Online Appendix: Identification and Estimation of Partial Effects in Nonlinear Semiparametric Panel Models

Laura Liu Alexandre Poirier Ji-Liang Shiu

July 29, 2024

This online appendix is organized as follows. In Appendix C we establish the asymptotic properties of the proposed ASF and APE estimators. In Appendix D we discuss a number of implementation details for the estimators. In Appendix E we extend our main identification results to models with sequential exogeneity, allowing for lagged outcomes as regressors. In Appendix F we conduct Monte Carlo experiments to study the finite-sample properties of our estimators. In Appendix G we present additional figures and tables that supplement the empirical results in the main text. Appendix H contains proofs for results in appendices C and E.

C Estimation and Inference

We propose three-step semiparametric estimators for the ASF, APE, and AME. The first step estimates the common parameters β_0, the second step conducts a nonparametric regression with a generated regressor, and the third step marginalizes over a subset of the regressors. Such estimators are called partial means.[1]

We show that the rate of convergence of the ASF estimator is similar to that of a kernel regression estimator with one continuous regressor. The APE estimator converges at a similar rate as a derivative of a kernel regression estimator with one continuous regressor. In particular, we show the ASF converges at the rate $\sqrt{Nb_N}$ and the APE at the rate $\sqrt{Nb_N^3}$ where b_N is a scalar bandwidth used in the estimation of the conditional expectation of Y_t. We describe below in Assumption B6 what assumptions b_N must satisfy. These rates of convergence are obtained from our estimators being partial means, where we average over all components of the conditional expectation of $\mathbb{E}[Y_t|X_t'\beta_0, V]$, except for one. These convergence rates do not depend on T or d_X,

[1] See Newey (1994) for seminal work on the estimation of partial means without generated regressors. The estimation of partial means with generated regressors is studied in Mammen, Rothe, and Schienle (2012, 2016), and Lee (2018).

the dimensions of $\mathbf{X}$.

Throughout this section, we assume we observe a random sample of $(Y_i, \mathbf{X}_i)$ of size N.

Assumption B1 (Random sampling). $\{(Y_i, \mathbf{X}_i)\}_{i=1}^N$ are iid.

We start by considering the estimation of β_0, the first step of our semiparametric estimator.

C.1 Estimation of β_0

In Section 2.2.2 we discussed several previous identification approaches for the common parameters β_0. Due to the breadth of these approaches, we consider the following high-level assumption on the rate of convergence of a first-step estimator of β_0.

Assumption B2 (First-stage estimator). The estimator $\widehat{\beta}$ satisfies $a_N \|\widehat{\beta} - \beta_0\| = O_p(1)$ where $a_N = O(N^\epsilon)$ for some $\epsilon > 0$.

The rate of convergence of this preliminary estimator plays a role in Assumption B6 below. The convergence of $\widehat{\beta}$ needs to be relatively fast to establish the limiting distributions of the ASF and APE estimators. In particular, convergence rates equal to or slower than $N^{1/3}$ are incompatible with our rate assumption B6 below. This rules out the maximum score estimator of Manski (1987) for binary panels, but not the smoothed maximum score estimator of Charlier, Melenberg, and van Soest (1995) and Kyriazidou (1995). The smoothed maximum score estimator converges at the rate $N^{\nu/(2\nu+1)}$, where ν is the order of the kernel used to estimate $\widehat{\beta}$.

In binary panels, the rate of convergence is usually slower than $\sqrt{N}$. One exception is the $\sqrt{N}$-consistent conditional maximum likelihood estimator (Rasch (1960), Andersen (1970)) when U_t follows a logistic distribution. While $\sqrt{N}$-estimation of β_0 is generally not possible in binary panels without specifying U_t's distribution (Magnac (2004), Chamberlain (2010)), some alternative assumptions and estimators allow for it. In particular, Lee (1999) considers an "index increment sufficiency" assumption: $(X_t'\beta_0, C)|(X_t - X_s) \overset{d}{=} (X_t'\beta_0, C)|(X_t - X_s)'\beta_0$. Honoré and Lewbel (2002) assume the presence of a special regressor among X_t. Chen, Si, Zhang, and Zhou (2017) assume that $C = v(\mathbf{X}) + \zeta$, where ζ satisfies $(U_1, \ldots, U_T, \zeta) \perp\!\!\!\perp \mathbf{X}$. In all three papers, $\sqrt{N}$-consistent estimators for β_0 are proposed.

With continuous outcomes and the index function taking the form $v(\mathbf{X})'\gamma_0$, Ichimura and Lee (1991)'s approach can be used to estimate β_0 (and non-zero entries in the coefficient matrix γ_0) at a $\sqrt{N}$-rate: see Appendix A.2 for definitions of the notations. Abrevaya (1999) proposes a $\sqrt{N}$-consistent *leapfrog* estimator when $Y_{it} = g(X_{it}'\beta_0 + C_i + U_{it})$ and Y_{it} is continuous. Also see Abrevaya (2000) and Botosaru and Muris (2017) for other $\sqrt{N}$-consistent estimators of β_0 in related models.

C.2 A Semiparametric Estimator of the ASF

We now present the ASF estimator and show its consistency and asymptotic normality under our assumptions. As mentioned earlier, this estimator is a three-step estimator. Appendix C.1 discussed the first step, which consists of estimating β_0 using an existing method. We now describe the second and third steps, which estimate the ASF using a sample analog of equation (2.3). In the second step, we nonparametrically estimate the conditional expectation $\mathbb{E}[Y_t | X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]$ using a local polynomial regression of Y_t on generated regressor $X_t'\widehat{\beta}$ and V. In the third and final step, we evaluate the estimated conditional expectation at $(\underline{x}_t'\widehat{\beta}, V_i)$ for $i = 1, \ldots, N$, and then average over the empirical marginal distribution of V_i. To define this estimator, let $Z_t(\beta) = (X_t'\beta, V) \in \mathbb{R}^{1+d_V}$ and denote $Z_t = Z_t(\beta_0)$. Throughout the paper, we use z to denote $z = (u, v) \in \mathbb{R}^{1+d_V}$ where $u \in \mathbb{R}$ and $v \in \mathbb{R}^{d_V}$.

In the rest of this section, we assume that V's components are all continuously distributed. We omit the discrete case for notational simplicity. In our analysis, the number of discrete components of V does not affect the convergence rate. When the number of support points for the discrete components is sufficiently small, we can handle these discrete components by performing a cell-by-cell analysis. Alternatively, they can be accommodated through a discrete kernel, for example, as in Racine and Li (2004) equation (2.3).

We consider a local polynomial regression of order $\ell \geq 0$. The following notation is similar to that in Masry (1996). For $s \in \{0, 1, \ldots, \ell\}$, let $N_s = \binom{s+d_V}{d_V}$ be the number of distinct $(1+d_V)$-tuples $r \in \mathbb{N}^{1+d_V}$ such that $|r| \equiv \sum_{k=1}^{1+d_V} |r_k| = s$. We arrange these $(1+d_V)$-tuples in a lexicographical order with the highest priority given to the last position so that $(0, \ldots, 0, s)$ is the first element and $(s, 0, \ldots, 0)$ is the last element in this sequence. We let τ_s denote this one-to-one mapping. This mapping satisfies $\tau_s(1) = (0, \ldots, 0, s), \ldots, \tau_s(N_s) = (s, 0, \ldots, 0)$. For each $s \in \{0, 1, \ldots, \ell\}$, define $N_s \times 1$ vector $\xi_s(a)$ by its kth element $a^{\tau_s(k)}$, where $k \in \{1, \ldots, N_s\}$ and $a \in \mathbb{R}^{1+d_V}$. Here we used the notation $a^b = a_1^{b_1} \times \cdots \times a_{d_V}^{b_{d_V}}$. Let

$$\xi(a) = (1, \xi_1(a)', \ldots, \xi_\ell(a)')' \in \mathbb{R}^{\bar{N}},$$

where $\bar{N} = \sum_{s=0}^{\ell} N_s$.

Let $\mathcal{K} : \mathbb{R}^{1+d_V} \to \mathbb{R}$ denote a $(1+d_V)$-dimensional kernel. Let $\mathcal{K}_b(z) = b^{-(1+d_V)}\mathcal{K}(z)$, where $b > 0$ is a scalar bandwidth. Let b_N denote a sequence of bandwidths converging to zero. Let

$$\begin{aligned}
\widehat{h}(z; \widehat{\beta}) &= \underset{h \in \mathbb{R}^{\bar{N}}}{\operatorname{argmin}} \sum_{j=1}^{N} \left(Y_{jt} - \sum_{0 \leq |r| \leq \ell} \left(\frac{Z_{jt}(\widehat{\beta}) - z}{b_N} \right)^r h_r \right)^2 \mathcal{K}_{b_N}\left(\frac{Z_{jt}(\widehat{\beta}) - z}{b_N} \right) \\
&= \underset{h \in \mathbb{R}^{\bar{N}}}{\operatorname{argmin}} \sum_{j=1}^{N} \left(Y_{jt} - \xi\left(\frac{Z_{jt}(\widehat{\beta}) - z}{b_N} \right)' h \right)^2 \mathcal{K}_{b_N}\left(\frac{Z_{jt}(\widehat{\beta}) - z}{b_N} \right).
\end{aligned}$$

As $\widehat{\beta} \xrightarrow{p} \beta_0$, the vector $\widehat{h}(z;\widehat{\beta})$ estimates coefficients in a Taylor expansion of degree ℓ of the conditional expectation of Y_t given $Z_t(\beta_0) = z$. In particular, the first component of this vector, denoted by $\widehat{h}_1(z;\widehat{\beta}) = e_1'\widehat{h}(z;\widehat{\beta})$, is an estimator of the conditional mean of Y_t given $(X_t'\beta_0, V)$. The vector $\widehat{h}(z;\widehat{\beta})$ is a least-squares estimator and can be written as

$$\widehat{h}(z;\widehat{\beta}) = S_N(z;\widehat{\beta})^{-1}T_N(z;\widehat{\beta}),$$

where

$$\begin{aligned} S_N(z;\beta) &= \frac{1}{N}\sum_{j=1}^{N}\xi\left(\frac{Z_{jt}(\beta)-z}{b_N}\right)\xi\left(\frac{Z_{jt}(\beta)-z}{b_N}\right)'\mathcal{K}_{b_N}\left(\frac{Z_{jt}(\beta)-z}{b_N}\right) \\ T_N(z;\beta) &= \frac{1}{N}\sum_{j=1}^{N}\xi\left(\frac{Z_{jt}(\beta)-z}{b_N}\right)Y_{jt}\mathcal{K}_{b_N}\left(\frac{Z_{jt}(\beta)-z}{b_N}\right). \end{aligned}$$

In analogy to equation (2.3), we average this conditional mean over the empirical marginal distribution of V_i to obtain the ASF estimator:

$$\widehat{\text{ASF}}_t(\underline{x}_t) = \frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i;\widehat{\beta})\hat{\pi}_{it},$$

where $\widehat{h}_1(z;\widehat{\beta}) = e_1'\widehat{h}(z;\widehat{\beta})$ is the first component in $\widehat{h}(z;\widehat{\beta})$, $\hat{\pi}_{it} = \mathbb{1}((\underline{x}_t'\widehat{\beta}, V_i) \in \mathcal{Z}_t)$ is a trimming function, and $\mathcal{Z}_t$ is an appropriately selected compact set in which the density $f_{Z_t(\beta)}(z)$ is bounded away from zero. This trimming function prevents issues with the invertibility of $S_N(z;\widehat{\beta})$. Since $\mathcal{Z}_t$ is a fixed compact set, the parameter that is consistently estimated by $\widehat{\text{ASF}}_t$ is a trimmed ASF defined by

$$\begin{aligned} \text{ASF}_t^{\pi}(\underline{x}_t) &\equiv \mathbb{E}[\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V]\pi_t] \\ &= \int_{\mathcal{C}}\mathbb{E}[Y_t|X_t = \underline{x}_t, C = c]\mathbb{P}((\underline{x}_t'\beta_0, V) \in \mathcal{Z}_t|C = c)\,dF_C(c), \end{aligned}$$

where we let $\pi_{it} = \mathbb{1}((\underline{x}_t'\beta_0, V_i) \in \mathcal{Z}_t)$. Note that if $(\underline{x}_t'\beta_0, V) \in \mathcal{Z}_t$ with probability 1, $\text{ASF}_t^{\pi}(\underline{x}_t) = \text{ASF}_t(\underline{x}_t)$ and the trimming does not alter the estimand. By expanding $\mathcal{Z}_t$ along with the sample size at a slow enough rate, which is sometimes called a vanishing, or random, trimming approach, we expect that $\text{ASF}_t(\underline{x}_t)$ is consistently estimated by $\widehat{\text{ASF}}_t(\underline{x}_t)$. However, since fixed trimming is often employed in the partial mean literature,[2] we let $\mathcal{Z}_t$ be fixed.

To understand the effect of trimming on the estimand, we consider the scenario where $\mathbb{P}((\underline{x}_t'\beta_0, V) \in \mathcal{Z}_t|C = c)$ is bounded away from zero. Formally, assume that $\mathbb{P}((\underline{x}_t'\beta_0, V) \in \mathcal{Z}_t|C) \in [1-\underline{\varepsilon}, 1]$ with probability 1. Then, if $\text{ASF}_t^{\pi}(\underline{x}_t) \geq 0$, we can show that $\text{ASF}_t^{\pi}(\underline{x}_t) \in [(1-\underline{\varepsilon})\text{ASF}_t(\underline{x}_t), \text{ASF}_t(\underline{x}_t)]$,

[2] See, for example, Newey (1994) or more recently Lee (2018).

and thus

$$\text{ASF}_t(\underline{x}_t) \in \left[\text{ASF}_t^\pi(\underline{x}_t), \frac{\text{ASF}_t^\pi(\underline{x}_t)}{1-\underline{\varepsilon}}\right].$$

These bounds are reversed when $\text{ASF}_t^\pi(\underline{x}_t) < 0$. Note that the bounds collapse to a point as $\underline{\varepsilon}$ approaches zero, and are narrow when $\underline{\varepsilon}$ is small.

We make the following assumptions to obtain the limiting distribution of the ASF. We begin with a standard assumption on the kernel.

Assumption B3 (Kernel). The kernel $\mathcal{K}$ satisfies $\mathcal{K}(z) = K(u) \cdot \prod_{k=1}^{d_V} K(v_k)$ where $K : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that (i) $K(u)$ is equal to zero for all u outside of a compact set, (ii) K is twice continuously differentiable on $\mathbb{R}$ with all these derivatives being Lipschitz continuous, (iii) $\int_{-\infty}^{\infty} K(u)\,du = 1$, (iv) K is symmetric.

Note that we do not require the use of higher-order kernels in this local polynomial regression.

To state the next assumption precisely, let $\mathcal{C}_m(\mathcal{A})$ denote the set of m-times continuously differentiable functions $f : \mathcal{A} \to \mathbb{R}$. Here m is an integer and $\mathcal{A}$ is a subset of $\mathbb{R}^{1+d_V}$. Denote the differential operator by

$$\nabla^\lambda = \frac{\partial^{|\lambda|}}{\partial z_1^{\lambda_1} \cdots \partial z_{1+d_V}^{\lambda_{1+d_V}}},$$

where $\lambda = (\lambda_1, \ldots, \lambda_{1+d_V}) \in \{0, 1, \ldots\}^{1+d_V}$ is comprised of nonnegative integers such that $\sum_{k=1}^{1+d_V} \lambda_k = |\lambda|$. For a given set $\mathcal{A}$, let

$$\|f\|_m^{\mathcal{A}} = \max_{|\lambda| \leq m} \sup_{z \in \text{int}(\mathcal{A})} \|\nabla^\lambda f(z)\|.$$

We omit the $\mathcal{A}$ superscript when it does not cause confusion. Next, we impose smoothness and regularity conditions on the distribution of $(Y_t, Z_t(\beta))$ for β in a neighborhood of β_0.

Assumption B4 (Smoothness). Let $\mathcal{B}_\varepsilon = \{\beta \in \mathcal{B} : \|\beta - \beta_0\| \leq \varepsilon\}$.

(i) There exists $\varepsilon > 0$ such that for all $\beta \in \mathcal{B}_\varepsilon$, $Z_t(\beta)$ has a density $f_{Z_t(\beta)}(z)$ with respect to the Lebesgue measure;

(ii) $f_{Z_t(\beta)}(z)$ and $\left\|\frac{\partial}{\partial \beta} f_{Z_t(\beta)}(z)\right\|$ are uniformly bounded and uniformly bounded away from zero for $z \in \mathcal{Z}_t$ and $\beta \in \mathcal{B}_\varepsilon$, where $\mathcal{Z}_t$ is a compact set;

(iii) $\|f_{Z_t(\beta_0)}(z)\|_{\ell+2}^{\mathcal{Z}_t} < \infty$ and $\|\mathbb{E}[Y_t | Z_t(\beta_0) = z]\|_{\ell+2}^{\mathcal{Z}_t} < \infty$;

(iv) $\underline{x}_t' \beta_0$ is in the interior of $\mathcal{Z}_{1t} \equiv \{e_1' z : z \in \mathcal{Z}_t\}$;

(v) $f_{Z_t(\beta_0)|Y_t}(z|y)$ exists and is bounded for $y \in \text{supp}(Y_t)$.

Assumptions (i) and (ii) ensure the boundedness and sufficient smoothness of the distribution of $f_{Z_t(\beta)}$ as a function of β in a neighborhood of β_0. Assumption (iii) ensures additional smoothness in

z for the distribution of $Z_t(\beta_0)$. The degree of smoothness is linked to the degree of the polynomial in the local polynomial regression. Assumptions (iv) and (v) are standard technical assumptions. We also impose the following moment existence condition.

Assumption B5 (Moment existence). Let $\mathbb{E}[\|X_t\|^2] < \infty$ and $\mathbb{E}[|Y_t|^n] < \infty$ for all $n \in \mathbb{N}$.

We can relax the assumption that all moments of Y_t exist at the cost of some additional notation and derivations: see the proof of Lemma H.8.

The following rate conditions govern the bandwidth's convergence rate. Let $\lceil x \rceil$ denote the smallest integer greater than or equal to x.

Assumption B6 (Bandwidth). For some $\kappa, \delta > 0$, let $b_N = \kappa \cdot N^{-\delta}$ where δ satisfies

$$\max\left\{\frac{1}{4\left\lceil\frac{\ell+1}{2}\right\rceil + 1}, 1 - 2\epsilon\right\} < \delta < \min\left\{\frac{2\epsilon}{3 + 2d_V}, \frac{1}{1 + 2d_V}\right\}.$$

A consequence of this assumption is that ℓ must increase as d_V increases. In particular, we require $\ell > d_V$ when $\widehat{\beta}$ is $\sqrt{N}$-consistent.

We can now state the main convergence result for the ASF.

Theorem C.1 (ASF asymptotics). Suppose the assumptions for Part 1 of Theorem 2.1 hold. Suppose Assumptions B1–B6 hold. Then,

$$\sqrt{Nb_N}\left(\widehat{\text{ASF}}_t(\underline{x}_t) - \text{ASF}_t^{\pi}(\underline{x}_t)\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\text{ASF}_t}(\underline{x}_t'\beta_0)),$$

where

$$\sigma^2_{\text{ASF}_t}(u) = \mathbb{E}\left[\text{Var}(Y_t | X_t'\beta_0 = u, V)\frac{f_V(V)}{f_{Z_t(\beta_0)}(u, V)}\mathbb{1}((u, V) \in \mathcal{Z}_t)\right]$$
$$\cdot\, e_1'\left(\int \xi(z)\xi(z)'\mathcal{K}(z)\,dz\right)^{-1}\left[\int\left(\int \mathcal{K}\left(z\right)\xi\left(z\right)\,dv\right)\left(\int \mathcal{K}\left(z\right)\xi\left(z\right)\,dv\right)'\,du\right]\left(\int \xi(z)\xi(z)'\mathcal{K}(z)\,dz\right)^{-1} e_1.$$

To understand the limiting distribution of this estimator, we break down its sampling variation into four separate sources. The terms associated with three of these are asymptotically negligible under our assumptions. We can write

$$
\begin{aligned}
\sqrt{Nb_N}\left(\widehat{\text{ASF}}_t(\underline{x}_t) - \text{ASF}_t^{\pi}(\underline{x}_t)\right) &= \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - \widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right) \\
&+ \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)\right)\hat{\pi}_{it}\right) \\
&+ \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)(\hat{\pi}_{it} - \pi_{it})\right) \\
&+ \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)\pi_{it} - \mathbb{E}[h_1(\underline{x}_t'\beta_0, V; \beta_0)\pi_t]\right).
\end{aligned}
$$

The first term reflects the impact of the generated regressors $X_t'\widehat{\beta}$ being used instead of $X_t'\beta_0$. The bandwidth constraints involving ϵ—the rate of convergence of $\widehat{\beta}$ to β_0—ensure this term is asymptotically negligible. The second term reflects the impact of the approximation of the evaluation point $\underline{x}_t'\beta_0$ by $\underline{x}_t'\widehat{\beta}$. Once again, ϵ plays a crucial role and this term is asymptotically negligible as it is of asymptotic order $O_p(\sqrt{Nb_N}a_N^{-1}) = o_p(1)$ by our assumptions. The third term pertains to the estimation of the trimming function π_{it} by $\hat{\pi}_{it}$. This term is asymptotically dominated due to the superconsistency of $\hat{\pi}_{it}$ to π_{it} uniformly in $i = 1, \ldots, N$. The fourth and final term asymptotically dominates the other three and converges in distribution to a mean-zero Gaussian variable at the $\sqrt{Nb_N}$ rate. Some of the technical tools we use to show this convergence in distribution build on Masry (1996) and Kong, Linton, and Xia (2010).

The rate of convergence of $\widehat{\text{ASF}}_t(\underline{x}_t)$ when $\epsilon = 1/2$ is $N^{\delta_{\text{ASF}}}$, where δ_{ASF} ranges in the interval $\left(\frac{1+d_V}{3+2d_V}, \frac{1+\ell}{3+2\ell}\right)$. In the case where $d_V = 1$ and $\ell = 2$, this range corresponds to $\left(\frac{2}{5}, \frac{3}{7}\right)$. Recall that $2/5$ is the standard rate of convergence of univariate kernel estimation when using second-order kernels. Again, we note that this rate of convergence does not depend on either T or d_X. We discuss various implementation details of this estimator and others in Appendix D.

C.3 Semiparametric Estimation of the APE

We focus here on the case where $X_t^{(k)}$ is continuously distributed. When $X_t^{(k)}$ is discretely distributed, the APE is a difference between two ASFs, in which case Theorem C.1 can be used to obtain its limiting distribution.

Let $\widehat{h}_2(z; \widehat{\beta}) = \frac{1}{b_N} e_{2+d_V}' \widehat{h}(z; \widehat{\beta})$ denote the $(2 + d_V)$-th component of the local polynomial regression coefficient vector. By the definition of the above lexicographical order, this is an estimator of the derivative of the conditional mean of Y_t given $(X_t'\beta_0, V) = (u, v)$ with respect to u. This estimated derivative is used in the APE estimator, which is defined as

$$\widehat{\mathrm{APE}}_{k,t}(\underline{x}_t) = \widehat{\beta}^{(k)} \cdot \frac{1}{N}\sum_{i=1}^{N} \widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})\hat{\pi}_{it},$$

where $\widehat{\beta}^{(k)}$ denotes the kth component of $\widehat{\beta}$.

As for the ASF, we use a trimming function in the estimator for technical reasons. Therefore, the estimator is consistent for a trimmed APE defined by $\mathrm{APE}^{\pi}_{k,t}(\underline{x}_t) \equiv \mathbb{E}\left[\frac{\partial}{\partial \underline{x}_t^{(k)}}\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V] \cdot \pi_t\right]$. As for the ASF, the untrimmed APE is bounded by $\mathrm{APE}_{k,t}(\underline{x}_t) \in \left[\mathrm{APE}^{\pi}_{k,t}(\underline{x}_t), (1-\underline{\varepsilon})^{-1}\mathrm{APE}^{\pi}_{k,t}(\underline{x}_t)\right]$ when $\mathbb{P}(\underline{x}_t'\beta_0, V) \in \mathcal{Z}_t|C) \in [1-\underline{\varepsilon}, 1]$ with probability 1 and the APE is positive: the bounds are reversed when it is negative.

The following theorem shows that the APE is $\sqrt{Nb_N^3}$-consistent, where b_N is a bandwidth satisfying Assumption B6. Like the ASF, the APE's rate of convergence does not depend on the dimensions of $\mathbf{X}$.

