Identification and Estimation of Partial Effects in Nonlinear Semiparametric Panel Models*

Laura Liu   Alexandre Poirier   Ji-Liang Shiu

December 31, 2022

Abstract

Average partial effects (APEs) are often not point identified in panel models with unrestricted unobserved heterogeneity, such as binary response panel model with fixed effects and logistic errors. We show the point identification of various partial effects in a wide class of nonlinear semiparametric models under an index sufficiency assumption on the unobserved heterogeneity, even when the error distribution is unspecified and non-stationary. This assumption does not impose parametric restrictions on the unobserved heterogeneity. We then construct a three-step semiparametric estimator for the APE. In the first step, we estimate the model's common parameters. In the second step, we estimate the conditional expectation of the outcomes given the index and a generated regressor that depends on first-step estimates. In the third step, we average derivatives of this conditional expectation to obtain a partial mean that estimates the APE. We show that this proposed three-step APE estimator is consistent and asymptotically normal. We also propose estimators for the average structural function and average marginal effects. We evaluate its finite-sample properties in Monte Carlo simulations. We then illustrate our estimator in a study of determinants of married women's labor supply.

Keywords: Average partial effects, panel data, semiparametric estimation, unobserved heterogeneity, binary response models

JEL classification: C13, C14, C23, C25

*Liu: Department of Economics, Indiana University, lauraliu@iu.edu. Poirier: Department of Economics, Georgetown University, alexandre.poirier@georgetown.edu. Shiu: Institute for Economic and Social Research, Jinan University, jishiu.econ@gmail.com. We thank Stéphane Bonhomme, Iván Fernández-Val, Jiaying Gu, Chris Hansen, Louise Laage, Sokbae (Simon) Lee, Ying-Ying Lee, Joon Park, Francis Vella, and seminar participants at Simon Fraser, Harvard/MIT, University of Chicago, OSU, Xiamen University, Cambridge, UCL, Oxford, Penn, FRB Philadelphia, Pittsburgh, USC, Rochester, Indiana University, Taiwan Economics Research conference, China Meeting of the Econometric Society, DMV Econometrics Conference, North American Winter Meeting of the Econometric Society, the Conference in Honor of Jim Powell, Workshop on Recent Advances in Econometrics at University of Glasgow, North American Summer Meeting of the Econometric Society, NBER-NSF CEME Conference for Young Econometricians, and Women in Econometrics Conference for helpful comments and discussions. The authors are solely responsible for any remaining errors.

1 Introduction

Nonlinear panel models with unobserved individual heterogeneity are commonly used in empirical research. This paper is concerned with panels where the outcome is generated from the nonlinear semiparametric model

$$Y_{it} = g_t(X_{it}'\beta_0, C_i, U_{it}), \tag{1.1}$$

for units $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$. Here, X_{it} are covariates, C_i is possibly multi-dimensional unobserved heterogeneity, and U_{it} are unobserved idiosyncratic errors. Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{iT})$. The function g_t is potentially unknown and may vary across time. We assume N is large, but that T is small and fixed, as is the case in many microeconomic datasets. This class of models includes fixed effects, random effects, and intermediate levels of structure on the conditional distribution of $C_i|\mathbf{X}_i$. A leading example of this class of models are binary outcome panel models generated by

$$Y_{it} = \mathbb{1}\left(X_{it}'\beta_0 + C_i - U_{it} \geq 0\right). \tag{1.2}$$

See Wooldridge (2010) Chapter 15.8 for an exposition of such models. We will use the binary model to illustrate some results from the more general model in (1.1).

Identification results for the common parameters β_0 in (1.2) are well-known and go back to the work of Rasch (1960) in the case where U_{it} is assumed to be logistic and Manski (1987) when its distribution is unspecified. Identification results for β_0 in many other special cases of (1.1) have been derived. However, in this paper we focus on features of the distribution of the potential outcome

$$Y_{it}(\underline{x}_t) \equiv g_t(\underline{x}_t'\beta_0, C_i, U_{it}) \tag{1.3}$$

where $\underline{x}_t$ is an element of the support of X_t. They include the *average structural function* (ASF), *average partial effects* (APE), and *average marginal effects* (AME). Studied in Blundell and Powell (2003), the ASF at potential value $\underline{x}_t$ is the unconditional expectation of potential outcome $Y_{it}(\underline{x}_t)$. In the binary response model, it is the conditional response probability $\mathbb{P}(Y_{it} = 1|X_{it} = \underline{x}_t, C_i = c)$ averaged over the marginal unobserved heterogeneity distribution F_C. It is used to assess the average impact of interventions in which the value of X_{it} is manipulated. Both the ASF and

APE can be averaged over various distributions for $\underline{x}_t$ to evaluate their averages.[1] The APE is a derivative of the ASF with respect to one covariate, hence measuring the partial effect of this covariate on the conditional response probability averaged over the marginal distribution of C_i. The AME measures the impact on the average potential outcome of a marginal increase in a covariate across the entire population. These measures are commonly used to evaluate the causal impact of policies. See Abrevaya and Hsu (2021) for a survey of various partial effects in panels.[2]

When $C_i|\mathbf{X}_i$ is unrestricted, i.e., fixed effects are assumed, the ASF, APE, and AME are often not point identified. This can be the case even when the error distribution is known: see Davezies, D'Haultfoeuille, and Laage (2021) who show their partial identification in a binary panel logit model with fixed effects.

The first contribution of this paper is to show the point identification of these features under an index sufficiency assumption on the unobserved heterogeneity. This assumption restricts this conditional distribution to depend on covariates only through $v(\mathbf{X}_i)$, a (multiple) index of $\mathbf{X}_i$. It is related to an assumption of Altonji and Matzkin (2005) and Bester and Hansen (2009) that they use to show the identification of the *local average response* (LAR). While the APE averages partial effects over the unconditional distribution of the unobservables, the LAR differs by conditioning on the covariates: see the discussion below and in Section 2.1 for a comparison of these estimands. Under our assumption, $v(\mathbf{X}_i)$ acts as a control function that does not require the specification of a first-stage or the existence of an instrument. As we discuss in Section 2.2.2, that $v(\mathbf{X}_i)$ might be a suitable control variable with enough variation depends crucially on the panel structure of the data. We also allow for estimated indices of the form $v(\mathbf{X}_i)'\gamma_0$, as in multiple index models (Ichimura and Lee, 1991). As in Imbens and Newey (2009), the support of this index variable plays an important role, which we study in detail.

Note that the identification results in this paper do not rely on parametric assumptions on the conditional distribution of $C_i|\mathbf{X}_i$, nor on the distribution of U_{it}. The ASF, APE, and AME depend directly on the distribution of C_i, which is not specified or identified, but these partial effects are identified despite this dependence. Our approach can be viewed as a unified framework for identifying various partial effects under this assumption in a wide class of panel models.

In our second main contribution, we construct three-step semiparametric estimators for the ASF, APE, and AME. In particular, we show the ASF is the partial mean of the conditional expectation

[1] For example, we can average the ASF or APE across all covariates except for one "treatment" variable.

[2] We note that some of their terminology differs from the one used in this paper.

of Y_{it} given $(X_{it}'\beta_0, v(\mathbf{X}_i))$, integrated over the marginal distribution of $v(\mathbf{X}_i)$. The APE and AME can be expressed as average derivatives of this conditional expectation. In a preliminary step, we estimate β_0 using one of the many available estimators in the literature. For example, in binary panels one can use the conditional maximum likelihood estimator (Rasch (1960), Andersen (1970)) when U_{it} is logistic, or the smoothed panel maximum score estimator of Charlier, Melenberg, and van Soest (1995) and Kyriazidou (1995) when U_{it} is unspecified. In a second step, we estimate the above conditional expectation, replacing the unobserved $X_{it}'\beta_0$ by generated regressor $X_{it}'\widehat{\beta}$. We then use local polynomial regression to recover this conditional mean. In the final step, we average this estimated conditional mean over the empirical distribution of $v(\mathbf{X}_i)$. The APE estimator is analogous, replacing the conditional expectation estimate with an estimate of its derivative, which is obtained directly via the local polynomial regression. The AME is similarly estimated.

Next, we provide rate conditions on bandwidths, the convergence rate of $\widehat{\beta}$, and the order of the polynomial regression that allow us to establish the consistency and asymptotic normality of the ASF and APE estimators.[3] Moreover, their convergence rates are fast relative to other nonparametric estimators since, after integrating over the distribution of $v(\mathbf{X}_i)$, the ASF and APE are functions of one-dimensional $X_{it}'\beta_0$. Their convergence rates do not depend on the dimension of $\mathbf{X}_i$. For example, when the index is one-dimensional, the ASF's and APE's rates of convergence are similar to standard rates of convergence of univariate nonparametric kernel regression estimators, which are fast within the class of nonparametric estimators.[4]

We then discuss an extension of our identification results to cases where sequential exogeneity of U_{it} is assumed. This includes models with lagged dependent variables, see, e.g., Honoré and Kyriazidou (2000). The associated estimators are the same under strict or sequential exogeneity.

In the Monte Carlo simulation experiments, we compare the proposed semiparametric estimator with a random effects (RE) estimator and a correlated random effects (CRE) estimator in a binary outcome model. The RE and CRE estimators we consider are commonly used parametric estimators that assume $C_i|V_i$ is Gaussian and that U_{it} is logistic (see the definitions at the beginning of Section 5). Our results show that the semiparametric estimator yields smaller biases but larger standard deviations, and the former channel dominates when the true distribution of the unobserved heterogeneity is non-Gaussian and the true distribution of the idiosyncratic errors is non-logistic.

In our empirical illustration, we study women's labor force participation using our semiparamet-

[3]We leave a complete asymptotic analysis of the AME estimator for future work.

[4]In particular, we show the ASF can converge at a rate faster than $N^{2/5}$ when the index is one-dimensional.

ric estimator. Relative to their parametric RE/CRE counterparts, we see that our APE estimates are closer to zero for lower husband's incomes and more negative for higher ones, while their parametric counterparts vary less with respect to husband's incomes. Additionally, the effects of the husband's income are no longer significant once we allow for the flexibility in the distributions of the unobserved heterogeneity and of the idiosyncratic errors.

Related Literature

We now review the related literature. While we focus on functionals of $Y_t(\underline{x}_t)$, our work builds on a large literature on the identification and estimation of β_0 in model (1.1). This literature can further be subdivided based on its distributional assumptions on $C_i|\mathbf{X}_i$ and those on $U_{it}|C_i, \mathbf{X}_i$.

In the case where g_t is known and both $C_i|\mathbf{X}_i$ and $U_{it}|C_i, \mathbf{X}_i$ are parametrized, the distribution of $Y_i|\mathbf{X}_i$ is fully parametrized and β_0 can be estimated via integrated maximum likelihood. The binary case is studied in Chamberlain (1980). This case includes random effects, where $C_i|\mathbf{X}_i \overset{d}{=} C_i$ and C_i follows a parametric distribution. See Chapter 3 in Chamberlain (1984) and Chapter 15.8 in Wooldridge (2010) for a review of this approach.

Under fixed effects, Manski (1987) shows the identification of β_0 in binary panels when X_t contains a regressor with large support, and when $U_{it}|C_i, \mathbf{X}_i$ is stationary. Abrevaya (1999) considers the identification of β_0 in the model $Y_{it} = g(X_{it}'\beta_0 + C_i + U_{it})$ when U_{it} is nonparametric. The identification argument generalizes the one from Manski (1987) for binary panels. Also see Abrevaya (2000) and Botosaru, Muris, and Pendakur (2021) for identification results under weaker assumptions on the link function g. These papers mainly focused on the identification of β_0 and $g(\cdot)$, but Botosaru and Muris (2022) show a range of point and partial identification results for the distribution of $Y_t(\underline{x}_t)$ under the above assumptions. In contrast, we achieve point identification even when Y_{it} is binary and without assuming stationarity, but we impose restrictions on the conditional distribution of the heterogeneity.

The restriction we consider is an intermediate assumptions on $F_{C_i|\mathbf{X}_i}$, where we assume it depends on $\mathbf{X}_i$ only through a potentially multivariate index $v(\mathbf{X}_i)$. We do not parametrize the distribution of $C_i|\mathbf{X}_i$ nor restrict how it depends on this index. As mentioned above, our primary focus is on aspects of the distribution of $Y_{it}(\underline{x}_t)$, such as the ASF, APE, and AME, rather than β_0.

Due to our conditional independence assumption between the heterogeneity and covariates conditional on an index, our work is also related to a large literature on control functions. Newey, Powell, and Vella (1999) show the identification of structural functions in a triangular model, where a control variable V_i is identified from a first stage. Blundell and Powell (2004) consider a binary response model with endogeneity and focus on the identification and estimation of the ASF. Imbens and Newey (2009) consider a nonseparable triangular model and, like us, focus on identifying functionals of the structural function, such as the ASF. The multiple index structure we obtain for the conditional expectation $\mathbb{E}[Y_{it}|\mathbf{X}_i]$ is related to the work of Ichimura and Lee (1991) and Escanciano, Jacho-Chávez, and Lewbel (2016) among others. For results on the ASF in binary panels, Maurer, Klein, and Vella (2011) use a semiparametric maximum likelihood approach and a control function assumption to identify and estimate the ASF. Also see Laage (2020) for a panel data model with triangular endogeneity.

Although they focus on a different estimand, the work of Altonji and Matzkin (2005) and Bester and Hansen (2009) is closely related to ours. In Altonji and Matzkin (2005), they consider an exchangeability assumption, where $F_{C|X_1,\ldots,X_T}$ is invariant to relabeling of the time indices on the regressors. They then assume that $C_i|X_{it}, v(\mathbf{X}_i) \overset{d}{=} C_i|v(\mathbf{X}_i)$ where $v(\mathbf{X}_i)$ are known symmetric functions of $(X_{i1}, \ldots, X_{iT})$. They consider a nonparametric outcome equation and show the identification of the LAR. The LAR averages changes in the conditional response probability over the *conditional* distribution of the heterogeneity. This object differs from the APE since it integrates over the conditional distribution of $C_i|X_{it}$ rather than its marginal distribution. We discuss in more detail in Section 2.1 the difference in estimands, and related differences in assumptions on the support of the index are illustrated in the discussion after Theorem 2.2. Unlike Altonji and Matzkin (2005), our structural equation (1.1) depends on index $X_{it}'\beta_0$ which allows for much faster rates of convergence for our APE when compared to the rates obtained for the LAR in their nonparametric outcome equation. In particular, their rate of convergence for their LAR estimator decreases with the dimension of X_{it} while the rate of convergence of our APE estimator does not, since it depends on the dimension of $X_{it}'\beta_0$, which is fixed. In Bester and Hansen (2009) they also consider an index sufficiency assumption where $v(\mathbf{X}_i) = (v_1(\mathbf{X}_i^{(1)}), \ldots, v_{d_X}(\mathbf{X}_i^{(d_X)}))$, but where the indices $\{v_j(\cdot)\}_{j=1}^{d_X}$ are allowed to be unknown.[5] We also achieve point identification of the LARs: see Theorem 2.2. Because of the single-index structure of outcome equation $g_t(\mathbf{X}_{it}'\beta_0, C_i, U_{it})$, this identification is achieved under weaker support conditions on $\mathbf{X}_i$ and the indices than in their models.

Other identification approaches in these models have also been proposed. Chernozhukov,

[5]The dimension of X_{it} is denoted by d_X, and $\mathbf{X}_i^{(k)}$ denotes a $T \times 1$ vector with the kth components of X_{it} for $t = 1, \ldots, T$.

