} \mathcal{N}(0, \sigma_\gamma^2) = \mathcal{N}(0, \gamma' \Omega \gamma).$$

The proofs of Theorem C.1, C.2, and C.3 follow the same arguments as Chen, Liao, and Sun (2014) and Hahn, Liao, and Ridder (2018), so we omit the details. By the Cramér-Wold theorem, Theorem C.3 implies Theorem 3.1.

C.2. A Consistent Estimator for Ω

To make inference, one would need a consistent estimator for Ω when the sieve MLE is implemented based on a Hermite polynomial series. The estimator is a sample analog. To be specific, for $j = 1,\ldots,d_\alpha$, define the empirical Riesz representer $\widehat{v}_j^*$ as

$$d\rho_j(\widehat{g})[v] = \langle \widehat{v}_j^*, v\rangle_{N,\psi}, \text{ where}$$
$$\langle v_1, v_2\rangle_{N,\psi} = \frac{1}{N}\sum_{i=1}^{N} -\frac{d\Delta_\psi(Z_i,\widehat{g}+\tau v_1)[v_2]}{d\tau}\Big|_{\tau=0}.$$

A consistent estimator for $\widehat{\Omega}$ is given as

$$\widehat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\big(\Delta_\psi(Z_i,\widehat{g})[\widehat{v}_1^*],\ldots,\Delta_\psi(Z_i,\widehat{g})[\widehat{v}_{d_\alpha}^*]\big)'\big(\Delta_\psi(Z_i,\widehat{g})[\widehat{v}_1^*],\ldots,\Delta_\psi(Z_i,\widehat{g})[\widehat{v}_{d_\alpha}^*]\big). \tag{C.8}$$

The consistency of $\widehat{\Omega}$ can be proven by Theorem 4.1 of Hahn, Liao, and Ridder (2018). More important, even if (C.8) appears complicated, one can apply Theorem 6.1 and Remark 6.1 of Hahn, Liao, and Ridder (2018) to use the variance estimator of the parametric MLE model when the sieve space is generated by a finite number of basis functions such as a Hermite polynomial series. Specifically, suppose that f_1 is approximated by $f_1(v;\beta)$ where β has $K(N)$ dimensions. We then write

$$f_{Y|X,\overline{W}}(y|x,\overline{w};\alpha,\beta) = \int_{\mathscr{C}} f_{Y|X,C}(y|x,c;\theta) f_1(c-\overline{w}\lambda;\beta))dc,$$

and the sieve estimator is given by

$$(\widehat{\alpha},\widehat{\beta}) \equiv \arg\max_{\alpha,\beta} \frac{1}{N}\sum_{i=1}^{N}\log\big(f_{Y|X,\overline{W}}(Y_i|X_i,\overline{W}_i;\theta,\lambda,\beta)\big), \text{ and}$$
$$\widehat{f}_1(v) = f_1(v;\widehat{\beta}).$$

Let $\widehat{s}_i = \nabla\log\big(f_{Y|X,\overline{W}}(Y_i|X_i,\overline{W}_i;\alpha,\beta)\big)$ and $\widehat{H}_i = \nabla^2\log\big(f_{Y|X,\overline{W}}(Y_i|X_i,\overline{W}_i;\alpha,\beta)\big)$ be the gradient and the Hessian of $\log\big(f_{Y|X,\overline{W}}(Y_i|X_i,\overline{W}_i;\alpha,\beta)\big)$, respectively, evaluated at $(\widehat{\alpha},\widehat{\beta})$. Let $\widehat{V}$ be the variance matrix estimator of the MLE estimator that is given by

$$\widehat{V} = \Big(\frac{1}{N}\sum_{i=1}^{N} -\widehat{H}_i\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\widehat{s}_i\widehat{s}_i'\Big)\Big(\frac{1}{N}\sum_{i=1}^{N} -\widehat{H}_i\Big)^{-1}.$$

The expression for $\widehat{\Omega}$ will then be the upper-left $d_\alpha \times d_\alpha$ matrix of $\widehat{V}$. Equivalently, let $\widehat{s}_{\alpha,i}$ be the first d_α row of the vector of $\widehat{\mathscr{H}}^{-1}\cdot\widehat{s}_i$ with $\widehat{\mathscr{H}} = N^{-1}\sum_{i=1}^{N} -\widehat{H}_i$, and $\widehat{\Omega}$ is given by $N^{-1}\sum_{i=1}^{N}\widehat{s}_{\alpha,i}\widehat{s}_{\alpha,i}'$.