Theorem C.2 (APE asymptotics)**.** Suppose the assumptions for Part 2 of Theorem 2.1 hold. Suppose Assumptions B1–B6 hold. Suppose $X_t^{(k)}$ is continuously distributed. Then,

$$\sqrt{Nb_N^3}\left(\widehat{\mathrm{APE}}_{k,t}(\underline{x}_t) - \mathrm{APE}^{\pi}_{k,t}(\underline{x}_t)\right) \xrightarrow{d} \mathcal{N}\left(0, (\beta_0^{(k)})^2 \cdot \sigma^2_{\mathrm{APE}_t}(\underline{x}_t'\beta_0)\right),$$

where

$$\sigma^2_{\mathrm{APE}_t}(u) = \mathbb{E}\left[\mathrm{Var}(Y_t|X_t'\beta_0 = u, V)\frac{f_V(V)}{f_{Z_t(\beta_0)}(u,V)}\mathbb{1}((u,V)\in\mathcal{Z}_t)\right] e_{2+d_V}'\left(\int \xi(z)\xi(z)'\mathcal{K}(z)\,dz\right)^{-1}$$
$$\cdot\left[\int\left(\int \mathcal{K}(z)\,\xi(z)\;dv\right)\left(\int \mathcal{K}(z)\,\xi(z)\;dv\right)'du\right]\left(\int \xi(z)\xi(z)'\mathcal{K}(z)\,dz\right)^{-1} e_{2+d_V}.$$

We can decompose the APE's sample variation into five components. The first four components are analogous to those in the earlier ASF decomposition. In particular, the fourth component is

$$\widehat{\beta}^{(k)} \cdot \sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(\underline{x}_t'\beta_0, V_i; \beta_0)\pi_{it} - \mathbb{E}[h_2(\underline{x}_t'\beta_0, V; \beta_0)\pi_t]\right).$$

This component converges in distribution to a mean-zero Gaussian distribution while dominating the other components. The fifth component is due to the presence of $\widehat{\beta}^{(k)}$ and is of the same order as $\sqrt{Nb_N^3}(\widehat{\beta}^{(k)} - \beta_0^{(k)}) = O_p\left(\sqrt{Nb_N^3}a_N\right) = o_p(1)$ by B6.

The rate of convergence of $\widehat{\mathrm{APE}}_{k,t}(\underline{x}_t)$ when $\epsilon = 1/2$ is $N^{\delta_{\mathrm{APE}}}$, where δ_{APE} ranges in the interval $\left(\frac{d_V}{3+2d_V}, \frac{\ell}{3+2\ell}\right)$. When $d_V = 1$ and $\ell = 2$, this range equals $\left(\frac{1}{5}, \frac{2}{7}\right)$. Recall that 2/7 is the standard rate of convergence of derivatives of univariate kernel estimators when using second-order kernels. Our estimator can approach this rate whenever $\ell \geq 2$, i.e., the local polynomial contains quadratic terms.

C.4 Estimation of the LAR and AME

The previous analysis focused on the estimation and inference for the ASF and APE using sample analog estimators. Under the assumptions of Theorem 2.2, the LAR and AME are also point identified via a function of the distribution of $(Y, \mathbf{X})$. Here are their sample analogs:

$$\widehat{\text{LAR}}_{k,t}(\underline{\mathbf{x}}) = \widehat{\beta}^{(k)} \cdot \widehat{h}_2(\underline{x}_t'\widehat{\beta}, v(\underline{\mathbf{x}}); \widehat{\beta})$$

$$\widehat{\text{AME}}_{k,t} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\text{LAR}}_{k,t}(\mathbf{X}_i)\hat{\pi}_{it} = \widehat{\beta}^{(k)} \cdot \frac{1}{N}\sum_{i=1}^{N} \widehat{h}_2(X_{it}'\widehat{\beta}, v(\mathbf{X}_i); \widehat{\beta})\hat{\pi}_{it}.$$

Establishing their consistency and asymptotic distribution can be done using the same tools used to establish the same properties for the ASF and APE. Since their proofs are likely similar to those for the ASF and APE, we leave formal asymptotic analyses for future work. The rate of convergence of the LAR estimator should be the same as the nonparametric rate used to estimate h_2, while we expect the rate of convergence of the AME to be $\sqrt{N}$ when $\widehat{\beta}^{(k)}$ is $\sqrt{N}$-consistent. This is because the AME averages over all conditioning variables in the local regression of Y_t on $Z_t(\widehat{\beta})$.

C.5 Estimation with Estimated Indices

We now briefly consider the estimation of these partial effects under the assumption that $C \perp\!\!\!\perp \mathbf{X}|V'\gamma_0$. Following the notations in Appendix A.2, V is $d_V \times 1$, γ_0 is $d_V \times d_{V^*}$, and the new index $V'\gamma_0$ is $1 \times d_{V^*}$. Let $\gamma_1^*, \cdots, \gamma_{d_{V^*}}^*$ be the *non-zero elements* in each column of γ_0, and $V_1^*, \ldots, V_{d_{V^*}}^*$ be the corresponding elements in V, and we have $V'\gamma_0 = \left(V_1^{*\prime}\gamma_1^*, \ldots, V_{d_{V^*}}^{*\prime}\gamma_{d_{V^*}}^*\right)$. Then, $\gamma^* = \left(\gamma_1^{*\prime}, \cdots, \gamma_{d_{V^*}}^{*\prime}\right)'$ contains all unknown elements in γ_0.

Suppose $\theta_0 \equiv (\beta_0', \gamma^{*\prime})'$ is consistently estimated. For example, Ichimura and Lee (1991)'s estimator is $\sqrt{N}$-consistent for θ_0 under their regularity conditions. Let $Z_t(\theta) = (X_t'\beta, V'\gamma) \in \mathbb{R}^{1+d_{V^*}}$ and let

$$\widehat{h}(z; \widehat{\theta}) = \underset{h \in \mathbb{R}^{\bar{N}}}{\operatorname{argmin}} \sum_{j=1}^{N} \left(Y_{jt} - \xi\left(\frac{Z_{jt}(\widehat{\theta}) - z}{b_N}\right)' h\right)^2 \mathcal{K}_{b_N}\left(\frac{Z_{jt}(\widehat{\theta}) - z}{b_N}\right).$$

Then, we can propose the following estimators:

$$\widehat{\text{ASF}}_t(\underline{x}_t) = \frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i'\widehat{\gamma}; \widehat{\theta})\hat{\pi}_{it}$$

$$\widehat{\text{APE}}_{k,t}(\underline{x}_t) = \widehat{\beta}^{(k)} \cdot \frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i'\widehat{\gamma}; \widehat{\theta})\hat{\pi}_{it}$$

$$\widehat{\text{AME}}_{k,t} = \widehat{\beta}^{(k)} \cdot \frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(X_{it}'\widehat{\beta}, V_i'\widehat{\gamma}; \widehat{\theta})\hat{\pi}_{it}.$$

The indices $V'\gamma_0$ are usually of lower dimension than V, which means the function $\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V'\gamma_0 = v'\gamma_0]$ has lower dimension than $\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]$, which helps satisfy the rate assumption B6. This comes at the cost of an additional generated regressor of the form $V_i'\widehat{\gamma}$. From examining Lemmas H.1–H.7, we expect these additional generated regressors do not impact the estimators' limiting distributions, but we leave a detailed asymptotic analysis for future work.

D Implementation Details

D.1 General Choices

Here are a few practical concerns related to the implementation of our estimators. We explored some of these in more detail in our simulations (Appendix F) and empirical illustration (Section 4).

Local polynomial regression. First, a common practice in kernel-based methods is the standardization and orthogonalization of the conditioning variables, in our case $Z_t(\widehat{\beta}) = (X_t'\widehat{\beta}, V)$, before the nonparametric estimation step. The standardization leads to more comparable scales across different components of $Z_t(\widehat{\beta})$. The orthogonalization facilitates the practical implementation of using a product of one-dimensional kernels as our joint kernel, as formulated in Assumption B3. Also note that the orthogonalization, which can be done via a Cholesky decomposition, is performed on V alone rather than all of $Z_t(\widehat{\beta})$. This is for technical reasons that $\underline{x}_t'\widehat{\beta}$ and V should enter in the kernel as a product since the latter is averaged out based on its empirical distribution: see the proofs in Appendix H.1, such as the proof of Lemma H.1.

Second, according to Assumption B6, the required polynomial order increases with d_V, the number of continuous index variables. When d_V is 1 or 2, as in our Monte Carlo and empirical illustration, any $\ell \geq 2$ is sufficient. Larger values of ℓ reduce the bias in the nonparametric approximation but may cause overfitting, especially in small samples. Our estimates are generally not sensitive to ℓ around 2 to 4 in our Monte Carlo simulations and empirical illustration. We use $\ell = 3$ in the Monte Carlo simulations and for estimators conditioning on $V'\gamma_0$ in the empirical illustration, and use $\ell = 2$ for estimators conditioning on V in the empirical illustration. The smaller ℓ is adopted for the latter case because we divide the observations into cells for discrete index variables, resulting in fewer observations in each cell: see Section 4.2.

Third, we modified the Gaussian kernel as follows to satisfy Assumption B3:

$$K(u) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) & \text{for } |u| \leq 5, \\ \frac{1}{\sqrt{2\pi}} \exp(-5^2/2) \cdot \left(4(6-|u|)^5 - 6(6-|u|)^4 + 3(6-|u|)^3\right) & \text{for } 5 < |u| \leq 6, \\ 0 & \text{for } |u| > 6. \end{cases}$$

This kernel is equivalent to the Gaussian kernel for $|u| \leq 5$ and their results are generally indistinguishable. The truncation at ± 6 ensures the compact support assumed in B3.(i). The quintic polynomial for $5 < |u| \leq 6$ guarantees the twice continuous differentiability assumed in B3.(ii).

Fourth, one needs to select a bandwidth $b_N = \kappa \cdot N^{-\delta}$. In principle, we could choose different bandwidths for $\underline{x}_t'\widehat{\beta}$ and V (after standardization and orthogonalization), but for simplicity, we keep it the same in the numerical exercises.

Here we first choose δ that satisfies our rate conditions in Assumption B6. We then find the scaling constant κ^* using a bootstrap over a finite grid: see Appendix D.2 for details. In our simulations and empirical illustration, κ^* usually ranges from 0.6 to 4, and the estimated ASF and APE are generally robust with respect to the scaling constant κ within a neighborhood of κ^*, such as within the range $[\kappa^* - 0.2,\, \kappa^* + 0.2]$.

Lastly, the compact set $\mathcal{Z}_t$ in the trimming function $\hat{\pi}_{it} = \mathbb{1}((\underline{x}_t'\widehat{\beta}, V_i) \in \mathcal{Z}_t)$ helps bound $f_{Z_t(\widehat{\beta})}(z)$ away from zero. Candidate criteria could be: a lower bound directly on $\widehat{f}_{Z_t(\widehat{\beta})}(z) = \frac{1}{N}\sum_{j=1}^N \mathcal{K}_{b_N}\left(\frac{Z_{jt}(\widehat{\beta}) - z}{b_N}\right)$, an upper bound on the condition number of $S_N(z; \widehat{\beta})$, and a lower bound on its determinant. We incorporate all three criteria to construct the trimming set in the numerical implementation.

Asymptotic variance estimation. To conduct inference on the ASF and APE, one could, in principle, estimate $\sigma_{\mathrm{ASF}_t}(\underline{x}_t'\beta_0)$ and $\sigma_{\mathrm{APE}_t}(\underline{x}_t'\beta_0)$ analytically. This can be done by estimating $\mathrm{Var}(Y_t | X_t'\beta_0 = \underline{x}_t'\beta_0, V_i)$ via local polynomial regressions of (Y_t, Y_t^2) on $(X_t'\widehat{\beta}, V)$, and replacing $f_{Z_t}(\underline{x}_t'\beta_0, V_i)$ by $\frac{1}{N}\sum_{j=1}^N \mathcal{K}_{b_N}\left(\frac{Z_{jt}(\widehat{\beta}) - (\underline{x}_t'\widehat{\beta}, V_i)}{b_N}\right)$, and $f_V(V_i)$ by $\frac{1}{N}\sum_{j=1}^N \mathcal{K}_{b_N}^V\left(\frac{V_j - V_i}{b_N}\right)$. In the numerical implementation, we instead focus on bootstrap-based inference, which may better capture higher-order terms in the asymptotic expansion of our estimator.

Multiple time periods. Finally, note that the above estimator is for the ASF (or APE/AME), at period t, which may vary with t in the population. If stationarity is further assumed, i.e., $(g_t, F_{U_t|C}) = (g_{t'}, F_{U_{t'}|C})$ for all $t, t' \in \{1, \ldots, T\}$, then $\mathrm{ASF}_t(\underline{x}_*) = \mathrm{ASF}_{t'}(\underline{x}_*)$ for any pair of time periods assuming that $\underline{x}_* \in \mathrm{supp}(X_t) \cap \mathrm{supp}(X_{t'})$. Then, the ASF does not depend on t, and we can combine ASF estimates from multiple time periods to obtain a more precise estimate. For example, we can average the estimated ASFs over time:

$$\overline{\widehat{\text{ASF}}}(\underline{x}_*) = \frac{1}{T}\sum_{t=1}^{T} \widehat{\text{ASF}}_t(\underline{x}_*).$$

We can further reduce its asymptotic variance by selecting weights that depend on t. However, weights that minimize the asymptotic variance of the weighted ASF depend on the inverse of an estimate of the joint asymptotic covariance matrix of $\widehat{\text{ASF}}_t(\underline{x}_*)$ across all $t = 1, \cdots, T$. For simplicity, we propose the simple time average as our rule of thumb.

D.2 Bandwidth Selection via Bootstrap

Let us take the APE as an example. The bandwidth selection for the ASF and AME can be implemented in a similar fashion. Recall that the bandwidth b_N equals $\kappa \cdot N^{-\delta}$ for a given $\delta > 0$ satisfying our rate conditions. Then, we want to select the tuning parameter κ by minimizing the integrated mean squared error

$$\text{IMSE}(\kappa) = \int_{\text{supp}(X_t)} \mathbb{E}\left[\left(\widehat{\text{APE}}_{k,t}(\underline{x}_t; \kappa N^{-\delta}) - \text{APE}_{k,t}(\underline{x}_t; 0)\right)^2\right] dF_{X_t}(\underline{x}_t),$$

where $\widehat{\text{APE}}_{k,t}(\underline{x}_t; b)$ is our estimated APE with bandwidth b, and $\text{APE}_{k,t}(\underline{x}_t; b)$ denotes the probability limit of $\widehat{\text{APE}}_{k,t}(\underline{x}_t; b)$ for a fixed bandwidth b. Note that $\text{APE}_{k,t}(\underline{x}_t; 0)$ is the true APE.

Since the IMSE depends on unknown population quantities, we first approximate

$$\mathbb{E}\left[\left(\widehat{\text{APE}}_{k,t}(\underline{x}_t; \kappa N^{-\delta}) - \text{APE}_{k,t}(\underline{x}_t; 0)\right)^2\right]$$

via

$$\frac{1}{S}\sum_{s=1}^{S}\left(\widehat{\text{APE}}_{k,t}^{*(s)}(\underline{x}_t; \kappa N^{-\delta}) - \widehat{\text{APE}}_{k,t}(\underline{x}_t; \kappa_0 N^{-\delta})\right)^2.$$

Here $\left\{\widehat{\text{APE}}_{k,t}^{*(s)}(\underline{x}_t; b)\right\}_{s=1}^{S}$ denote S draws of the estimated APE according to its bootstrap distribution. We let κ_0 be a constant close to 0 and small relative to potential choices of κ. Note that we cannot set $\kappa_0 = 0$ since the estimated APE is defined only when $\kappa > 0$. We also approximate $F_{X_t}(\underline{x}_t)$ via the empirical distribution of X_t.

More specifically, the bandwidth constant κ can be selected according to the following procedure.

Implementation procedure.

1. Generate a range of evaluation points $\underline{x}_{t,j}$, $j = 1, \ldots, J$, with weights $\hat{w}(\underline{x}_{t,j})$ determined from

the empirical distribution of X_t.

2. Choose κ_0 to be a small value and estimate $\widehat{\text{APE}}_{k,t}(\underline{x}_{t,j}; \kappa_0 N^{-\delta})$, $j = 1, \ldots, J$, based on the original data $\{Y_i, \mathbf{X}_i\}_{i=1}^N$.
3. Generate bootstrap samples $\{Y_i^{(s)}, \mathbf{X}_i^{(s)}\}_{i=1}^N$ for $s = 1, \ldots, S$.
4. For each bootstrap sample $s = 1, \ldots, S$ and each bandwidth κ on grid $\{\kappa_1, \ldots, \kappa_K\}$, calculate $\widehat{\text{APE}}_{k,t}^{*(s)}(\underline{x}_{t,j}; \kappa N^{-\delta})$ for $j = 1, \ldots, J$.
5. Choose $\kappa \in \{\kappa_1, \ldots, \kappa_K\}$ that minimizes

$$\widehat{\text{IMSE}}(\kappa; w) = \sum_{j=1}^{J} \left(\frac{1}{S} \sum_{s=1}^{S} \left(\widehat{\text{APE}}_{k,t}^{*(s)}(\underline{x}_{t,j}; \kappa N^{-\delta}) - \widehat{\text{APE}}_{k,t}(\underline{x}_{t,j}; \kappa_0 N^{-\delta}) \right)^2 \right) \hat{w}(\underline{x}_{t,j}).$$

In the Monte Carlo simulations and empirical illustration, we choose the number of bootstrap samples to be $S = 100$. We initialize κ_0 at 0.6 and increase it by 0.1 if a numerical issue occurs. The bandwidth grid ranges from κ_0 to 4 with increments of 0.1.

D.3 Estimated Indices

When the conditioning variable(s) take the form $V'\gamma_0$, we can implement the following three variations of the semiparametric estimator:

1. SP: the original three-step estimator.
   (a) First, estimate β_0 (possibly with smoothed maximum score if Y_t is binary).
   (b) Second, perform a local polynomial regression of Y_{it} on $(X_{it}'\widehat{\beta},\ V_i)$.
   (c) Third, average over V_i.
2. SP ($V'\gamma_0$): a three-step estimator for estimated indices.
   (a) First, estimate (β_0, γ_0) using Ichimura and Lee (1991).
   (b) Second, perform a local polynomial regression of Y_{it} on $(X_{it}'\widehat{\beta},\ V_i'\widehat{\gamma})$.
   (c) Third, average over V_i.
3. SP ($V'\gamma_0$, iter.): a four-step estimator for estimated indices.
   (a) First, estimate β_0 (possibly with smoothed maximum score if Y_t is binary).
   (b) Second, plug in $\widehat{\beta}$ into the objective function in Ichimura and Lee (1991) to estimate γ_0.
   (c) Third, perform a local polynomial regression of Y_{it} on $(X_{it}'\widehat{\beta},\ V_i'\widehat{\gamma})$.
   (d) Fourth, average over V_i.

Note that: (i) SP ($V'\gamma_0$) and SP ($V'\gamma_0$, iter.) assume the multiple index structure, which is more efficient when the assumption holds but is less robust to misspecification. (ii) SP ($V'\gamma_0$, iter.) reduces the dimension of numerical optimization in Ichimura and Lee (1991) and can achieve better numerical performance than SP ($V'\gamma_0$) for applications with higher dimensions of parameters.

E Extension to a Dynamic Model

We now present an extension of our identification results to a dynamic panel model.

E.1 Related Literature

There is a large literature on dynamic binary response models going back to Cox (1958). In particular, see Chamberlain (1985), Magnac (2000), Honoré and Kyriazidou (2000), and Honoré and Tamer (2006) for results on the identification of common coefficients. For recent results under a logistic error distribution, see Honoré and Weidner (2023), and Kitazawa (2021) for identification results for common coefficients, and Aguirregabiria and Carro (2021) and Dobronyi, Gu, and Kim (2021) for other functionals such as AMEs. Khan, Ponomareva, and Tamer (2023) establish a number of identification results for β_0 without assuming logistic errors, ranging from point identification to sharp bounds under minimal assumptions. Torgovitsky (2019) also obtains partial identification results without parametric restrictions. See Aristodemou (2021) and Khan, Ouyang, and Tamer (2021) for results on dynamic discrete response models. Also see Arellano and Bonhomme (2017) for a review of nonlinear dynamic panel data models.

E.2 Model and Identification

Our previous Assumption A1.(ii) in the main text rules out the dependence of X_t on past $U_{t'}$, thus preventing $\mathbf{X}$ from containing lagged outcome variables. We consider a model that assumes weak or sequential exogeneity. We distinguish between predetermined and exogenous regressors and denote them by $X_t \equiv \begin{pmatrix} X_{t,\text{pre}} & X_{t,\text{exog}} \end{pmatrix}$. Let $\mathbf{X}_{\text{exog}} = (X_{1,\text{exog}}, \ldots, X_{T,\text{exog}})$ denote all past, current, and future values of the exogenous regressor, and let $\mathbf{X}^t_{\text{pre}} = (X_{1,\text{pre}}, \ldots, X_{t,\text{pre}})$ denote all current and past values of the predetermined regressors. We assume that errors are conditionally independent of past, current, and future values of the exogenous regressors, as well as past and current values of the predetermined regressors.

Assumption A1†.(ii) (Sequential exogeneity) For each $t = 1, \ldots, T$, $U_t \perp\!\!\!\perp (\mathbf{X}_{\text{exog}}, \mathbf{X}^t_{\text{pre}})|C$.

This assumption replaces A1.(ii) and allows the future covariates to depend on the current error term U_t. In particular, it allows for the inclusion of lagged dependent variables in $\mathbf{X}$.

We also maintain Assumption A2 that states β_0 is point-identified. When compared to strict exogeneity, point-identification of the common coefficients under sequential exogeneity can be more challenging: see the literature in Appendix E.1. To fix ideas, let us consider a relatively simple dynamic binary model as a running example, where the only predetermined regressor is the lagged dependent variable, i.e., $X_{t,\text{pre}} = Y_{t-1}$. For notational simplicity, denote $\widetilde{X}_t = X_{t,\text{exog}}$. Let

$$Y_{it} = \mathbb{1}(\widetilde{X}_{it}'\widetilde{\beta}_0 + \rho_0 Y_{i\,t-1} + C_i - U_{it} \geq 0), \tag{E.1}$$

and define $\beta_0 = (\widetilde{\beta}_0, \rho_0)$. Versions of this binary outcome model with lagged dependent variables have been studied in Chamberlain (1985) and Honoré and Kyriazidou (2000), where they study the identification of β_0. Its identification generally requires the presence of units whose covariate values do not change over time, known as "stayers," which rules out the inclusion of time dummies in $\widetilde{X}_t$. It also requires a minimum number of time periods which is usually greater than 2. As shown in Honoré and Kyriazidou (2000), identification of β_0 can be achieved when U_t follows a logistic distribution. Furthermore, identification of β_0 can still be attained when U_t does not follow a logistic distribution, given additional conditions, such as the existence of a continuous regressor with full support. For example, Khan, Ponomareva, and Tamer (2023) provide such a result in their Theorem 3.

Given the identification of β_0, we can make a modified index assumption to help identify partial effects.

Assumption A3† (Dynamic index sufficiency) For $t \in \{1, \ldots, T\}$, given $V^t = v_t(\mathbf{X}_{\text{exog}}, \mathbf{X}_{\text{pre}}^t) \in \mathbb{R}^{d_V}$, where v_t is known, let $C|(\mathbf{X}_{\text{exog}}, \mathbf{X}_{\text{pre}}^t) \overset{d}{=} C|V^t$.

This assumption replaces A3 and allows the index to depend on all regressors except for future values of the predetermined regressor. The following theorem shows the identification of our partial effects in these models.

Theorem E.1 (Identification under weak exogeneity)**.** Let $t \in \{1, \ldots, T\}$, $\underline{x}_t \in \text{supp}(X_t)$, and $\underline{\mathbf{x}}^t \in \text{supp}(\mathbf{X}_{\text{exog}}, \mathbf{X}_{\text{pre}}^t)$. Let Assumptions A1–A3 hold with A1†.(ii) replacing A1.(ii), and A3† replacing A3. Then,

1. $\text{ASF}_t(\underline{x}_t) = \mathbb{E}[\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V^t]]$ is point identified when $\text{supp}(V^t|X_t'\beta_0 = \underline{x}_t'\beta_0) = \text{supp}(V^t)$;
2. Let the partial derivative of $\text{ASF}_t(\underline{x}_t)$ with respect to $\underline{x}_t^{(k)}$ exist. $\text{APE}_{k,t}(\underline{x}_t) = \mathbb{E}[\frac{\partial}{\partial \underline{x}_t^{(k)}}\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V^t]]$ is point identified when $\text{supp}(V^t|X_t'\beta_0 = u) = \text{supp}(V^t)$ for all u in a neighborhood of $\underline{x}_t'\beta_0$;
3. $\text{LAR}_{k,t}(\underline{\mathbf{x}}^t) = \frac{\partial}{\partial \underline{x}_t^{(k)}}\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]|_{v=v_t(\underline{\mathbf{x}}^t)}$ is point identified when the derivative exists and when $v_t(\underline{\mathbf{x}}^t) \in \text{supp}(V^t|X_t'\beta_0 = u)$ for all u in a neighborhood of $\underline{x}_t'\beta_0$;
4. $\text{AME}_{k,t} = \mathbb{E}[\frac{\partial}{\partial X_t^{(k)}}\mathbb{E}[Y_t|X_t'\beta_0, V^t]]$ is point identified if the above condition on $\text{LAR}_{k,t}(\underline{\mathbf{x}}^t)$ holds for all $\underline{\mathbf{x}}^t \in \text{supp}(\mathbf{X}_{\text{exog}}, \mathbf{X}_{\text{pre}}^t)$ up to a $\mathbb{P}_{\mathbf{X}_{\text{exog}}, \mathbf{X}_{\text{pre}}^t}$-measure zero set.