Fernández-Val, Hahn, and Newey (2013) derive bounds on the ASF in fixed effects binary response models with nonparametric distributions of $C_i|\mathbf{X}_i$ and of U_{it}. Also see Chernozhukov, Fernández-Val, Hoderlein, Holzmann, and Newey (2015). Graham and Powell (2012) consider a correlated random coefficients model and estimate averages of these coefficients, which correspond to APEs. Hoderlein and White (2012) consider the identification of the LAR for a subpopulation of stayers in a nonseparable model. Bonhomme, Lamadon, and Manresa (2017) consider a discretization of the unobserved heterogeneity as an intermediate assumption between fixed and random effects.

Several results are obtained under parametric assumptions on the error U_{it}. We review results on our leading example, the class binary outcome panel models, for brevity. Fernández-Val (2009) proposes bias-corrected estimators of marginal effects when T is large and when U_{it} follows a normal distribution. Davezies, D'Haultfoeuille, and Laage (2021) derive bounds for the AME and ASF when U_{it} is assumed to be logistic. This builds on a literature studying the identification and estimation of β_0 under logistic errors, which goes back to the work of Rasch (Rasch, 1960, 1961). Also see Andersen (1970), Chamberlain (1980), Magnac (2004), and Chamberlain (2010).

Our estimator is a three-step semiparametric estimator. The first step involves estimating the common parameters β_0, the second step is a nonparametric regression including a generated regressor, and the third step marginalizes over a subset of the regressors. Such estimators are called partial means. 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).

Finally, while we focus on the static case, 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 (2020), Khan, Ponomareva, and Tamer (2020), and Kitazawa (2021) for identification results for common coefficients, and Aguirregabiria and Carro (2020) and Dobronyi, Gu, and Kim (2021) for other functionals such as AMEs. Torgovitsky (2019) obtains partial identification results without parametric restrictions.

The remainder of this paper is organized as follows. In Section 2 we present the baseline model and provide our main identification results. In Section 3 we establish the asymptotic properties of our proposed ASF and APE estimators. Section 4 extends our main identification results to models with sequential exogeneity, allowing for lagged outcomes as regressors. In Section 5 we conduct

Monte Carlo experiments to study the finite-sample properties of our estimators. Section 6 applies our APE estimator to an empirical illustration on female labor force participation. Finally, Section 7 concludes. In Appendix A, we show the identification of the distribution of F_{C_i} under additional support conditions, and Appendix B contains additional details regarding the implementation of our estimator. The appendix also contains the proofs for all propositions and theorems, implementation details, as well as supplemental tables and figures.

2 Model and Identification

In this section, we describe the panel model of interest. Then, we show the identification of the ASF, APE, and AME under an index sufficiency assumption on the conditional distribution of the heterogeneity.

2.1 Model and Estimands

Recall the baseline model in equation (1.1)

$$Y_{it} = g_t(X_{it}'\beta_0, C_i, U_{it}),$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$. Here, $X_{it} \in \mathcal{X}_t \subseteq \mathbb{R}^{d_X}$ are covariates and $\beta_0 \in \mathcal{B} \subseteq \mathbb{R}^{d_X}$ are unknown parameters. Let $\mathbf{X}_i \in \mathcal{X} \subseteq \mathbb{R}^{T \times d_X}$ denote the observed covariate matrix which has X_{it}' as its tth row. Let $C_i \in \mathcal{C} \subseteq \mathbb{R}^{d_C}$ denote the unobserved individual-specific heterogeneity, and $U_{it} \in \mathcal{U}_t \subseteq \mathbb{R}^{d_U}$ are idiosyncratic errors. Let $Y_i = (Y_{i1}, \ldots, Y_{iT})$ denote the vector of outcomes for unit i. Note that the outcome function g_t can depend on the time period. For example, this allows for $g_t(X_{it}'\beta_0, C_i, U_{it}) = g(X_{it}'\beta_0 + \delta_t, C_i, U_{it})$, where δ_t is an additive time-effect, or for $g_t(X_{it}'\beta_0, C_i, U_{it}) = g(X_{it}'\beta_0 + \delta_t C_i, U_{it})$, an interactive effect.

The i subscript is suppressed in the remainder of this section and when there is no confusion. We maintain the following assumptions on the baseline model.

Assumption A1 (Model assumptions). For each $t = 1, \ldots, T$,

(i) Y_t is generated according to equation (1.1);

(ii) $U_t \perp\!\!\!\perp \mathbf{X} | C$;

(iii) $\mathbb{E}[g_t(a, c, U_t) | C = c]$ is bounded over $(a, c) \in \operatorname{supp}(X_t'\beta_0, C)$ and differentiable in a for all $a \in \operatorname{supp}(X_t'\beta_0)$.

Besides assuming model equation (1.1) holds, A1 also imposes that unobserved variables U_t are independent from covariates conditionally on the individual-specific heterogeneity. This strict exogeneity assumption rules out the presence of lagged dependent variables in $\mathbf{X}$. We relax this assumption and consider models with lagged dependent variables in Section 4. The relationship between C and $\mathbf{X}$ is unrestricted by A1. This assumption does not rule out serial correlation or nonstationarity. Assumption A1.(iii) ensures the APE and AME are well defined.

Let $Y_t(\underline{x}_t) \equiv g_t(\underline{x}_t'\beta_0, C, U_t)$ denote the potential outcome at time t evaluated at covariate value $\underline{x}_t \in \mathcal{X}_t$. We define the ASF at time t evaluated at $\underline{x}_t$ by

$$\text{ASF}_t(\underline{x}_t) \equiv \mathbb{E}[Y_t(\underline{x}_t)] = \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)]. \tag{2.1}$$

It is the average outcome if X_t were set to $\underline{x}_t$ in an exogenous manner. The ASF differs from the identified conditional expectation $\mathbb{E}[Y_t|X_t = \underline{x}_t]$ due to the dependence between C and X_t. Note that they are equal under $C \perp\!\!\!\perp X_t$, a random effects assumption.

In the binary response model, it can alternatively be defined as a function of the *conditional response probability*, defined by $\mathbb{P}(Y_t = 1|X_t = \underline{x}, C = c)$. The ASF is then defined as the conditional response probability integrated over the marginal distribution of the unobserved effect C:

$$\text{ASF}_t(\underline{x}_t) = \int_{\mathcal{C}} \mathbb{P}\left(Y_t = 1|X_t = \underline{x}_t, C = c\right) \, dF_C(c).$$

By Assumption A1.(ii), this definition coincides with ours.

Our second object of interest is the APE, which measures the partial effect of changing one covariate, averaged over the marginal distribution of C. If this covariate is continuously distributed, the APE is the derivative of the ASF with respect to this covariate. Formally, define the APE of the kth element of $\underline{x}_t \in \mathcal{X}_t$, denoted by $\underline{x}_t^{(k)}$, as follows:

$$\text{APE}_{k,t}(\underline{x}_t) \equiv \frac{\partial \mathbb{E}[Y_t(\underline{x}_t)]}{\partial \underline{x}_t^{(k)}} = \beta_0^{(k)} \cdot \frac{\partial}{\partial a}\, \mathbb{E}[g_t(a, C, U_t)]|_{a=\underline{x}_t'\beta_0}\,, \tag{2.2}$$

where $\beta_0^{(k)}$ is the kth element of β_0.

In the case where $X_t^{(k)}$ is discretely distributed, the APE be the difference between the ASF at two values, which can be interpreted as an average treatment effect. We let

$$\text{APE}_{k,t}(\underline{x}_t, \underline{x}_{k,t}) \equiv \mathbb{E}[Y_t(\underline{x}_{k,t}) - Y_t(\underline{x}_t)] = \text{ASF}_t(\underline{x}_{k,t}) - \text{ASF}_t(\underline{x}_t). \tag{2.3}$$

where $\underline{x}_{k,t}$ is a vector that differs from $\underline{x}_t$ in its kth position.

Finally, define the *local average response* as

$$\text{LAR}_{k,t}(\underline{x}) \equiv \left.\frac{\partial \mathbb{E}[Y_t(\underline{x}_t)|\mathbf{X}=\tilde{x}]}{\partial \underline{x}_t^{(k)}}\right|_{\tilde{x}=\underline{x}} = \beta_0^{(k)} \cdot \frac{\partial}{\partial a}\,\mathbb{E}[g_t(a,C,U_t)|\mathbf{X}=\underline{x}]|_{a=\underline{x}_t'\beta_0},$$

and its average over the distribution of $\mathbf{X}$ as the *average marginal effect*

$$\text{AME}_{k,t} \equiv \mathbb{E}[\text{LAR}_{k,t}(\mathbf{X})] = \mathbb{E}\left[\frac{\partial}{\partial X_t^{(k)}}\mathbb{E}[Y_t|X_t,C]\right].$$

The local average response can be viewed as a average causal response on the treated (ACRT) since it conditions on the subpopulation with covariate values $\underline{x}$, see Callaway, Goodman-Bacon, and Sant'Anna (2021).[6]

Compared to the LAR, the APE averages this response over the entire population. Thus the APE is analogous to average treatment effects (ATE) in the causal inference literature, which averages the difference between two potential outcomes over its unconditional distribution; meanwhile, the LAR is analogous to a local treatment effect, where the averaging occurs over the conditional distribution of the heterogeneity given $\mathbf{X}=\underline{x}$. Neither estimand is more general since knowledge of the LAR for all $\underline{x} \in \text{supp}(X)$ does not imply knowledge of APEs, and vice-versa. We later show that the LAR is identified under weaker index support assumptions than the APE.

The AME averages this local response over the distribution of covariates and essentially measures the impact of a small change in covariate $X_t^{(k)}$ for all units on average outcomes.

Remark 2.1 (Integrated Estimands)**.** It may also be of interest to consider averages of the ASF or APE over certain covariate values. For example, one can consider the APE's average over the marginal distribution of X_t, or over the distribution of all covariates except for one:

$$\widetilde{\text{APE}}_{k,t} \equiv \int_{\text{supp}(X_t)} \text{ASF}_t(\underline{x}_t)\, dF_{X_t}(\underline{x}_t)$$
$$\widetilde{\text{APE}}_{k,t}(\underline{x}_t^{(k)}) \equiv \int_{\text{supp}(X_t^{(-k)})} \text{APE}_{k,t}(\underline{x}_t)\, dF_{X_t^{(-k)}}(\underline{x}_t^{(-k)})$$

where the $(-k)$ superscript denotes removal of the kth entry.

[6]Note that we can also define an alternative LAR that conditions on covariate values at time t only: $\partial\mathbb{E}[Y_t(\underline{x}_t)|X_t = \tilde{x}_t]/\partial\underline{x}_t^{(k)}|_{\tilde{x}_t=\underline{x}_t}$. This alternative LAR can be obtained from $\mathbb{E}[\text{LAR}_{k,t}(\mathbf{X})|X_t=\underline{x}_t]$.

2.2 Identifying Assumptions

Without further assumptions, it is generally impossible to point identify these partial effects, even under parametric assumptions on U_t. Under fixed effects, the ASF, APE, and AME are generally partially identified. Point identification is obtained if random effects are assumed: $C \perp\!\!\!\perp \mathbf{X}$. It can also be achieved, for example, under the assumptions of Botosaru and Muris (2022), which include g_t being invertible in $X_t'\beta_0 + C - U_t$, Y_t being continuous, and (β_0, g_t) being identified.

2.2.1 (Non-)Identification under Fixed Effects

To fix ideas, we illustrate this identification failure in the binary response model with scalar individual effects, a special case of our general model.

Denote by $G_t(\underline{x}_t'\beta_0, \mathbf{x})$ the counterfactual conditional probability

$$G_t(\underline{x}_t'\beta_0, \mathbf{x}) \equiv \mathbb{P}(Y_t(\underline{x}_t) = 1 | \mathbf{X} = \mathbf{x}) = \int_{\text{supp}(C|\mathbf{X}=\mathbf{x})} F_{U_t|C}(\underline{x}_t'\beta_0 + c|c)\, dF_{C|\mathbf{X}}(c|\mathbf{x}).$$

Note that we observe the conditional probabilities $\mathbb{P}(Y_t = 1|\mathbf{X} = \mathbf{x}) \equiv G_t(x_t'\beta_0, \mathbf{x})$ for all $\mathbf{x} \in \text{supp}(X)$ and $t \in \{1, \ldots, T\}$. By the law of total probability, the ASF for covariate value $\underline{x}_t$ is

$$\begin{aligned}
\text{ASF}_t(\underline{x}_t) &= \int_{\mathcal{C}} F_{U_t|C}(\underline{x}_t'\beta_0 + c|c)\, dF_C(c) \\
&= \int_{\mathcal{X}} G_t(\underline{x}_t'\beta_0, \mathbf{x})\, dF_{\mathbf{X}}(\mathbf{x}) && (2.4)\\
&= \int_{\mathcal{X}} \underbrace{G_t(\underline{x}_t'\beta_0, \mathbf{x})\mathbb{1}(x_t'\beta_0 = \underline{x}_t'\beta_0)}_{=\mathbb{P}(Y_t=1|\mathbf{X}=\mathbf{x})\mathbb{1}(x_t'\beta_0=\underline{x}_t'\beta_0)} dF_{\mathbf{X}}(\mathbf{x}) + \int_{\mathcal{X}} \underbrace{G_t(\underline{x}_t'\beta_0, \mathbf{x})\mathbb{1}(x_t'\beta_0 \neq \underline{x}_t'\beta_0)}_{\text{not point identified}} dF_{\mathbf{X}}(\mathbf{x}). && (2.5)
\end{aligned}$$

We can see from equation (2.4) that the ASF is an average over the distribution of $\mathbf{X}$ of conditional probability $G_t(\underline{x}_t'\beta_0, \mathbf{X})$. For the ASF to be point identified, we need $G_t(\underline{x}_t'\beta_0, \mathbf{x})$ to be identified for $\mathbf{x} \in \text{supp}(\mathbf{X})$, but this generally fails since, given $X_t'\beta_0 = \underline{x}_t'\beta_0$, the support of $\mathbf{X}$ does not equal its marginal support. In equation (2.5), the $G_t(\underline{x}_t'\beta_0, \mathbf{x})$ where $x_t'\beta_0 \neq \underline{x}_t'\beta_0$ are counterfactual probabilities that are not point identified from the data since they do not correspond to any conditional probability of Y_t given $\mathbf{X} = \mathbf{x}$. Unless restrictions are imposed on the distribution of $C|\mathbf{X}$ or other aspects of the model, this causes the ASF, and therefore the APE too, to be partially identified. For the logit case, see Davezies, D'Haultfoeuille, and Laage (2021) for partial identification results for the ASF and AME. Without the assumption that the error distribution is known and logistic, these bounds become weakly larger. Thus, point identification is not achieved in other binary response models. In the nonparametric case, bounds on the ASF are obtained in

Chernozhukov, Fernández-Val, Hahn, and Newey (2013).

2.2.2 An Index Assumption

To achieve point identification, in this paper we instead consider an index sufficiency restriction that imposes additional structure on the conditional distribution of the heterogeneity. In the above example, we assume that $G_t(\underline{x}_t'\beta_0, \mathbf{x})$ depends on $\mathbf{x}$ only through index functions.

Assumption A2 (Index sufficiency). Let $V \equiv v(\mathbf{X})$, where $v : \mathbb{R}^{T \times d_X} \to \mathbb{R}^{d_V}$ is known. Let $C \mid \mathbf{X} \overset{d}{=} C \mid V$.

This assumption is a correlated random effects assumption that restricts the conditional distribution of $C|\mathbf{X}$ to depend solely on $v(\mathbf{X})$, which are indices of $\mathbf{X}$. The conditional distribution of $C|v(\mathbf{X})$ remains nonparametric though.

Motivation for Assumption A2

Such an index assumption is considered in Altonji and Matzkin (2005), and in Bester and Hansen (2009), and can be justified from several perspectives.