D. A Consistent Estimator for Ω_{di}

In this section, we provide a consistent estimator for Ω_{di}. Recall that the parameter MLE estimator given in Section 4.1 is

$$(\widehat{\alpha}_{pa}, \widehat{\tau}_{pa}) \equiv \arg\max_{\alpha \in \mathscr{A}, \tau \in \mathscr{T}} \frac{1}{N} \sum_{i=1}^{N} f_{pa}(Y_i | X_i, \overline{W}_i; \alpha, \tau),$$

$$f_{pa}(y|x, \overline{w}; \alpha, \tau) = \int_{\mathscr{C}} f_{Y|X,C}(y|x, c; \theta) f_V(c - \overline{W}\lambda; \tau) dc.$$

Let $\widehat{s}_{pa,i} = \nabla \log\left(f_{pa}(Y_i|X_i, \overline{W}_i; \alpha, \beta)\right)$ and $\widehat{H}_{pa,i} = \nabla^2 \log\left(f_{pa}(Y_i|X_i, \overline{W}_i; \alpha, \beta)\right)$ be the gradient and the Hessian of $\log\left(f_{pa}(Y_i|X_i, \overline{W}_i; \alpha, \beta)\right)$, respectively, evaluated at $(\widehat{\alpha}_{pa}, \widehat{\tau}_{pa})$. Let $\widehat{V}_{pa}$ be the variance matrix estimator of the MLE estimator that is given by

$$\widehat{V}_{pa} = \left(\frac{1}{N} \sum_{i=1}^{N} -\widehat{H}_{pa,i}\right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{s}_{pa,i} \widehat{s}'_{pa,i}\right) \left(\frac{1}{N} \sum_{i=1}^{N} -\widehat{H}_{pa,i}\right)^{-1}.$$

The expression for $\widehat{\Omega}_{pa}$ will then be the upper-left $d_\alpha \times d_\alpha$ matrix of $\widehat{V}_{pa}$. Equivalently, let $\widehat{s}_{pa,\alpha,i}$ be the first d_α row of the vector of $\widehat{\mathscr{H}}_{pa}^{-1} \cdot \widehat{s}_{pa,i}$ with $\widehat{\mathscr{H}}_{pa} = N^{-1} \sum_{i=1}^{N} -\widehat{H}_{pa,i}$, and $\widehat{\Omega}_{pa}$ is given by $N^{-1} \sum_{i=1}^{N} \widehat{s}_{pa,\alpha,i} \widehat{s}'_{pa,\alpha,i}$.

Given these results, a consistent estimator for Ω_{di} is given by

$$\widehat{\Omega}_{di} = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{s}_{pa,\alpha,i} - \widehat{s}_{\alpha,i}\right) \cdot \left(\widehat{s}_{pa,\alpha,i} - \widehat{s}_{\alpha,i}\right)'.$$

E. Monte Carlo Simulation

In this section, we present simulation results to illustrate the finite-sample performance of the proposed sieve ML estimation procedure of a panel data model in Section 3. We consider both panel data probit models and panel data Poisson models.

E.1. Panel Data Probit Models

We demonstrate our simulation results through static and dynamic settings. The *static* data generating process (DGP) is defined as follows:

$$
\begin{aligned}
&Y_t = \mathbf{1}(\theta X_t + C + \varepsilon_t \geq 0), \quad \text{for } t = 1,2, \\
&C = \lambda \overline{W} + V, \quad \overline{W} = \frac{1}{2}\sum_{t=1}^{2} X_t, \\
&X_2 = 0.5X_1 + \xi, \quad X_1 \sim U(0,2), \quad \xi \sim N(0,1), \\
&(\varepsilon_1, \varepsilon_2) \sim N(0, I_2),
\end{aligned}
$$

where I_2 is the 2×2 identity matrix and $(\theta, \lambda) = (-0.5, 0.5)$. For a random variable Q, we denote the corresponding truncated random variable over interval $[a,b]$ as $Trun(Q,[a,b])$.[3] Let μ_ω be the mean of ω. Three specifications of V are considered:

(E.1) DGP I: $V \sim Trun(N(0,1),[-1,1])$,

(E.2) DGP II: $V = \omega - \mu_\omega$ with $\omega \sim Trun\big(H,[-2,2]\big)$ and $\ln H = N(0,5)$,

(E.3) DGP III: $V = \omega - \mu_\omega$ with $\sqrt{\omega} \sim Trun(Rayleigh(1),[-2,15])$.