In the dynamic binary outcome model in equation (E.1) above, to identify the ASF at time $t = 1$, one can consider an index that depends on $(\widetilde{X}_1, \ldots, \widetilde{X}_T, Y_0)$, where Y_0 is the initial time period outcome. Specifically, let $V^1 = (\widetilde{v}(\widetilde{\mathbf{X}}), Y_0)$, where $\widetilde{\mathbf{X}} = \mathbf{X}_{\text{exog}}$ for notational simplicity. Assume that β_0 is identified, for example, from the identification results in Honoré and Kyriazidou (2000). Define $\mathcal{V}^1 = \text{supp}(V^1)$ and $\mathcal{V}^1(\underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0) = \text{supp}(V^1|\widetilde{X}_1'\beta_0 + Y_0\rho_0 = \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0)$. Then,

$$
\begin{aligned}
&\text{ASF}_1(\underline{\widetilde{x}}_1, \underline{y}_0) = \mathbb{E}[\mathbb{1}(U_1 \leq \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0 + C)] \\
&= \int_{\mathcal{V}^1} \mathbb{E}[\mathbb{1}(U_1 \leq \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0 + C)|\tilde{v}(\widetilde{\mathbf{X}}) = v, Y_0 = y_0]\, dF_{\widetilde{v}(\widetilde{\mathbf{X}}),Y_0}(v, y_0) \\
&= \int_{\mathcal{V}^1(\underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0)} \mathbb{E}[\mathbb{1}(U_1 \leq \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0 + C)|\tilde{v}(\widetilde{\mathbf{X}}) = v, Y_0 = y_0]\, dF_{\widetilde{v}(\widetilde{\mathbf{X}}),Y_0}(v, y_0) \\
&= \int_{\mathcal{V}^1(\underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0)} \mathbb{E}[\mathbb{1}(U_1 \leq \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0 + C)|\widetilde{X}_1'\beta_0 + Y_0\rho_0 = \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0, \tilde{v}(\widetilde{\mathbf{X}}) = v, Y_0 = y_0]\, dF_{\widetilde{v}(\widetilde{\mathbf{X}}),Y_0}(v, y_0) \\
&= \int_{\mathcal{V}^1(\underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0)} \mathbb{E}[\mathbb{1}(U_1 \leq \widetilde{X}_1'\widetilde{\beta}_0 + Y_0\rho_0 + C)|\widetilde{X}_1'\beta_0 + Y_0\rho_0 = \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0, \tilde{v}(\widetilde{\mathbf{X}}) = v, Y_0 = y_0]\, dF_{\widetilde{v}(\widetilde{\mathbf{X}}),Y_0}(v, y_0) \\
&= \mathbb{E}[\mathbb{E}[Y_1|\widetilde{X}_1'\beta_0 + Y_0\rho_0 = \underline{\widetilde{x}}_1'\widetilde{\beta}_0 + \underline{y}_0\rho_0, \tilde{v}(\widetilde{\mathbf{X}}), Y_0]].
\end{aligned}
$$

The first equality follows from $U_1 \perp\!\!\!\perp (\widetilde{X}_1, Y_0)|C$, the second from iterated expectations, and the third from the support assumption in the statement of Theorem E.1. The fourth follows from $(C, U_1) \perp\!\!\!\perp (\widetilde{X}_1'\beta_0 + Y_0\rho_0)|V^1$, which is implied by Assumptions A1†.(ii) and A3†, and the proof is similar to step 1 in the proof of Theorem 2.1. Finally, the last two equalities follow directly.

One can also identify the APE or AME under the appropriate support conditions on the index variables. Finding sufficient index variables in dynamic models is potentially more delicate than in static ones because the exchangeability of covariates across time is an unlikely justification in dynamic models. We leave an analysis of this task for future work.

F Monte Carlo Simulations

We conduct two sets of Monte Carlo simulation experiments based on binary panel data models with the conditioning variable being V in Case 1 and $V'\gamma_0$ in Case 2. We focus on the former while streamlining the discussion of the latter, as their main messages are similar. Both cases account for two key features: multidimensional index variables and flexible error distributions. For Monte Carlo simulations with logistic errors, please see the previous version of this paper (Liu, Poirier, and Shiu, 2021).

F.1 Case 1: Conditioning on V

The Monte Carlo design is summarized in Table F.1. Note that both X_t and V are 2-by-1 vectors. Covariates $X_t^{(k)},\ k = 1, 2$, are drawn from a bivariate standard normal distribution, which satisfies the support conditions in Theorem 2.1. Our choices of $N = 1500$ and $T = 10$ are directly comparable with the dataset in our empirical illustration on female labor force participation, in which $N = 1461$ and $T = 9$. We use "DGP xy" to indicate the data-generating process (DGP) with $f_{C|V}$ being type x and f_{U_t} being type y. The distribution of individual effects, $f_{C|V}$, is skewed in DGP 1y and bimodal in DGP 2y.[3] For the error term, we consider error distributions f_{U_t} that exhibit skewness in DGP x1 and fat-tails in DGP x2.

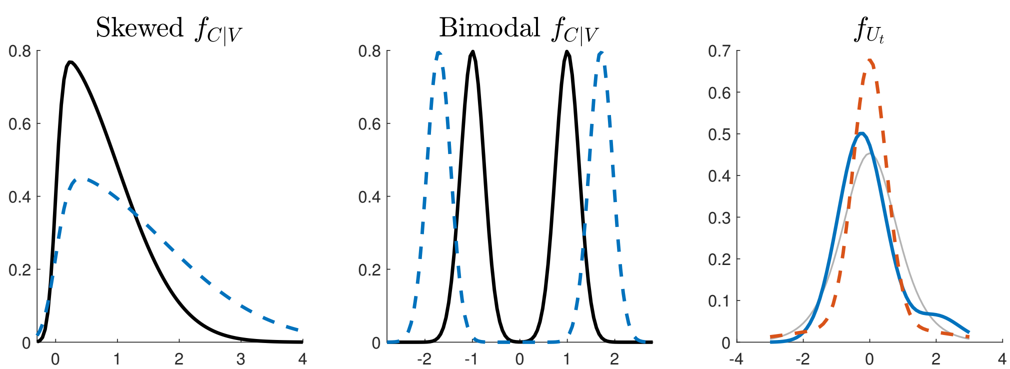
We evaluate the estimated ASF and APE based on a collection of $\underline{x}_t = \left(\underline{x}_t^{(1)}, \underline{x}_t^{(2)}\right)'$. We fix $\underline{x}_t^{(1)}$ at its population mean (i.e., $\underline{x}_t^{(1)} = 0$) and vary $\underline{x}_t^{(2)} \in [-1,\ 1]$, which covers 68% of the distribution of $X_t^{(2)}$. Given the non-logistic error distributions, we estimate β_0 using a smoothed maximum score estimator as in Charlier, Melenberg, and van Soest (1995) and Kyriazidou (1995), employing a fourth-order cdf kernel to satisfy the bandwidth requirement in Assumption B6. We normalize $|\widehat{\beta}^{(1)}| = 1$ since the identification of β_0 is up to scale. We use a local cubic regression (i.e., polynomial order $\ell = 3$) to estimate the conditional expectation of Y_t evaluated at $(\underline{x}_t'\widehat{\beta}, V)$. Finally, given the DGPs, the ASFs and APEs do not change over time, so we average the estimated ASFs and APEs across time periods. See Section D.1 for more details.

Figure F.1 compares the estimated APEs to the true APEs based on 100 Monte Carlo repetitions in each setup, and Figure F.2 plots the biases, standard deviations, and root mean square errors (RMSEs). Figures F.3 and F.4 show corresponding graphs for the ASF estimates. We see that the proposed semiparametric estimator better captures the peak in the skewed case and the valley in the bimodal case, whereas the RE and CRE reverse the valley in the bimodal case due to their parametric restrictions. As expected, the semiparametric estimator generates smaller biases

[3]Many empirical applications feature skewed and/or multimodal distributions of unobserved individual heterogeneity. For example, Liu (2023) estimated the latent productivity distribution of young firms, which exhibits a long right tail since good ideas are scarce. Also, Fisher and Jensen (2022) found two modes in the underlying skill distribution of mutual fund management—a primary mode with average ability and a secondary mode with poor performance.

Table F.1: Monte Carlo Design - Case 1

Model:	$Y_{it} = \mathbb{1}\left(X_{it}'\beta_0 + C_i - U_{it} \geq 0\right)$
Common param.:	$\beta_0 = (1,2)'$
Covariates:	$X_{it} \sim \mathcal{N}\left(0_{2\times 1}, I_2\right)$
Index:	$V_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$
Sample Size:	$N = 1500$, $T = 10$
# Repetitions:	$N_{sim} = 100$
$f_{C\|V}$:	
DGP 1y, skewed:	$C_i \| V_i \sim \left(\sum_{k=1}^{2}\left(V_i^{(k)}\right)^2 + 1\right) \cdot \mathcal{SN}(0,1,10)$
DGP 2y, bimodal:	$C_i \| V_i \sim \frac{1}{2}\mathcal{N}\left(\sum_{k=1}^{2}\left(V_i^{(k)}\right)^2 + 2,\ 1\right) + \frac{1}{2}\mathcal{N}\left(-\sum_{k=1}^{2}\left(V_i^{(k)}\right)^2 - 2,\ 1\right)$
f_{U_t}, with $\mathbb{E}\left(U_t\right) = 0$ and $\mathrm{Var}\left(U_t\right) = 1$:	
DGP x1, skewed:	$U_{it} \sim \frac{1}{9}\mathcal{N}\left(2, \frac{1}{2}\right) + \frac{8}{9}\mathcal{N}\left(-\frac{1}{4}, \frac{1}{2}\right)$
DGP x2, fat-tailed:	$U_{it} \sim \frac{1}{5}\mathcal{N}\left(0, 4\right) + \frac{4}{5}\mathcal{N}\left(0, \frac{1}{4}\right)$



Notes: $\mathcal{SN}(\xi,\omega,\alpha)$ denotes a skewed normal distribution with location parameter ξ, scale parameter ω, and shape parameter α, and its pdf is given by $f(x) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of a standard normal distribution. The two left panels depict $f_{C|V}$. The black solid and blue dashed lines are conditional on $\sqrt{\sum_{k=1}^{2}\left(V_i^{(k)}\right)^2} = 0$ and 0.5, respectively. The rightmost panel depicts f_{U_t}. The blue solid and red dashed lines are f_{U_t} in DGPs G.x1 (skewed) and G.x2 (fat-tailed), respectively. For reference, the thin gray line plots a rescaled logistic distribution with zero mean and unit variance.

and larger standard deviations than the RE and CRE. The improvement in bias dominates the deterioration in standard deviation for most covariate values in all setups. The difference between the RE and CRE is relatively negligible—their parametric assumptions in $f_{C|V}$ seem too restrictive and lead to considerable misspecification biases given current DGPs.

In Table F.2, the first three columns summarize the APE estimator's performance by computing weighted averages of biases, standard deviations, and RMSEs across the collection of evaluation points $\underline{x}_t$ with weights proportional to $f_{X_t}(\underline{x}_t)$. Similar to what we observed in Figures F.1 and F.2, the semiparametric estimator yields the smallest RMSE in all cases. The last three columns present the minimum, median, and maximum of the ratios of RMSE($\underline{x}_t$) to the true APE($\underline{x}_t$). The minimum, median, and maximum are taken over the collection of evaluation points $\underline{x}_t$. We see that the ratios range between 2.5% and 120% across all setups. Therefore, the RMSEs are generally sizeable compared to the true APEs, and thus the more precise semiparametric estimator would indeed make a significant difference. The RE and CRE have lower *minimal* ratios, which occurs at $\underline{x}_t$'s where the grey bands "intersect" with true APE curves; at the same time, the semiparametric estimator largely reduces the *median* and *maximal* ratios. For example, in DGP 22, the median (maximal) ratio of the semiparametric estimator is less than 1/3 (1/4) of its RE and CRE counterparts.

We also examine the estimation of the common parameter and the ASF in Table F.3. The structure of the ASF part of the table is the same as Table F.2 for the APE. The ratios of RMSE($\underline{x}_t$) to the true ASF($\underline{x}_t$) are generally smaller than their APE counterparts, and the semiparametric estimator again dominates the RE and CRE.

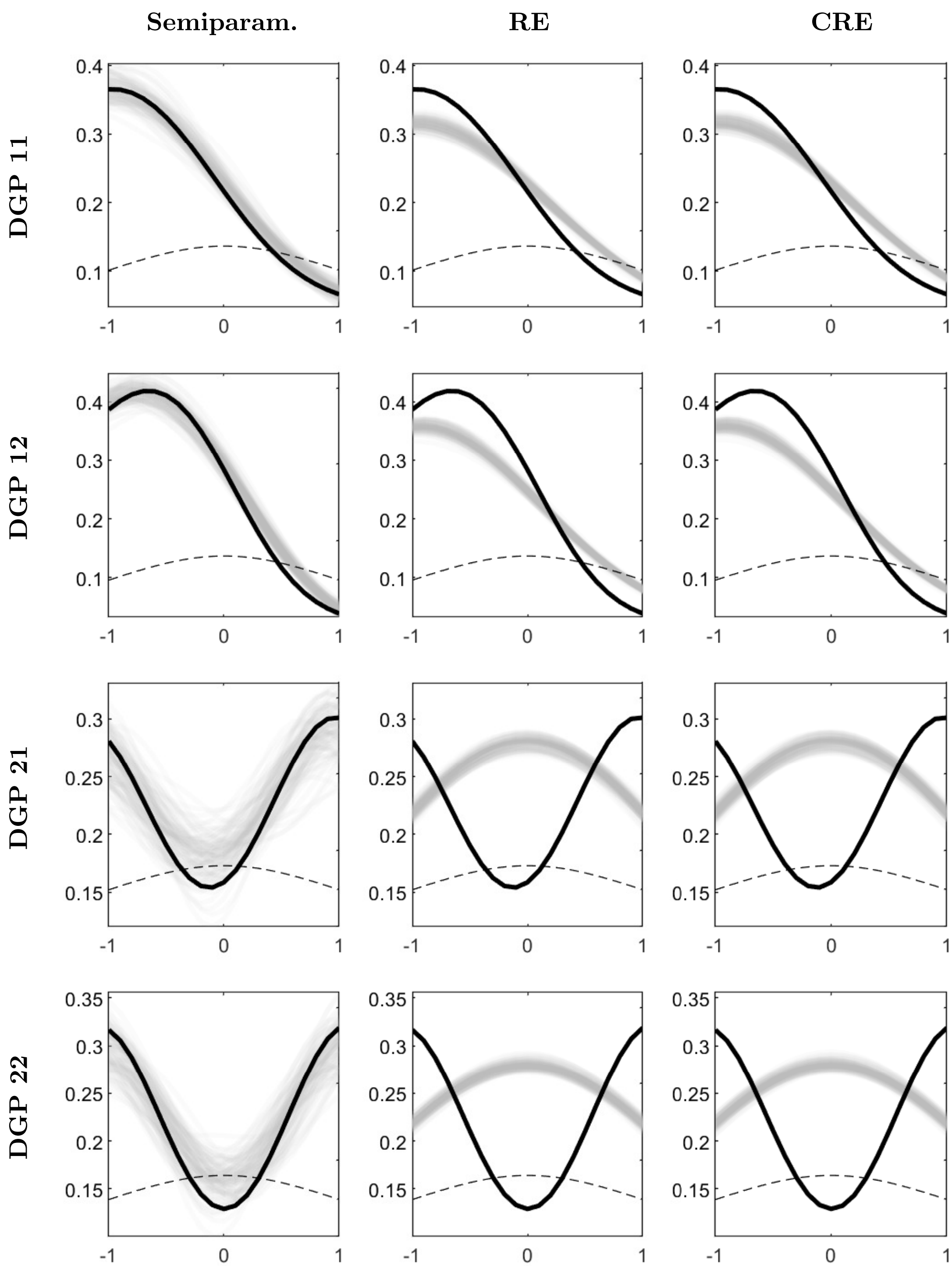
For $\widehat{\beta}$, the nonparametric smoothed maximum score estimator produces less biased but noisier estimates, and their RMSEs are larger than those of the RE and CRE. Nevertheless, the semiparametric estimator still better traces the shapes of the ASFs, hence providing the most accurate ASF estimates. Its RMSEs are around or less than half that of the RE and CRE. To take a closer look at how the β_0 estimation affects the APE estimation, we further examine an infeasible semiparametric estimator with known β_0 (see Table 8 in the previous version of this paper, Liu, Poirier, and Shiu (2021)). Results show that the smoothed maximum score estimates of β_0 slightly increase the absolute value of the bias, the standard deviation, and the RMSE, but the difference is minor—the flexible semiparametric estimator of the APE partially absorbs the effect of the slightly imprecisely estimated β_0.

Table F.2: APE Estimation - Monte Carlo Case 1

		|Bias|	SD	RMSE	Min	Med.	Max
	Semiparam.	0.013	0.012	**0.016**	4.2%	8.4%	15.7%
DGP 11	RE	0.028	0.005	0.029	2.7%	13.6%	39.1%
	CRE	0.028	0.005	0.029	2.7%	13.3%	39.1%
	Semiparam.	0.018	0.012	**0.020**	3.3%	6.0%	35.2%
DGP 12	RE	0.047	0.006	0.047	2.6%	18.3%	107.5%
	CRE	0.046	0.006	0.047	2.5%	18.4%	107.3%
	Semiparam.	0.019	0.018	**0.023**	7.2%	8.5%	20.5%
DGP 21	RE	0.071	0.004	0.071	3.1%	23.8%	81.5%
	CRE	0.071	0.004	0.071	3.0%	23.7%	81.7%
	Semiparam.	0.022	0.019	**0.026**	7.4%	9.3%	26.6%
DGP 22	RE	0.086	0.004	0.086	6.3%	31.1%	116.9%
	CRE	0.086	0.004	0.086	6.2%	31.0%	117.2%

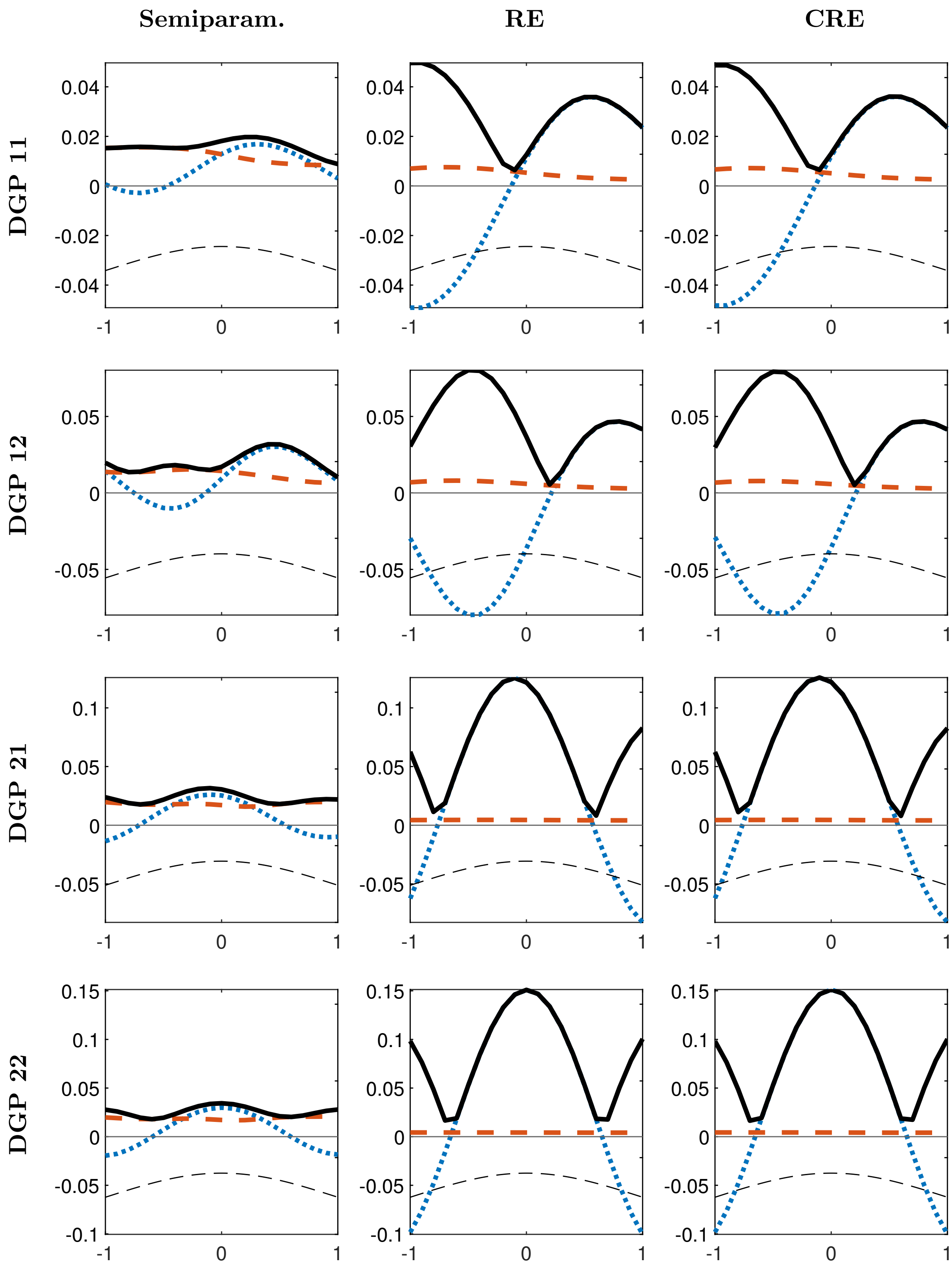
Notes: |Bias| indicates the absolute value of the bias. The reported |Bias|, SD, and RMSE are weighted averages across the collection of evaluation points $\underline{x}_t$, where the weights are proportional to $f_{X_t}(\underline{x}_t)$. The bold entries indicate the best estimator (i.e., with the smallest RMSE) for each DGP. The last three columns are the minimum/median/maximum of $\text{RMSE}(\underline{x}_t)/\text{APE}(\underline{x}_t) \times 100\%$ over $\underline{x}_t$.