In Altonji and Matzkin (2005), the exchangeability of $f_{C|\mathbf{X}}(c|x_1, \ldots, x_T)$ in $(x_1, \ldots, x_T)$ is assumed. They consider symmetric polynomials as candidates for the index function, e.g., $v(\mathbf{X}) = \left(\sum_{t=1}^T X_t, \sum_{1 \leq t_1 < t_2 \leq T} X_{t_1} X_{t_2}\right)$ when the indices are the first two elementary symmetric functions and X_t is scalar. Unlike us, Bester and Hansen (2009) do not assume $v(\cdot)$ is known, but they do not allow for the indices to be arbitrary functions of $\mathbf{X}$: each component on the index may only depend on one component of $\mathbf{X}_t$. Note that if their assumptions are met, it is possible to relax A2 and assume the index function is unknown. This requires that $T \geq 3$, that covariates are continuously distributed, and that the index v satisfies their separability requirement. The focus of these two papers is also different from ours: they identify the LAR rather than the ASF or APE, and its identification is shown for continuous covariates.

Treatment assignment models in panel data can also be used to find candidate indices. This is explored in Arkhangelsky and Imbens (2019). For example, consider the following treatment assignment model when X_t is binary: Assume that $X_t = \mathbb{1}(E_t \leq \nu(C))$ where E_t are iid and independent of C, and ν is an arbitrary function. In this case, $(X_1, \ldots, X_T)|C$ are iid Bernoulli variables, hence $v(\mathbf{X}) = \sum_{t=1}^T X_t$ is a sufficient statistic that satisfies $\mathbf{X}|C, v(\mathbf{X}) \overset{d}{=} \mathbf{X}|v(\mathbf{X})$ which implies A2 holds. This fact can be easily derived from the Fisher-Neyman factorization theorem.

This result is generalized in Arkhangelsky and Imbens (2019) to cases where the distribution of $\mathbf{X}|C$ is from an exponential family with known sufficient statistic. For example, if $(X_1, \ldots, X_T)|C$ are assumed iid Gaussian, then $v(\mathbf{X}) = (\sum_{t=1}^T X_t, \sum_{t=1}^T X_t^2)$ forms a sufficient statistics for $\mathbf{X}$.

In a special case where the indices are time-averages, i.e., $v(\mathbf{X}) = \frac{1}{T}\sum_{t=1}^T X_t$, the index assumption is consistent with $C = \zeta\left(\left(\frac{1}{T}\sum_{t=1}^T X_t\right)' \gamma_0, \eta\right)$ where $\eta \perp\!\!\!\perp \mathbf{X}$ and $\zeta(\cdot, \cdot)$ is any function. This is a relaxation of the specification of the conditional distribution of C given $\mathbf{X}$ in Mundlak (1978) since we do not specify the distribution of η nor restrict the functional form of $\zeta(\cdot, \cdot)$. In this specification for C, the one-dimensional index $\tilde{v}(\mathbf{X}) = \sum_{t=1}^T X_t'\gamma_0$ also satisfies A2, but is unknown due to its dependence on unknown γ_0. Proposition 2.2 below shows how the unknown index parameter can be identified using the work of Ichimura and Lee (1991) on the identification and estimation of multiple index models.

2.2.3 Identification of Common Parameters

While β_0 is not the object of interest, its identification facilitates the identification of partial effects. In what follows, we also take the identification of β_0 as given.

Assumption A3 (Identification of Coefficients). β_0 is point identified or point identified up to scale.

This assumption can be justified by the fact that β_0's identification can be established for many special cases of models (1.1) satisfying Assumption (1). For example, when Y_t is binary, $g_t(X_t'\beta_0, C, U_t) = \mathbb{1}(X_t'\beta_0 + C - U_t \geq 0)$, and U_t follows a standard logistic distribution, Rasch (1960) showed that β_0 is point identified under minimal assumptions requiring variation in X_t over time. Still, in the binary outcome model, Manski (1987) showed that β_0 is identified up to scale when $U_t|C, \mathbf{X}$ is stationary and when a regressor has large support. Unlike the previous result, this does not require knowledge that U_t follows a logistic distribution. Zhu (2022) recently showed this identification holds under weaker support assumptions on the regressors.

Manski's result is generalized to non-binary outcomes in Abrevaya (1999) where he considers $g_t(X_t'\beta_0 + C + U_t) = h(C + X_t'\beta_0 - U_t)$, where h is weakly increasing. He assumes the existence of a regressor with large support and shows β_0 is point identified up to scale. Abrevaya (2000) and Botosaru, Muris, and Pendakur (2021) generalize these results to nonseparable and time-varying models. A variety of other identification approaches can also be used. These include special regressors, as in Honoré and Lewbel (2002). See also Lee (1999) and Chen, Si, Zhang, and Zhou

(2017) for alternative assumptions that yield the identification of β_0.

Identification of β_0 under the Index Condition

The works cited above do not use assumption A2 to deliver the identification of β_0. However, it is possible to use A2 to show β_0's identification under minimal additional assumptions on the model.

To illustrate this, first consider the case with two time periods and known index function.

Proposition 2.1. Let assumptions A1 and A2 hold. Suppose there exists $s, t \in \{1, \ldots, T\}$ such that $s \neq t$, $\Psi_s(u, v) \equiv \mathbb{E}[Y_s | X_s'\beta_0 = u, V = v]$ is differentiable with $\frac{\partial}{\partial u}\Psi_s(u, v) \neq 0$, and that $W_t(\mathbf{x}) \equiv \frac{\partial}{\partial x_t} v(\mathbf{x})'$ has rank d_V. Normalize $\beta_0 = (1, \tilde{\beta}_0)$, where $\widetilde{\beta}_0 \in \mathbb{R}^{d_X - 1}$. Then, $\widetilde{\beta}_0$ is identified.

This result shows that β_0 can be identified (up to scale) with as few as two time periods, assuming that derivatives of $v(\mathbf{x})'$ have rank d_V. This rank condition requires $d_V \leq \dim(X_t)$, thus that the indices are of lower dimension than $\mathbf{X}$.

We also consider the following generalization of Assumption A2.

Assumption A4 (Unknown index sufficiency). Let $V \equiv v(\mathbf{X})$, where $v : \mathbb{R}^{T \times d_X} \to \mathbb{R}^{d_V \times d_\gamma}$ is known. Let $C \mid \mathbf{X} \overset{d}{=} C \mid V'\gamma_0$, where $\gamma_0 \in \mathbb{R}^{d_\gamma}$.

This assumption generalizes A2 since it allows the index $V'\gamma_0$ to be unknown. Note that $\dim(V'\gamma_0)$ can be multivariate. Since $C \perp\!\!\!\perp \mathbf{X} | V'\gamma_0$ implies that $C \perp\!\!\!\perp \mathbf{X} | V$, A2 is weaker than A4. The lower dimension of $V'\gamma_0$ relative to V helps estimate the conditional expectation of the outcomes given both $X_t'\beta_0$ and the sufficient statistics, which is done nonparametrically in Section 3.

Under this assumption we can show that

$$\begin{aligned} \mathbb{E}[Y_t | \mathbf{X} = \underline{x}] &= \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t) | V'\gamma_0 = v(\underline{x})'\gamma_0] \\ &\equiv \Psi_t(\underline{x}_t'\gamma_0, v(\underline{x})'\gamma_0), \end{aligned}$$

an unknown function containing $d_V + 1$ linear indices of $\underline{x}$. Ichimura and Lee (1991) show that under conditions on the indices $(\underline{x}_t'\gamma_0, v(\underline{x})'\gamma_0)$, the parameters (β_0, γ_0) are point identified up to scale and are estimable at a $\sqrt{N}$-rate. These conditions require that each index contains a continuous component, that none of the vectors $(X_t, v_1(\mathbf{X}), \ldots, v_{d_\gamma}(\mathbf{X}))$ are contained in one another, and the linear independence of the $1 + d_V$ partial derivatives of Ψ_t with respect to the indices. Because of the panel structure of the data, the indices $v(\mathbf{X})$ and X_t generally contain different elements since

$v(\mathbf{X})$ can depend on $(X_1, \ldots, X_T)$, whereas X_t is a single element of this vector. We now state a version of their Lemma 3 when regressors are continuous.

Proposition 2.2 (Lemma 3 in Ichimura and Lee (1991))**.** Let $V = v(\mathbf{X})'\gamma_0 \equiv (v_1(\mathbf{X})'\gamma_0, \ldots, v_{d_V}(\mathbf{X})'\gamma_0) \in \mathbb{R}^{d_V}$, where $\gamma_0 \in \mathbb{R}^{d_\gamma}$ is unknown. Assume that

1. Assumptions A1 and A4 hold;
2. Each of $(X_t, v_1(\mathbf{X}), \ldots, v_{d_V}(\mathbf{X}))$ contains a continuous component with nonzero coefficient which is not contained in any of the other variables;
3. For some $t \in \{1, \ldots, T\}$, the function $\Psi_t(u, v) \equiv \mathbb{E}[Y_t | X_t'\beta_0 = u, V'\gamma_0 = v]$ is differentiable;
4. The partial derivatives $\left(\frac{\partial \Psi_t(u,v)}{\partial u}, \frac{\partial \Psi_t(u,v)}{\partial v^{(1)}}, \cdots, \frac{\partial \Psi_t(u,v)}{\partial v^{(d_V)}}\right)|_{(u,v)=(X_t'\beta_0, v(\mathbf{X})'\gamma_0)}$ are not linearly dependent with probability 1.

Then, (β_0, γ_0) are identified up to scale.

To summarize, we have highlighted a few identification approaches that result in the point identification (up to scale) of β_0. This list is not exhaustive, and it contains a new approach that uses the index sufficiency assumption.

2.3 Identification of Partial Effects

We can now state our two main identification theorems.

Theorem 2.1. Let $\underline{x}_t \in \text{supp}(X_t)$ and let assumptions A1–A3 hold. Let $\text{supp}(V | X_t'\beta_0 = u) = \text{supp}(V)$ for all u in a neighborhood of $\underline{x}_t'\beta_0$. Then $\text{ASF}_t(\underline{x}_t)$ and $\text{APE}_{k,t}(\underline{x}_t)$ are point identified from the distribution of $(Y, \mathbf{X})$ for all $t \in \{1, \ldots, T\}$.

This theorem assumes that $X_t^{(k)}$ is continuously distributed. For discretely distributed $X_t^{(k)}$, the APE is a difference between two ASFs, and its identification is achieved when the two corresponding ASFs are point identified. We omit this case for brevity.

Intuitively, point identification occurs because we can write the ASF as follows:

$$\begin{aligned}
\text{ASF}_t(\underline{x}_t) &\equiv \mathbb{E}[Y_t(\underline{x}_t)] \\
&= \int_{\text{supp}(V)} \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|V = v]\, dF_V(v) \\
&= \int_{\text{supp}(V)} \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]\, dF_V(v) \\
&= \mathbb{E}[\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V]].
\end{aligned} \tag{2.6}$$

The second equality follows from the law of iterated expectations and the third one from the index restriction. Equation (2.6) depends only on $\{\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v] : v \in \text{supp}(V)\}$ and on the marginal distribution of V, which are both identified from the data. All identification results in this subsection still hold if V is replaced by $V'\gamma_0$, and Assumption A4 holds.

Remark 2.2. Note that Theorem 2.1 can be used to point identify $\mathbb{E}[m(Y_t(\underline{x}_t))]$ for any known function m without having to modify any of its assumptions. In particular, one can identify $F_{Y_t(\underline{x}_t)}(y) = \mathbb{E}[\mathbb{1}(Y_t(\underline{x}_t) \leq y)]$, the potential outcomes cdf, by setting $m(a) = \mathbb{1}(a \leq y)$ under the assumptions of Theorem 2.1. Therefore, one can also identify the Quantile Structural Function (QSF),[7] $\text{QSF}_t(\tau; \underline{x}_t) \equiv F^{-1}_{Y_t(\underline{x}_t)}(\tau)$ whenever the ASF is identified.

To identify the APE for a continuous regressor, we note that the support assumption implies the ASF is point identified for values of X_t near $\underline{x}_t$. Since the APE is a derivative of the ASF, we can identify the APE as a limit of finite differences between identified ASFs. Formally, we can write

$$\text{APE}_{k,t}(\underline{x}_t) = \mathbb{E}\left[\frac{\partial}{\partial \underline{x}_t^{(k)}} \mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V]\right]. \tag{2.7}$$

All quantities in equation (2.7) are identified, hence the APE is identified. Note that the identification of the ASF and APE bypasses the need to identify F_C, the distribution of the heterogeneity. In Appendix A, we show that F_C is identified under stronger support assumptions on $(X_t'\beta_0, V)$ when outcomes are binary and C is scalar.

We now contrast these with identification results for the LAR and AME.

Theorem 2.2. Let $\underline{x} \in \text{supp}(\mathbf{X})$ and let assumptions A1–A3 hold. Let $v(\underline{x}) \in \text{supp}(V|X_t'\beta_0 = u)$ for all u in a neighborhood of $\underline{x}_t'\beta_0$. Then $\text{LAR}_{k,t}(\underline{x})$ is point identified from the distribution of

[7] See Imbens and Newey (2009) or Chernozhukov, Fernández-Val, Hoderlein, Holzmann, and Newey (2015) for example.

$(Y, \mathbf{X})$ for all $t \in \{1, \ldots, T\}$. If the support condition holds for all $\underline{x} \in \text{supp}(\mathbf{X})$ up to a $P_{\mathbf{X}}$-measure zero set, then $\text{AME}_{k,t}$ is point identified.

These equations help explain how the LAR and AME are identified:

$$\text{LAR}_{k,t}(\underline{x}) = \frac{\partial}{\partial \underline{x}_t^{(k)}} \mathbb{E}[Y_t | X_t'\beta_0 = \underline{x}_t'\beta_0, V = \tilde{v}]|_{\tilde{v}=v(\underline{x})} \tag{2.8}$$

$$\text{AME}_{k,t} = \mathbb{E}[\text{LAR}_{k,t}(\mathbf{X})] = \mathbb{E}\left[\frac{\partial}{\partial X_t^{(k)}} \mathbb{E}[Y_t | X_t'\beta_0, V]\right]. \tag{2.9}$$

Note that the condition for the identification of the LAR is weaker than that for the APE. The APE requires that $\text{supp}(V) = \text{supp}(V | X_t'\beta_0 = u)$ for u in a neighborhood of $\underline{x}_t'\beta_0$, while the LAR requires that $v(\underline{x}) \in \text{supp}(V | X_t'\beta_0 = u)$ for u in a neighborhood of $\underline{x}_t'\beta_0$. This is weaker because $v(\underline{x}) \in \text{supp}(V)$ by construction.

These support conditions indirectly but critically rely on the panel structure of the data and the scalar nature of the index $X_t'\beta_0$. We discuss these support conditions below.

Discussion of the Support Conditions in Theorems 2.1–2.2

The support condition in Theorem 2.1 requires $X_t'\beta_0$ be continuously distributed in a neighborhood of $\underline{x}_t'\beta_0$. This allows for some components of X_t to be discretely distributed. Note that we do not require $X_t'\beta_0$ to be supported on the entire real line. We do require that the support of the sufficient statistic is independent of the value of $X_t'\beta_0$ in a neighborhood of $\underline{x}_t'\beta_0$. This support assumption is related to the common support assumption of Imbens and Newey (2009), although we only restrict the support of $V | X_t'\beta_0$ rather than the support of $V | X_t$. This is a key benefit of considering the index $X_t'\beta_0$ in the outcome equation, as the support of $V | X_t'\beta_0$ is by construction a superset of the support of $V | X_t$. As opposed to Imbens and Newey (2009), we do not posit the existence of a first stage or of exogenous excluded variables since our indices are functions of $\mathbf{X}$ only.