The unobserved heterogeneities in all the simulation designs have bounded supports, so Assumption 2.4(ii) is satisfied in all cases. We consider sample sizes of 500 and 1,000, and for each case, we consider 1000 simulation replications. For comparison, we also consider two other estimators. The first is an infeasible estimator that treats V as known. The second is the conventional random effects estimator, which specifies the unobserved heterogeneity to be normally distributed. The simulation results for parameters and APE are presented in Tables 1–2 and 3–4, respectively.

The estimation results of the parameters in DGP I show a small bias in all three estimators. In this case, the normal specification in the conventional random effects estimator is close to the true distribution of the data, and the estimation does not suffer from the misspecification of the estimator. The proposed sieve ML estimator exhibits small degrees of bias in DGPs II &

[3] $Trun(Q,[a,b])$ is a random variable generated by $F_Q^{-1}(u \cdot (F_Q(b) - F_Q(a)) + F_Q(a))$, where F_Q is the CDF of the Q random variable, F_Q^{-1} is the inverse of F_Q, and u is a uniform random variable on $[0,1]$.

III, but the conventional random effects estimator exhibits conspicuous bias in θ, λ, and σ for all sample sizes.

The simulation results overall show that the proposed sieve ML estimator works well in simulation designs. As expected, the infeasible estimator outperforms the proposed estimator in RMSE. The conventional estimator performs well at estimating θ and λ in DGP I, but causes bias in DGPs II & III. The estimation results for APEs in Tables 3–4 present a similar pattern. While the infeasible estimator and the proposed sieve ML estimator perform well in all simulations, the conventional estimator performs well only in DGP I.

We also consider the Hausman-type test proposed in Section 4.1 for the normality assumption of V. The results are summarized in Table 9. For DGP I, the rejection rates are 0.051 and 0.033, which are close to the nominal size of 5% given that the normality assumption holds.[4] For DGPs II and III, the rejection rates are much higher than the nominal size of 5% and increase with sample size, indicating that our test is consistent when the normality assumption is violated.

The plots of the simulation results for the remaining error in the CRE specification f_V of DGPs I, II, and III in the static probit models appear in Figures 1–3. While the distribution of V in DGP I is symmetric, the distributions of V in DGPs II and III are nonsymmetric. These plots show that the sieve ML estimators perform poorly in DGPs II and III simulation designs because not only do the average results miss the shape of the function, but the confidence bands also do not offer a view of accuracy. This implies that to obtain a better nonparametric estimate of f_V, a sample size larger than 1000 may be needed. However, the visual displays of the average estimator capture the basic features of the true f_V such as the location of the mode and symmetry or nonsymmetry. The black dashed lines are constructed by 1000 sieve ML estimated curves from 1000 replications in the 10th and 90th percentiles.

The simulation design for dynamic panel data probit models is close to the static panel data probit models. We define the DGP for *dynamic* models as

$$Y_t = \mathbf{1}\left(\gamma Y_{t-1} + \theta X_t + C + \varepsilon_t \geq 0\right), \quad \text{for } t = 1, \ldots, 7,$$

[4] We use an empirical covariance matrix, which is the average of the 1000 simulated estimators, to conduct the Hausman test in the simulations.

where $(\gamma, \theta, \lambda) = (-0.5, 0.5, 0.5)$ and DGPs for X_t and C are the same as those in the static models.

Tables 6–7 and 8–9 present the estimation results for parameters and the magnitudes of state dependence. We reach the same conclusion as the estimation results of the static models. While the proposed sieve ML estimator performs well in all DGPs, the conventional random effects estimator cannot deliver a consistent estimation for the parameters γ, λ, and σ in DGPs II & III. The simulation results for the Hausman-type test are also similar to the static case.