Figure F.1: Estimated APEs vs True APEs - Monte Carlo Case 1



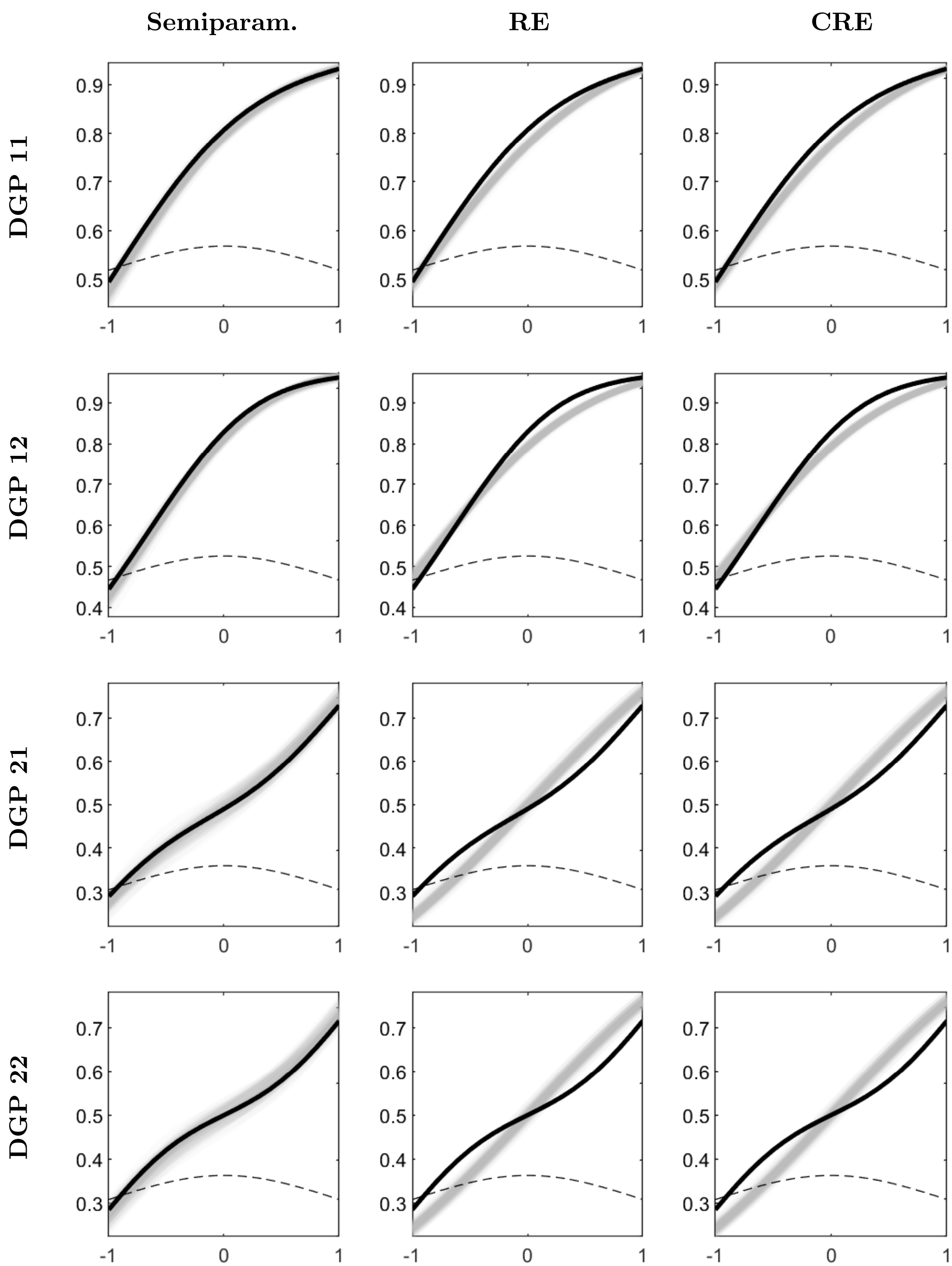
Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid lines are the true APEs. The gray bands are collections of lines where each line corresponds to the estimated APE based on one simulation repetition. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Figure F.2: Bias, Standard Deviation, and RMSE in APE Estimation - Monte Carlo Case 1



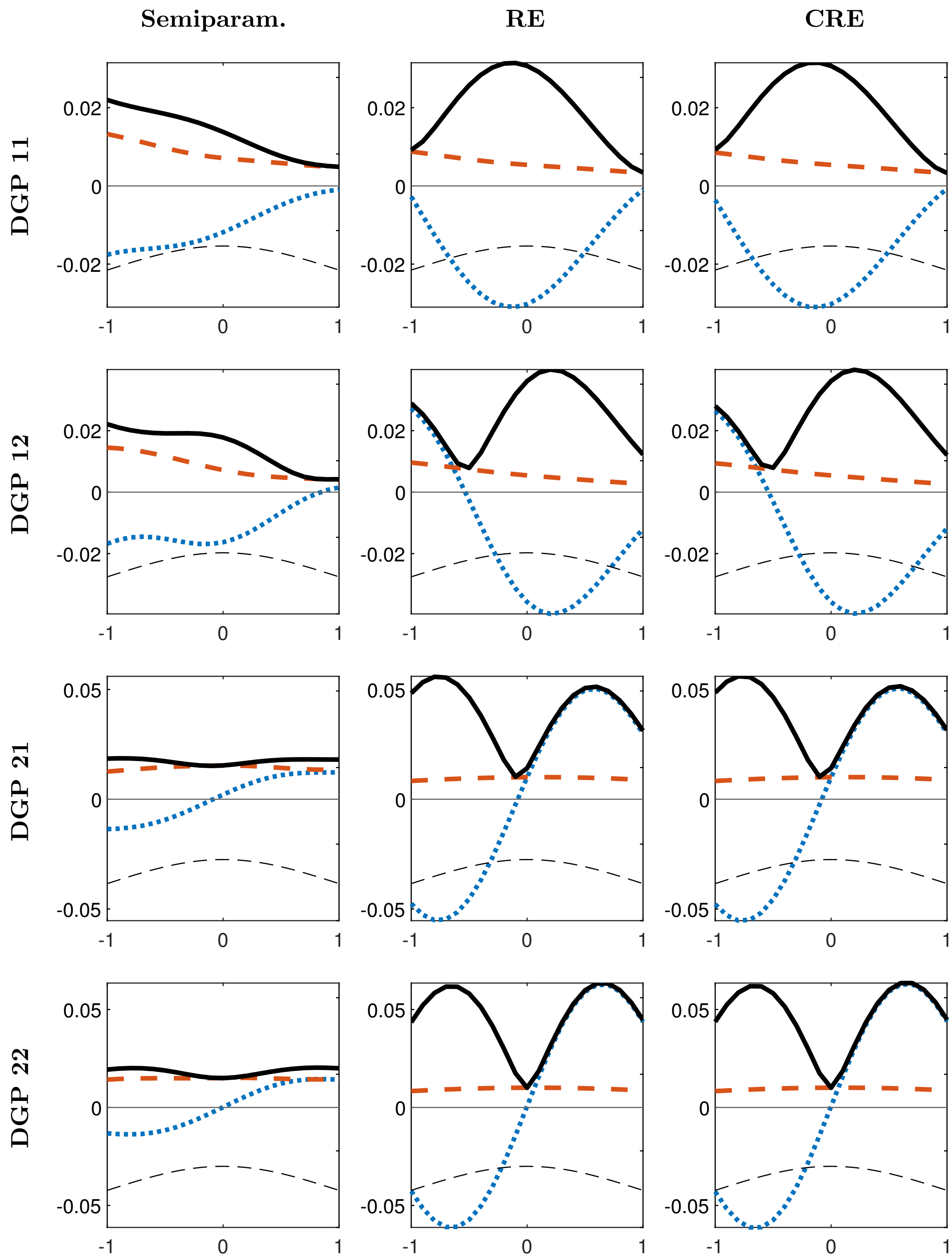
Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid / blue dotted / red dashed lines represent the RMSEs / biases / standard deviations of the APE estimates. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Figure F.3: Estimated ASFs vs True ASFs - Monte Carlo Case 1



Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid lines are the true ASFs. The gray bands are collections of lines where each line corresponds to the estimated ASF based on one simulation repetition. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Figure F.4: Bias, Standard Deviation, and RMSE in ASF Estimation - Monte Carlo Case 1



Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid / blue dotted / red dashed lines represent the RMSEs / biases / standard deviations of the ASF estimates. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Table F.3: Estimation of Common Parameter and ASF - Monte Carlo Case 1

		$\widehat{\beta}^{(2)}$			ASF					
		Bias	SD	RMSE	|Bias|	SD	RMSE	Min	Med.	Max
DGP 11	Semiparam.	0.011	0.031	0.033	0.011	0.008	**0.013**	0.5%	1.7%	4.4%
	RE	0.004	0.023	0.023	0.020	0.006	0.021	0.4%	2.8%	4.1%
	CRE	0.005	0.022	0.023	0.020	0.006	0.021	0.4%	2.8%	4.2%
DGP 12	Semiparam.	0.012	0.026	0.028	0.013	0.008	**0.015**	0.4%	2.1%	5.0%
	RE	0.005	0.019	0.020	0.025	0.006	0.026	1.2%	3.4%	6.5%
	CRE	0.006	0.019	0.020	0.025	0.006	0.026	1.2%	3.4%	6.3%
DGP 21	Semiparam.	0.015	0.064	0.065	0.014	0.014	**0.017**	2.5%	3.2%	6.4%
	RE	0.007	0.041	0.042	0.037	0.010	0.038	2.2%	7.7%	16.8%
	CRE	0.008	0.043	0.043	0.037	0.010	0.039	2.2%	7.7%	16.9%
DGP 22	Semiparam.	0.011	0.072	0.073	0.014	0.015	**0.018**	2.8%	3.3%	6.8%
	RE	0.004	0.043	0.043	0.044	0.010	0.045	2.0%	9.5%	16.9%
	CRE	0.005	0.044	0.044	0.044	0.010	0.045	2.0%	9.5%	17.0%

Notes: For the RE and CRE, we normalize $\widehat{\beta}$ such that $|\widehat{\beta}^{(1)}| = 1$ to ensure comparability across estimators. |Bias| indicates the absolute value of the bias. The reported |Bias|, SD, and RMSE of the ASF are weighted averages across the collection of evaluation points $\underline{x}_t$, where the weights are proportional to $f_{X_t}(\underline{x}_t)$. The bold entries indicate the best ASF estimator (i.e., with the smallest RMSE) for each DGP. The last three columns are the minimum/median/maximum of $\text{RMSE}(\underline{x}_t)/\text{ASF}(\underline{x}_t) \times 100\%$ over $\underline{x}_t$.

Table F.4: Monte Carlo Design - Case 2

Model:	$Y_{it} = \mathbb{1}\left(X_{it}'\beta_0 + C_i - U_{it} > 0\right)$
Common param.:	$\beta_0 = (1,2)'$, $\gamma_0 = (1,1)'$
Covariates:	$X_{it} \sim \mathcal{N}\left(0_{2\times 1}, I_2\right)$
Index:	$V_i'\gamma_0 = \frac{1}{T}\sum_{t=1}^{T} X_{it}'\gamma_0$
Sample Size:	$N = 1500$, $T = 10$
# Repetitions:	$N_{sim} = 100$
$f_{C\|V}$:	
DGP 1y, skewed:	$C_i\|V_i \sim V_i'\gamma_0 + \left((V_i'\gamma_0)^2 + 1\right) \cdot \mathcal{SN}(0,1,10)$
DGP 2y, bimodal:	$C_i\|V_i \sim V_i'\gamma_0 + \frac{1}{2}\mathcal{N}\left((V_i'\gamma_0)^2 + 2, 1\right) + \frac{1}{2}\mathcal{N}\left(-(V_i'\gamma_0)^2 - 2, 1\right)$
f_{U_t}, with $\mathbb{E}(U_{it}) = 0$ and $\mathrm{Var}(U_{it}) = 1$:	
DGP x1, skewed:	$U_{it} \sim \frac{1}{9}\mathcal{N}\left(2, \frac{1}{2}\right) + \frac{8}{9}\mathcal{N}\left(-\frac{1}{4}, \frac{1}{2}\right)$
DGP x2, fat-tailed:	$U_{it} \sim \frac{1}{5}\mathcal{N}(0,4) + \frac{4}{5}\mathcal{N}\left(0, \frac{1}{4}\right)$

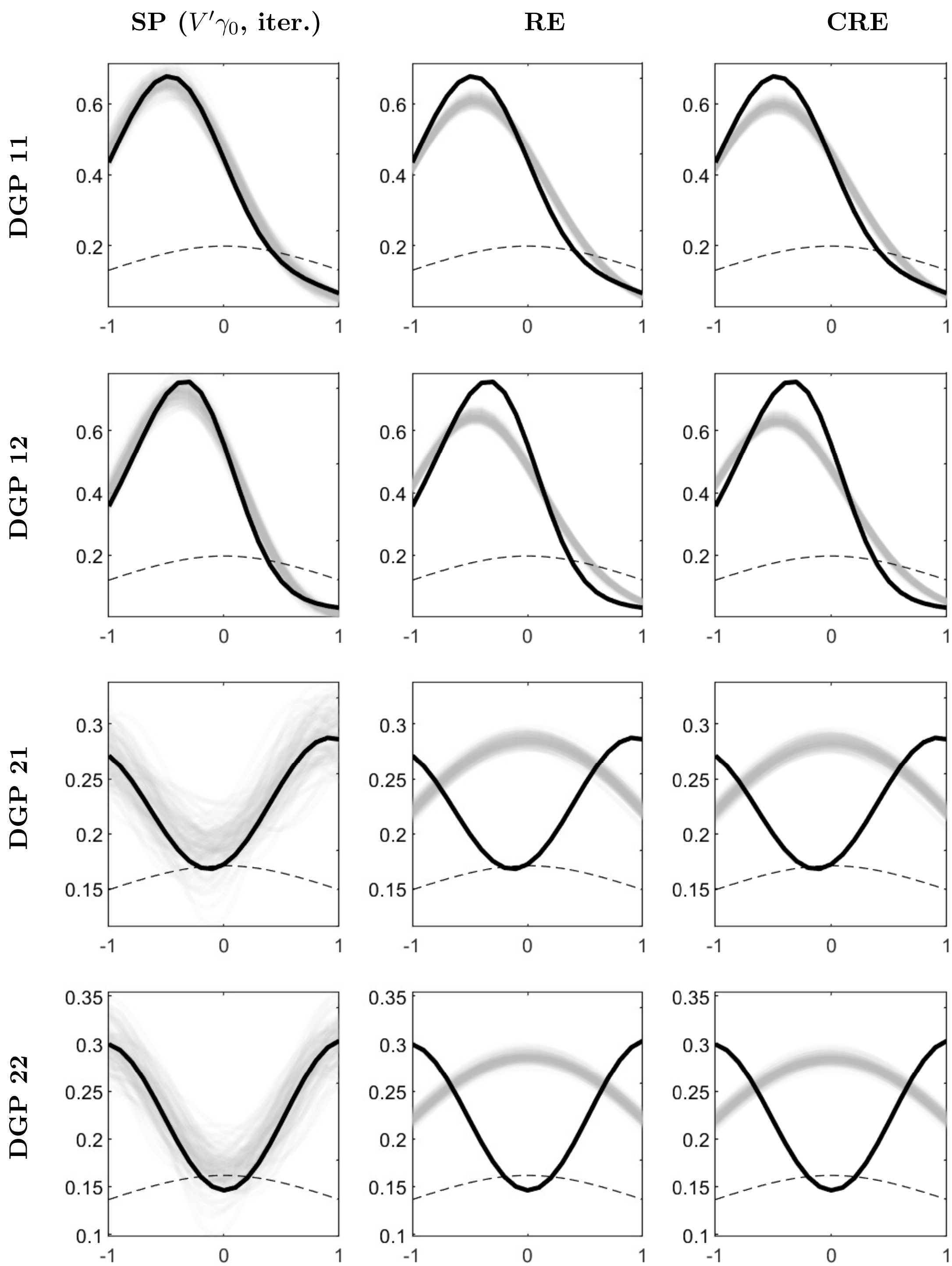
Notes: See the description in Table F.1.

F.2 Conditioning on $V'\gamma_0$, Estimated Indices

The Monte Carlo design is described in Table F.4, which is modified from Case 1. Now the distributions of individual effects, $f_{C|V}$, depend on a linear combination of V. Here we consider the three variations of the semiparametric estimator discussed in Appendix D.3: SP, SP $(V'\gamma_0)$, and SP $(V'\gamma_0$, iter.). In the current setup, there is no misspecification for all three variations of the semiparametric estimator.

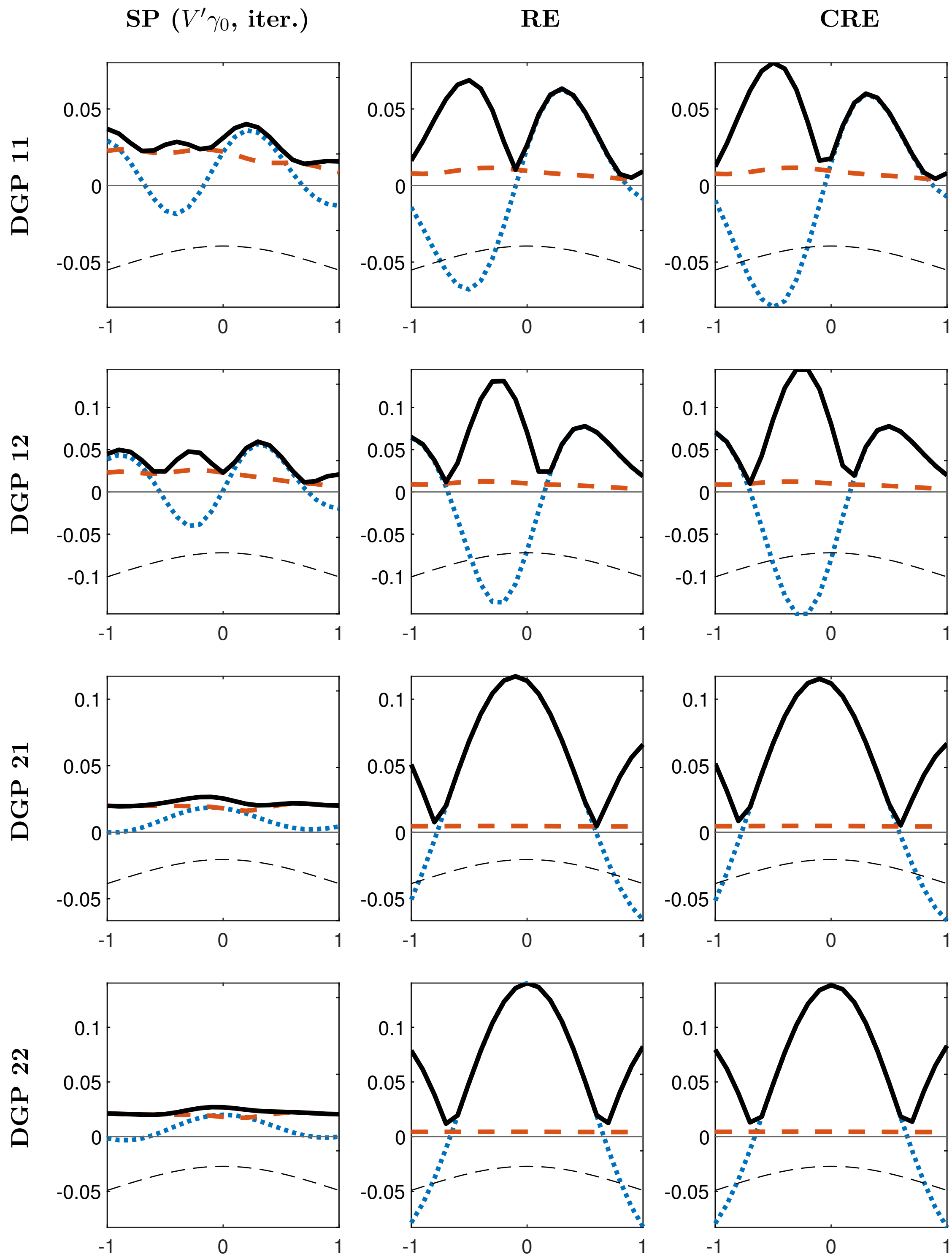
Figure F.5 shows the estimated APEs, and Figure F.6 depicts their corresponding biases, standard deviations, and RMSEs. Similarly, Figures F.7 and F.8 present the estimated ASFs and their corresponding statistics. Table F.5 reports these statistics for the APE estimators, and Table F.6 for the common parameter and the ASF. In terms of estimation performance, the differences across the three variations of the semiparametric estimator are relatively small and, similar to Case 1, they dominate the RE and CRE.

Figure F.5: Estimated APEs vs True APEs - Monte Carlo Case 2



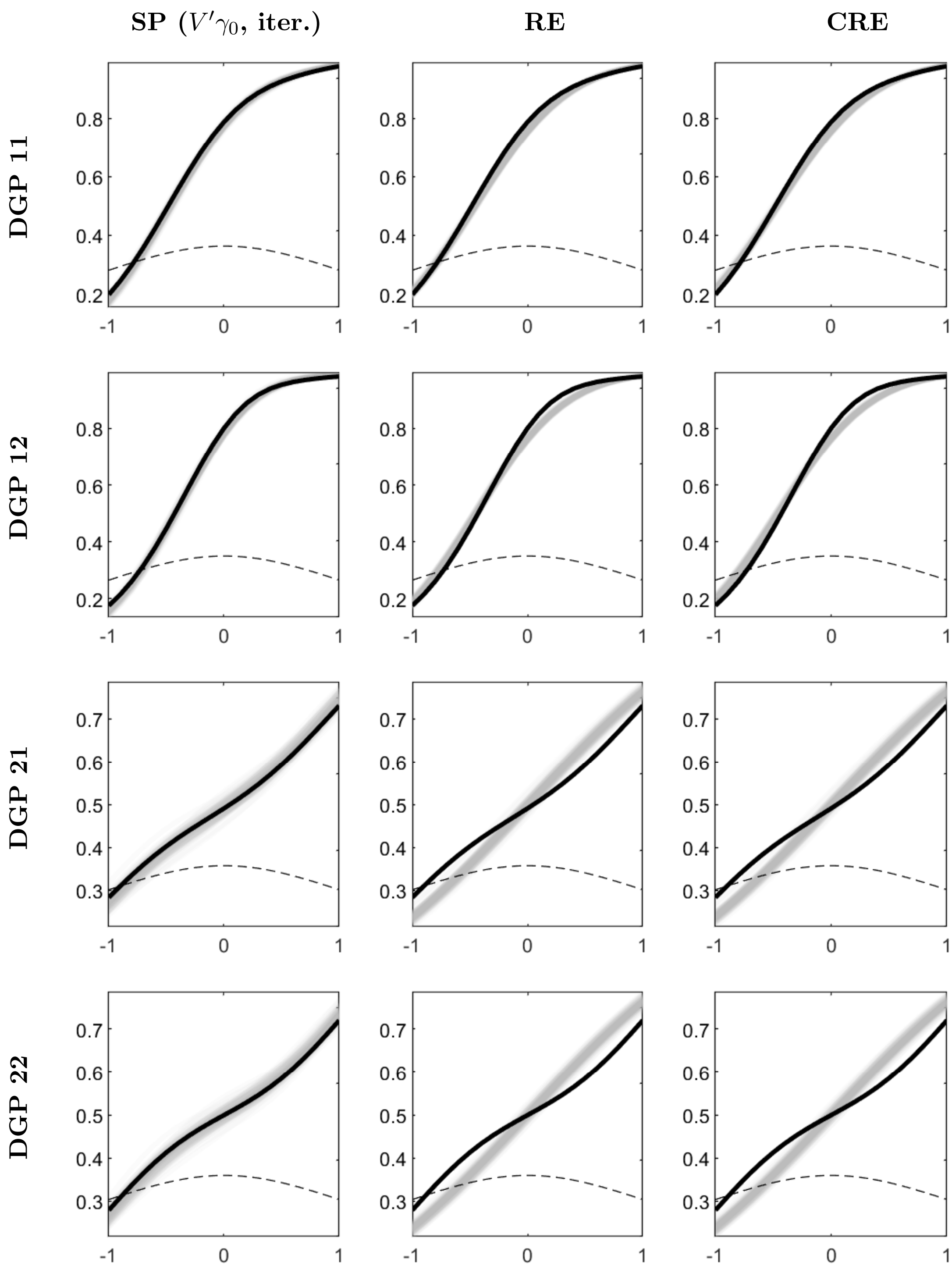
Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid lines are the true APEs. The gray bands are collections of lines where each line corresponds to the estimated APE based on one simulation repetition. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Figure F.6: Bias, Standard Deviation, and RMSE in APE Estimation - Monte Carlo Case 2



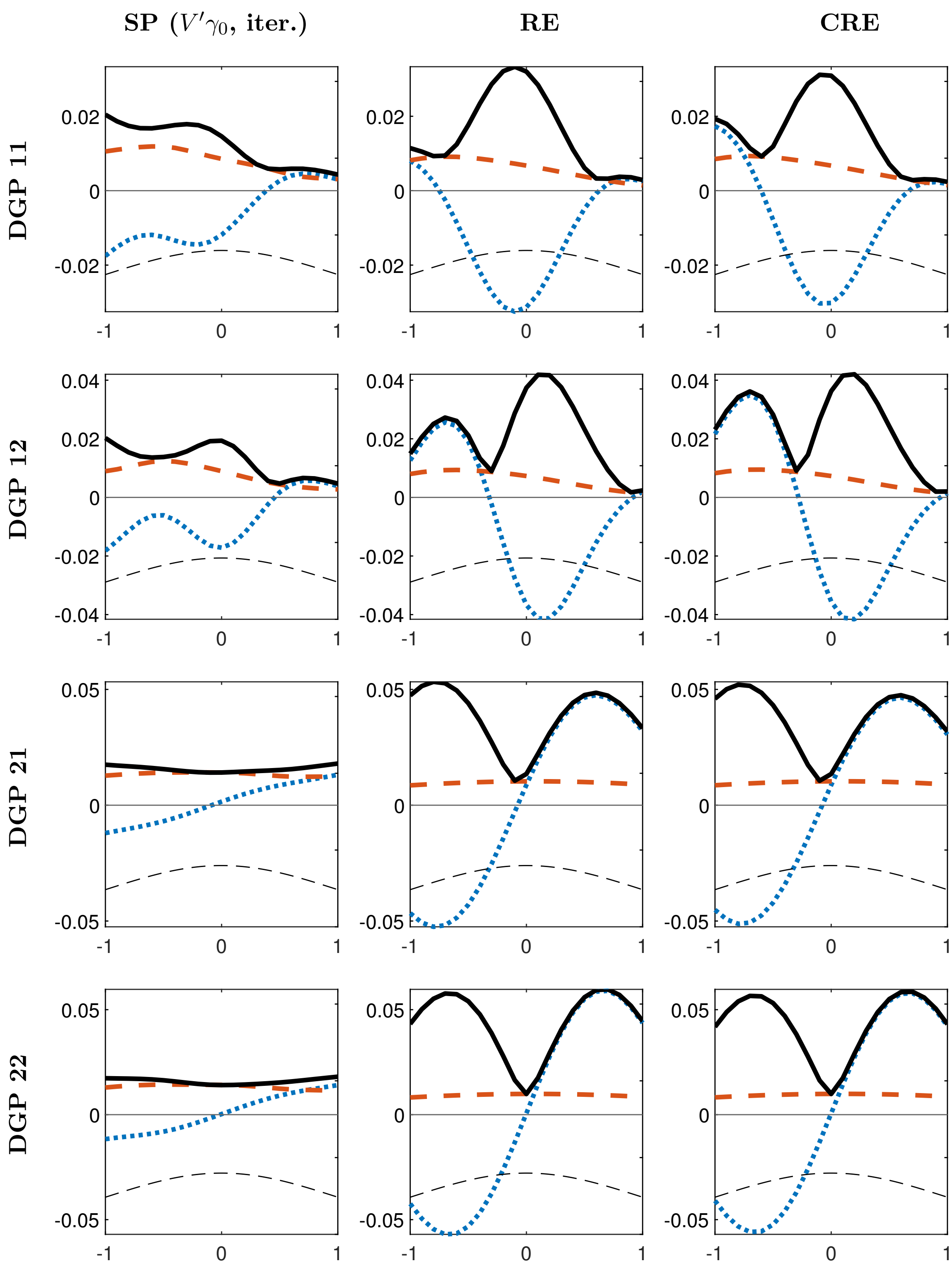
Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid / blue dotted / red dashed lines represent the RMSEs / biases / standard deviations of the APE estimates. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Figure F.7: Estimated ASFs vs True ASFs - Monte Carlo Case 2



Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid lines are the true ASF. The gray bands are collections of lines where each line corresponds to the estimated ASF based on one simulation repetition. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Figure F.8: Bias, Standard Deviation, and RMSE in ASF Estimation - Monte Carlo Case 2



Notes: The x-axes are potential values $\underline{x}_t^{(2)}$. The black solid / blue dotted / red dashed lines represent the RMSEs / biases / standard deviations of the ASF estimates. The thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}_t^{(2)}\right)$.