Altonji and Matzkin (2005) do not consider the identification of the ASF/APE and instead focus on the LAR. To understand the difference in identifying assumptions, in their nonparametric setting, identification of the ASF or APE would require $\text{supp}(v(\mathbf{X}) | X_t = \underline{x}) = \text{supp}(v(\mathbf{X}))$. This is significantly stronger than our condition whenever more than one covariate is present. To see this, assume $d_X = T = 2$, $X_t^{(1)}$ is continuously distributed on $\mathbb{R}$, and that $X_t^{(2)} \in \{0, 1\}$ is binary. Let $v(\mathbf{X}) = \sum_{t=1}^2 X_t = (\sum_{t=1}^2 X_t^{(1)}, \sum_{t=1}^2 X_t^{(2)})$. Then, under minimal assumptions, $\text{supp}(v(\mathbf{X})) = \mathbb{R} \times \{0, 1, 2\}$ but $\text{supp}(v(\mathbf{X}) | X_1 = \underline{x}) = \mathbb{R} \times \{\underline{x}^{(2)}, \underline{x}^{(2)} + 1\} \neq \text{supp}(v(\mathbf{X}))$. On the other hand, the conditional support of $v(\mathbf{X})$ given $\{X_t'\beta_0 = x_t'\beta_0\}$ equals $\text{supp}(v(\mathbf{X}))$ when $\beta_0^{(1)} \neq 0$. Therefore, in this example, the ASF/APE will be identified under our assumptions in the semiparametric model.

This important condition also has implications on the dimension of $v(\mathbf{X})$. For example, this condition is violated when $v(\mathbf{X}) = \mathbf{X}$, i.e., no index restrictions are imposed and, equivalently, we have fixed effects. This is because the support of $\mathbf{X}$ does not equal its conditional support given $X_t'\beta_0$: $\text{supp}(\mathbf{X}|X_t'\beta_0 = \underline{x}_t'\beta_0) \neq \text{supp}(\mathbf{X})$. In the case where $v(\mathbf{X}) = \sum_{t=1}^T X_t \in \mathbb{R}^{d_X}$, this condition is written as $\text{supp}(\sum_{t=1}^T X_t | X_t'\beta_0 = u) = \text{supp}(\sum_{t=1}^T X_t)$. For example, we can see that this holds in the simple case where $(X_1, \ldots, X_T)$ are jointly normally distributed. In Remark 2.3 below, we show that while this support condition may not always be warranted, the ASF and APE are partially identified when it fails.

Finally, although not studied here, we note that the support conditions are potentially testable since they only depend on the observed variables $\mathbf{X}$ and identified parameter β_0.

Remark 2.3 (Relaxing support assumptions)**.** The validity of the support assumptions in theorems 2.1 and 2.2 depends intricately on the support of $\mathbf{X}$, and on the considered value $\underline{x}_t$. For a given value of $\underline{x}_t$, this assumption may fail. When it does, we can show the ASF, APE, and AME are partially identified instead. Let $\mathcal{V}_t(u) = \text{supp}(V|X_t'\beta_0 = u)$, and assume that $\mathcal{V}_t(\underline{x}_t'\beta_0) \subsetneq \mathcal{V} \equiv \text{supp}(V)$. Then, the conditional expectation $\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]$ is identified for all $v \in \mathcal{V}_t(\underline{x}_t'\beta_0)$. Therefore, following Theorem 4 in Imbens and Newey (2009), the ASF is contained in

$$\begin{aligned}
\text{ASF}_t(\underline{x}_t) &\in \left[\underline{\text{ASF}}_t(\underline{x}_t), \overline{\text{ASF}}_t(\underline{x}_t)\right] \\
&\equiv \Bigg[\int_{\mathcal{V}_t(\underline{x}_t'\beta_0)} \mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]\, dF_V(v) + \underline{g}\, \mathbb{P}(V \in \mathcal{V} \setminus \mathcal{V}_t(\underline{x}_t'\beta_0)), \\
&\qquad \int_{\mathcal{V}_t(\underline{x}_t'\beta_0)} \mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]\, dF_V(v) + \overline{g}\, \mathbb{P}(V \in \mathcal{V} \setminus \mathcal{V}_t(\underline{x}_t'\beta_0))\Bigg]
\end{aligned}$$

provided that $g_t(\underline{x}_t'\beta_0, C, U_t) \in [\underline{g}, \overline{g}]$ with probability 1. In the case where Y_t is binary, $[\underline{g}, \overline{g}] = [0, 1]$ and the bounds take on a simpler form. The width of these bounds depends only on $\overline{g} - \underline{g}$ and the probability V lies outside of $\mathcal{V}_t(\underline{x}_t'\beta_0)$, which is small if V is continuously distributed and the measure of $\mathcal{V} \setminus \mathcal{V}_t(\underline{x}_t'\beta_0)$ is close to zero. Hence, small violations of the support condition yield a narrow identified set. These bounds can be used to construct bounds on the APE with discrete covariates as well:

$$\text{APE}_{k,t}(\underline{x}_t, \underline{x}_{k,t}) \in \left[\underline{\text{ASF}}_t(\underline{x}_{k,t}) - \overline{\text{ASF}}_t(\underline{x}_t), \overline{\text{ASF}}_t(\underline{x}_{k,t}) - \underline{\text{ASF}}_t(\underline{x}_t)\right].$$

To obtain bounds on the APE for a continuous covariate, bounds on $\frac{\partial}{\partial a}\mathbb{E}[g_t(a, C, U_t)|C = c]$ are needed. In the case where $\frac{\partial}{\partial a}\mathbb{E}[g_t(a, C, U_t)|C = c] \in [\underline{g}', \overline{g}']$ and $\beta_0^{(k)} > 0$, the APE bounds are given by

$$\begin{aligned}
\text{APE}_{k,t}(\underline{x}_t) &\in \left[\underline{\text{APE}}_{k,t}(\underline{x}_t), \overline{\text{APE}}_{k,t}(\underline{x}_t)\right] \\
&\equiv \Bigg[\int_{\mathcal{V}_t(\underline{x}_t'\beta_0)} \frac{\partial}{\partial \underline{x}_t^{(k)}} \mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]\, dF_V(v) + \underline{g}'\beta_0^{(k)}\, \mathbb{P}(V \in \mathcal{V} \setminus \mathcal{V}_t(\underline{x}_t'\beta_0)), \\
&\qquad \int_{\mathcal{V}_t(\underline{x}_t'\beta_0)} \frac{\partial}{\partial \underline{x}_t^{(k)}} \mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]\, dF_V(v) + \overline{g}'\beta_0^{(k)}\, \mathbb{P}(V \in \mathcal{V} \setminus \mathcal{V}_t(\underline{x}_t'\beta_0))\Bigg].
\end{aligned}$$

The bounds are reversed if $\beta_0^{(k)} < 0$. With binary outcomes and under the assumption that $U_t \perp\!\!\!\perp C$, the partial derivative $\frac{\partial}{\partial a}\mathbb{E}[g_t(a, C, U_t)|C = c]$ is the density f_{U_t}. Its lower bound is trivially 0, but it is harder to postulate an upper bound. However, in the logit case this density attains a maximal value of 1/4 at the origin. Therefore, substituting $[\underline{g}', \overline{g}'] = [0, 1/4]$ gives us bounds on the APE in binary panel logits when common support fails.

Bounds for the AME can be similarly obtained. The estimation methods we provide below for the point identified ASF, APE, or AME can be adapted to estimate these bounds under support condition violations.

Remark 2.4 (Excluded control variable)**.** A more general version of A2 is that we can *identify* a variable V such that $C|\mathbf{X}, V \overset{d}{=} C|V$. This is a control variable assumption, where V may be an unobservable that is not functionally related to X, as is the case in A2. For example, it could be the residual in a first-stage equation relating X to some excluded instruments, whose existence we do not assume in this paper. See, for example, Imbens and Newey (2009) in the nonseparable cross-sectional case or Laage (2020) for a panel model with triangular endogeneity and control functions. If this control variable satisfies the support conditions in either Theorem 2.1 or 2.2, the corresponding identification result still applies provided that $U_t \perp\!\!\!\perp \mathbf{X}|(C, V)$, a modification of A1.(ii). As for estimation, if V is identified from a first-stage equation, we should substitute $\widehat{V}_i$ for V_i, where $\widehat{V}_i$ is a suitable estimator for the control variable. This additional generated regressor's impact on the limiting distribution would then have to be taken into account.[8] In this paper, we focus on the case where no such V is observed or identified from a first-stage model, and instead where V is an index of $\mathbf{X}$.

[8]Note that we take into account the generated regressor $X_t'\widehat{\beta}$'s impact in this paper.

3 Estimation

We now propose estimators for the ASF and APE, and establish their limiting distributions. We also propose estimators for the LAR and AME. The estimators we construct are sample analogs of (2.6) for the ASF, and of (2.7) for the APE. 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 the same 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 estimator being a partial mean, 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 does not depend on d_X or T, the dimensions of $\mathbf{X}$.

Throughout this section, we assume that 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.

3.1 Estimation of β_0

In section 2.2.3 we discussed a number of prior 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 not compatible 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). It converges at the rate $N^{\nu/(2\nu+1)}$, where ν is the order of the kernel used to estimate $\widehat{\beta}$.

With continuous outcomes, Ichimura and Lee (1991)'s approach can be used to estimate β_0, and any estimated index parameter γ_0 when $V = v(\mathbf{X})'\gamma_0$, at a $\sqrt{N}$-rate. 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.

In binary panels, the rate of convergence is usually slower. 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, some alternative assumptions and estimators allow for it. In particular, Lee (1999) considers an "index increment sufficiency" assumption: $(X_t'\beta_0, C)|X_t - X_s \stackrel{d}{=} (X_t'\beta_0, C)|(X_t - X_s)'\beta_0$. 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.

3.2 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. Section 3.1 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 (2.6). 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 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 notation that follows 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

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

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 $(X_t'\beta_0, V)$. This estimator 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 (2.6), 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

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

Here 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, $\text{ASF}_t^{\pi}(\underline{x}) = \text{ASF}_t(\underline{x}_t)$ and the trimming does not alter the estimand. By expanding $\mathcal{Z}_t$ along with the sample size at a slow enough rate,[9] 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,[10] 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, when $\text{ASF}_t^{\pi}(\underline{x}_t) \geq 0$, we can show that $\text{ASF}_t^{\pi}(\underline{x}_t) \in [(1-\underline{\varepsilon})\text{ASF}_t(\underline{x}_t), \text{ASF}_t(\underline{x}_t)]$, and thus

$$\text{ASF}_t(\underline{x}_t) \in \left[\text{ASF}_t^{\pi}(\underline{x}_t), \frac{\text{ASF}_t^{\pi}(\underline{x}_t)}{1-\underline{\varepsilon}}\right]. \tag{3.1}$$

These bounds are reversed when $\text{ASF}_t^{\pi}(\underline{x}_t) < 0$. Note that bounds on the ASF that 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 $K : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is 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.

[9]This is sometimes called a vanishing, or random, trimming approach.

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

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

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

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 $\mathcal{B}_\varepsilon = \{\beta \in \mathcal{B} : \|\beta - \beta_0\| \le \varepsilon\}$.

(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) $\left\| f_{Z_t(\beta_0)}(z) \right\|_{\ell+2}^{\mathcal{Z}_t} < \infty$ 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.[11]

[11] See the proof of Lemma D.8.

The following rate conditions govern the bandwidth's convergence rate.

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 3.1 (ASF Asymptotics). Suppose the assumptions of Theorem 2.1 hold. Suppose Assumptions B1–B6 hold. Then,

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

where

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

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

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

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 $N^{\tilde{\delta}_{\text{ASF}}}$, where $\tilde{\delta}_{\text{ASF}}$ 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. We again 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 B.

3.3 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 3.1 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

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

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 $\text{APE}_{k,t}^{\pi}(\underline{x}_t) \equiv \mathbb{E}\left[\frac{\partial}{\partial \underline{x}^{(k)}} \mathbb{E}[Y_t | X_t'\beta_0 = \underline{x}_t'\beta_0, V] \cdot \pi_t\right]$. As for the ASF, the untrimmed APE is bounded by $\text{APE}_{k,t}(\underline{x}_t) \in \left[\text{APE}_{k,t}^{\pi}(\underline{x}_t), (1 - \underline{\varepsilon})^{-1} \text{APE}_{k,t}^{\pi}(\underline{x}_t)\right]$ 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 3.2 (APE Asymptotics)**.** Suppose the assumptions 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

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

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

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

and 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 $\sqrt{Nb_N^3}(\widehat{\beta}^{(k)} - \beta_0^{(k)}) = O_p(\sqrt{Nb_N^3}a_N) = o_p(1)$ by B6.

The rate of convergence of $\widehat{\mathrm{APE}}_{k,t}(\underline{x}_t)$ when $\epsilon = 1/2$ is $N^{\tilde{\delta}_{\mathrm{APE}}}$, where $\tilde{\delta}_{\mathrm{APE}}$ 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 for 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.

3.4 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:

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

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 proof is 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 $Z_t(\widehat{\beta})$.

3.5 Estimation with Estimated Indices

Now consider the estimation of these partial effects under the assumption that $C \perp\!\!\!\perp \mathbf{X} | V'\gamma_0$, where γ_0 is unknown. In this case, suppose $\theta_0 \equiv (\beta_0, \gamma_0) \in \mathbb{R}^{d_X + d_\gamma}$ are consistently estimated. For example, Ichimura and Lee (1991)'s estimator is $\sqrt{N}$-consistent for θ_0 under their regularity conditions. Let $Z_t(\theta) = (X_t'\beta, V'\gamma) \in \mathbb{R}^{1+\dim(V'\gamma)}$ 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 of lower dimension than V, which helps satisfy the rate assumption B6. This is because the function $\mathbb{E}[Y_t | X_t'\beta_0 = \underline{x}_t'\beta_0, V'\gamma_0 = v'\gamma_0]$ has lower dimension that $\mathbb{E}[Y_t | X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]$. This comes at the cost of an additional generated regressor of the form $V_i'\widehat{\gamma}$. From examining Lemmas D.1–D.7, we expect these additional generated regressors do not impact the estimators' limiting distributions, but we leave a detailed asymptotic analysis for future work.

4 Extension to a Dynamic Model

We now present an extension of our identification results to a dynamic panel model.[12] We use the dynamic binary model as a running example, but our Theorem 4.1 below holds for general dynamic nonlinear panels. Our assumption A1.(ii) rules out the dependence of U_t on future $X_{t'}$, 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 $X^t_{\text{pre}} = (X_{1,\text{pre}}, \ldots, X_{t,\text{pre}})$ denote all current and past values of the predetermined regressors. We will assume that errors are conditionally independent of past and future values of the exogenous regressors but may depend on future values of predetermined regressors.

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

This assumption replaces A1.(ii) and allows some of the future regressor values to depend on the current error term U_t. In particular, it allows for the inclusion of lagged dependent variables in $\mathbf{X}$. For example, let $\beta_0 = (\widetilde{\beta}_0, \rho_0)$ and assume that

$$Y_{it} = \mathbb{1}(\widetilde{X}'_{it}\widetilde{\beta}_0 + \rho_0 Y_{i\,t-1} + C_i - U_{it} \geq 0). \tag{4.1}$$

and that $U_t \perp\!\!\!\perp (\mathbf{X}_{\text{exog}}, X^t_{\text{pre}}, U_1, \ldots, U_{t-1})|C$. Here $X_{t,\text{exog}} = \widetilde{X}_t$ and $X_{t,\text{pre}} = Y_{t-1}$.