E.2. Dynamic Panel Data Poisson Models

We consider a data generating process for dynamic panel data Poisson models as follows:

$$m(y_{t-1}, x_t, c; \theta) = \gamma Y_{t-1} + \theta X_t + C \text{ with } \quad y_t = 0, 1, ..., \tag{E.4}$$

where $m(y_{t-1}, x_t, c; \theta) = \mathrm{E}(Y_t | y_{t-1}, x_t, c)$ and $(\gamma, \theta, \lambda) = (-0.5, 0.5)$. While the DGP for X_t is $X_1 \sim U(0,2)$, $X_t \sim N(0,1)$ for $t > 1$, the DGP for C is $C = 0.5\overline{W} + V$, $\overline{W} = \frac{1}{T}\sum_{t=1}^{T} X_t + z$ with $z \sim N(0,1)$. In this case, we consider unbounded support for V and three specifications:

(E.5) DGP IV: $V \sim N(0,1)$,

(E.6) DGP V: $V = \omega - \mu_\omega$ with $\omega \sim$ Student's t with 100 degrees of freedom,

(E.7) DGP VI: $V = \omega - \mu_\omega$ with $\omega = \omega_1 + \omega_2, \omega_1 \sim N(0,1), \omega_2 \sim Trun(Rayleigh(1), [-2, 12])$.

The finite-sample performance is evaluated over two different time dimensions: $T = 2$ and $T = 4$.

While Tables 10–13 report the estimated results for panel data for two periods, Tables 14–17 report the estimated results for panel data for four periods. From the estimation results, we find that in the designs of panel data for two periods, the proposed estimator outperforms the conventional estimator in the parameter estimation. However, for longer periods, such as four periods, the conventional estimator performs well and is close to the proposed estimator. Table 18 reports the Hausman-type test proposed for the normality assumption of V. In the simulated data for two periods, the rejection rates of the proposed Hausman-type test are much higher than the nominal size of 5% and increase with sample size in all designs. This implies that the conventional estimator performs poorly even for the case in which V is normally distributed. However, in the simulated data for four periods with a sample size of 1000, the estimation results of the conventional estimator are close to those of the proposed estimator in DGP I, and the rejection rate is 0.080. This suggests under an unbounded assumption of V that the conventional estimator may perform better for panel data with a longer period and with a larger sample size.

References

CHEN, X., Z. LIAO, AND Y. SUN (2014): "Sieve Inference on Possibly Misspecified Semi-nonparametric Time Series Models," *Journal of Econometrics*, 178(1), 639–658.

HAHN, J., Z. LIAO, AND G. RIDDER (2018): "Nonparametric Two-step Sieve M Estimation and Inference," *Econometric Theory*, 34(6), 1281–1324.

HORN, R., AND C. JOHNSON (1985): *Matrix Analysis*. Cambridge University Press.

Table 1: Simulations of Static Probit Models

	Infeasible		Conventional			Sieve ML	
N=500	θ	λ	θ	λ	σ	θ	λ
True	-0.5	0.5	-0.5	0.5	1	-0.5	0.5
DGP I:							
Mean	-0.500	0.500	-0.502	0.505	0.920	-0.423	0.404
Std.dev.	-0.499	0.500	-0.500	0.503	0.919	-0.420	0.408
RMSE	0.049	0.057	0.065	0.079	0.141	0.165	0.143
DGP II:							
Mean	-0.499	0.496	-0.607	0.197	1.879	-0.419	0.403
Std.dev.	-0.497	0.496	-0.600	0.200	1.868	-0.415	0.416
RMSE	0.060	0.062	0.141	0.321	0.911	0.176	0.158
DGP III:							
Mean	-0.499	0.496	-0.594	0.228	1.768	-0.417	0.405
Std.dev.	-0.496	0.496	-0.590	0.229	1.755	-0.410	0.421
RMSE	0.057	0.061	0.129	0.292	0.797	0.184	0.163

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 2: Simulations of Static Probit Models

	Infeasible		Conventional			Sieve ML	
N=1000	θ	λ	θ	λ	σ	θ	λ
True	-0.5	0.5	-0.5	0.5	1	-0.5	0.5
DGP I:							
Mean	-0.501	0.500	-0.504	0.506	0.923	-0.489	0.558
Std.dev.	-0.502	0.500	-0.503	0.504	0.926	-0.491	0.552
RMSE	0.034	0.038	0.045	0.055	0.114	0.116	0.161
DGP II:							
Mean	-0.502	0.501	-0.608	0.206	1.866	-0.474	0.543
Std.dev.	-0.500	0.500	-0.606	0.207	1.853	-0.465	0.546
RMSE	0.041	0.044	0.128	0.304	0.883	0.114	0.111
DGP III:							
Mean	-0.501	0.501	-0.594	0.235	1.758	-0.469	0.545
Std.dev.	-0.500	0.499	-0.592	0.234	1.749	-0.462	0.551
RMSE	0.039	0.043	0.114	0.275	0.773	0.114	0.112