Table F.5: APE Estimation - Monte Carlo Case 2

		|Bias|	SD	RMSE	Min	Med.	Max
	SP ($V'\gamma_0$, iter.)	0.023	0.019	0.027	3.5%	8.5%	23.6%
	SP ($V'\gamma_0$)	0.021	0.023	**0.026**	4.1%	7.8%	23.3%
DGP 11	SP	0.024	0.019	0.028	3.1%	8.4%	30.3%
	RE	0.039	0.008	0.040	2.0%	9.4%	31.3%
	CRE	0.041	0.008	0.042	2.8%	10.6%	31.1%
	SP ($V'\gamma_0$, iter.)	0.032	0.019	0.036	3.4%	11.6%	66.5%
	SP ($V'\gamma_0$)	0.024	0.024	**0.030**	3.9%	8.7%	59.2%
DGP 12	SP	0.033	0.019	0.037	2.8%	10.2%	78.9%
	RE	0.064	0.009	0.065	2.0%	17.3%	98.6%
	CRE	0.069	0.009	0.070	1.7%	19.2%	100.6%
	SP ($V'\gamma_0$, iter.)	0.018	0.019	0.022	7.0%	9.0%	15.8%
	SP ($V'\gamma_0$)	0.017	0.020	**0.021**	7.6%	9.0%	13.5%
DGP 21	SP	0.018	0.018	0.022	6.9%	7.8%	17.1%
	RE	0.064	0.005	0.065	1.7%	20.9%	69.7%
	CRE	0.063	0.005	0.064	2.0%	20.2%	68.6%
	SP ($V'\gamma_0$, iter.)	0.019	0.020	**0.023**	6.8%	9.4%	18.4%
	SP ($V'\gamma_0$)	0.019	0.023	0.024	7.9%	11.1%	16.4%
DGP 22	SP	0.019	0.019	**0.023**	7.3%	8.7%	20.9%
	RE	0.078	0.004	0.078	4.5%	26.3%	96.1%
	CRE	0.077	0.004	0.077	5.0%	26.6%	94.8%

Notes: |Bias| indicates the absolute value of the bias. The reported |Bias|, SD, and RMSE are weighted averages across the collection of evaluation points $\underline{x}_t$, where the weights are proportional to $f_{X_t}(\underline{x}_t)$. The bold entries indicate the best estimator (i.e., with the smallest RMSE) for each DGP. The last three columns are the minimum/median/maximum of RMSE($\underline{x}_t$)/APE($\underline{x}_t$) $\times$ 100% over $\underline{x}_t$.

Table F.6: Estimation of Common Parameter and ASF - Monte Carlo Case 2

		$\widehat{\beta}^{(2)}$			ASF					
		Bias	SD	RMSE	|Bias|	SD	RMSE	Min	Med.	Max
	SP ($V'\gamma_0$, iter.)	0.019	0.050	0.053	0.011	0.008	**0.012**	0.4%	1.9%	10.3%
	SP ($V'\gamma_0$)	0.006	0.068	0.068	0.018	0.010	0.020	0.5%	1.7%	24.7%
DGP 11	SP	0.019	0.050	0.053	0.019	0.009	0.021	0.6%	2.4%	24.8%
	RE	-0.030	0.036	0.047	0.016	0.006	0.017	0.3%	2.9%	5.8%
	CRE	0.013	0.038	0.040	0.015	0.006	0.016	0.2%	2.7%	9.8%
	SP ($V'\gamma_0$, iter.)	0.006	0.042	0.042	0.011	0.008	**0.013**	0.5%	2.4%	11.5%
	SP ($V'\gamma_0$)	0.006	0.067	0.067	0.019	0.010	0.021	0.4%	2.1%	26.9%
DGP 12	SP	0.006	0.042	0.042	0.020	0.009	0.022	0.6%	3.1%	27.9%
	RE	-0.036	0.030	0.047	0.022	0.006	0.023	0.2%	3.9%	9.6%
	CRE	0.005	0.032	0.032	0.024	0.006	0.025	0.2%	3.6%	13.7%
	SP ($V'\gamma_0$, iter.)	0.010	0.063	0.064	0.012	0.013	0.015	2.5%	2.9%	6.2%
	SP ($V'\gamma_0$)	-0.057	0.120	0.132	0.011	0.013	**0.014**	2.0%	2.9%	4.5%
DGP 21	SP	0.010	0.063	0.064	0.014	0.015	0.018	2.9%	3.1%	7.7%
	RE	-0.007	0.041	0.042	0.035	0.010	0.037	2.2%	7.3%	16.7%
	CRE	0.003	0.042	0.042	0.035	0.010	0.036	2.2%	7.1%	16.2%
	SP ($V'\gamma_0$, iter.)	0.002	0.070	0.069	0.012	0.013	0.016	2.5%	2.8%	6.2%
	SP ($V'\gamma_0$)	-0.058	0.110	0.123	0.011	0.014	**0.015**	2.2%	2.8%	5.3%
DGP 22	SP	0.002	0.070	0.069	0.014	0.015	0.018	3.0%	3.1%	7.7%
	RE	-0.010	0.043	0.044	0.041	0.009	0.043	2.0%	8.8%	16.4%
	CRE	0.000	0.044	0.044	0.041	0.009	0.042	2.0%	8.6%	16.0%

Notes: For the RE and CRE, we normalize $\widehat{\beta}$ such that $|\widehat{\beta}^{(1)}| = 1$ to ensure comparability across estimators. |Bias| indicates the absolute value of the bias. The reported |Bias|, SD, and RMSE of the ASF are weighted averages across the collection of evaluation points $\underline{x}_t$, where the weights are proportional to $f_{X_t}(\underline{x}_t)$. The bold entries indicate the best ASF estimator (i.e., with the smallest RMSE) for each DGP. The last three columns are the minimum/median/maximum of RMSE$(\underline{x}_t)$/ASF$(\underline{x}_t) \times 100\%$ over $\underline{x}_t$.

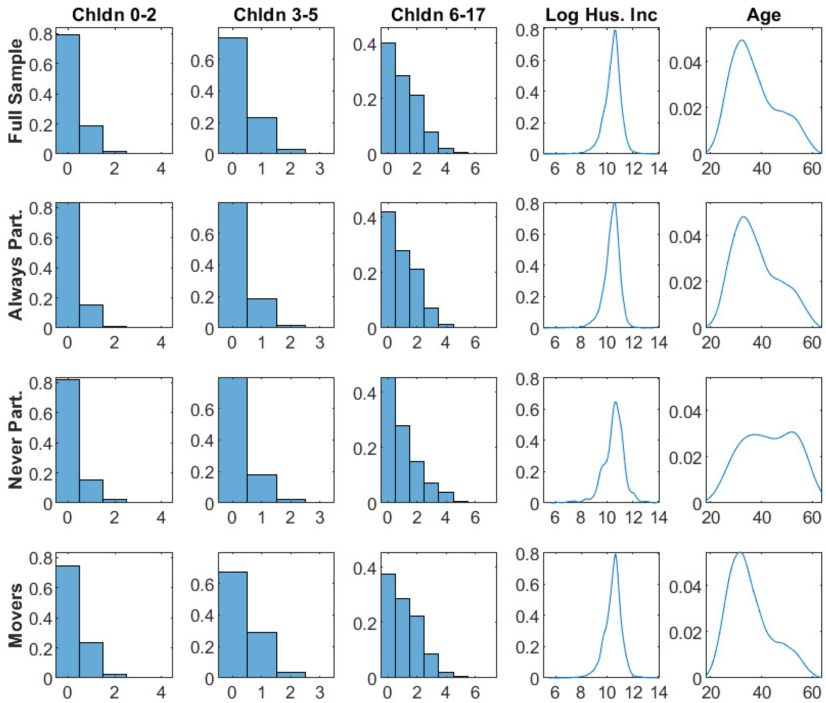
G Additional Figures and Tables for the Empirical Illustration

Figure G.1 plots the distributions of the covariates, and Table G.1 summarizes the corresponding descriptive statistics.

Figure G.2 depicts the estimated coefficients on time dummies which capture time-variation in aggregate participation rates. Point estimates of the time profiles are generally parallel to each other (from top to bottom: the smoothed maximum score, RE, and CRE) and yield higher participation rates after 1983, which coincides with the beginning of the Great Moderation. Most of the time-variation within each estimator and differences across estimators are insignificant at the 10% level, and standard errors generally increase with time for all three estimators. The smoothed maximum score yields the widest confidence band, as expected.

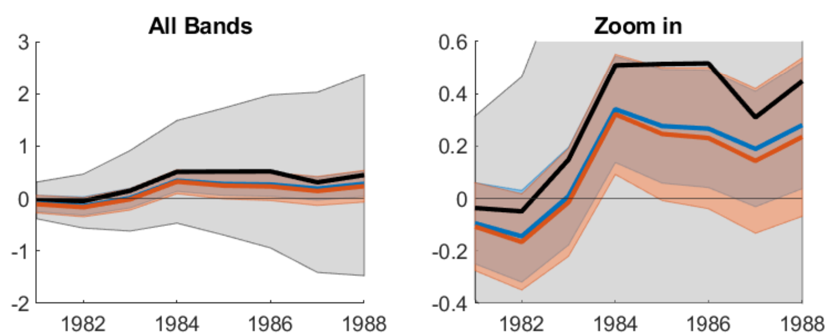
Figure G.3 plots the estimated ASF and APE based on alternative specifications. In the benchmark specification for the results in the main paper, we construct the indices based on the initial value of the covariates X_{i1} and use our original three-step semiparametric estimator without estimated indices (see Supplemental Appendix D.3 for a detailed comparison across different variations of the semiparametric estimator). To assess the sensitivity of our empirical findings to these choices, we conduct robustness checks with various alternative specifications. Specifically, we consider (i) V_i constructed from X_{i1} or $\overline{X}_i = \frac{1}{T}\sum_t X_{it}$, and (ii) with or without estimated indices ($V'\gamma_0$). Comparing with the benchmark specification in Figure 1, we see that, in general, the estimates do not change much as we vary the timing of V or incorporate estimated indices. For robustness checks with alternative coarsening schemes and the local logit estimator, see the previous version of this paper (Liu, Poirier, and Shiu, 2021).

Figure G.1: Distribution of Observables - Female Labor Force Participation



Notes: The sample consists of $N = 1461$ married women observed for $T = 9$ years from the PSID between 1980–1988. See Fernández-Val (2009) for details.

Figure G.2: Estimated Coefficients on Time Dummies - Female Labor Force Participation



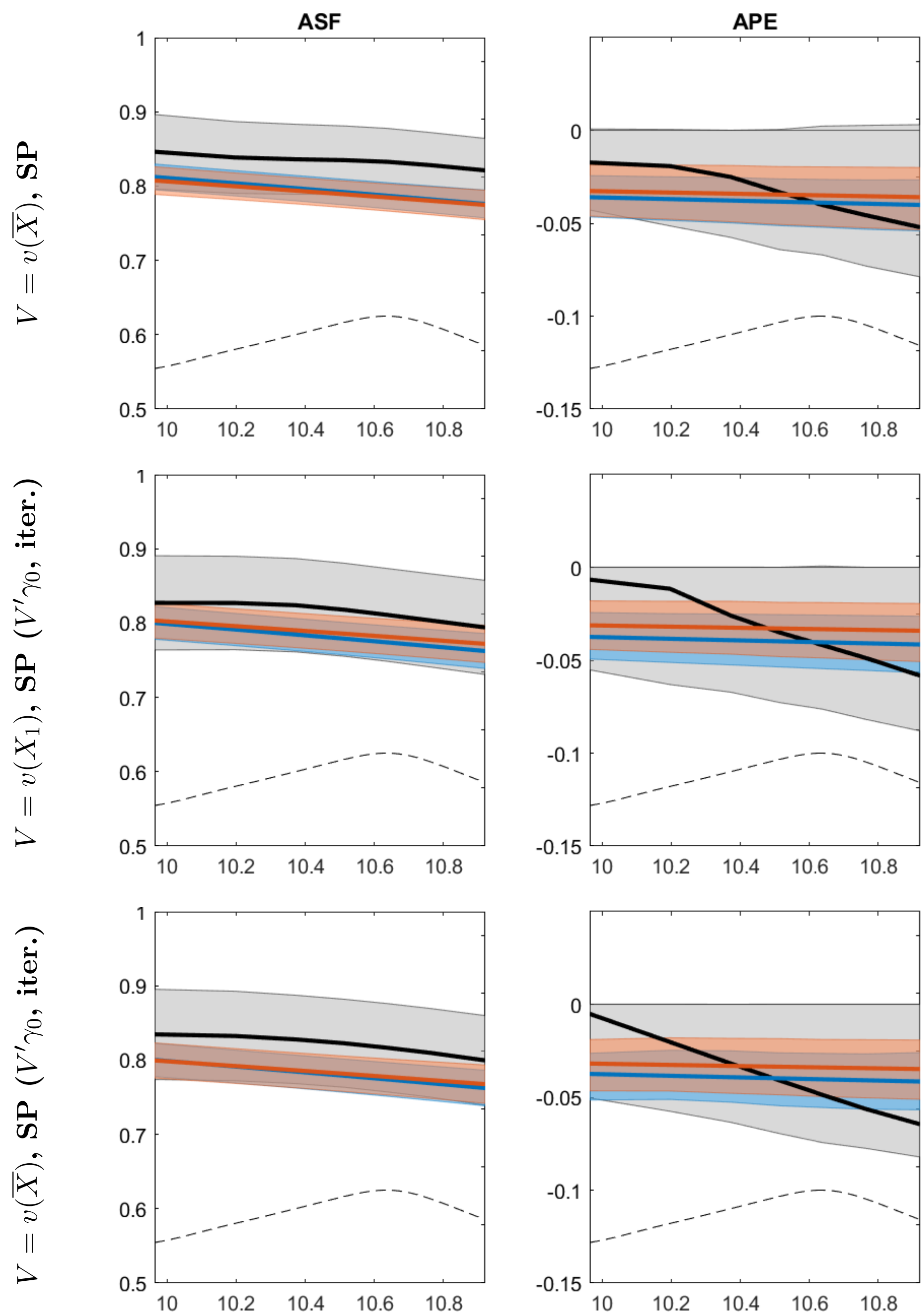
Notes: The black/blue/orange solid lines represent point estimates of the coefficients on time dummies using the smoothed maximum score/RE/CRE. The bands with corresponding colors indicate the 90% symmetric percentile-t confidence intervals based on bootstrap standard deviations. The right panel further zooms in on y-axis values between -0.4 and 0.6.

Table G.1: Descriptive Statistics - Female Labor Force Participation

	25%	Med.	75%	Mean	SD	Skew.	Kurt.
(a) Full Sample, #obs = $N \times T$ = 13,149							
Participate	-	-	-	0.72	0.45	-	-
Children 0–2	0	0	0	0.23	0.47	1.99	6.79
Children 3–5	0	0	1	0.29	0.51	1.60	4.85
Children 6–17	0	1	2	1.05	1.10	0.91	3.46
Log Husband's Income	10.09	10.51	10.83	10.43	0.69	-0.89	7.27
Age	30.00	35.00	43.00	37.30	9.22	0.56	2.50
(b) Always Participate, %obs = 46.27%							
Children 0–2	0	0	0	0.18	0.41	2.25	7.56
Children 3–5	0	0	0	0.23	0.46	1.93	6.12
Children 6–17	0	1	2	1.00	1.06	0.91	3.47
Log Husband's Income	10.08	10.47	10.77	10.37	0.65	-1.36	8.89
Age	31.00	36.00	44.00	37.98	9.04	0.51	2.45
(c) Never Participate, %obs = 8.28%							
Children 0–2	0	0	0	0.21	0.47	2.35	8.50
Children 3–5	0	0	0	0.23	0.48	2.05	6.79
Children 6–17	0	1	2	0.99	1.19	1.30	4.54
Log Husband's Income	10.13	10.62	11.04	10.53	0.85	-0.74	6.52
Age	35.00	43.00	52.00	42.98	10.09	-0.06	1.90
(d) Movers, %obs = 45.45%							
Participate	-	-	-	0.57	0.49	-	-
Children 0–2	0	0	1	0.28	0.51	1.70	5.74
Children 3–5	0	0	1	0.36	0.56	1.27	3.82
Children 6–17	0	1	2	1.11	1.11	0.83	3.18
Log Husband's Income	10.11	10.55	10.87	10.47	0.69	-0.59	5.81
Age	29.00	34.00	40.00	35.57	8.71	0.73	2.88

Notes: The sample consists of $N = 1461$ married women observed for $T = 9$ years from the PSID between 1980–1988. "Movers" refers to women who have participated in the labor market in some years but not all years. See Fernández-Val (2009) for details.

Figure G.3: Estimated ASF and APE - Female Labor Force Participation, Alternative Specifications



Notes: The x-axes are potential values of log husband's income. The blue/orange solid lines represent point estimates of the ASF and APE using the RE/CRE. The bands with corresponding colors indicate the 90% bootstrap confidence intervals. The thin dashed lines at the bottom of all panels show the distribution of log husband's income.

H Proofs

H.1 Proofs for Appendix C

We now present a sequence of lemmas that are used to prove our two main theorems of Appendix C: Theorem C.1 and Theorem C.2. When applied to matrices, let $\|\cdot\|$ denote the spectral norm.

Lemma H.1 (Convergence of S_N)**.** Suppose B1–B6 hold. Then,

$$\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})-S_N(z;\beta_0)\right\|=o_p\left(\frac{1}{\sqrt{Nb_N}}\right).$$

Proof of Lemma H.1. Select the same generic entry from matrices $S_N(z;\widehat{\beta})$ and $S_N(z;\beta_0)$. These entries can respectively be written as

$$S_N^{\tau,\tau'}(z;\widehat{\beta})\equiv\frac{1}{N}\sum_{j=1}^{N}\left(\frac{Z_{jt}(\widehat{\beta})-z}{b_N}\right)^{\tau}\left(\frac{Z_{jt}(\widehat{\beta})-z}{b_N}\right)^{\tau'}\mathcal{K}_{b_N}\left(\frac{Z_{jt}(\widehat{\beta})-z}{b_N}\right)$$

and

$$S_N^{\tau,\tau'}(z;\beta_0)\equiv\frac{1}{N}\sum_{j=1}^{N}\left(\frac{Z_{jt}(\beta_0)-z}{b_N}\right)^{\tau}\left(\frac{Z_{jt}(\beta_0)-z}{b_N}\right)^{\tau'}\mathcal{K}_{b_N}\left(\frac{Z_{jt}(\beta_0)-z}{b_N}\right),$$

where τ, τ' are vectors of exponents which satisfy $0\le|\tau|,|\tau'|\le\ell$. Let τ_1 and τ_1' denote the first components of τ and τ', and let τ_{-1} and τ_{-1}' denote vectors with all other components of τ and τ'. We can write

$$\begin{aligned}
&S_N^{\tau,\tau'}(z;\widehat{\beta})-S_N^{\tau,\tau'}(z;\beta_0)\\
&=\frac{1}{N}\sum_{j=1}^{N}\left[\left(\frac{X_{jt}'\widehat{\beta}-u}{b_N}\right)^{\tau_1+\tau_1'}\frac{1}{b_N}K\left(\frac{X_{jt}'\widehat{\beta}-u}{b_N}\right)-\left(\frac{X_{jt}'\beta_0-u}{b_N}\right)^{\tau_1+\tau_1'}\frac{1}{b_N}K\left(\frac{X_{jt}'\beta_0-u}{b_N}\right)\right]\\
&\cdot\left(\frac{V_j-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_{b_N}^{V}\left(\frac{V_j-v}{b_N}\right)\\
&=\frac{1}{N}\sum_{j=1}^{N}\left[\frac{1}{b_N}\Gamma\left(\frac{X_{jt}'\widehat{\beta}-u}{b_N}\right)-\frac{1}{b_N}\Gamma\left(\frac{X_{jt}'\beta_0-u}{b_N}\right)\right]\left(\frac{V_j-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_{b_N}^{V}\left(\frac{V_j-v}{b_N}\right)
\end{aligned}$$

where $\mathcal{K}_{b_N}^{V}(v)=b_N^{-d_V}\cdot\prod_{k=1}^{d_V}K(v_k)$, and $\Gamma(u)\equiv u^{\tau_1+\tau_1'}K(u)$ for generic $u\in\mathbb{R}$.

By B3, Γ is continuously differentiable. A first-order Taylor expansion yields

$$S_N^{\tau,\tau'}(z;\widehat{\beta})-S_N^{\tau,\tau'}(z;\beta_0)=\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b_N^2}\gamma\left(\frac{X_{jt}'\widetilde{\beta}-u}{b_N}\right)\left(\frac{V_j-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_{b_N}^{V}\left(\frac{V_j-v}{b_N}\right)X_{jt}'(\widehat{\beta}-\beta_0)$$

where $\widetilde{\beta}$ is such that $X_{jt}'\widetilde{\beta}$ is between $X_{jt}'\widehat{\beta}$ and $X_{jt}'\beta_0$, and where $\gamma(u) \equiv \Gamma'(u) = (\tau_1+\tau_1')u^{\tau_1+\tau_1'-1}K(u)+u^{\tau_1+\tau_1'}K'(u)$.