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". As shown in Honoré and Kyriazidou (2000), identification of β_0 can be achieved even when U_t does not follow a logistic distribution.

Given the identification of β_0, a modified index assumption can be made to identify the ASF/APE/AME.

Assumption A2† [Dynamic Index Sufficiency] For $t \in \{1, \ldots, T\}$, $C|(\mathbf{X}_{\text{exog}}, X^t_{\text{pre}}) \stackrel{d}{=} C|V^t$ where $V^t = v_t(\mathbf{X}_{\text{exog}}, X^t_{\text{pre}}) \in \mathbb{R}^{d_V}$ and v_t is known.

This assumption replaces A2 and allows the index to depend on all regressors except for future values

[12] See Arellano and Bonhomme (2017) for a review of nonlinear dynamic panel data models.

of the predetermined regressor. The following theorem shows the identification of our partial effects in these models.

Theorem 4.1 (Identification under Weak Exogeneity)**.** Let assumptions A1–A3 hold with A1†.(ii) replacing A1.(ii), and A2† replacing A2. 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. $\text{APE}_{k,t}(\underline{x}_t) = \mathbb{E}[\frac{\partial}{\partial \underline{x}_t^{(k)}}\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V^t]]$ 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. $\text{AME}_{k,t} = \mathbb{E}[\frac{\partial}{\partial X_t^{(k)}}\mathbb{E}[Y_t|X_t'\beta_0, V^t]]$ is point identified when $v_t(\underline{x}_{\text{exog}}, \underline{x}_{\text{pre}}^t) \in \text{supp}(V^t|X_t'\beta_0 = u)$ all u in a neighborhood of $\underline{x}_t'\beta_0$ holds for all $(\underline{x}_{\text{exog}}, \underline{x}_{\text{pre}}^t) \in \text{supp}(\mathbf{X}_{\text{exog}}, X_{\text{pre}}^t)$ up to a $P_{\mathbf{X}_{\text{exog}}, X_{\text{pre}}^t}$-measure zero set.

In the binary outcome model above, to identify the ASF at time $t = 1$, one can consider an index that depends on $(\widetilde{X}_1, \dots, \widetilde{X}_T, Y_0)$, where Y_0 is the initial time period outcome. Specificially, let $V^1 = (\widetilde{v}(\mathbf{X}_{\text{exog}}), Y_0)$. For simplicity, assume U_t is stationary and that β_0 is identified, perhaps from the identification results in Honoré and Kyriazidou (2000). Then,

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

The second line follows from $U_1 \perp\!\!\!\perp (\widetilde{X}_1, Y_0)|C$, the third from the law of iterated expectations, the fourth from Assumption A2†, and the last two 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 exchangeability of covariates across time is an unlikely justification in dynamic models. We leave an analysis of this task for future work.

5 Monte Carlo Simulations

We conduct two sets of Monte Carlo simulation experiments based on binary panel data models with the conditioning variable being V or $V'\gamma_0$. We focus on the former (Case 1) in this section and defer the latter (Case 2) to Appendix G.2, as their main messages are similar. Both cases account for two key features: multidimensional index variables and general error distributions.[13]

5.1 Alternative Estimators

We compare the proposed semiparametric estimator with two commonly used parametric alternatives: a random effects (RE) and a correlated random effects (CRE) estimator. See, for example, Wooldridge (2010). Both assume a standard logistic distribution for the error term U_t. They are characterized by different assumptions on the distribution of individual effects C. For the RE,

$$C \sim \mathcal{N}(\mu_c, \sigma_c^2)$$

and is independent of V. For the CRE,

$$C|V \sim \mathcal{N}(\mu_{c0} + \mu_{c1}'V, \sigma_c^2).$$

Then, the CRE is equivalent to an augmented RE with V being additional regressors. Following standard practice in the literature, we use the MLE to jointly estimate β_0 and the distribution parameters (μ_c, σ_c^2) or $(\mu_{c0}, \mu_{c1}, \sigma_c^2)$.

In the same spirit as the semiparametric estimator in Sections 3.2 and 3.3, we allow the marginal distribution of V to be unrestricted. The conditional expectation of the binary outcome and its derivative are calculated based on the MLE estimates, and the ASF and APE are obtained by averaging out V.

5.2 Case 1: Conditioning on V

The Monte Carlo design is summarized in Table 1. Note that both X_t and V are 2-by-1 vectors. Covariates $X_t^{(k)}$, $k = 1, 2$, are drawn from a 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 $N = 1461$

[13] For Monte Carlo simulations with logistic errors, please see the previous version of this paper (Liu, Poirier, and Shiu, 2021).

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.[14] For the error term, we consider error distributions f_{U_t} that exhibit skewness (DGP x1) or fat-tails (DGP x2).

We evaluate the estimated ASF and APE based on a collection of $\underline{x} = (\underline{x}^{(1)}, \underline{x}^{(2)})'$. We fix $\underline{x}^{(1)}$ at its population mean (i.e., $\underline{x}^{(1)} = 0$) and vary $\underline{x}^{(2)} \in [-1,\ 1]$, which covers 68% of the distribution of $X_t^{(2)}$. Given these 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), and adopt 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. Finally, we employ a local cubic regression (i.e., polynomial order $\ell = 3$) to estimate the conditional expectation of Y_t evaluated at $(\underline{x}'\widehat{\beta}, V)$.

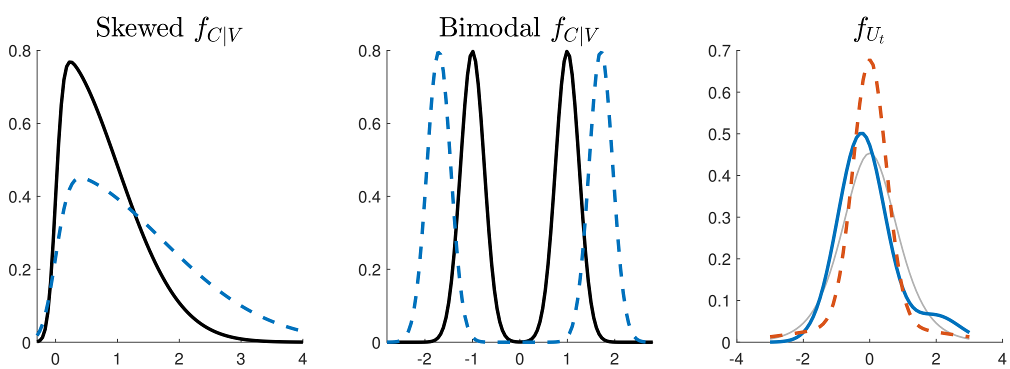
Figure 1 compares the estimated APE to the true APE based on 100 Monte Carlo repetitions in each setup, and Figure 2 plots the bias, standard deviation, and root mean square error (RMSE). Figures 5 and 6 in the appendix show corresponding graphs for the ASF estimates. We see that the semiparametric estimator 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 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 these 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 2, the first three columns summarize the APE estimator's performance by computing weighted average performance measures across the collection of evaluation points $\underline{x}$ with weights proportional to $f_{X_t}(\underline{x})$. Similar to what we observed in Figures 1 and 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($\underline{x}$) to the true APE($\underline{x}$). The minimum, median, and maximum are taken over the collection of evaluation points $\underline{x}$. 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, so the more precise semiparametric estimator is preferable. The RE and CRE

[14] Many empirical applications feature skewed and/or multimodal distributions of unobserved individual heterogeneity. For example, Liu (2021) estimated the latent productivity distribution of young firms, which exhibits a long right tail as 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.

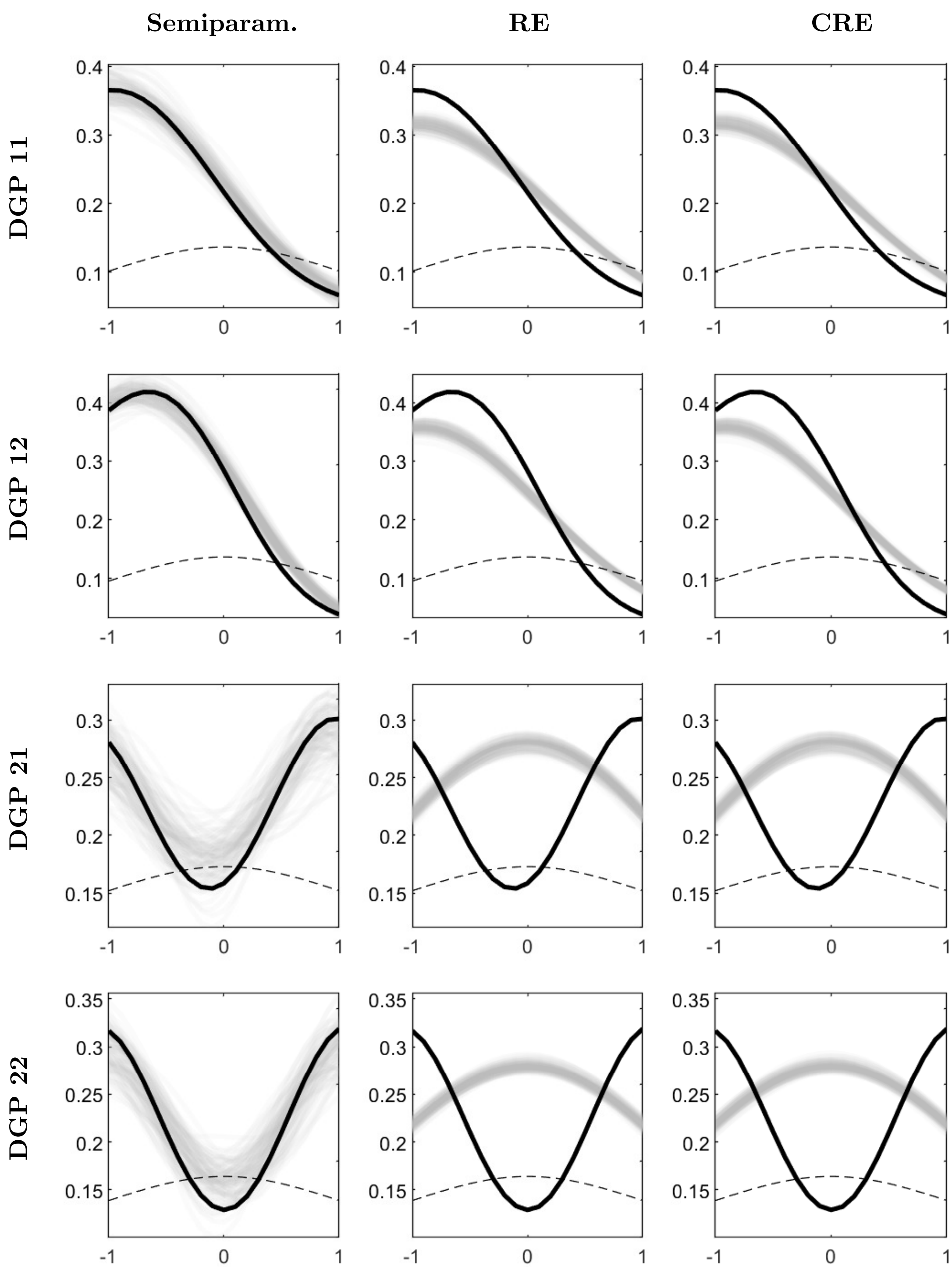
Table 1: Monte Carlo Design - Case 1

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 $\mathbb{E}\left(U_{it}\right) = 0$ and $\mathrm{Var}\left(U_{it}\right) = 1$:	
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:	$U_{it} \sim \frac{1}{5}\mathcal{N}\left(0, 4\right) + \frac{4}{5}\mathcal{N}\left(0, \frac{1}{4}\right)$



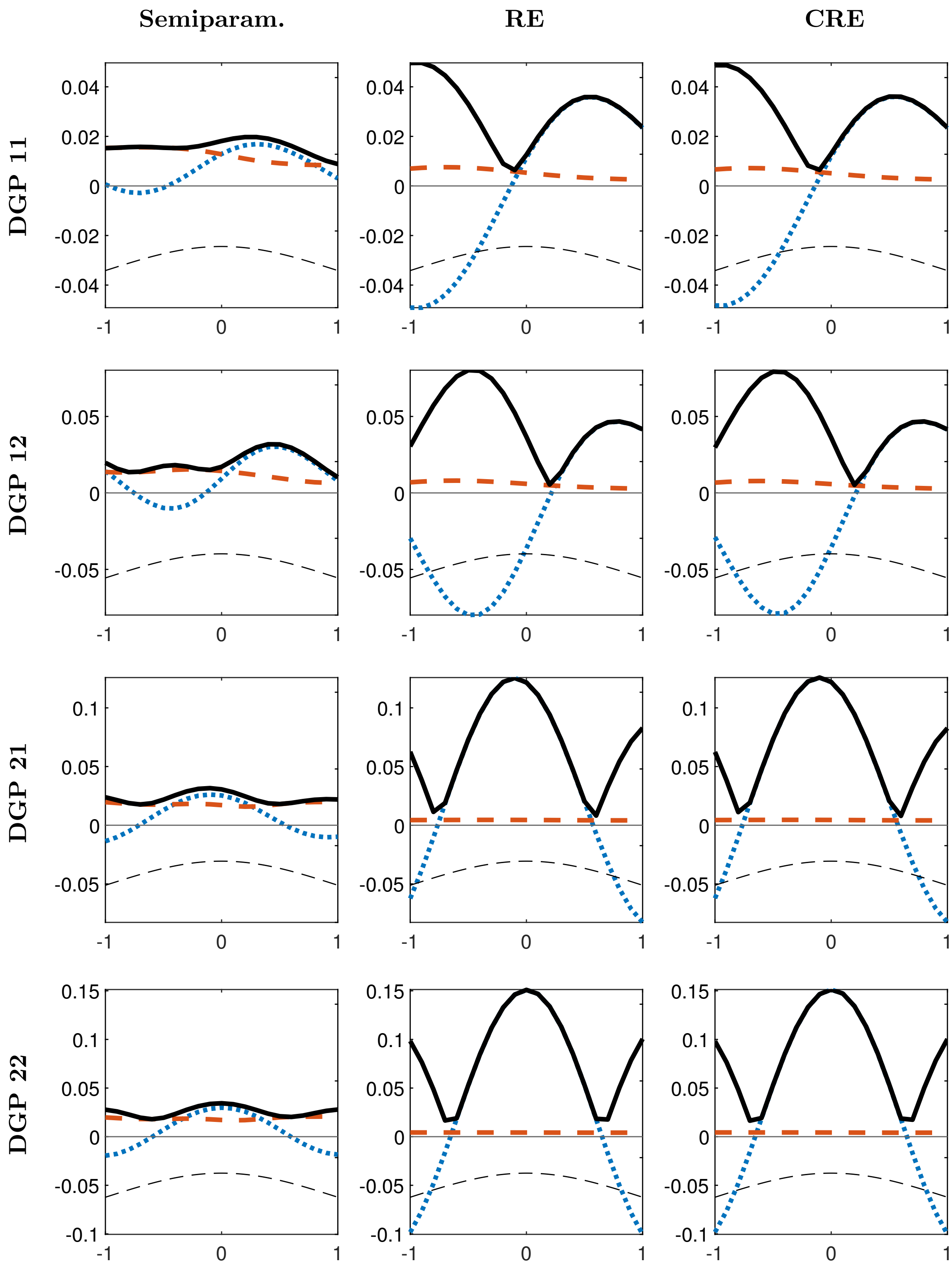
Notes: $\mathcal{SN}(\xi, \omega, \alpha)$ denotes a skewed normal distribution with location parameter ξ, scale parameter ω, and shape parameter α, and its pdf is given by $f(x) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of a standard normal distribution. The two left panels depict $f_{C|V}$. 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.