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 3: Simulation of APE($\bar{x}$) in Static Probit Models

N=500	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP I:			
Mean	-0.168	-0.143	-0.126
Std. dev.	0.001	0.017	0.042
RMSE	–	0.017	0.059
DGP II:			
Mean	-0.188	-0.113	-0.125
Std. dev.	0.001	0.017	0.045
RMSE	–	0.026	0.072
DGP III:			
Mean	-0.188	-0.116	-0.124
Std. dev.	0.001	0.016	0.046
RMSE	–	0.025	0.079

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 4: Simulation of APE($\bar{x}$) in Static Probit Models

N=1000	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP I:			
Mean	-0.168	-0.144	-0.141
Std. dev.	0.001	0.011	0.034
RMSE	–	0.011	0.043
DGP II:			
Mean	-0.188	-0.113	-0.136
Std. dev.	0.001	0.011	0.032
RMSE	–	0.024	0.055
DGP III:			
Mean	-0.188	-0.116	-0.135
Std. dev.	0.001	0.011	0.031
RMSE	–	0.022	0.062

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 1000 simulations. The notation DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

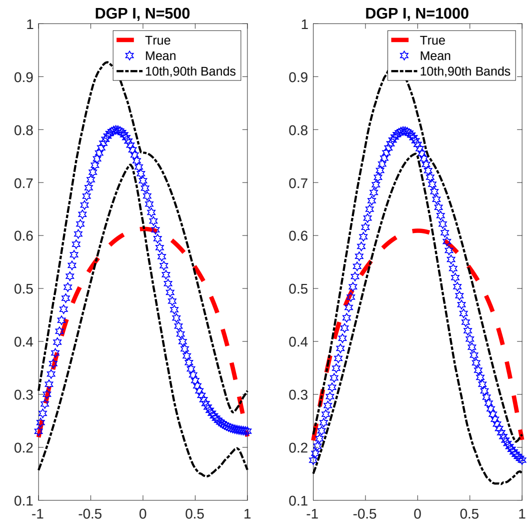

Figure 1: Average Estimates of f_V in the CRE Specification of DGP I in Eq. (E.1) in Section E.1

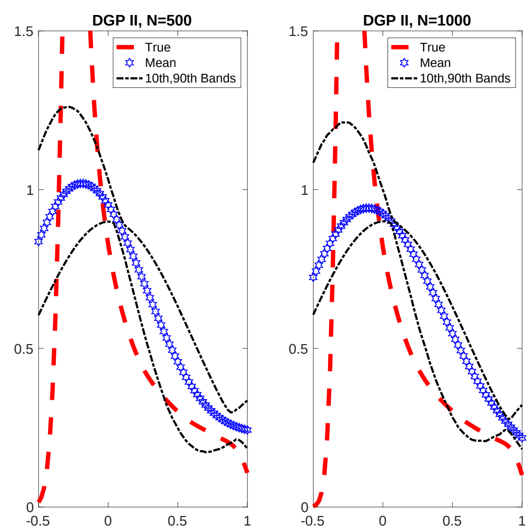

Figure 2: Average Estimates of f_V in the CRE Specification of DGP II in Eq. (E.2) in Section E.1

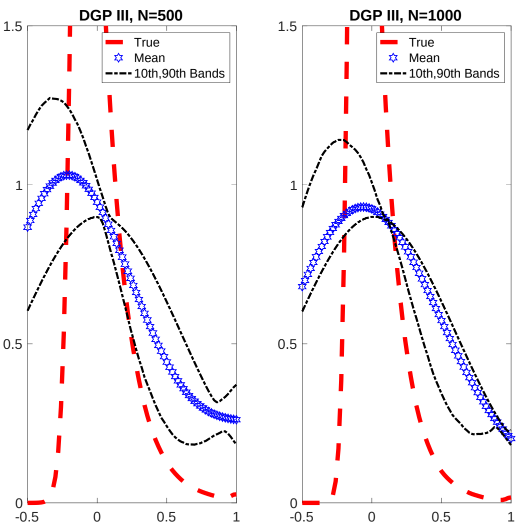

Figure 3: Average Estimates of f_V in the CRE Specification of DGP III in Eq. (E.3) in Section E.1