Since $\mathbb{P}(\widehat{\beta} \in \mathcal{B}_\varepsilon) \to 1$ as $N \to \infty$, with probability arbitrarily close to 1, we have that

$$
\begin{aligned}
&\sup_{z\in\mathcal{Z}_t}\left|S_N^{\tau,\tau'}(z;\widehat{\beta}) - S_N^{\tau,\tau'}(z;\beta_0)\right| \\
&\leq \frac{1}{b_N^2}\sup_{z\in\mathcal{Z}_t}\left\|\frac{1}{N}\sum_{j=1}^N \gamma\left(\frac{X_{jt}'\widetilde{\beta}-u}{b_N}\right)\left(\frac{V_j-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_{b_N}^V\left(\frac{V_j-v}{b_N}\right)X_{jt}\right. \\
&\left.-\mathbb{E}\left[\gamma\left(\frac{X_t'\widetilde{\beta}-u}{b_N}\right)\left(\frac{V-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_{b_N}^V\left(\frac{V-v}{b_N}\right)X_t\right]\right\|\|\widehat{\beta}-\beta_0\| \\
&+\sup_{z\in\mathcal{Z}_t}\frac{1}{b_N^2}\left\|\mathbb{E}\left[\mathcal{K}_{b_N}^V\left(\frac{V-v}{b_N}\right)\left(\frac{V-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\gamma\left(\frac{X_t'\widetilde{\beta}-u}{b_N}\right)X_t\right]\right\|\|\widehat{\beta}-\beta_0\| \\
&\leq \frac{1}{b_N^2}\sup_{z\in\mathcal{Z}_t,\beta\in\mathcal{B}_\varepsilon,b\in(0,\bar{b}]}\left\|\frac{1}{N}\sum_{j=1}^N \gamma\left(\frac{X_{jt}'\beta-u}{b}\right)\left(\frac{V_j-v}{b}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_b^V\left(\frac{V_j-v}{b}\right)X_{jt}\right. \\
&\left.-\mathbb{E}\left[\gamma\left(\frac{X_t'\beta-u}{b}\right)\left(\frac{V-v}{b}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_b^V\left(\frac{V-v}{b}\right)X_t\right]\right\|\|\widehat{\beta}-\beta_0\| \qquad \text{(H.1)} \\
&+\sup_{z\in\mathcal{Z}_t,\beta\in\mathcal{B}_\varepsilon}\frac{1}{b_N^2}\left\|\mathbb{E}\left[\gamma\left(\frac{X_t'\beta-u}{b_N}\right)\left(\frac{V-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}_{b_N}^V\left(\frac{V-v}{b_N}\right)X_t\right]\right\|\|\widehat{\beta}-\beta_0\|, \qquad \text{(H.2)}
\end{aligned}
$$

where $\bar{b} > 0$. To obtain the stochastic order of term (H.1), define the class of functions

$$\widetilde{\mathcal{F}} = \left\{\gamma\left(\frac{X_t'\beta-u}{b}\right) : u \in \mathbb{R}, \beta \in \mathcal{B}_\varepsilon, b \in (0,\bar{b}]\right\}.$$

These functions are of the form $\gamma(X_t'c+d)$ where $c = \beta/b$ and $d = -u/b$. Since K has a bounded domain and is twice continuously differentiable with bounded derivatives (Assumption B3), the function $\gamma(u)$ is of bounded variation on $\mathbb{R}$. By Nolan and Pollard (1987) Lemma 22.(ii), the above class of functions is Euclidean. It is also bounded since K is bounded. Similarly, the classes

$$\mathcal{F}_{V_k} = \left\{\left(\frac{V_k-v_k}{b}\right)^{\tau_{k+1}+\tau_{k+1}'}K\left(\frac{V_k-v_k}{b}\right) : v_k \in \mathbb{R}, b \in (0,\bar{b}]\right\}$$

are Euclidean and bounded for $k = 1,\ldots,d_V$ by the same argument as above. Here τ_{k+1} and τ_{k+1}' denote the $(k+1)$th components of τ and τ'. The product of bounded Euclidean classes is also bounded and Euclidean, hence

$$\mathcal{F}_V = \left\{\gamma\left(\frac{X_t'\beta-u}{b}\right)\left(\frac{V-v}{b}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}^V\left(\frac{V-v}{b}\right) : z \in \mathcal{Z}_t, \beta \in \mathcal{B}_\varepsilon, b \in (0,\bar{b}]\right\}$$

is bounded and Euclidean. By B5, $\mathbb{E}[\|X_t\|^2] < \infty$. Hence, by Lemma 2.14 (ii) in Pakes and Pollard (1989), the class

$$\mathcal{F} = \left\{ \gamma \left(\frac{X_t'\beta - u}{b} \right) \left(\frac{V - v}{b} \right)^{\tau_{-1} + \tau'_{-1}} \mathcal{K}^V \left(\frac{V - v}{b} \right) X_t : z \in \mathcal{Z}_t, \beta \in \mathcal{B}_\varepsilon, b \in (0, \bar{b}] \right\}$$

is also Euclidean, and hence Donsker. Therefore, by the continuous mapping theorem,

$$\begin{aligned} &\frac{1}{\sqrt{N} b_N^{2+d_V}} \sup_{z \in \mathcal{Z}_t, \beta \in \mathcal{B}_\varepsilon, b \in (0, \bar{b}]} \left\| \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left\{ \gamma \left(\frac{X_{jt}'\beta - u}{b} \right) \left(\frac{V_j - v}{b} \right)^{\tau_{-1} + \tau'_{-1}} \mathcal{K}^V \left(\frac{V_j - v}{b} \right) X_{jt} \right. \right. \\ &\left. \left. - \mathbb{E} \left[\gamma \left(\frac{X_t'\beta - u}{b} \right) \left(\frac{V - v}{b} \right)^{\tau_{-1} + \tau'_{-1}} \mathcal{K}^V \left(\frac{V - v}{b} \right) X_t \right] \right\} \right\| \\ &= \frac{1}{\sqrt{N b_N^{4+2d_V}}} \cdot O_p(1) \\ &= O_p \left((N b_N^{4+2d_V})^{-1/2} \right). \end{aligned}$$

Thus, term (H.1) can be written as

$$\begin{aligned} &\frac{1}{b_N^2} \sup_{z \in \mathcal{Z}_t, \beta \in \mathcal{B}_\varepsilon, b \in (0, \bar{b}]} \left\| \frac{1}{N} \sum_{j=1}^{N} \gamma \left(\frac{X_{jt}'\beta - u}{b} \right) \left(\frac{V_j - v}{b} \right)^{\tau_{-1} + \tau'_{-1}} \mathcal{K}_b^V \left(\frac{V_j - v}{b} \right) X_{jt} \right. \\ &\left. - \mathbb{E} \left[\gamma \left(\frac{X_t'\beta - u}{b} \right) \left(\frac{V - v}{b} \right)^{\tau_{-1} + \tau'_{-1}} \mathcal{K}_b^V \left(\frac{V - v}{b} \right) X_t \right] \right\| \|\widehat{\beta} - \beta_0\| \\ &= O_p \left((N b_N^{4+2d_V})^{-1/2} \right) \cdot O_p(a_N^{-1}) \\ &= o_p \left((N b_N)^{-1/2} \right), \end{aligned}$$

where the last line follows from $a_N^2 b_N^{3+2d_V} \to \infty$ as $N \to \infty$ (Assumption B6).

To bound term (H.2), we first note that

$$\begin{aligned} &\frac{1}{b_N^2} \left\| \mathbb{E} \left[\gamma \left(\frac{X_t'\beta - u}{b_N} \right) \left(\frac{V - v}{b_N} \right)^{\tau_{-1} + \tau'_{-1}} \mathcal{K}_{b_N}^V \left(\frac{V - v}{b_N} \right) X_t \right] \right\| \\ &= \left\| \mathbb{E} \left[\frac{\partial}{\partial \beta} \left(\frac{Z_t(\beta) - z}{b_N} \right)^{\tau + \tau'} \mathcal{K}_{b_N} \left(\frac{Z_t(\beta) - z}{b_N} \right) \right] \right\| \\ &= \left\| \int \frac{\partial}{\partial \beta} \left(\frac{\widetilde{z} - z}{b_N} \right)^{\tau + \tau'} \mathcal{K}_{b_N} \left(\frac{\widetilde{z} - z}{b_N} \right) f_{Z_t(\beta)}(\widetilde{z}) \, d\widetilde{z} \right\| \\ &= \left\| \int a^{\tau + \tau'} \mathcal{K}(a) \frac{\partial}{\partial \beta} f_{Z_t(\beta)}(z + a b_N) \, da \right\|. \end{aligned}$$

The last equality follows from the change of variables $\widetilde{z} = z + ab_N$. We then have that

$$\sup_{z\in\mathcal{Z}_t,\beta\in\mathcal{B}_\varepsilon}\left\|\int a^{\tau+\tau'}\mathcal{K}\left(a\right)\frac{\partial}{\partial\beta}f_{Z_t(\beta)}(z+ab_N)\,da\right\| \leq \sup_{z\in\mathcal{Z}_t,\beta\in\mathcal{B}_\varepsilon}\left\|\frac{\partial}{\partial\beta}f_{Z_t(\beta)}(z)\right\|\left|\int a^{\tau+\tau'}\mathcal{K}\left(a\right)\,da\right|$$

$$< \infty.$$

To see that the last inequality holds, recall Assumption B4.(ii), and that $\mathcal{K}$ is a bounded function with compact support, hence $a^{\tau+\tau'}\mathcal{K}(a)$ is bounded with compact support. Therefore, term (H.2) is of order $O(1)\cdot\|\widehat{\beta}-\beta_0\| = O_p(a_N^{-1}) = o_p\left((Nb_N)^{-1/2}\right)$ since, by B6, $Nb_Na_N^{-2}\to 0$ as $N\to\infty$.

Combining the rates of convergence of terms (H.1) and (H.2), we obtain

$$\sup_{z\in\mathcal{Z}_t}\left|S_N^{\tau,\tau'}(z;\widehat{\beta}) - S_N^{\tau,\tau'}(z;\beta_0)\right| = o_p\left(\frac{1}{\sqrt{Nb_N}}\right)$$

Since this rate of convergence applies uniformly in $z\in\mathcal{Z}_t$ to a generic element of $S_N^{\tau,\tau'}(z;\widehat{\beta}) - S_N^{\tau,\tau'}(z;\beta_0)$, it also applies uniformly in $z\in\mathcal{Z}_t$ to the matrix norm of $S_N(z;\widehat{\beta}) - S_N(z;\beta_0)$, which concludes the proof. □

Define

$$S(z;\beta_0) = \int \xi(a)\xi(a)'\mathcal{K}(a)\,da\cdot f_{Z_t(\beta_0)}(z).$$

Lemma H.2 (Convergence of S_N to S)**.** Suppose B1–B6 hold. Then,

$$\sup_{z\in\mathcal{Z}_t}\|S_N(z;\beta_0) - S(z;\beta_0)\| = O_p\left(\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)^{1/2}\right) + O(b_N).$$

Proof of Lemma H.2. This is Corollary 1.(ii) in Masry (1996) with $\theta = 1$ (in his notation), therefore we verify its assumptions. His condition 1(b) holds by B4.(iv). His conditions 2 and 3 hold by B3 and B4.(iii). Finally, the rate conditions of Theorem 2 in Masry (1996) hold by B6. Therefore, all assumptions of his corollary hold and the above result holds. □

Lemma H.3 (Convergence of T_N)**.** Suppose B1–B6 hold. Then,

$$\sup_{z\in\mathcal{Z}_t}\left\|T_N(z;\widehat{\beta}) - T_N(z;\beta_0)\right\| = o_p\left(\frac{1}{\sqrt{Nb_N}}\right).$$

Proof of Lemma H.3. Select the same generic component from $T_N(z;\widehat{\beta})$ and $T_N(z;\beta_0)$. These components can respectively be written as

$$T_N^\tau(z;\widehat{\beta}) \equiv \frac{1}{N}\sum_{j=1}^{N}\left(\frac{Z_{jt}(\widehat{\beta})-z}{b_N}\right)^{\tau} Y_{jt}\mathcal{K}_{b_N}\left(\frac{Z_{jt}(\widehat{\beta})-z}{b_N}\right)$$

$$T_N^\tau(z;\beta_0) \equiv \frac{1}{N}\sum_{j=1}^{N}\left(\frac{Z_{jt}(\beta_0)-z}{b_N}\right)^{\tau} Y_{jt}\mathcal{K}_{b_N}\left(\frac{Z_{jt}(\beta_0)-z}{b_N}\right),$$

where τ is a vector of exponents which satisfies $0 \le |\tau| \le \ell$. Again let τ_1 denote the first component of τ and let τ_{-1} denote all other components of τ. Let $\Gamma(u) \equiv u^{\tau_1}K(u)$ and $\gamma(u) \equiv \Gamma'(u) = \tau_1 u^{\tau_1-1}K(u) + u^{\tau_1}K'(u)$. As in the proof of Lemma H.1, we write

$$\begin{aligned}
&T_N^\tau(z;\widehat{\beta}) - T_N^\tau(z;\beta_0)\\
&= \frac{1}{N}\sum_{j=1}^{N} Y_{jt}\left[\frac{1}{b_N}\Gamma\left(\frac{X_{jt}'\widehat{\beta}-u}{b_N}\right) - \frac{1}{b_N}\Gamma\left(\frac{X_{jt}'\beta_0-u}{b_N}\right)\right]\left(\frac{V_j-v}{b_N}\right)^{\tau_{-1}}\mathcal{K}_{b_N}^V\left(\frac{V_j-v}{b_N}\right)\\
&= \frac{1}{N}\sum_{j=1}^{N} Y_{jt}\frac{1}{b_N^2}\gamma\left(\frac{X_{jt}'\widetilde{\beta}-u}{b_N}\right)\left(\frac{V_j-v}{b_N}\right)^{\tau_{-1}}\mathcal{K}_{b_N}^V\left(\frac{V_j-v}{b_N}\right)X_{jt}'(\widehat{\beta}-\beta_0)
\end{aligned}$$

By the same arguments as in the proof of Lemma H.1, and by $\mathbb{E}[Y_{jt}^2] < \infty$, we can show that

$$\begin{aligned}
&\sup_{z\in\mathcal{Z}_t}\left|T_N^\tau(z;\widehat{\beta}) - T_N^\tau(z;\beta_0)\right|\\
&\le \frac{1}{b_N^2}\sup_{z\in\mathcal{Z}_t,\beta\in\mathcal{B}_\varepsilon,b\in(0,\bar{b}]}\left\|\frac{1}{N}\sum_{j=1}^{N}Y_{jt}X_{jt}\gamma\left(\frac{X_{jt}'\beta-u}{b}\right)\left(\frac{V_j-v}{b}\right)^{\tau_{-1}}\mathcal{K}_b^V\left(\frac{V_j-v}{b}\right)\right.\\
&\left.-\mathbb{E}\left[Y_tX_t\gamma\left(\frac{X_t'\beta-u}{b}\right)\left(\frac{V-v}{b}\right)^{\tau_{-1}}\mathcal{K}_b^V\left(\frac{V-v}{b}\right)\right]\right\|\|\widehat{\beta}-\beta_0\|\\
&+\sup_{z\in\mathcal{Z}_t,\beta\in\mathcal{B}_\varepsilon}\frac{1}{b_N^2}\left\|\mathbb{E}\left[Y_tX_t\gamma\left(\frac{X_t'\beta-u}{b_N}\right)\left(\frac{V-v}{b_N}\right)^{\tau_{-1}}\mathcal{K}_{b_N}^V\left(\frac{V-v}{b_N}\right)\right]\right\|\|\widehat{\beta}-\beta_0\|\\
&= O_p\left(\frac{1}{\sqrt{Nb_N^{4+2d_V}}}\right)\cdot O_p(a_N^{-1}) + O(1)\cdot O_p(a_N^{-1})\\
&= o_p\left(\frac{1}{\sqrt{Nb_N}}\right)
\end{aligned}$$

holds with probability arbitrarily close to 1 as $N \to \infty$ since $\mathbb{P}(\widehat{\beta} \in \mathcal{B}_\varepsilon) \to 1$. The last equality follows from B6.

Since this rate of convergence applies uniformly in $z \in \mathcal{Z}_t$ to generic components of the vector $T_N(z;\widehat{\beta}) - T_N(z;\beta_0)$, it applies to its vector norm uniformly in $z \in \mathcal{Z}_t$ as well, which concludes the proof. □

Let

$$T(z;\beta_0) = \int \xi(a)\mathcal{K}(a)\, da \cdot \mathbb{E}[Y_t|Z_t(\beta_0) = z] f_{Z_t(\beta_0)}(z).$$

Also, recall that $Z_t \equiv Z_t(\beta_0)$.

Lemma H.4 (Convergence of T_N to T)**.** Suppose B1–B6 hold. Then,

$$\sup_{z\in\mathcal{Z}_t} \|T_N(z;\beta_0) - T(z;\beta_0)\| = O_p\left(\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)^{1/2}\right) + O(b_N).$$

Proof of Lemma H.4. By the triangle inequality,

$$\sup_{z\in\mathcal{Z}_t} \|T_N(z;\beta_0) - T(z;\beta_0)\| \leq \sup_{z\in\mathcal{Z}_t} \|T_N(z;\beta_0) - \mathbb{E}[T_N(z;\beta_0)]\| + \sup_{z\in\mathcal{Z}_t} \|\mathbb{E}[T_N(z;\beta_0)] - T(z;\beta_0)\|.$$

Generic components of $T_N(z;\beta_0) - \mathbb{E}[T_N(z;\beta_0)]$ can be written as

$$\sup_{z\in\mathcal{Z}_t} \left| \frac{1}{N}\sum_{j=1}^{N} \left(\frac{Z_{jt}-z}{b_N}\right)^{\tau} Y_{jt}\mathcal{K}_{b_N}\left(\frac{Z_{jt}-z}{b_N}\right) - \mathbb{E}\left[\left(\frac{Z_t-z}{b_N}\right)^{\tau} Y_t\mathcal{K}_{b_N}\left(\frac{Z_t-z}{b_N}\right)\right]\right|.$$

By an argument similar to that used in Corollary 1.(ii) in Masry (1996) or in Lemma B.ii.(2) in Rothe and Firpo (2019), this term is of order $O_p\left(\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)^{1/2}\right)$.

Next, note that generic elements of $\mathbb{E}[T_N(z;\beta_0)]$ are of the form

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{Z_t-z}{b_N}\right)^{\tau} Y_t\mathcal{K}_{b_N}\left(\frac{Z_t-z}{b_N}\right)\right] \\
&= \int \left(\frac{\tilde{z}-z}{b_N}\right)^{\tau} \mathbb{E}[Y_t|Z_t=\tilde{z}]\mathcal{K}_{b_N}\left(\frac{\tilde{z}-z}{b_N}\right) f_{Z_t}(\tilde{z})\, d\tilde{z} \\
&= \int a^{\tau}\mathcal{K}(a)\mathbb{E}[Y_t|Z_t = z + ab_N] f_{Z_t}(z+ab_N)\, da \\
&\leq \mathbb{E}[Y_t|Z_t=z] f_{Z_t}(z) \int a^{\tau}\mathcal{K}(a)\, da + b_N \sup_{z\in\mathcal{Z}_t}\left\|\frac{\partial}{\partial z}\left(\mathbb{E}[Y_t|Z_t=z]f_{Z_t}(z)\right)\right\| \cdot \left\|\int a^{\tau}\mathcal{K}(a)\cdot a\, da\right\|.
\end{aligned}$$

The second equality follows from a change in variables. Note that $\mathbb{E}[Y_t|Z_t = z]f_{Z_t}(z)\int a^{\tau}\mathcal{K}(a)\, da$ is the corresponding element of $T(z;\beta_0)$. Therefore,

$$\begin{aligned}
&\sup_{z\in\mathcal{Z}_t}\left|\int a^{\tau}\mathcal{K}(a)\mathbb{E}[Y_t|Z_t = z+ab_N] f_{Z_t}(z+ab_N)\, da - \mathbb{E}[Y_t|Z_t=z]f_{Z_t}(z)\int a^{\tau}\mathcal{K}(a)\, da\right| \\
&\leq b_N \sup_{z\in\mathcal{Z}_t}\left\|\frac{\partial}{\partial z}\left(\mathbb{E}[Y_t|Z_t=z]f_{Z_t}(z)\right)\right\| \cdot \left\|\int a^{\tau}\mathcal{K}(a)\cdot a\, da\right\|.
\end{aligned}$$

By B3, $\left\|\int a^{\tau}\mathcal{K}(a)\cdot a\, da\right\| < \infty$. By B4.(iii), we have that $\sup_{z\in\mathcal{Z}_t}\left\|\frac{\partial}{\partial z}\left(\mathbb{E}[Y_t|Z_t=z]f_{Z_t}(z)\right)\right\| <$

∞. Therefore,

$$\sup_{z\in\mathcal{Z}_t} \|\mathbb{E}[T_N(z;\beta_0)] - T(z;\beta_0)\| = O(b_N)$$

and

$$\sup_{z\in\mathcal{Z}_t} \|T_N(z;\beta_0) - T(z;\beta_0)\| = O_p\left(\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)^{1/2}\right) + O(b_N).$$

□

Lemma H.5 (Convergence of S_N part 2)**.** Suppose B1–B6 hold. Then,

$$\sup_{z\in\mathcal{Z}_t} \left\|\frac{\partial}{\partial u} S_N(z;\beta_0)\right\| = o_p\left(\frac{a_N}{\sqrt{Nb_N}}\right).$$

Proof of Lemma H.5. As in the proof of Lemma H.1, consider a generic entry of $S_N(z;\beta_0)$, which we write as

$$S_N^{\tau,\tau'}(z;\beta_0) = \frac{1}{N}\sum_{j=1}^N \left(\frac{Z_{jt}-z}{b_N}\right)^{\tau+\tau'} \mathcal{K}_{b_N}\left(\frac{Z_{jt}-z}{b_N}\right).$$

Its derivative with respect to u, the first element of z, is

$$\frac{\partial}{\partial u} S_N^{\tau,\tau'}(z;\beta_0) = \frac{-1}{b_N^{2+d_V}}\frac{1}{N}\sum_{j=1}^N \gamma\left(\frac{X_{jt}'\beta_0 - u}{b_N}\right)\left(\frac{V_j - v}{b_N}\right)^{\tau_{-1}+\tau'_{-1}} \mathcal{K}^V\left(\frac{V_j-v}{b_N}\right)$$

where $\gamma(u) = (\tau_1+\tau_1')u^{\tau_1+\tau_1'-1}K(u) + u^{\tau_1+\tau_1'}K'(u)$.

Therefore, we have that

$$\begin{aligned}
\sup_{z\in\mathcal{Z}_t}\left|\frac{\partial}{\partial u} S_N^{\tau,\tau'}(z;\beta_0)\right| &= \sup_{z\in\mathcal{Z}_t}\left|\frac{-1}{b_N^{2+d_V}}\frac{1}{N}\sum_{j=1}^N \gamma\left(\frac{X_{jt}'\beta_0 - u}{b_N}\right)\left(\frac{V_j - v}{b_N}\right)^{\tau_{-1}+\tau'_{-1}} \mathcal{K}^V\left(\frac{V_j-v}{b_N}\right)\right| \\
&\le \sup_{z\in\mathcal{Z}_t, b\in(0,\bar{b}]} \frac{1}{\sqrt{N}b_N^{2+d_V}}\left|\frac{1}{\sqrt{N}}\sum_{j=1}^N\left\{\gamma\left(\frac{X_{jt}'\beta_0 - u}{b}\right)\left(\frac{V_j - v}{b}\right)^{\tau_{-1}+\tau'_{-1}} \mathcal{K}^V\left(\frac{V_j-v}{b}\right)\right.\right. \\
&\left.\left. -\mathbb{E}\left[\gamma\left(\frac{X_t'\beta_0 - u}{b}\right)\left(\frac{V - v}{b}\right)^{\tau_{-1}+\tau'_{-1}} \mathcal{K}^V\left(\frac{V-v}{b}\right)\right]\right\}\right| \qquad \text{(H.3)} \\
&+ \sup_{z\in\mathcal{Z}_t}\frac{1}{b_N^2}\left|\mathbb{E}\left[\gamma\left(\frac{X_t'\beta_0 - u}{b}\right)\left(\frac{V - v}{b}\right)^{\tau_{-1}+\tau'_{-1}} \mathcal{K}_{b_N}^V\left(\frac{V-v}{b}\right)\right]\right|. \qquad \text{(H.4)}
\end{aligned}$$

The class

$$\left\{\gamma\left(\frac{X_t'\beta_0 - u}{b}\right)\left(\frac{V-v}{b}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}^V\left(\frac{V-v}{b}\right) : z \in \mathcal{Z}_t, b \in (0,\bar{b}]\right\}$$

is a subset of $\mathcal{F}_V$ which is Euclidean, therefore it is also Euclidean and hence Donsker. We therefore have that term (H.3) is of order $O_p\left(\frac{1}{\sqrt{Nb_N^{4+2d_V}}}\right)$.

We can bound term (H.4) as follows,

$$\begin{aligned}
&\sup_{z\in\mathcal{Z}_t}\frac{1}{b_N^2}\left|\mathbb{E}\left[\gamma\left(\frac{X_t'\beta_0 - u}{b_N}\right)\left(\frac{V-v}{b_N}\right)^{\tau_{-1}+\tau_{-1}'}\mathcal{K}^V_{b_N}\left(\frac{V-v}{b_N}\right)\right]\right| \\
&= \sup_{z\in\mathcal{Z}_t}\left|\mathbb{E}\left[\frac{\partial}{\partial u}\left(\frac{Z_t - z}{b_N}\right)^{\tau+\tau'}\mathcal{K}_{b_N}\left(\frac{Z_t-z}{b_N}\right)\right]\right| \\
&= \sup_{z\in\mathcal{Z}_t}\left|\int\frac{\partial}{\partial u}\left(\frac{\tilde{z}-z}{b_N}\right)^{\tau+\tau'}\mathcal{K}_{b_N}\left(\frac{\tilde{z}-z}{b_N}\right)f_{Z_t(\beta_0)}(\tilde{z})\,d\tilde{z}\right| \\
&= \sup_{z\in\mathcal{Z}_t}\left|\int\frac{\partial}{\partial u}a^{\tau+\tau'}\mathcal{K}\left(a\right)f_{Z_t}(z+ab_N)\,da\right| \\
&\leq \sup_{z\in\mathcal{Z}_t}\left|\frac{\partial}{\partial u}f_{Z_t}(z)\right|\left|\int a^{\tau+\tau'}\mathcal{K}\left(a\right)\,da\right| \\
&= O(1).
\end{aligned}$$

The third equality follows from the change of variables $\tilde{z} = z + ab_N$. The final line follows from B3 and B4.(iii).