Figure 1: Estimated APE vs True APE - Monte Carlo Case 1



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

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



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

Table 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%

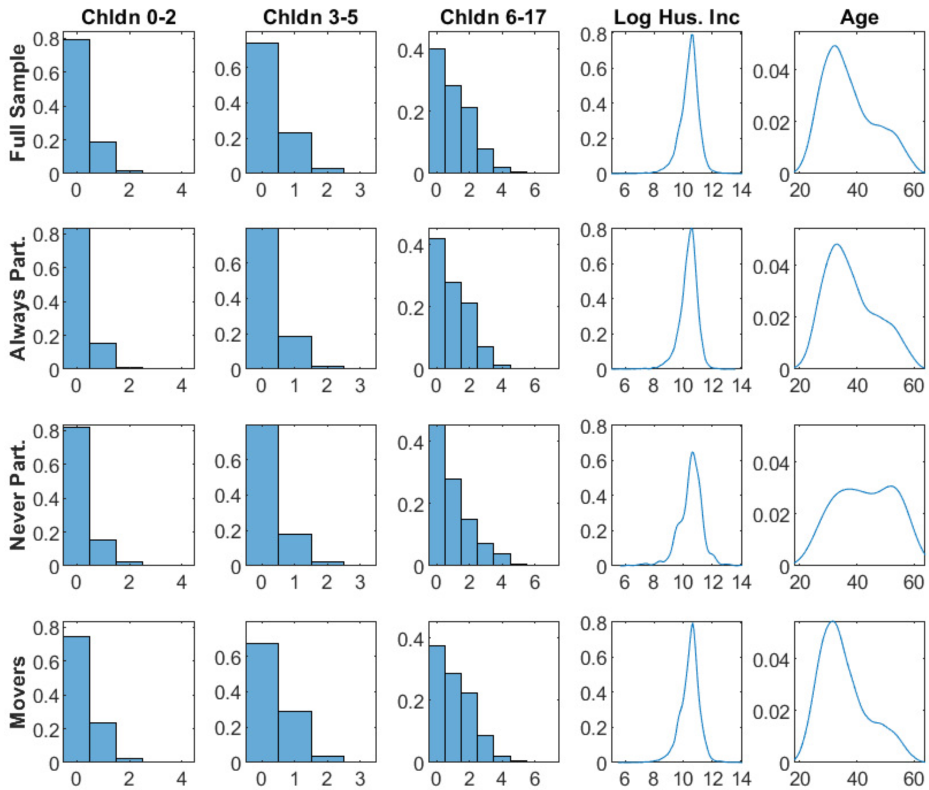
Notes: |Bias| indicates the absolute value of the bias. The reported |Bias|, SD, and RMSE are weighted averages across the collection of evaluation points $\underline{x}$, where the weights are proportional to $f_{X_t}(\underline{x})$. Bold entries indicate the best estimator (i.e., with the smallest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

have lower *minimal* ratios, which occurs at $\underline{x}$'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 performance for the common parameter and ASF in Table 4 in the appendix. The structure of the ASF part of the table is the same as Table 2 for the APE. The ratios of RMSE($\underline{x}$) to the true ASF($\underline{x}$) are generally smaller than their APE counterparts, and the semiparametric estimator 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 shape of the ASF and APE, and hence provides the most accurate ASF/APE estimates. Its RMSEs are around or less than half that of the RE and CRE.[15]

[15]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

Figure 3: Distributions 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.

6 Empirical Illustration

6.1 Background and Specification

In this empirical illustration, we examine women's participation in the labor market using our semiparametric approach. See the handbook chapter by Killingsworth and Heckman (1986) for an extensive review of the literature on female labor supply. For illustrative purposes, our analysis is based on the static setup of Fernández-Val (2009), where covariates X_t include numbers of children in three age categories, log husband's income, a quadratic function of age, as well as time dummies.[16]

the APE partially absorbs the effect of the slightly imprecisely estimated β_0.

[16] Charlier, Melenberg, and van Soest (1995) and Chen, Si, Zhang, and Zhou (2017), among others, also considered female labor force participation in their empirical applications. They used similar model specifications, but most of these papers focused on the estimation of common parameters β_0 instead of the ASF or APE.

The sample consists of $N = 1461$ married women observed for $T = 9$ years from the PSID between 1980–1988. We use the dataset kindly made available on Iván Fernández-Val's website. Figure 3 plots the distributions of the covariates, and Table 8 in the appendix summarizes the corresponding descriptive statistics. Roughly 45% of the women in the sample always participated in the labor market, less than 10% never participated, and around 45% changed their status during the sample period. Movers tended to be younger and have more children in all children's age categories. Never participants were relatively uniformly distributed between ages 30 to 50, whereas the women in other subgroups were generally younger. All subgroups exhibited heavy tails in log husband's income.

The unobserved individual effects C could be interpreted as an individual's willingness to work. In the benchmark specification, we construct indices V based on the initial values of the covariates X_{i1}. Women's ages and numbers of children are discrete variables, and we consider a cell-by-cell analysis.[17] These covariates generate over 1000 cells in this sample, thus some cells do not contain sufficient observations to use a semiparametric estimator within them. Therefore, we collapse the discrete index variables as follows. First, we sum over children's age categories and obtain the total number of children under 18 in the initial period. Then, the total number of children is collapsed into a trinary variable depending on whether it is below the 33rd quantile, between the 33rd and 67th quantiles, or above the 67th quantile, and the initial age is collapsed into a binary indicator depending on whether it is above or below the median. This coarsening scheme results in 6 cells, and the number of observations in each cell ranges from 156 to 314. Thus, we have three index variables: a trinary fertility variable, a binary age variable, and a continuously distributed average log husband's income. The number of continuous index variables is $d_V = 1$.

Various robustness checks regarding alternative choices of V_i (e.g., constructed from X_{i1} or $\overline{X}_i = \frac{1}{T}\sum_t X_{it}$), alternative estimators (e.g., multiple indices and local logit), and alternative coarsening schemes are explored in Appendix G.3 as well as the previous version of this paper (Liu, Poirier, and Shiu, 2021). The semiparametric estimator is generally robust with respect to these variations.

[17] For a more comprehensive empirical analysis, one could handle discrete index variables using a discrete kernel as suggested in Racine and Li (2004), which would be outside the scope of the current empirical illustration.

Table 3: Estimated β_0 - Female Labor Force Participation

	Smoothed Max. Score		RE		CRE	
	$\widehat{\beta}$	SD	$\widehat{\beta}$	SD	$\widehat{\beta}$	SD
Children 0–2	*-1*	*0*	*-1*	*0*	*-1*	*0*
Children 3–5	-0.83***	0.18	-0.60***	0.08	-0.60***	0.08
Children 6–17	-0.19	0.17	-0.19***	0.06	-0.17***	0.06
Log Husband's Income	-0.54**	0.25	-0.38***	0.08	-0.34***	0.10
Age/10	3.45*	1.88	2.34***	0.64	2.63***	0.68
$(\text{Age}/10)^2$	-0.51***	0.18	-0.35***	0.08	-0.37***	0.08

Notes: Standard deviations are calculated via the bootstrap. Significance levels are indicated by *: 10%, **: 5%, and ***: 1%. The first row follows from scale normalization $|\widehat{\beta}^{(1)}| = 1$, and we rescale the RE and CRE estimates to allow comparisons across estimators. $\widehat{\beta}^{(1)}$ is negative in all bootstrap samples for all three estimators so, after rescaling, their bootstrap standard deviations all equal to 0. Considering that the support of $\widehat{\beta}^{(1)}$ is ± 1, we do not put asterisks in the first row.

6.2 Results

Table 3 reports the estimated common coefficients on key covariates.[18] We see that women are more inclined to withdraw from the labor force when they have more children, especially younger ones, and when their husbands earn a higher income. Compared to the RE and CRE, the flexible smoothed maximum score estimator provides slightly larger (in magnitude) estimates with larger standard errors.

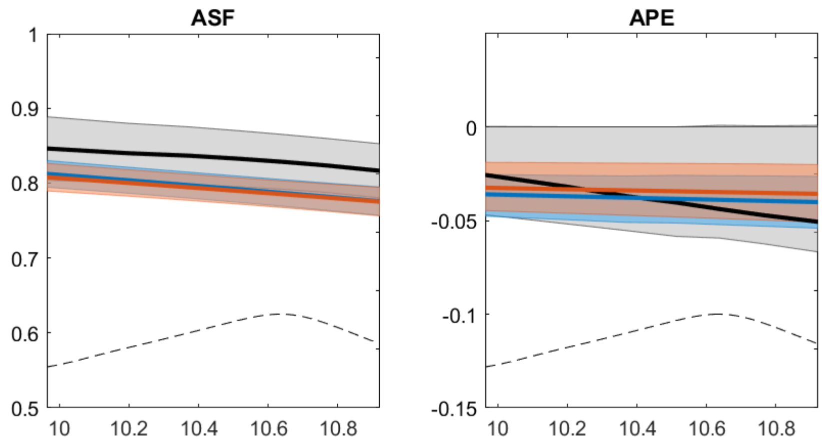
In our empirical example, we focus on the effects of the husband's income, which is linked to the wife's reservation wage. We select evaluation points $\underline{x}$ such that the log husband's income ranges from its 20th to 80th quantiles, and other variables are equal to their medians. These choices correspond to a hypothetical woman who is 35 years old, has 0 children between 0 and 2, 0 children between 3 and 5, 1 child between 6 and 17, and whose husband's income ranges from \$21K to \$55K. All time dummies are set to zero in this counterfactual.

Figure 4 shows estimates of the ASF and APE across $\underline{x}$ together with the 90% bootstrap confidence intervals based on 500 bootstrap samples.[19] For the ASF, all point estimates are downward

[18]Figure 11 in the appendix also plots the estimated coefficients on time dummies, which capture the time-variation in aggregate participation rates.

[19]For the ASF, all bootstrap estimates are between 0 and 1, and so is the symmetric percentile-t confidence band based on bootstrap standard deviations. For the APE, the smoothed maximum score in the first step requires monotonicity, i.e., $\frac{d}{du}\mathbb{P}(Y_t = 1|X_t'\beta_0 = u)|_{u=\underline{x}'\beta_0} \geq 0$. In the bootstrap, this constraint occasionally binds so we censor it at zero and employ the percentile bootstrap to account for the possible non-standard distribution due to censoring. Note that in principle, the bootstrap band for the APE could still contain positive values since the estimated coefficient for the log husband's income could be positive in some bootstrap samples. However, this incidence is rare in our

Figure 4: Estimated ASF and APE - Female Labor Force Participation



Notes: X-axes are potential values of log husband's income. Black/blue/orange solid lines represent point estimates of the ASF and APE using the semiparametric/RE/CRE estimators. Bands with corresponding colors indicate the 90% bootstrap confidence intervals. Thin dashed lines at the bottom of both panels show the distribution of log husband's income.

sloping with respect to the husband's income. The semiparametric estimator yields slightly higher participation probabilities compared to the RE and CRE.

For the APE, the semiparametric estimates are closer to zero for lower husband's incomes and more negative for higher ones, while their RE and CRE counterparts are rather flat. Note that for continuous $\underline{x}^{(k)}$,

$$\text{APE}_{k,t}(\underline{x}) = \beta_0^{(k)} \cdot f_{U_t - C}\left(\underline{x}'\beta_0\right) = \beta_0^{(k)} \cdot \int_{\mathcal{C}} f_{U_t}(\underline{x}'\beta_0 + c)\, dF_C(c),$$

where $f_{U_t - C}$ denotes the pdf of $U_t - C$, i.e., a convolution of $-C$ and U_t. Thus, the slope of the APE with respect to $\underline{x}^{(k)}$ reflects the shapes of f_C and f_{U_t} as well as the magnitude of $\beta_0^{(k)}$. In this sense, the flatter APE profile with respect to the husband's incomes in the RE and CRE could be due to the following three sources: (i) The RE and CRE feature a Gaussian $f_{C|V}$ and estimate the mean and variance of the Gaussian distribution. The estimated Gaussian variance could be fairly large to accommodate some non-Gaussian heterogeneity in $C|V$, and the resulting $\widehat{f}_{U_t - C}$ could be flatter (around the peak) than the true distribution. (ii) The RE and CRE assume a logistic f_{U_t}, which may deviate from the true data generating process. (iii) The smaller magnitudes of $\widehat{\beta}_0^{(k)}$ for RE and CRE could be due to misspecification of the distributions of U_t and C, and in turn further

empirical example.

lead to a milder slope of the APE profile.[20] In contrast, the semiparametric estimator does not require the parametrization of $f_{C|V}$ or f_{U_t}, thus reducing potential biases due to misspecification.

Moreover, when using our flexible semiparametric estimator which does not constrain the distributions of C or U_t, APEs with respect to the husband's income are no longer significant. Highly significant APEs estimated via RE and CRE could partly be an artifact of their parametric restrictions. This is consistent with the empirical observation that married women's labor supply choices became less sensitive to their husbands' income around 1980 when baby boomers started constituting a larger portion of the labor force, and both partners contribute to housework and earnings more equally. Hence fewer married women were at the margin of labor force participation that could be nudged by temporary fluctuations in husbands' income.

7 Conclusion

The distributions of the unobserved heterogeneity and the idiosyncratic errors play a crucial role in identifying partial effects in nonlinear panel models. In this paper, we first show the identification of the ASF, APE, and AME in a nonlinear semiparametric panel model with potentially unspecified distributions of the unobserved heterogeneity and of the idiosyncratic errors. To achieve this identification, we assume that units with the same value of the index V have correspondingly similar distributions of their unobserved heterogeneity C. We then develop three-step semiparametric estimators for the ASF and APE, and show their consistency and asymptotic normality. After conducting simulation experiments, we illustrate our semiparametric estimator in a study of determinants of women's labor supply.

In Section 4 we provided an identification result that applies to dynamic panel models. The generalization of our results to a broader class of dynamic models appears promising. We hope to pursue these ideas further in future work.

[20] The discrepancy in $\left|\widehat{\beta}_0^{(k)}\right|$ alone cannot explain all differences in the slopes of the APE profiles.

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.

Abrevaya, J., and Y.-C. Hsu (2021): "Partial effects in non-linear panel data models with correlated random effects," *The Econometrics Journal*, 24(3), 519–535.

Aguirregabiria, V., and J. M. Carro (2020): "Identification of Average Marginal Effects in Fixed Effects Dynamic Discrete Choice Models," *Working Paper*.

Altonji, J., and R. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73(4), 1053–1102.

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.

Arkhangelsky, D., and G. Imbens (2019): "The Role of the Propensity Score in Fixed Effect Models," *arXiv preprint arXiv:1807.02099v6*.

Bester, C., and C. Hansen (2009): "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Business & Economic Statistics*, 27(2), 235–250.

Blundell, R., and J. L. Powell (2003): *Endogeneity in Nonparametric and Semiparametric Regression Models*vol. 2 of *Econometric Society Monographs*, p. 312–357. Cambridge University Press.

Blundell, R. W., and J. L. Powell (2004): "Endogeneity in semiparametric binary response models," *The Review of Economic Studies*, 71(3), 655–679.

Bonhomme, S., T. Lamadon, and E. Manresa (2017): "Discretizing unobserved heterogeneity," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2019-16).

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

——— (2022): "Identification of time-varying counterfactual parameters in nonlinear panel models," *arXiv preprint arXiv.2212.09193*.

Botosaru, I., C. Muris, and K. Pendakur (2021): "Identification of time-varying transformation models with fixed effects, with an application to unobserved heterogeneity in resource shares," *Journal of Econometrics*.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2021): "Difference-in-Differences with a Continuous Treatment," *arXiv preprint arXiv.2107.02637*.