Table 5: Simulations of Dynamic Probit Models

	Infeasible			Conventional				Sieve ML		
N=500	γ	θ	λ	γ	θ	λ	σ	γ	θ	λ
True	-0.5	0.5	0.5	-0.5	0.5	0.5	1	-0.5	0.5	0.5
DGP I:										
Mean	-0.502	0.498	0.498	-0.382	0.378	0.198	0.301	-0.555	0.444	0.490
Std.dev.	-0.501	0.500	0.496	-0.379	0.376	0.198	0.334	-0.551	0.453	0.464
RMSE	0.034	0.066	0.057	0.125	0.138	0.307	0.722	0.158	0.156	0.191
DGP II:										
Mean	-0.503	0.497	0.500	-0.383	0.376	0.198	0.306	-0.553	0.449	0.501
Std.dev.	-0.501	0.498	0.500	-0.381	0.378	0.195	0.338	-0.545	0.459	0.460
RMSE	0.033	0.067	0.059	0.124	0.140	0.306	0.718	0.153	0.177	0.220
DGP III:										
Mean	-0.505	0.500	0.500	-0.372	0.352	0.155	0.317	-0.543	0.454	0.512
Std.dev.	-0.502	0.501	0.500	-0.370	0.349	0.154	0.351	-0.536	0.463	0.467
RMSE	0.034	0.068	0.062	0.135	0.163	0.349	0.710	0.160	0.172	0.232

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 6: Simulations of Dynamic Probit Models

	Infeasible			Conventional				Sieve ML		
N=1000	γ	θ	λ	γ	θ	λ	σ	γ	θ	λ
True	-0.5	0.5	0.5	-0.5	0.5	0.5	1	-0.5	0.5	0.5
DGP I:										
Mean	-0.501	0.498	0.498	-0.372	0.352	0.155	0.317	-0.510	0.500	0.512
Std.dev.	-0.500	0.499	0.500	-0.370	0.349	0.154	0.351	-0.513	0.495	0.513
RMSE	0.024	0.045	0.042	0.135	0.163	0.349	0.710	0.105	0.102	0.098
DGP II:										
Mean	-0.501	0.499	0.499	-0.380	0.381	0.198	0.328	-0.506	0.505	0.507
Std.dev.	-0.500	0.497	0.499	-0.379	0.382	0.198	0.354	-0.505	0.505	0.507
RMSE	0.023	0.045	0.043	0.123	0.127	0.304	0.687	0.104	0.103	0.103
DGP III:										
Mean	-0.501	0.497	0.497	-0.371	0.360	0.159	0.339	-0.508	0.503	0.508
Std.dev.	-0.500	0.499	0.495	-0.368	0.360	0.158	0.367	-0.506	0.504	0.508
RMSE	0.023	0.050	0.044	0.133	0.147	0.343	0.679	0.102	0.099	0.107

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 7: Simulations of State Dependence in the Dynamic Probit Models

N=500	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP I:			
Mean	-0.127	-0.138	-0.162
Std. dev.	0.003	0.013	0.042
RMSE	–	0.019	0.055
DGP II:			
Mean	-0.126	-0.138	-0.161
Std. dev.	0.003	0.013	0.041
RMSE	–	0.019	0.054
DGP III:			
Mean	-0.117	-0.135	-0.158
Std. dev.	0.003	0.013	0.044
RMSE	–	0.024	0.060

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 8: Simulations of State Dependence in the Dynamic Probit Models

N=1000	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP I:			
Mean	-0.127	-0.137	-0.141
Std. dev.	0.002	0.009	0.029
RMSE	–	0.016	0.033
DGP II:			
Mean	-0.126	-0.137	-0.140
Std. dev.	0.002	0.009	0.029
RMSE	–	0.016	0.032
DGP III:			
Mean	-0.117	-0.135	-0.141
Std. dev.	0.002	0.010	0.029
RMSE	–	0.022	0.037

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 9: Hausman-type Test for Normality: Empirical Size

	Static Probit Models		Dynamic Probit Models	
	N=500	N=1000	N=500	N=1000
DGP I:	0.051	0.033	0.039	0.062
DGP II:	0.145	0.977	0.420	0.598
DGP III:	0.139	0.957	0.368	0.545

Note: The p-values of 0.05 of Chi-distributions for static models and dynamic models are 5.991 and 7.815, respectively. Empirical size refers to the fraction of rejections when using these values as the critical values. The notations DGP I, DGP II, and DGP III represent the data generating processes in Eqs. (E.1), (E.2), and (E.3), respectively.