Therefore,

$$\begin{aligned}
\sup_{z\in\mathcal{Z}_t}\left|\frac{\partial}{\partial z_1}S_N^{\tau,\tau'}(z;\beta_0)\right| &= O_p\left(\frac{1}{\sqrt{Nb_N^{4+2d_V}}}\right) + O(1) \\
&= o_p\left(\frac{a_N}{\sqrt{Nb_N}}\right)
\end{aligned}$$

since, as $N\to\infty$, $\frac{1}{\sqrt{Nb_N^{4+2d_V}}}\cdot\frac{\sqrt{Nb_N}}{a_N} = O(N^{\epsilon-\delta(3/2+d_V)}) = o(1)$ by B6, and since $\frac{\sqrt{Nb_N}}{a_N}\cdot O(1) = O(N^{1/2-\epsilon-\delta/2}) = o(1)$, also by B6. Since this holds for a generic entry of the matrix $\frac{\partial}{\partial u}S_N(z;\beta_0)$, it holds for its matrix norm as well, which concludes this lemma. □

Lemma H.6 (Convergence of T_N part 2)**.** Suppose B1–B6 hold. Then,

$$\sup_{z\in\mathcal{Z}_t}\left\|\frac{\partial}{\partial u}T_N(z;\beta_0)\right\| = o_p\left(\frac{a_N}{\sqrt{Nb_N}}\right).$$

Proof of Lemma H.6. As in the proof of Lemma H.3, consider a generic component of the vector $T_N(z;\beta_0)$. Denote this element by

$$T_N^{\tau}(z;\beta_0) = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{Z_{jt}-z}{b_N}\right)^{\tau} Y_{jt}\mathcal{K}_{b_N}\left(\frac{Z_{jt}-z}{b_N}\right).$$

Its derivative with respect to u is

$$\frac{\partial}{\partial u}T_N^{\tau}(z;\beta_0) = \frac{-1}{b_N^{2+d_V}}\frac{1}{N}\sum_{j=1}^{N} Y_{jt}\gamma\left(\frac{X_{jt}'\beta-u}{b}\right)\left(\frac{V_j-v}{b}\right)^{\tau_{-1}}\mathcal{K}^V\left(\frac{V_j-v}{b}\right).$$

where $\gamma(u) = \tau_1 u^{\tau_1-1}K(u) + u^{\tau_1}K'(u)$. The rest of the proof follows directly from the arguments used in the proofs of Lemmas H.3 and H.5. □

Lemma H.7 (Convergence of indicators)**.** Suppose B1–B6 hold. Suppose $\widetilde{\beta} \xrightarrow{p} \beta_0$. Let $\pi_{it}(\beta) \equiv \mathbb{1}((\underline{x}_t'\beta, V_i) \in \mathcal{Z}_t)$. Then,

$$\mathbb{P}\left(\sup_{i=1,\ldots,N}\left|\pi_{it}(\widetilde{\beta}) - \pi_{it}(\beta_0)\right| = 0\right) \to 1$$

as $N \to \infty$.

Proof of Lemma H.7. We note that

$$\begin{aligned}\sup_{i=1,\ldots,N}|\pi_{it}(\widetilde{\beta}) - \pi_{it}(\beta_0)| &= \sup_{i=1,\ldots,N}\Big(\mathbb{1}((\underline{x}_t'\widetilde{\beta}, V_i) \in \mathcal{Z}_t, (\underline{x}_t'\beta_0, V_i) \notin \mathcal{Z}_t) + \mathbb{1}((\underline{x}_t'\widetilde{\beta}, V_i) \notin \mathcal{Z}_t, (\underline{x}_t'\beta_0, V_i) \in \mathcal{Z}_t)\Big)\\ &\leq \sup_{i=1,\ldots,N}\Big(\mathbb{1}(\underline{x}_t'\widetilde{\beta} \in \mathcal{Z}_{1t}, \underline{x}_t'\beta_0 \notin \mathcal{Z}_{1t}) + \mathbb{1}(\underline{x}_t'\widetilde{\beta} \notin \mathcal{Z}_{1t}, \underline{x}_t'\beta_0 \in \mathcal{Z}_{1t})\Big)\\ &= \mathbb{1}(\underline{x}_t'\widetilde{\beta} \in \mathcal{Z}_{1t}, \underline{x}_t'\beta_0 \notin \mathcal{Z}_{1t}) + \mathbb{1}(\underline{x}_t'\widetilde{\beta} \notin \mathcal{Z}_{1t}, \underline{x}_t'\beta_0 \in \mathcal{Z}_{1t}),\end{aligned}$$

where $\mathcal{Z}_{1t} = \{z_1 = e_1'z : z \in \mathcal{Z}_t\}$. By B4.(v), $\underline{x}_t'\beta_0 \in \mathcal{Z}_{1t}$, and therefore $\mathbb{1}(\underline{x}_t'\widetilde{\beta} \in \mathcal{Z}_{1t}, \underline{x}_t'\beta_0 \notin \mathcal{Z}_{1t}) = 0$, and $\mathbb{1}(\underline{x}_t'\widetilde{\beta} \notin \mathcal{Z}_{1t}, \underline{x}_t'\beta_0 \in \mathcal{Z}_{1t}) = \mathbb{1}(\underline{x}_t'\widetilde{\beta} \notin \mathcal{Z}_{1t})$.

By assumption, $\widetilde{\beta}$ converges in probability to β_0. By Theorem 18.9.(v) in Vaart (1998), $\mathbb{P}(\underline{x}_t'\widetilde{\beta} \in \mathcal{Z}_{1t}) \to \mathbb{1}(\underline{x}_t'\beta_0 \in \mathcal{Z}_{1t}) = 1$ since $\underline{x}_t'\beta_0$ is not in the boundary of $\mathcal{Z}_{1t}$ by B4.(v).

Therefore,

$$\mathbb{P}\left(\sup_{i=1,\ldots,N}|\pi_{it}(\widetilde{\beta}) - \pi_{it}(\beta_0)| = 0\right) \geq \mathbb{P}(\mathbb{1}(\underline{x}_t'\widetilde{\beta} \notin \mathcal{Z}_{1t}) = 0) = \mathbb{P}(\underline{x}_t'\widetilde{\beta} \in \mathcal{Z}_{1t}) \to \mathbb{P}(\underline{x}_t'\beta_0 \in \mathcal{Z}_{1t}) = 1$$

as $N \to \infty$. □

Lemma H.8 (ASF convergence in distribution)**.** Suppose B1–B6 hold. Then,

$$\sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_1(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\text{ASF}_t}(\underline{x}_t'\beta_0)).$$

Proof of Lemma H.8. This proof builds on the proof of Corollary 2 in Kong, Linton, and Xia (2010) (KLX hereafter). First, we verify that Assumptions A1–A7 of KLX hold under ours. Their A1 holds with our squared loss function, and we note that $\psi(\varepsilon_i) \equiv -2(Y_{it} - \mathbb{E}[Y_t|Z_{it}])$ in their notation. By our A5, $\mathbb{E}[|\psi(\varepsilon_i)|^{\nu_1}] < \infty$ holds for arbitrary large ν_1. Their A2 holds immediately. Their A3 holds by our B3. Their A4 and A5 hold by our B4.(iii). Their A6 holds if

$$\begin{aligned} Nb_N^{1+d_V}/\log(N) &\to \infty \\ Nb_N^{1+d_V+2(\ell+1)}/\log(N) &= O(1) \\ N^{\nu_2/8-\lambda_1-1/4} b_N^{(1+d_V)(\nu_2/8-\lambda_1+3/4)} \log(N)^{-\nu_2/8+5/4+\lambda_1} &\to \infty, \end{aligned}$$

for some $2 < \nu_2 \leq \nu_1$. Since $b_N = \kappa \cdot N^{-\delta}$, these conditions are equivalent to

$$\begin{aligned} 1 - \delta(1 + d_V) &> 0 \\ 1 - \delta(3 + 2\ell + d_V) &\leq 0 \\ \nu_2/8 - \lambda_1 - 1/4 - \delta(1 + d_V)(\nu_2/8 - \lambda_1 + 3/4) &> 0. \end{aligned}$$

Since ν_1 can be made arbitrarily large, ν_2 can also taken to be arbitrarily large, and the last inequality is equivalent to

$$\delta < \frac{1}{1 + d_V}.$$

By our B6, these rate conditions all hold. Finally, their A7 holds by our B4.(v). Since these assumptions hold for $\lambda_1 = 1$, we can use equation (13) in KLX and their Corollary 1 to write

$$\widehat{h}_1(z;\beta_0) = h_1(z;\beta_0) + B_{1,N}(z) + \frac{1}{N}\sum_{j=1}^{N}\phi_{1,jN}(z) + R_{1,N}(z)$$

where $B_{1,N}(z)$ is a bias term satisfying $\sup_{z\in\mathcal{Z}_t}|B_{1,N}(z)| = O(b_N^{\ell+1})$ if ℓ is odd or $O(b_N^{\ell+2})$ if ℓ is even, where $\phi_{1,jN}(z)$ are mean-zero random variables, and where $R_{1,N}(z)$ is a higher-order term satisfying $\sup_{z\in\mathcal{Z}_t}|R_{1,N}(z)| = O_p\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)$.

Second, we note that

$$\sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_1(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right)$$

$$= \sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0) - h_1(\underline{x}_t'\beta_0, V_i;\beta_0)\right)\pi_{it} \tag{H.5}$$

$$+ \sqrt{b_N}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(h_1(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_1(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right). \tag{H.6}$$

To analyze term (H.5), we use the fact that

$$\sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0) - h_1(\underline{x}_t'\beta_0, V_i;\beta_0)\right)\pi_{it}$$

$$= \sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}B_{1,N}(\underline{x}_t'\beta_0, V_i)\pi_{it}$$

$$+ \sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{1,jN}(\underline{x}_t'\beta_0, V_i)\pi_{it} + \sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}R_{1,N}(\underline{x}_t'\beta_0, V_i)\pi_{it}.$$

When ℓ is odd, $\sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}B_{1,N}(\underline{x}_t'\beta_0, V_i)\pi_{it}$ is $o(1)$ because

$$\begin{aligned}\left|\sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}B_{1,N}(\underline{x}_t'\beta_0, V_i)\pi_{it}\right| &\leq \sqrt{Nb_N}\cdot\sup_{z\in\mathcal{Z}_t}|B_{1,N}(z)| \\ &= \sqrt{Nb_N}\cdot O(b_N^{\ell+1}) \\ &= O(\sqrt{Nb_N^{2\ell+3}})\end{aligned}$$

and by B6. A similar derivation applies when ℓ is even.

We now show that term $\sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{1,jN}(\underline{x}_t'\beta_0, V_i)\pi_{it}$ converges in distribution to a normal distribution. By standard arguments from Masry (1996), which are also referred to in the proof of Corollary 2 in KLX, we have that

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{1,jN}(\underline{x}_t'\beta_0,V_i)\pi_{it}$$

$$= \frac{-1}{Nb_N}\sum_{i=1}^{N}(Y_{it}-\mathbb{E}[Y_t|Z_{it}])f_V(V_i)\mathbb{1}((\underline{x}_t'\beta_0,V_i)\in\mathcal{Z}_t)$$

$$\cdot\, e_1'S_N(\underline{x}_t'\beta_0,V_i;\beta_0)^{-1}\int\mathcal{K}\left(\frac{X_{it}'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_{it}'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\left(1+O_p\left(\left(\frac{\log(N)}{Nb_N^{d_V}}\right)^{1/2}\right)\right)$$

$$= \frac{-1}{Nb_N}\sum_{i=1}^{N}(Y_{it}-\mathbb{E}[Y_t|Z_{it}])f_V(V_i)\mathbb{1}((\underline{x}_t'\beta_0,V_i)\in\mathcal{Z}_t)$$

$$\cdot\, e_1'S_N(\underline{x}_t'\beta_0,V_i;\beta_0)^{-1}\int\mathcal{K}\left(\frac{X_{it}'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_{it}'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv+o_p(1).$$

We now calculate the asymptotic variance of

$$\frac{-1}{Nb_N}\sum_{i=1}^{N}(Y_{it}-\mathbb{E}[Y_t|Z_{it}])f_V(V_i)\mathbb{1}((\underline{x}_t'\beta_0,V_i)\in\mathcal{Z}_t)\cdot e_1'S_N(\underline{x}_t'\beta_0,V_i;\beta_0)^{-1}\int\mathcal{K}\left(\frac{X_{it}'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_{it}'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv.$$

We have that

$$\mathrm{Var}\left(\frac{-1}{Nb_N}\sum_{i=1}^{N}(Y_t-\mathbb{E}[Y_t|Z_t])f_V(V)\mathbb{1}((\underline{x}_t'\beta_0,V)\in\mathcal{Z}_t)\cdot e_1'S_N(\underline{x}_t'\beta_0,V;\beta_0)^{-1}\int\mathcal{K}\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\right)$$
$$=\frac{1}{Nb_N^2}\mathbb{E}\left[(Y_t-\mathbb{E}[Y_t|Z_t])^2f_V(V)^2\mathbb{1}((\underline{x}_t'\beta_0,V)\in\mathcal{Z}_t)e_1'S_N(\underline{x}_t'\beta_0,V;\beta_0)^{-1}\left(\int\mathcal{K}\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\right)\right.$$
$$\left.\left(\int\mathcal{K}\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\right)'S_N(\underline{x}_t'\beta_0,V;\beta_0)^{-1}e_1\right]$$

Recall that $S_N(z;\beta_0)=S(z;\beta_0)+o_p(1)=\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\cdot f_{Z_t(\beta_0)}(z)+o_p(1)$ uniformly in $z\in\mathcal{Z}_t$

by Lemma H.2. Therefore, the above expression

$$= \frac{1}{Nb_N^2}\mathbb{E}\Bigg[\mathrm{Var}(Y_t|Z_t(\beta_0))\frac{f_V(V)^2}{f_{Z_t(\beta_0)}(\underline{x}_t'\beta_0,V)^2}\mathbb{1}((\underline{x}_t'\beta_0,V)\in\mathcal{Z}_t)e_1'\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}$$
$$\left(\int\mathcal{K}\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\right)\cdot\left(\int\mathcal{K}\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{X_t'\beta_0-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\right)'$$
$$\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}e_1\Bigg]+o((Nb_N)^{-1})$$
$$= \frac{1}{Nb_N^2}\mathbb{E}\Bigg[\int\mathrm{Var}(Y_t|X_t'\beta_0=\tilde{u},V)\frac{f_V(V)^2}{f_{Z_t(\beta_0)}(\underline{x}_t'\beta_0,V)^2}\mathbb{1}((\underline{x}_t'\beta_0,V)\in\mathcal{Z}_t)e_1'\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}$$
$$\left(\int\mathcal{K}\left(\frac{\tilde{u}-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{\tilde{u}-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\right)\cdot\left(\int\mathcal{K}\left(\frac{\tilde{u}-\underline{x}_t'\beta_0}{b_N},v\right)\xi\left(\frac{\tilde{u}-\underline{x}_t'\beta_0}{b_N},v\right)\,dv\right)'$$
$$\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}e_1\,f_{X_t'\beta_0|V}(\tilde{u}|V)\,d\tilde{u}\Bigg]+o((Nb_N)^{-1})$$
$$= \frac{1}{Nb_N}\mathbb{E}\Bigg[\int\mathrm{Var}(Y_t|X_t'\beta_0=\underline{x}_t'\beta_0+b_Nu,V)\frac{f_V(V)^2}{f_{Z_t(\beta_0)}(\underline{x}_t'\beta_0,V)^2}\mathbb{1}((\underline{x}_t'\beta_0,V)\in\mathcal{Z}_t)e_1'\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}$$
$$\cdot\left(\int\mathcal{K}(z)\,\xi(z)\,dv\right)\left(\int\mathcal{K}(z)\,\xi(z)\,dv\right)'\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}e_1f_{X_t'\beta_0|V}(\underline{x}_t'\beta_0+b_Nu|V)\,du\Bigg]+o((Nb_N)^{-1})$$
$$= \frac{1}{Nb_N}\mathbb{E}\left[\mathrm{Var}(Y_t|X_t'\beta_0=\underline{x}_t'\beta_0,V)\frac{f_V(V)}{f_{Z_t(\beta_0)}(\underline{x}_t'\beta_0,V)}\mathbb{1}((\underline{x}_t'\beta_0,V)\in\mathcal{Z}_t)\right]$$
$$\cdot\,e_1'\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}\int\left(\int\mathcal{K}(z)\,\xi(z)\,dv\right)\left(\int\mathcal{K}(z)\,\xi(z)\,dv\right)'du\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)^{-1}e_1+o((Nb_N)^{-1})$$
$$= \frac{1}{Nb_N}\sigma^2_{\mathrm{ASF}_t}(\underline{x}_t'\beta_0)+o((Nb_N)^{-1}).$$

The third equality follows from the change of variables $\tilde{u}=\underline{x}_t'\beta_0+b_Nu$. The above equations re-derive and fix a minor typo in equation (A.42) in KLX. By the proof of Corollary 2 in KLX, we have that

$$\sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\phi_{1,jN}(\underline{x}_t'\beta_0,V_i)\pi_{it}\xrightarrow{d}\mathcal{N}(0,\sigma^2_{\mathrm{ASF}_t}(\underline{x}_t'\beta_0)).$$

Also, the term $\sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^N R_{1,N}(\underline{x}_t'\beta_0,V_i)\pi_{it}$ is $o_p(1)$ because

$$\begin{aligned}\left|\sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^N R_{1,N}(\underline{x}_t'\beta_0,V_i)\pi_{it}\right| &\le \sqrt{Nb_N}\cdot\sup_{z\in\mathcal{Z}_t}|R_{1,N}(z)|\\ &= \sqrt{Nb_N}\cdot O_p\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)\\ &= O_p\left(\frac{\log(N)}{\sqrt{Nb_N^{1+2d_V}}}\right)\\ &= o_p(1)\end{aligned}$$

by B6.

Third, term (H.6) above is of order $O_p(\sqrt{b_N}) = o_p(1)$ by an application of the central limit theorem.

Finally, we obtain that

$$\sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)\pi_{it} - \mathbb{E}[h_1(\underline{x}_t'\beta_0, V; \beta_0)\pi_t]\right) = \sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{jN}(\underline{x}_t'\beta_0, V_i)\pi_{it} + o_p(1)$$

$$\xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{ASF}_t}(\underline{x}_t'\beta_0)).$$

□

We use the following technical lemma in the proof of Theorem C.1.

Lemma H.9. Let A and B be positive-definite, symmetric matrices. Let $\lambda_{\min}(A)$ denote the minimum eigenvalue of A. Then,

$$|\lambda_{\min}(A) - \lambda_{\min}(B)| \leq \|A - B\|.$$

Proof of Lemma H.9. Since A and B are positive-definite and symmetric, they are invertible and $\lambda_{\min}(A) = \|A^{-1}\|^{-1} > 0$ and $\lambda_{\min}(B) = \|B^{-1}\|^{-1} > 0$. We then have

$$\begin{aligned}
|\lambda_{\min}(A) - \lambda_{\min}(B)| &= |\|A^{-1}\|^{-1} - \|B^{-1}\|^{-1}| \\
&= |\|A^{-1}\| - \|B^{-1}\|| \cdot \frac{1}{\|A^{-1}\|\|B^{-1}\|} \\
&\leq \|A^{-1} - B^{-1}\| \cdot \frac{1}{\|A^{-1}\|\|B^{-1}\|} \\
&= \|B^{-1}(B - A)A^{-1}\| \cdot \frac{1}{\|A^{-1}\|\|B^{-1}\|} \\
&\leq \|B^{-1}\|\|A - B\|\|A^{-1}\| \cdot \frac{1}{\|A^{-1}\|\|B^{-1}\|} \\
&= \|A - B\|.
\end{aligned}$$

The first inequality follows from the triangle inequality, and the second inequality is from $\|CD\| \leq \|C\|\|D\|$ for the spectral norm and square matrices C and D. □

Proof of Theorem C.1. We have the following decomposition:

$$\sqrt{Nb_N}\left(\widehat{\text{ASF}}_t(\underline{x}_t) - \text{ASF}^\pi_t(\underline{x}_t)\right) = \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^N \left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - \widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right) \tag{H.7}$$

$$+ \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^N \left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)\right)\hat{\pi}_{it}\right) \tag{H.8}$$

$$+ \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^N \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)(\hat{\pi}_{it} - \pi_{it})\right) \tag{H.9}$$

$$+ \sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^N \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)\pi_{it} - \mathbb{E}[h_1(\underline{x}_t'\beta_0, V; \beta_0)\pi_t]\right). \tag{H.10}$$

We break down the proof into four parts. In the first three parts, we show that terms (H.7)–(H.9) are $o_p(1)$. In the fourth and last part, we show that term (H.10) converges in distribution.

Part 1: Convergence of Term (H.7)

We have that

$$\begin{aligned}
&\sqrt{Nb_N}\cdot\left|\frac{1}{N}\sum_{i=1}^N \left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - \widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right| \\
&= \sqrt{Nb_N}\cdot\left|\frac{1}{N}\sum_{j=1}^N e_1'\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right| \\
&= \sqrt{Nb_N}\cdot\left|\frac{1}{N}\sum_{j=1}^N e_1'\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})^{-1}(T_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0))\right.\right. \\
&\left.\left.+S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})^{-1}\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})\right)S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\mathbb{1}((\underline{x}_t'\widehat{\beta}, V_i) \in \mathcal{Z}_t)\right| \\
&\leq \sqrt{Nb_N}\cdot\|e_1\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})^{-1}\right\|\sup_{z\in\mathcal{Z}_t}\left\|T_N(z;\widehat{\beta}) - T_N(z;\beta_0)\right\| \\
&+ \sqrt{Nb_N}\cdot\|e_1\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})^{-1}\right\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta}) - S_N(z;\beta_0)\right\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\beta_0)^{-1}\right\|\sup_{z\in\mathcal{Z}_t}\|T_N(z;\beta_0)\|.
\end{aligned}$$

The terms in the previous expressions are of these asymptotic orders:

- $\|e_1\| = 1$.
- $\left\|S_N(z;\widehat{\beta})^{-1}\right\| = \lambda_{\min}\left(S_N(z;\widehat{\beta})\right)^{-1}$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a sym-

metric matrix. We have that

$$
\begin{aligned}
\sup_{z\in\mathcal{Z}_t}\left|\lambda_{\min}\left(S_N(z;\widehat{\beta})\right)-\lambda_{\min}\left(S(z;\beta_0)\right)\right| &\leq \sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})-S(z;\beta_0)\right\| \\
&\leq \sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})-S_N(z;\beta_0)\right\|+\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\beta_0)-S(z;\beta_0)\right\| \\
&= o_p\left(\frac{1}{\sqrt{Nb_N}}\right)+O_p\left(\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)^{1/2}\right)+O(b_N) \\
&= o_p(1).
\end{aligned}
$$

The first line follows from Lemma H.9. The second line follows from the triangle inequality. The third line follows from Lemmas H.1 and H.2. The last line follows from B6. Also note that

$$
\inf_{z\in\mathcal{Z}_t}\lambda_{\min}\left(S(z;\beta_0)\right)=\inf_{z\in\mathcal{Z}_t}f_{Z_t}(z)\cdot\lambda_{\min}\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)>0.
$$

This follows from the definition of the set $\mathcal{Z}_t$, which is such that $\inf_{z\in\mathcal{Z}_t}f_{Z_t}(z)>0$: see B4.(ii). $\int\xi(a)\xi(a)'\mathcal{K}(a)\,da$ is positive definite since, for $c\in\mathbb{R}^{\bar{N}}$ such that $c\neq\mathbf{0}$,

$$
c'\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)c=\int(c'\xi(a))^2\mathcal{K}(a)\,da=0
$$

implies that $c'\xi(a)=0$ for all a in the support of $\mathcal{K}(a)$. Since $\xi(a)$ is comprised of products of powers of components of a, $c'\xi(a)=0$ over this entire support implies $c=\mathbf{0}$, a contradiction. Therefore $\lambda_{\min}\left(\int\xi(a)\xi(a)'\mathcal{K}(a)\,da\right)>0$ and $\inf_{z\in\mathcal{Z}_t}\lambda_{\min}\left(S(z;\beta_0)\right)>0$.