Chamberlain, G. (1980): “Analysis of Covariance with Qualitative Data,” *Review of Economic Studies*, 47, 225–238.

——— (1984): “Panel Data,” *Handbook of Econometrics*, 2, 1247–1318.

——— (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.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): “Average and Quantile Effects in Nonseparable Panel Models,” *Econometrica*, 81(2), 535–580.

Chernozhukov, V., I. Fernández-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015): “Nonparametric identification in panels using quantiles,” *Journal of Econometrics*, 188(2), 378–392, Heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao.

Chernozhukov, V., I. Fernández-Val, and Y. Luo (2018): “The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages,” *Econometrica*, 86(6), 1911–1938.

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

Davezies, L., X. D’Haultfoeuille, and L. Laage (2021): “Identification and Estimation of Average Marginal Effects in Fixed Effect Logit Models,” *arXiv preprint arXiv:2105.00879*.

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

Escanciano, J. C., D. Jacho-Chávez, and A. Lewbel (2016): “Identification and estimation of semiparametric two-step models,” *Quantitative Economics*, 7(2), 561–589.

Evdokimov, K. (2009): “Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity,” Mimeo.

Evdokimov, K., and H. White (2012): “Some Extensions of a Lemma of Kotlarski,” *Econometric Theory*, 28(4), 925–932.

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.

Graham, B. S., and J. L. Powell (2012): "Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models," *Econometrica*, 80(5), 2105–2152.

Hoderlein, S., and H. White (2012): "Nonparametric identification in nonseparable panel data models with generalized fixed effects," *Journal of Econometrics*, 168(2), 300–314.

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 (2020): "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.

Imbens, G. W., and W. K. Newey (2009): "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77(5), 1481–1512.

Khan, S., M. Ponomareva, and E. Tamer (2020): "Identification of dynamic binary response models," *W*.

Killingsworth, M. R., and J. J. Heckman (1986): "Female labor supply: A survey," *Handbook of labor economics*, 1, 103–204.

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*, pp. 1529–1564.

Kotlarski, I. (1967): "On characterizing the gamma and the normal distribution," *Pacific Journal of Mathematics*, 20(1), 69–76.

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

Laage, L. (2020): "A Correlated Random Coefficient Panel Model with Time-Varying Endogeneity," *arXiv preprint arXiv:2003.09367*.

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. (2021): "Density forecasts in panel data models: A semiparametric Bayesian perspective," *Journal of Business & Economic Statistics*, forthcoming.

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.

Maurer, J., R. Klein, and F. Vella (2011): "Subjective Health Assessments and Active Labor Market Participation of Older Men: Evidence From a Semiparametric Binary Choice Model With Nonadditive Correlated Individual-Specific Effects," *The Review of Economics and Statistics*, 93(3), 764–774.

Mundlak, Y. (1978): "On the pooling of time series and cross section data," *Econometrica*, 46(1), 69–85.

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

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67(3), 565–603.

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.

——— (1961): "On general laws and the meaning of measurement in psychology," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 4, pp. 321–333.

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.

Wooldridge, J. (2010): *Econometric Analysis of Cross Section and Panel Data.* MIT press.

Zhu, Y. (2022): "A sufficient and necessary condition for identification of binary choice models with fixed effects," *arXiv preprint arXiv.2206.10475.*

Appendix

This appendix is organized as follows. Appendix A contains additional identification results on the heterogeneity distribution. Appendix B discusses a number of implementation details for our estimators. Appendices C-D-E-F contain proofs for results in Sections 2-3-4 and Appendix A respectively. Finally, Appendix G contains additional figures and tables that supplement the results in the main text for the Monte Carlo simulation and empirical illustration.

A Identification of the Heterogeneity Distribution

In this appendix, we consider additional identification results in the binary outcome panel model $Y_{it} = \mathbb{1}(X_{it}'\beta_0 + C_i - U_{it} \geq 0)$.

In the logit case, we show that full support assumptions on $X_{it}'\beta_0$ lead to the identification of the marginal distribution of the unobserved heterogeneity. We also show that further support assumptions point identify F_C even when U_t is not specified. Therefore, we can recover additional functionals of this unconditional distribution. For example, the entire distribution of C or other nonlinear functionals of the conditional response probability $\mathbb{P}(Y_t = 1|\mathbf{X} = \mathbf{x}, C) = F_{U_t}(\underline{x}_t'\beta_0 + C)$ could be of interest, such as its variance or its quantiles: see Chernozhukov, Fernández-Val, and Luo (2018). The following proposition formalizes the logit result.

Proposition A.1 (Logit case)**.** Let A1–A3 hold and let $U_t|(C, \mathbf{X})$ follow a standard logistic distribution. Suppose $\text{supp}(X_t'\beta_0, V) = \mathbb{R} \times \text{supp}(V)$. Suppose the distribution of $(Y, \mathbf{X})$ is known. Then, $F_{C|V}$ and F_C are point identified.

Note that this proposition applies only to the logit case. The assumption that $U_t|(C, \mathbf{X})$ follows a logistic distribution is standard in this literature. This result relies on a deconvolution, which requires stronger support conditions than required by Theorem 2.1. In particular, the support restriction implies that $X_t'\beta_0$ has full support on the real line. This is the case when at least one regressor, say $X_t^{(k)}$, has full support conditional on $X_t^{(-k)}$, and when $\beta_0^{(k)} \neq 0$. Under this support assumption, it is possible to recover the conditional distribution of $U_t - C|\{V = v\}$ over its entire support for all $v \in \text{supp}(V)$. Given that the distribution of $U_t|V \stackrel{d}{=} U_t$ is known, a conditional deconvolution argument shows the point identification of the conditional distribution of $C|V$.

Since the distribution of V is identified from $F_{\mathbf{X}}$, the marginal distribution of C is also identified. This implies the identification of all functionals of $(F_C, F_{Y,\mathbf{X}}, \beta_0)$, such as quantiles of the conditional response probability. The estimation of F_C and its functionals is beyond the scope of this paper.

Going further, under stronger support restrictions, we can also show the identification of the distributions of U_t and C when neither are specified. This is described in the following proposition.

Proposition A.2 (General case)**.** Let A1–A3 hold. For $1 \leq s < t \leq T$, assume that $\text{supp}(X_s'\beta_0, X_t'\beta_0, V) = \mathbb{R}^2 \times \text{supp}(V)$. Assume the conditional characteristic functions of U_s, U_t, and C have no zeros and that (U_s, U_t, C) are mutually independent. Also assume that $(U_s, U_t) \perp\!\!\!\perp \mathbf{X}$. Suppose the distribution of $(Y, \mathbf{X})$ is known. Then, F_C, F_{U_s} and F_{U_t} are point identified.

This result uses Kotlarski's lemma (Kotlarski, 1967), which has also been used in panel data models by Evdokimov (2009). It requires stronger support assumptions on the joint distribution of indices $X_s'\beta_0$ and $X_t'\beta_0$ to ensure the nonparametric identification of the distribution of $(U_s - C, U_t - C)|V$. The conditions for applying Kotlarski's lemma, such as the characteristic function restrictions, can be relaxed following Evdokimov and White (2012).

B Implementation Details

B.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 (Section 5) and empirical illustration (Section 6).

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, which can be done via a Cholesky decomposition, is performed on V alone rather than all of $Z_t(\widehat{\beta})$.[21] This orthogonalization makes it sensible to use a product of one-dimensional kernels as our joint kernel, as is done in Assumption B3.

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 improve the accuracy of the nonparametric approximation but may cause overfitting, especially in small samples. In general, our estimates are 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 because there are discrete index variables dividing the observations into cells, resulting in fewer observations in each cell: see Section 6.1.

[21]This is for technical reasons that ensure that $\underline{x}_t'\widehat{\beta}$ and V enter in the kernel as a product since the latter is averaged out based on its empirical distribution: see the proofs in Appendix D, such as the proof of Lemma D.1.

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 assumption B3.(i) holds. The quintic polynomial for $5 < |u| \leq 6$ guarantees the twice continuous differentiability assumed in B3.(ii).

Bandwidth selection. In practice, one needs to select a bandwidth $b_N = \kappa \cdot \sigma_z \cdot N^{-\delta}$.[22] First, we choose δ that satisfies our rate conditions in Assumption B6. We choose σ_z using the standard deviation of the regressors. We then find a scaling constant κ using the bootstrap over a finite grid: see Appendix B.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 stable for scaling constants κ ranging in $[\kappa^* - 0.2,\, \kappa^* + 0.2]$.

Trimming set. 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 specify a threshold for each of the three criteria to construct the trimming set in our Monte Carlo simulations and empirical illustration.

Asymptotic variance estimation. To conduct inference on the ASF and APE, one could, in principle, estimate $\sigma_{\text{ASF}_t}(\underline{x}_t'\beta_0)$ and $\sigma_{\text{APE}_t}(\underline{x}_t'\beta_0)$ analytically. This can be done by estimating $\text{Var}(Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V_i)$ by via local polynomial regressions of (Y_t, Y_t^2) on $(X_t'\widehat{\beta}, V)$, $f_{Z_t}(\underline{x}_t'\beta_0, V_i)$ 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 $\frac{1}{N}\sum_{j=1}^N \mathcal{K}_{b_N}^V\left(\frac{V_j - V_i}{b_N}\right)$. For simplicity, we focus on bootstrap-based inference instead. Another benefit of the bootstrap is that it 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. Under a stationarity assumption, i.e., $(g_t, F_{U_t}) = (g_{t'}, F_{U_{t'}})$ for all $t, t' \in \{1, \ldots, T\}$, then $\text{ASF}_t(\underline{x}) = \text{ASF}_{t'}(\underline{x})$ for any pair of time periods assuming that $\underline{x} \in \text{supp}(X_t) \cap \text{supp}(X_{t'})$. For example, this assumption is made in binary outcome models under a logit assumption. When the ASF does not depend on t, we can combine ASF estimators from multiple time periods to obtain a more precise estimator. A straightforward

[22] In principle, we could adopt different κ's for $\underline{x}_t'\widehat{\beta}$ and V but we choose a single one for simplicity.

combination consists of averaging the estimated ASFs over time:

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

We can reduce the asymptotic variance by selecting weights that depend on t. Weights that minimize the asymptotic variance of the weighted ASF depend on the inverse of an estimate of the joint asymptotic covariance matrix of all T ASF estimators. For simplicity, we propose the simple time average as our rule of thumb.

B.2 Bandwidth Selection via Bootstrap

To select the bandwidth b_N, we want to minimize the integrated mean squared error

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

where $\widehat{\mathrm{APE}}_{k,t}(\underline{x}_t;\kappa N^{-\delta})$ is our estimated APE with bandwidth $b = \kappa N^{-\delta}$ for a fixed $\delta > 0$ satisfying our rate conditions.[23] $\mathrm{APE}_{k,t}(\underline{x}_t, b)$ denotes the probability limit of $\widehat{\mathrm{APE}}_{k,t}(\underline{x}_t;b)$ for a fixed bandwidth b. Note that $\mathrm{APE}_{k,t}(\underline{x}_t;0)$ is the true APE.

Since the IMSE depends on the unknown distribution of the data, we can approximate it via

$$\widehat{\mathrm{IMSE}}(\kappa) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{S}\sum_{s=1}^{S}\left(\widehat{\mathrm{APE}}^{*(s)}_{k,t}(X_{it};\kappa N^{-\delta}) - \widehat{\mathrm{APE}}_{k,t}(X_{it};\kappa_0 N^{-\delta})\right)^2\right).$$

Here $\{\widehat{\mathrm{APE}}^{*(s)}_{k,t}(X_{it};\kappa N^{-\delta})\}_{s=1}^{S}$ denote S draws of the estimated APE according to its bootstrap distribution using bandwidth $b = \kappa N^{-\delta}$. We let κ_0 be a constant that is 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 can choose the bandwidth constant κ through the following procedure.

Implementation procedure.

1. Generate a range of evaluation points $\underline{x}_j$, $j = 1, \ldots, J$, with weights $w(\underline{x}_j)$ determined from the empirical distribution of X_t.

2. Choose κ_0 to be a small value and estimate $\widehat{\mathrm{APE}}_{k,t}(\underline{x}_j;\kappa_0 N^{-\delta})$, $j = 1, \ldots, J$, 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 $s = 1, \ldots, S$.

[23] In this description, we let the regressors be normalized so that the bandwidth does not depend on their (unnormalized) standard deviations.

4. For each bootstrap sample $s = 1, \ldots, S$ and each bandwidth κ on grid $\{\kappa^{(1)}, \ldots, \kappa^{(K)}\}$, calculate $\widehat{\text{APE}}^{*(s)}_{k,t}(\underline{x}_j; \kappa N^{-\delta})$ for $j = 1, \ldots, J$.

5. Choose $\kappa \in \{\kappa^{(1)}, \ldots, \kappa^{(K)}\}$ that minimizes

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

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.

B.3 Estimated Indices

When the conditioning variable(s) take the form $V'\gamma_0$, we can implement the following three semi-parametric estimators:

1. SP: the original three-step estimator.
   (a) First, estimate β_0 based on the smoothed maximum score.
   (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 based on the smoothed maximum score.
   (b) Second, plug in $\widehat{\beta}$ into the objective function in Ichimura and Lee (1991) to estimate γ_0 using $\widehat{\gamma}$.
   (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 and SP ($V'\gamma_0$, iter.) use the smoothed maximum score in the first step and thus require monotonicity of g_t in $X_{it}'\beta_0$, which is a commonly assumed though not completely innocuous. (ii) SP ($V'\gamma_0$) and SP ($V'\gamma_0$, iter.) assume the multiple index structure, which is more efficient when the assumption holds but less robust to misspecification. (iii) 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 problems with higher dimensions of parameters.

C Proofs for Section 2

Proof of Theorem 2.1. To show the ASF is identified, note that

$$\begin{aligned}\text{ASF}_t(\underline{x}_t) &= \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)] \\ &= \mathbb{E}[\mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|V]] \\ &= \mathbb{E}[\mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|X_t'\beta_0 = \underline{x}_t'\beta_0, V]] \\ &= \int_{\text{supp}(V)} \mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]\, dF_V(v).\end{aligned}$$

The first equality is by definition. The second follows from the iterated expectations.

The third equality holds if (a) $U_t \perp\!\!\!\perp X_t'\beta_0|(C, V)$ and (b) $C \perp\!\!\!\perp X_t'\beta_0|V$. To show (a) holds, note that $U_t|(X_t'\beta_0, V, C) \stackrel{d}{=} U_t|C \stackrel{d}{=} U_t|(V, C)$ from $U_t \perp\!\!\!\perp X|C$, (Assumption A1.(ii)), and from $(X_t'\beta_0, V)$ being functions of $\mathbf{X}$. To show (b) holds, note that $C \perp\!\!\!\perp \mathbf{X}|V$ by A2, and from $X_t'\beta_0$ being a function of $\mathbf{X}$. Therefore the third equality follows. The fourth equality follows immediately.

The expression in the fourth equality depends on $\{\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v] : v \in \text{supp}(V)\}$, and F_V, the marginal distribution of V. By A3, β_0 is identified up to scale, thus the conditioning set $\{X_t'\beta_0 = \underline{x}_t'\beta_0\}$ is identified since it is invariant to the scale of β_0.

By definition, $\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v]$ is identified for $v \in \text{supp}(V|X_t'\beta_0 = \underline{x}_t'\beta_0)$. By the support assumption, $\text{supp}(V|X_t'\beta_0 = \underline{x}_t'\beta_0) = \text{supp}(V)$, so $\{\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0, V = v] : v \in \mathit{supp}(V)\}$ is identified. The marginal distribution of V is identified since it is a known function of observed $\mathbf{X}$. Hence, $\text{ASF}_t(\underline{x}_t)$ is identified from the distribution of $(Y, \mathbf{X})$ for all $t \in \{1, \ldots, T\}$.