Table 10: Simulations of Two-Period Dynamic Poisson Models

	Infeasible			Conventional				Sieve ML		
N=500	γ	θ	λ	γ	θ	λ	σ	γ	θ	λ
True	−0.5	0.5	0.5	−0.5	0.5	0.5	1	−0.5	0.5	0.5
DGP IV:										
Mean	-0.502	0.498	0.498	-0.382	0.378	0.198	0.301	-0.555	0.444	0.490
Std.dev.	-0.501	0.500	0.496	-0.379	0.376	0.198	0.334	-0.551	0.453	0.464
RMSE	0.034	0.066	0.057	0.125	0.138	0.307	0.722	0.158	0.156	0.191
DGP V:										
Mean	-0.503	0.497	0.500	-0.383	0.376	0.198	0.306	-0.553	0.449	0.501
Std.dev.	-0.501	0.498	0.500	-0.381	0.378	0.195	0.338	-0.545	0.459	0.460
RMSE	0.033	0.067	0.059	0.124	0.140	0.306	0.718	0.153	0.177	0.220
DGP VI:										
Mean	-0.505	0.500	0.500	-0.372	0.352	0.155	0.317	-0.543	0.454	0.512
Std.dev.	-0.502	0.501	0.500	-0.370	0.349	0.154	0.351	-0.536	0.463	0.467
RMSE	0.034	0.068	0.062	0.135	0.163	0.349	0.710	0.160	0.172	0.232

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 11: Simulations of Two-Period Dynamic Poisson Models

	Infeasible			Conventional				Sieve ML		
N=1000	γ	θ	λ	γ	θ	λ	σ	γ	θ	λ
True	−0.5	0.5	0.5	−0.5	0.5	0.5	1	−0.5	0.5	0.5
DGP IV:										
Mean	-0.501	0.498	0.498	-0.372	0.352	0.155	0.317	-0.510	0.500	0.512
Std.dev.	-0.500	0.499	0.500	-0.370	0.349	0.154	0.351	-0.513	0.495	0.513
RMSE	0.024	0.045	0.042	0.135	0.163	0.349	0.710	0.105	0.102	0.098
DGP V:										
Mean	-0.501	0.499	0.499	-0.380	0.381	0.198	0.328	-0.506	0.505	0.507
Std.dev.	-0.500	0.497	0.499	-0.379	0.382	0.198	0.354	-0.505	0.505	0.507
RMSE	0.023	0.045	0.043	0.123	0.127	0.304	0.687	0.104	0.103	0.103
DGP VI:										
Mean	-0.501	0.497	0.497	-0.371	0.360	0.159	0.339	-0.508	0.503	0.508
Std.dev.	-0.500	0.499	0.495	-0.368	0.360	0.158	0.367	-0.506	0.504	0.508
RMSE	0.023	0.050	0.044	0.133	0.147	0.343	0.679	0.102	0.099	0.107

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 12: Simulations of APE($\bar{x}$) in the Two-Period Dynamic Poisson Models

N=500	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP IV:			
Mean	-0.127	-0.138	-0.162
Std. dev.	0.003	0.013	0.042
RMSE	–	0.019	0.055
DGP V:			
Mean	-0.126	-0.138	-0.161
Std. dev.	0.003	0.013	0.041
RMSE	–	0.019	0.054
DGP VI:			
Mean	-0.117	-0.135	-0.158
Std. dev.	0.003	0.013	0.044
RMSE	–	0.024	0.060

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 13: Simulations of APE($\bar{x}$) in the Two-Period Dynamic Poisson Models

N=1000	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP IV:			
Mean	-0.127	-0.137	-0.141
Std. dev.	0.002	0.009	0.029
RMSE	–	0.016	0.033
DGP V:			
Mean	-0.126	-0.137	-0.140
Std. dev.	0.002	0.009	0.029
RMSE	–	0.016	0.032
DGP VI:			
Mean	-0.117	-0.135	-0.141
Std. dev.	0.002	0.010	0.029
RMSE	–	0.022	0.037

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 14: Simulations of Four-Period Dynamic Poisson Models