This implies that,

$$
\begin{aligned}
\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})^{-1}\right\| &= \frac{1}{\inf_{z\in\mathcal{Z}_t}\lambda_{\min}\left(S_N(z;\widehat{\beta})\right)} \\
&\leq \frac{1}{\inf_{z\in\mathcal{Z}_t}\lambda_{\min}\left(S(z;\beta_0)\right)-\sup_{z\in\mathcal{Z}_t}\left|\lambda_{\min}\left(S_N(z;\widehat{\beta})\right)-\lambda_{\min}\left(S(z;\beta_0)\right)\right|} \\
&= \frac{1}{\inf_{z\in\mathcal{Z}_t}\lambda_{\min}\left(S(z;\beta_0)\right)-o_p(1)} \\
&= O_p(1).
\end{aligned}
$$

- By Lemma H.1, we have that $\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})-S_N(z;\beta_0)\right\|=o_p\left(\frac{1}{\sqrt{Nb_N}}\right)$.
- By Lemma H.3, we have that $\sup_{z\in\mathcal{Z}_t}\left\|T_N(z;\widehat{\beta})-T_N(z;\beta_0)\right\|=o_p\left(\frac{1}{\sqrt{Nb_N}}\right)$
- As above, we have that $\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\beta_0)^{-1}\right\|=O_p(1)$.

- We have that

$$\sup_{z \in \mathcal{Z}_t} \|T_N(z;\beta_0)\| \le \sup_{z \in \mathcal{Z}_t} \|T_N(z;\beta_0) - T(z;\beta_0)\| + \sup_{z \in \mathcal{Z}_t} \|T(z;\beta_0)\|$$

where

$$\sup_{z \in \mathcal{Z}_t} \|T_N(z;\beta_0) - T(z;\beta_0)\| = O_p\left(\left(\frac{\log(N)}{N b_N^{1+d_V}}\right)^{1/2}\right) + O(b_N)$$

by Lemma H.4. We also have that

$$\begin{aligned}
\sup_{z \in \mathcal{Z}_t} \|T(z;\beta_0)\| &= \sup_{z \in \mathcal{Z}_t} |\mathbb{E}[Y_t|Z_t = z] f_{Z_t}(z)| \cdot \left\| \int \xi(a)\mathcal{K}(a)\, da \right\| \\
&\le \sup_{z \in \mathcal{Z}_t} |\mathbb{E}[Y_t|Z_t = z]| \cdot \sup_{z \in \mathcal{Z}_t} f_{Z_t}(z) \cdot O(1) \\
&= O(1)
\end{aligned}$$

by $\sup_{z \in \mathcal{Z}_t} |\mathbb{E}[Y_t|Z_t = z]| < \infty$ (Assumption B4.(iii)), $\sup_{z \in \mathcal{Z}_t} f_{Z_t}(z) < \infty$ (Assumption B4.(iii)), and $\left\| \int \xi(a)\mathcal{K}(a)\, da \right\| < \infty$ (Assumption B3). Therefore,

$$\begin{aligned}
\sup_{z \in \mathcal{Z}_t} \|T_N(z;\beta_0)\| &= O_p\left(\left(\frac{\log(N)}{N b_N^{1+d_V}}\right)^{1/2}\right) + O(b_N) + O(1) \\
&= O_p(1),
\end{aligned}$$

by B6.

Combining the asymptotic orders of the above six terms, we have

$$\begin{aligned}
\sqrt{N b_N} \cdot \left| \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - \widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) \right) \hat{\pi}_{it} \right| &\le \sqrt{N b_N} \cdot O_p(1) \cdot o_p\left(\frac{1}{\sqrt{N b_N}}\right) \\
&\quad + \sqrt{N b_N} \cdot O_p(1) \cdot o_p\left(\frac{1}{\sqrt{N b_N}}\right) \cdot O_p(1) \cdot O_p(1) \\
&= o_p(1).
\end{aligned}$$

Part 2: Convergence of Term (H.8)

We have that

$$
\begin{aligned}
&\left| \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0) \right) \hat{\pi}_{it} \right| \\
&= \left| \frac{1}{N} \sum_{j=1}^{N} e_1' \left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1} T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - S_N(\underline{x}_t'\beta_0, V_i; \beta_0)^{-1} T_N(\underline{x}_t'\beta_0, V_i; \beta_0) \right) \hat{\pi}_{it} \right| \\
&= \left| \frac{1}{N} \sum_{j=1}^{N} e_1' \left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1} (T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - T_N(\underline{x}_t'\beta_0, V_i; \beta_0)) \right.\right. \\
&\left.\left. + S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1} \left(S_N(\underline{x}_t'\beta_0, V_i; \beta_0) - S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) \right) S_N(\underline{x}_t'\beta_0, V_i; \beta_0)^{-1} T_N(\underline{x}_t'\beta_0, V_i; \beta_0) \right) \hat{\pi}_{it} \right| \\
&\leq \|e_1\| \sup_{z \in \mathcal{Z}_t} \left\| S_N(z; \beta_0)^{-1} \right\| \sup_{z \in \mathcal{Z}_t} \left\| \frac{\partial}{\partial u} T_N(z; \beta_0) \right\| \left\| \underline{x}_t'\widehat{\beta} - \underline{x}_t'\beta_0 \right\| \\
&+ \|e_1\| \sup_{z \in \mathcal{Z}_t} \left\| S_N(z; \beta_0)^{-1} \right\| \sup_{z \in \mathcal{Z}_t} \left\| \frac{\partial}{\partial u} S_N(z; \beta_0) \right\| \left\| \underline{x}_t'\widehat{\beta} - \underline{x}_t'\beta_0 \right\| \sup_{z \in \mathcal{Z}_t} \left\| S_N(z; \beta_0)^{-1} \right\| \sup_{z \in \mathcal{Z}_t} \| T_N(z; \beta_0) \|.
\end{aligned} \tag{H.11}
$$

The inequality follows from applications of the mean-value theorem and the Cauchy-Schwarz inequality. By Lemmas H.3 and H.5,

$$
\begin{aligned}
\sup_{z \in \mathcal{Z}_t} \left\| \frac{\partial}{\partial u} S_N(z; \beta_0) \right\| &= o_p\left(\frac{a_N}{\sqrt{N b_N}} \right) \\
\sup_{z \in \mathcal{Z}_t} \left\| \frac{\partial}{\partial u} T_N(z; \beta_0) \right\| &= o_p\left(\frac{a_N}{\sqrt{N b_N}} \right).
\end{aligned}
$$

By B2, $\|\underline{x}_t'\widehat{\beta} - \underline{x}_t'\beta_0\| \leq \|\underline{x}_t\| \|\widehat{\beta} - \beta_0\| = O_p(a_N^{-1})$. The asymptotic orders of all other terms in equation (H.11) were characterized in the analysis of the convergence of term (H.7). Therefore

$$
\begin{aligned}
&\sqrt{N b_N} \cdot \left| \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{h}_1(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0) \right) \hat{\pi}_{it} \right| \\
&= \sqrt{N b_N} \cdot O_p(1) \cdot o_p\left(\frac{a_N}{\sqrt{N b_N}} \right) \cdot O_p(a_N^{-1}) + \sqrt{N b_N} \cdot O_p(1) \cdot o_p\left(\frac{a_N}{\sqrt{N b_N}} \right) \cdot O_p(a_N^{-1}) \cdot O_p(1) \cdot O_p(1) \\
&= o_p(1).
\end{aligned}
$$

Part 3: Convergence of Term (H.9)

First note that

$$
\left| \frac{1}{N} \sum_{i=1}^{N} \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0)(\hat{\pi}_{it} - \pi_{it}) \right| \leq \frac{1}{N} \sum_{i=1}^{N} \left| \widehat{h}_1(\underline{x}_t'\beta_0, V_i; \beta_0) \right| \cdot \sup_{i=1,\ldots,N} |\hat{\pi}_{it} - \pi_{it}|.
$$

Therefore

$$
\begin{aligned}
&\mathbb{P}\left(\sqrt{Nb_N}\left|\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0)(\hat{\pi}_{it}-\pi_{it})\right| = 0\right) \\
&\geq \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0)\right| \cdot \sup_{i=1,\ldots,N}|\hat{\pi}_{it}-\pi_{it}| = 0\right) \\
&\geq \mathbb{P}\left(\sup_{i=1,\ldots,N}|\hat{\pi}_{it}-\pi_{it}| = 0\right) \\
&\to 1
\end{aligned}
$$

as $N\to\infty$ by Lemma H.7. Therefore

$$
\sqrt{Nb_N}\left|\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0)(\hat{\pi}_{it}-\pi_{it})\right| = o_p(1)
$$

Part 4: Convergence of Term (H.10)

By Lemma H.8, this term converges in distribution:

$$
\sqrt{Nb_N}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_1(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_1(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{ASF}_t}(\underline{x}_t'\beta_0)).
$$

The conclusion follows from an application of Slutsky's Theorem. □

Lemma H.10 (APE convergence in distribution)**.** Suppose B1–B6 hold. Then,

$$
\sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_2(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{APE}_t}(\underline{x}_t'\beta_0)).
$$

Proof of Lemma H.10. This proof builds on that of Corollary 2 in KLX and our Lemma H.8. Recall that Assumptions A1–A7 of KLX hold under ours. We can then use equation (13) in KLX and their Corollary 1 to write

$$
\begin{aligned}
b_N\widehat{h}_2(z;\beta_0) &= b_N h_2(z;\beta_0) + B_{2,N}(z) + \frac{1}{N}\sum_{j=1}^{N}\phi_{2,jN}(z) + R_{2,N}(z) \\
&= e_{2+d_V}' h(z;\beta_0) + B_{2,N}(z) + \frac{1}{N}\sum_{j=1}^{N}\phi_{2,jN}(z) + R_{2,N}(z),
\end{aligned}
$$

where $B_{2,N}(z)$ is a bias term satisfying $\sup_{z\in\mathcal{Z}_t}|B_{2,N}(z)| = O(b_N^{\ell+1})$ if ℓ is odd or $O(b_N^{\ell+2})$ if ℓ is even, where $\phi_{2,jN}(z)$ are mean-zero random variables, and where $R_{2,N}(z)$ is a higher-order term satisfying $\sup_{z\in\mathcal{Z}_t}|R_{2,N}(z)| = O_p\left(\frac{\log(N)}{Nb_N^{1+d_V}}\right)$.

Second, note that

$$\sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_2(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right)$$

$$= \sqrt{Nb_N^3}\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_2(\underline{x}_t'\beta_0, V_i;\beta_0) - h_2(\underline{x}_t'\beta_0, V_i;\beta_0)\right)\pi_{it} \tag{H.12}$$

$$+ \sqrt{b_N^3}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(h_2(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_2(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right) \tag{H.13}$$

To analyze term (H.12), we use the fact that

$$\sqrt{Nb_N^3}\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_2(\underline{x}_t'\beta_0, V_i;\beta_0) - h_2(\underline{x}_t'\beta_0, V_i;\beta_0)\right)\pi_{it}$$

$$= \sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}e_{2+d_V}'\left(\widehat{h}(\underline{x}_t'\beta_0, V_i;\beta_0) - h(\underline{x}_t'\beta_0, V_i;\beta_0)\right)\pi_{it}$$

$$= \sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}B_{2,N}(\underline{x}_t'\beta_0, V_i)\pi_{it} + \sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{2,jN}(\underline{x}_t'\beta_0, V_i)\pi_{it} + \sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}R_{2,N}(\underline{x}_t'\beta_0, V_i)\pi_{it}.$$

The terms $\sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}B_{2,N}(\underline{x}_t'\beta_0, V_i)\pi_{it}$ and $\sqrt{Nb_N}\frac{1}{N}\sum_{i=1}^{N}R_{2,N}(\underline{x}_t'\beta_0, V_i)\pi_{it}$ are $o_p(1)$ from the same arguments used in the proof of Lemma H.8.

The term $\sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{2,jN}(\underline{x}_t'\beta_0, V_i)\pi_{it}$ converges in distribution to

$$\sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{2,jN}(\underline{x}_t'\beta_0, V_i)\pi_{it} \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{APE}_t}(\underline{x}_t'\beta_0))$$

by standard arguments from Masry (1996) referred to in the proof of Corollary 2 in KLX.

Term (H.13) above is of order $O_p(b_N^{3/2}) = o_p(1)$ by an application of the central limit theorem. Therefore,

$$\sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_2(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right) = \sqrt{Nb_N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{2,jN}(\underline{x}_t'\beta_0, V_i)\pi_{it} + o_p(1)$$

$$\xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{APE}_t}(\underline{x}_t'\beta_0)).$$

□

Proof of Theorem C.2. First, we write

$$\sqrt{Nb_N^3}\left(\widehat{\mathrm{APE}}_{k,t}(\underline{x}_t) - \mathrm{APE}^{\pi}_{k,t}(\underline{x}_t)\right)$$

$$= \widehat{\beta}^{(k)} \cdot \sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - \widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right) \tag{H.14}$$

$$+ \widehat{\beta}^{(k)} \cdot \sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - \widehat{h}_2(\underline{x}_t'\beta_0, V_i; \beta_0)\right)\hat{\pi}_{it}\right) \tag{H.15}$$

$$+ \widehat{\beta}^{(k)} \cdot \sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(\underline{x}_t'\beta_0, V_i; \beta_0)(\hat{\pi}_{it} - \pi_{it})\right) \tag{H.16}$$

$$+ \widehat{\beta}^{(k)} \cdot \sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{h}_2(\underline{x}_t'\beta_0, V_i; \beta_0)\pi_{it} - \mathbb{E}[h_2(\underline{x}_t'\beta_0, V; \beta_0)\pi_t]\right) \tag{H.17}$$

$$+ \sqrt{Nb_N^3}(\widehat{\beta}^{(k)} - \beta_0^{(k)}) \cdot \mathbb{E}[h_2(\underline{x}_t'\beta_0, V; \beta_0)\pi_t]. \tag{H.18}$$

We will show that terms (H.14)–(H.16) and (H.18) are $o_p(1)$, and that term (H.17) converges in distribution.

Convergence of Term (H.14)

Note that

$$\sqrt{Nb_N^3} \cdot \left|\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - \widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right|$$
$$= \sqrt{Nb_N} \cdot \left|\frac{1}{N}\sum_{i=1}^{N} e_{2+d_V}'\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right|$$

by the definition of $\widehat{h}_2$. Also note that $\widehat{\beta}^{(k)} = O_p(1)$. Therefore, we can follow the same steps used in the proof of Theorem C.1 to show term (H.7) is $o_p(1)$.

Convergence of Term (H.15)

We have that

$$\sqrt{Nb_N^3} \cdot \left|\widehat{\beta}^{(k)}\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{h}_2(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - \widehat{h}_2(\underline{x}_t'\beta_0, V_i; \beta_0)\right)\hat{\pi}_{it}\right|$$
$$= \left|\widehat{\beta}^{(k)}\right| \cdot \sqrt{Nb_N}$$
$$\cdot \left|\frac{1}{N}\sum_{i=1}^{N} e_{2+d_V}'\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - S_N(\underline{x}_t'\beta_0, V_i; \beta_0)^{-1}T_N(\underline{x}_t'\beta_0, V_i; \beta_0)\right)\hat{\pi}_{it}\right|.$$

Again, we can follow the same steps used in the proof of Theorem C.1 to show term (H.8) is

$o_p(1)$.

Convergence of Term (H.16)

The convergence of this term is shown identically to that of term (H.9).

Convergence of Term (H.17)

By Lemma H.10, we have that $\sqrt{Nb_N^3}\left(\frac{1}{N}\sum_{i=1}^N \widehat{h}_2(\underline{x}_t'\beta_0, V_i;\beta_0)\pi_{it} - \mathbb{E}[h_2(\underline{x}_t'\beta_0, V;\beta_0)\pi_t]\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{APE}_t}(\underline{x}_t'\beta_0))$. By B2, $\widehat{\beta}^{(k)} \xrightarrow{p} \beta_0^{(k)}$. Therefore, by Slutsky's Theorem, term (H.17) converges in distribution to a mean-zero Gaussian distribution with variance $(\beta_0^{(k)})^2 \cdot \sigma^2_{\mathrm{APE}_t}(\underline{x}_t'\beta_0)$.

Convergence of Term (H.18)

Note that $\mathbb{E}[h_2(\underline{x}_t'\beta_0, V;\beta_0)\pi_t] = O(1)$. Term (H.18) is of order $\sqrt{Nb_N^3}(\widehat{\beta}^{(k)} - \beta_0^{(k)}) \cdot O(1) = O_p\left(\sqrt{Nb_N^3 a_N^{-1}}\right)$. By B2, the order of this term is

$$O_p\left(N^{\frac{1}{2}(1-3\delta-2\epsilon)}\right) = o_p(1).$$

This equality follows from $\delta > \frac{1-2\epsilon}{3}$, which can be seen from $\delta > 1-2\epsilon$ and $\delta > 0$: see B6.

Combining the convergence of terms (H.14)–(H.18) with Slutsky's Theorem, we obtain our result. $\square$

H.2 Proofs for Appendix E

Proof of Theorem E.1. This proof is similar to the proofs of Theorems 2.1 and 2.2 in the main paper. For the ASF, note that

$$\begin{aligned}
\mathrm{ASF}_t(\underline{x}_t) &= \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)] \\
&= \int_{\mathrm{supp}(V^t)} \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|V^t = v^t]\, dF_{V^t}(v^t) \\
&= \int_{\mathrm{supp}(V^t|X_t'\beta_0 = \underline{x}_t'\beta_0)} \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|V^t = v^t]\, dF_{V^t}(v^t) \\
&= \int_{\mathrm{supp}(V^t|X_t'\beta_0 = \underline{x}_t'\beta_0)} \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|X_t'\beta_0 = \underline{x}_t'\beta_0, V^t = v^t]\, dF_{V^t}(v^t) \\
&= \int_{\mathrm{supp}(V^t|X_t'\beta_0 = \underline{x}_t'\beta_0)} \mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V^t = v^t]\, dF_{V^t}(v^t).
\end{aligned}$$

The second equality follows from the law of iterated expectations. The third follows from the support condition. The fourth follows from $(C, U_t) \perp\!\!\!\perp X_t'\beta_0|V^t$, which can be shown similarly to step 1 in the proof of Theorem 2.1. Note that $U_t \perp\!\!\!\perp X_t'\beta_0|C, V^t$ is implied by $U_t \perp\!\!\!\perp (\mathbf{X}_{\mathrm{exog}}, \mathbf{X}_{\mathrm{pre}}^t)|C$ and by $X_t'\beta_0$ being a function of $(\mathbf{X}_{\mathrm{exog}}, \mathbf{X}_{\mathrm{pre}}^t)$. Also note that $C \perp\!\!\!\perp X_t'\beta_0|V^t$, which follows from $C \perp\!\!\!\perp (\mathbf{X}_{\mathrm{exog}}, \mathbf{X}_{\mathrm{pre}}^t)|V^t$ and from $X_t'\beta_0$ being a function of $(\mathbf{X}_{\mathrm{exog}}, \mathbf{X}_{\mathrm{pre}}^t)$. The last line follows directly.

Finally, the last expression is identified from the distribution of $(Y, \mathbf{X})$ using similar arguments as before.

Proofs for the identification of the APE, LAR, and AME proceed similarly. $\square$

References

Abrevaya, J. (1999): "Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable," *Journal of Econometrics*, 93(2), 203–228.

——— (2000): "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 95(1), 1–23.

Aguirregabiria, V., and J. M. Carro (2021): "Identification of Average Marginal Effects in Fixed Effects Dynamic Discrete Choice Models," *arXiv preprint arXiv:2107.06141*.

Andersen, E. B. (1970): "Asymptotic Properties of Conditional Maximum-Likelihood Estimators," *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(2), 283–301.

Arellano, M., and S. Bonhomme (2017): "Nonlinear Panel Data Methods for Dynamic Heterogeneous Agent Models," *Annual Review of Economics*, 9, 471–496.

Aristodemou, E. (2021): "Semiparametric identification in panel data discrete response models," *Journal of Econometrics*, 220(2), 253–271.

Botosaru, I., and C. Muris (2017): "Binarization for panel models with fixed effects," *cemmap Working Paper*.

Chamberlain, G. (1985): "Heterogeneity, omitted variable bias, and duration dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. S. Singer, Econometric Society Monographs, p. 3–38. Cambridge University Press.

——— (2010): "Binary Response Models for Panel Data: Identification and Information," *Econometrica*, 78(1), 159–168.

Charlier, E., B. Melenberg, and A. H. van Soest (1995): "A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation," *Statistica Neerlandica*, 49(3), 324–342.

Chen, S., J. Si, H. Zhang, and Y. Zhou (2017): "Root-N Consistent Estimation of a Panel Data Binary Response Model With Unknown Correlated Random Effects," *Journal of Business & Economic Statistics*, 35(4), 559–571.

Cox, D. R. (1958): "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 215–232.

Dobronyi, C., J. Gu, and K. I. Kim (2021): "Identification of Dynamic Panel Logit Models with Fixed Effects," *arXiv preprint arXiv:2104.04590*.

Fernández-Val, I. (2009): "Fixed effects estimation of structural parameters and marginal effects in panel probit models," *Journal of Econometrics*, 150(1), 71–85.

Fisher, M., and M. J. Jensen (2022): "Bayesian nonparametric learning of how skill is distributed across the mutual fund industry," *Journal of Econometrics*, 230(1), 131–153.

Honoré, B. E., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68(4), 839–874.

Honoré, B. E., and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models without Strictly Exogeneous Regressors," *Econometrica*, 70(5), 2053–2063.

Honoré, B. E., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74(3), 611–629.

Honoré, B. E., and M. Weidner (2023): "Moment Conditions for Dynamic Panel Logit Models with Fixed Effects," *arXiv preprint arXiv:2005.05942*.

Ichimura, H., and L.-F. Lee (1991): "Semiparametric least squares estimation of multiple index models: Single equation estimation," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics. Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, ed. by J. P. William A. Barnett, and G. E. Tauchen, pp. 3–49. Cambridge University Press.

Khan, S., F. Ouyang, and E. Tamer (2021): "Inference on semiparametric multinomial response models," *Quantitative Economics*, 12(3), 743–777.

Khan, S., M. Ponomareva, and E. Tamer (2023): "Identification of dynamic binary response models," *Journal of Econometrics*, 237(1), 105515.

Kitazawa, Y. (2021): "Transformations and moment conditions for dynamic fixed effects logit models," *Journal of Econometrics*, 229(2), 350–362.

Kong, E., O. Linton, and Y. Xia (2010): "Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model," *Econometric Theory*, 26(5), 1529–1564.

Kyriazidou, E. (1995): "Essays in estimation and testing of econometric models," Ph.D. thesis, Northwestern University.

Lee, M.-J. (1999): "A Root-N Consistent Semiparametric Estimator for Related-Effect Binary Response Panel Data," *Econometrica*, 67(2), 427–433.

Lee, Y.-Y. (2018): "Partial Mean Processes with Generated Regressors: Continuous Treatment Effects and Nonseparable Models," *arXiv preprint arXiv:1811.00157*.

Liu, L. (2023): "Density forecasts in panel data models: A semiparametric Bayesian perspective," *Journal of Business & Economic Statistics*, 41(2), 349–363.

Liu, L., A. Poirier, and J.-L. Shiu (2021): "Identification and estimation of average partial effects in semiparametric binary response panel models," *arXiv preprint arXiv:2105.12891*.

Magnac, T. (2000): "Subsidised Training and Youth Employment: Distinguishing Unobserved Heterogeneity from State Dependence in Labour Market Histories," *The Economic Journal*, 110(466), 805–837.

Magnac, T. (2004): "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica*, 72(6), 1859–1876.

Mammen, E., C. Rothe, and M. Schienle (2012): "Nonparametric regression with nonparametrically generated covariates," *The Annals of Statistics*, 40(2), 1132–1170.

——— (2016): "Semiparametric Estimation With Generated Covariates," *Econometric Theory*, 32(5), 1140–1177.

Manski, C. F. (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica*, 55(2), 357–362.

Masry, E. (1996): "Multivariate local polynomial regression for time series: uniform strong consistency and rates," *Journal of Time Series Analysis*, 17(6), 571–599.

Newey, W. K. (1994): "Kernel estimation of partial means and a general variance estimator," *Econometric Theory*, 10(2), 233–253.

Nolan, D., and D. Pollard (1987): "U-processes: rates of convergence," *The Annals of Statistics*, 15(2), 780–799.

Pakes, A., and D. Pollard (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica*, 57(5), 1027–1057.

Racine, J., and Q. Li (2004): "Nonparametric estimation of regression functions with both categorical and continuous data," *Journal of Econometrics*, 119(1), 99–130.

Rasch, G. (1960): *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* Nielsen & Lydiche.

Rothe, C., and S. Firpo (2019): "Properties of doubly robust estimators when nuisance functions are estimated nonparametrically," *Econometric Theory*, 35(5), 1048–1087.

Torgovitsky, A. (2019): "Nonparametric Inference on State Dependence in Unemployment," *Econometrica*, 87(5), 1475–1505.

Vaart, A. W. v. d. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.