	Infeasible			Conventional				Sieve ML		
N=500	γ	θ	λ	γ	θ	λ	σ	γ	θ	λ
True	-0.5	0.5	0.5	-0.5	0.5	0.5	1	-0.5	0.5	0.5
DGP IV:										
Mean	-0.500	0.499	0.500	-0.524	0.494	0.500	1.004	-0.418	0.445	0.358
Std. dev.	-0.500	0.499	0.499	-0.522	0.494	0.499	1.004	-0.429	0.458	0.351
RMSE	0.020	0.023	0.018	0.035	0.036	0.039	0.046	0.133	0.157	0.196
DGP V:										
Mean	-0.501	0.499	0.501	-0.525	0.493	0.499	1.013	-0.421	0.460	0.393
Std. dev.	-0.500	0.501	0.500	-0.525	0.492	0.500	1.011	-0.427	0.463	0.385
RMSE	0.020	0.023	0.018	0.036	0.040	0.040	0.047	0.121	0.130	0.152
DGP VI:										
Mean	-0.502	0.500	0.500	-0.528	0.487	0.497	1.200	-0.463	0.454	0.366
Std. dev.	-0.502	0.500	0.500	-0.527	0.488	0.499	1.199	-0.475	0.461	0.359
RMSE	0.020	0.021	0.017	0.039	0.048	0.046	0.206	0.101	0.150	0.183

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 15: Simulations of Four-Period Dynamic Poisson Models

	Infeasible			Conventional				Sieve ML		
N=1000	γ	θ	λ	γ	θ	λ	σ	γ	θ	λ
True	-0.5	0.5	0.5	-0.5	0.5	0.5	1	-0.5	0.5	0.5
DGP IV:										
Mean	-0.500	0.500	0.499	-0.524	0.491	0.502	1.004	-0.511	0.470	0.462
Std. dev.	-0.500	0.500	0.500	-0.524	0.491	0.502	1.005	-0.516	0.472	0.460
RMSE	0.015	0.015	0.012	0.031	0.028	0.029	0.032	0.073	0.089	0.098
DGP V:										
Mean	-0.500	0.499	0.500	-0.524	0.492	0.501	1.013	-0.421	0.472	0.389
Std. dev.	-0.501	0.499	0.500	-0.524	0.492	0.501	1.013	-0.426	0.484	0.382
RMSE	0.015	0.016	0.013	0.030	0.027	0.030	0.034	0.119	0.116	0.154
DGP VI:										
Mean	-0.501	0.500	0.500	-0.527	0.484	0.494	1.203	-0.471	0.448	0.365
Std. dev.	-0.501	0.501	0.499	-0.526	0.486	0.493	1.203	-0.484	0.452	0.357
RMSE	0.015	0.015	0.011	0.033	0.038	0.032	0.206	0.101	0.149	0.182

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 16: Simulations of State Dependence in the Four-Period Dynamic Poisson Models

N=500	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP IV:			
Mean	-0.130	-0.133	-0.119
Std. dev.	0.003	0.007	0.030
RMSE	–	0.009	0.032
DGP V:			
Mean	-0.129	-0.133	-0.120
Std. dev.	0.003	0.007	0.026
RMSE	–	0.009	0.027
DGP VI:			
Mean	-0.119	-0.124	-0.132
Std. dev.	0.003	0.007	0.027
RMSE	–	0.010	0.030

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 17: Simulations of State Dependence in the Four-Period Dynamic Poisson Models

N=1000	Infeasible Estimator	Conventional Estimator	Sieve ML Estimator
DGP IV:			
Mean	-0.130	-0.133	-0.144
Std. dev.	0.002	0.005	0.020
RMSE	–	0.008	0.025
DGP V:			
Mean	-0.129	-0.132	-0.120
Std. dev.	0.002	0.005	0.025
RMSE	–	0.007	0.027
DGP VI:			
Mean	-0.119	-0.123	-0.134
Std. dev.	0.002	0.005	0.028
RMSE	–	0.008	0.031

Note: Standard deviations of the parameters are computed by the standard deviations of the estimates across 1000 simulations. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.

Table 18: Hausman-type Test for Normality: Empirical Size

	Two-Period Dynamic Poisson Models		Four-Period Dynamic Poisson Models	
	N=500	N=1000	N=500	N=1000
DGP I:	0.313	0.833	0.190	0.080
DGP II:	0.287	0.807	0.210	0.200
DGP II:	0.340	0.867	0.166	0.139

Note: The *p*-values of 0.05 of Chi-distributions for static models and dynamic models are 5.991 and 7.815, respectively. Empirical size refers to the fraction of rejections when using these values as the critical values. The notations DGP IV, DGP V, and DGP VI represent the data generating processes in Eqs. (E.5), (E.6), and (E.7), respectively.