For the APE, by Assumption A1.(iii) write

$$\begin{aligned}\text{APE}_{k,t}(\underline{x}_t) &= \frac{\partial}{\partial \underline{x}_t^{(k)}} \text{ASF}_t(\underline{x}_t) \\ &= \lim_{\tilde{u} \to 0} \frac{\text{ASF}_t(\underline{x}_t + \tilde{u}e_k) - \text{ASF}_t(\underline{x}_t)}{\tilde{u}}.\end{aligned}$$

Thus, $\text{APE}_{k,t}(\underline{x}_t)$ is identified if $\text{ASF}_t(\underline{x}_t + \tilde{u}e_k)$ is identified for all $\tilde{u}$ in a neighborhood of 0. By the above result, this is the case if $\text{supp}(V|X_t'\beta_0 = (\underline{x}_t + \tilde{u}e_k)'\beta_0) = \text{supp}(V)$ for all $\tilde{u}$ in a neighborhood of 0. Note that $\text{supp}(V|X_t'\beta_0 = (\underline{x}_t + \tilde{u}e_k)'\beta_0) = \text{supp}(V|X_t'\beta_0 = \underline{x}_t'\beta_0 + \tilde{u}\beta_0^{(k)})$, and that $\underline{x}_t'\beta_0 + \tilde{u}\beta_0^{(k)}$ lies in an arbitrary neighborhood of $\underline{x}_t'\beta_0$ as $\tilde{u} \to 0$. Therefore $\text{APE}_{k,t}(\underline{x}_t)$ is identified. □

Proof of Theorem 2.2. We can write the LAR as

$$
\begin{aligned}
\mathrm{LAR}_{k,t}(\underline{x}) &= \left.\frac{\partial \mathbb{E}[Y_t(\underline{x}_t)|\mathbf{X}=\tilde{x}]}{\partial \underline{x}_t^{(k)}}\right|_{\tilde{x}=\underline{x}} \\
&= \left.\frac{\partial \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|\mathbf{X}=\tilde{x}]}{\partial \underline{x}_t^{(k)}}\right|_{\tilde{x}=\underline{x}} \\
&= \left.\frac{\partial \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|v(\mathbf{X})=v(\tilde{x})]}{\partial \underline{x}_t^{(k)}}\right|_{\tilde{x}=\underline{x}} \\
&= \left.\frac{\partial \mathbb{E}[g_t(\underline{x}_t'\beta_0, C, U_t)|X_t'\beta_0=\underline{x}_t'\beta_0, V=\tilde{v}]}{\partial \underline{x}_t^{(k)}}\right|_{\tilde{v}=v(\underline{x})} \\
&= \left.\frac{\partial \mathbb{E}[Y_t|X_t'\beta_0=\underline{x}_t'\beta_0, V=\tilde{v}]}{\partial \underline{x}_t^{(k)}}\right|_{\tilde{v}=v(\underline{x})}.
\end{aligned}
$$

The first two equalities are by definition. The third follows from A1 and A2. The fourth follows from $(C, U_t) \perp\!\!\!\perp X_t'\beta_0|V$, which was established in the proof of Theorem 2.1 under the same assumptions.

By A3, β_0 is identified up to scale, thus the conditioning set $\{X_t'\beta_0 = \underline{x}_t'\beta_0\}$ is identified. Note that

$$
\begin{aligned}
&\frac{\partial}{\partial \underline{x}_t^{(k)}}\mathbb{E}[Y_t|X_t'\beta_0=\underline{x}_t'\beta_0, V=\tilde{v}]|_{\tilde{v}=v(\underline{x})} \\
&= \lim_{\tilde{u}\to 0}\frac{\mathbb{E}[Y_t|X_t'\beta_0=\underline{x}_t'\beta_0+\tilde{u}\beta_0^{(k)}, V=v(\underline{x})] - \mathbb{E}[Y_t|X_t'\beta_0=\underline{x}_t'\beta_0, V=v(\underline{x})]}{\tilde{u}}.
\end{aligned}
$$

The conditional expectation $\mathbb{E}[Y_t|X_t'\beta_0 = \underline{x}_t'\beta_0 + \tilde{u}\beta_0^{(k)}, V = v(\underline{x})]$ is identified for sufficiently small $\tilde{u}$ if $v(\underline{x}) \in \mathrm{supp}(V|X_t'\beta_0 = \underline{x}_t'\beta_0 + \tilde{u}\beta_0^{(k)})$, which is implied by $v(\underline{x}) \in \mathrm{supp}(V|X_t'\beta_0 = u)$ for all u in a neighborhood of $\underline{x}_t'\beta_0$. By the theorem assumptions, this conditional expectation is identified and therefore $\mathrm{LAR}_{k,t}(\underline{x})$ is identified.

The proof of the identification of the AME follows from the identification of the LAR for all $\underline{x}$ in $\mathrm{supp}(\mathbf{X})$ (up to a $P_{\mathbf{X}}$-measure zero set) and from

$$
\mathrm{AME}_{k,t} = \int_{\mathrm{supp}(\mathbf{X})} \mathrm{LAR}_{k,t}(\underline{x})\, dF_{\mathbf{X}}(\underline{x}).
$$

□

Proof of Proposition (2.2). By Assumption A1, we have that

$$
\begin{aligned}
\mathbb{E}[Y_t|\mathbf{X}] &= \int_{\text{supp}(C,U_t|\mathbf{X})} g_t(X_t'\beta_0, c, u_t)\, dF_{C,U_t|\mathbf{X}}(c, u_t|\mathbf{X}) \\
&= \int_{\text{supp}(C,U_t|\mathbf{X})} g_t(X_t'\beta_0, c, u_t)\, dF_{C,U_t|X_t'\beta_0, V'\gamma_0}(c, u_t|X_t'\beta_0, v(\mathbf{X})'\gamma_0) \\
&= \Psi_t(X_t'\beta_0, v(\mathbf{X})'\gamma_0)
\end{aligned}
$$

where the second equality follows from $C \perp\!\!\!\perp \mathbf{X}|v(\mathbf{X})'\gamma_0$ (Assumption A4) and $U \perp\!\!\!\perp \mathbf{X}|C$ (Assumption A1.(ii)). The result then follows immediately from Lemma 3 in Ichimura and Lee (1991) with no additive index in the outcome equation and when applied to continuous regressors only. □

Proof of Proposition 2.1. Without loss of generality, let $s = 1$ and $t = 2$. We start by computing

$$
\begin{aligned}
&\frac{\partial}{\partial x_1}\mathbb{E}[Y_1|\mathbf{X} = \mathbf{x}] = \frac{\partial}{\partial x_1}\Psi_1(x_1'\beta_0, v(\mathbf{x})) = \frac{\partial}{\partial u}\Psi_1(u, v(\mathbf{x}))\Big|_{u=x_1'\beta_0}\begin{pmatrix}1\\ \widetilde{\beta}_0\end{pmatrix} + W_1(\mathbf{x})\frac{\partial}{\partial v'}\Psi_1(x_1'\beta_0, v)|_{v=v(\mathbf{x})} \\
&\frac{\partial}{\partial x_2}\mathbb{E}[Y_1|\mathbf{X} = \mathbf{x}] = \frac{\partial}{\partial x_2}\Psi_1(x_1'\beta_0, v(\mathbf{x})) = W_2(\mathbf{x})\frac{\partial}{\partial v'}\Psi_1(x_1'\beta_0, v)|_{v=v(\mathbf{x})}.
\end{aligned}
$$

where

$$
W_t(\mathbf{x}) = \frac{\partial}{\partial x_t}v(\mathbf{x})' \in \mathbb{R}^{d_X \times d_V}.
$$

From the second equation above, we can recover

$$
\frac{\partial}{\partial v'}\Psi_1(x_1'\beta_0, v)|_{v=v(\mathbf{x})} = \left(W_2(\mathbf{x})'W_2(\mathbf{x})\right)^{-1} W_2(\mathbf{x})'\frac{\partial}{\partial x_2}\mathbb{E}[Y_1|\mathbf{X} = \mathbf{x}]
$$

since $W_2(\mathbf{x})$ has rank d_V.

Therefore, we can identify

$$
\frac{\partial}{\partial u}\Psi_1(u, v(\mathbf{x}))\Big|_{u=x_1'\beta_0}\begin{pmatrix}1\\ \widetilde{\beta}_0\end{pmatrix} = \frac{\partial}{\partial x_1}\mathbb{E}[Y_1|\mathbf{X} = \mathbf{x}] - W_1(\mathbf{x})\left(W_2(\mathbf{x})'W_2(\mathbf{x})\right)^{-1} W_2(\mathbf{x})'\frac{\partial}{\partial x_2}\mathbb{E}[Y_1|\mathbf{X} = \mathbf{x}]. \tag{C.1}
$$

By the assumption that $\frac{\partial \Psi_1(u,v(\mathbf{x}))}{\partial u} \neq 0$, $\widetilde{\beta}_0$ is point identified as the ratio of the elements in equation (C.1). □

D Proofs for Section 3

We now present a sequence of lemmas that are used to prove our two main theorems of Section 3: Theorem 3.1 and Theorem 3.2. When applied to matrices, let $\|\cdot\|$ denote the spectral norm.

Lemma D.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 D.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

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

where $\bar{b} > 0$. To obtain the stochastic order of term (D.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,

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

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

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

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

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

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

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

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

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 (D.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 (D.1) and (D.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 D.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 D.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 D.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 D.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 D.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 D.1, and by $E[Y_{jt}^2] < \infty$, 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 D.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 D.4. By the triangle inequality,

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

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

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

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,

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

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 D.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 D.5. As in the proof of Lemma D.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

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

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 (D.3) is of order $O_p\left(\frac{1}{\sqrt{Nb_N^{4+2d_V}}}\right)$.

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

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

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 D.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 D.6. As in the proof of Lemma D.3, consider a generic component the vector of $T_N(z;\beta_0)$. Write this element as

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

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 Lemma D.3 and D.5. □

Lemma D.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,

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

as $N\to\infty$.

Proof of Lemma D.7. We note that

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

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,

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

as $N\to\infty$. □

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

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

Proof of Lemma D.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 A5, $\mathbb{E}[|\psi(\varepsilon_i)|^{\nu_1}] < \infty$ holds for arbitrary large ν_1. Their A2 holds immediately. A3 holds by Assumption B3. A4 and A5 holds by B4.(iii). 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 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

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

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

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

To analyze term (D.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

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

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

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 D.2. Therefore,

$$
\begin{aligned}
&= \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}).
\end{aligned}
$$

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

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

by B6.

Third, term (D.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 3.1.

Lemma D.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 D.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 3.1. We have the following decomposition:

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

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

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

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

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

Part 1: Convergence of Term (D.7)

We have that

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

$$= \sqrt{Nb_N}\cdot\left|\frac{1}{N}\sum_{j=1}^{N}e_1'\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\hat{\pi}_{it}\right|$$

$$= \sqrt{Nb_N}\cdot\left|\frac{1}{N}\sum_{j=1}^{N}e_1'\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})^{-1}(T_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta}) - T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0))\right.\right.$$

$$\left.\left.+S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})^{-1}\left(S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0) - S_N(\underline{x}_t'\widehat{\beta}, V_i; \widehat{\beta})\right)S_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)^{-1}T_N(\underline{x}_t'\widehat{\beta}, V_i; \beta_0)\right)\mathbb{1}((\underline{x}_t'\widehat{\beta}, V_i) \in \mathcal{Z}_t)\right|$$

$$\leq \sqrt{Nb_N}\cdot\|e_1\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})^{-1}\right\|\sup_{z\in\mathcal{Z}_t}\left\|T_N(z;\widehat{\beta}) - T_N(z;\beta_0)\right\|$$

$$+ \sqrt{Nb_N}\cdot\|e_1\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta})^{-1}\right\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\widehat{\beta}) - S_N(z;\beta_0)\right\|\sup_{z\in\mathcal{Z}_t}\left\|S_N(z;\beta_0)^{-1}\right\|\sup_{z\in\mathcal{Z}_t}\|T_N(z;\beta_0)\|.$$

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

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

The first line follows from Lemma D.9. The second line follows from the triangle inequality. The third line follows from Lemma D.1 and D.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,

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

- By Lemma D.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 D.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

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

where

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

by Lemma D.4. We also have that

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

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

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

by B6.

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

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

Part 2: Convergence of Term (D.8)

We have that

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

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

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

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

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

Part 3: Convergence of Term (D.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 D.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 (D.10)

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

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

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

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

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

Proof of Lemma D.10. This proof builds on that of Corollary 2 in KLX and our Lemma D.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

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

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

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

To analyze term (D.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 D.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 (D.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 3.2. First, we write

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

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

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

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

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

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

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

Convergence of Term (D.14)

Note that

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

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

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 3.1 to show term (D.7) is $o_p(1)$.

Convergence of Term (D.15)

We have that

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

$$= \left|\widehat{\beta}^{(k)}\right| \cdot \sqrt{Nb_N}$$

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

Again, we can follow the same steps used in the proof of Theorem 3.1 to show term (D.8) is

$o_p(1)$.

Convergence of Term (D.16)

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

Convergence of Term (D.17)

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

Convergence of Term (D.18)

Note that $\mathbb{E}[h_2(\underline{x}_t'\beta_0, V; \beta_0)\pi_t] = O(1)$. Term (D.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 (D.14)–(D.18) with Slutsky's Theorem, we obtain our result. □

E Proofs for Section 4

Proof of Theorem 4.1. For the ASF, note that

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

The second equality follows from the law of iterated expectations. The third follows from $X_t'\beta_0 \perp\!\!\!\perp (C, U_t)|V^t$. To see this, note that $U_t \perp\!\!\!\perp X_t'\beta_0|C, V^t$ is implied by $U_t \perp\!\!\!\perp (\mathbf{X}_{\text{exog}}, X^t_{\text{pre}})|C$ and from $X_t'\beta_0$ being a function of $(\mathbf{X}_{\text{exog}}, X^t_{\text{pre}})$. Also note that $X_t'\beta_0 \perp\!\!\!\perp C|V^t$ follows from $(\mathbf{X}_{\text{exog}}, X^t_{\text{pre}}) \perp\!\!\!\perp C|V^t$ and from $X_t'\beta_0$ being a function of $(\mathbf{X}_{\text{exog}}, X^t_{\text{pre}})$. The last line follows directly. The last expression is identified from similar arguments as before.

Proofs for the identification of the APE and AME proceed similarly. □

F Proofs for Appendix Section A

Proof of Proposition A.1. Under $\text{supp}(X_t'\beta_0, V) = \mathbb{R} \times \text{supp}(V)$, the conditional probability $\mathbb{P}(Y_t = 1|X_t'\beta_0 = u, V = v)$ is identified for all $(u, v) \in \mathbb{R} \times \text{supp}(V)$. Assumptions A1, A2, and basic manipulations show that this conditional probability equals $F_{U_t - C|V}(u|v)$. Therefore, the conditional distribution of $U_t - C$ conditional on V is point identified. By $U_t \perp\!\!\!\perp (C, \mathbf{X})$, U_t and C are independent given V, and

$$\mathbb{E}[\exp(i\zeta(U_t - C))|V = v] = \mathbb{E}[\exp(i\zeta U_t)|V = v] \cdot \mathbb{E}[\exp(-i\zeta C)|V = v]$$

for any $\zeta \in \mathbb{R}$, where $i = \sqrt{-1}$.

By $U_t \perp\!\!\!\perp (C, \mathbf{X})$, we have that $U_t|V \overset{d}{=} U_t$. The distribution of U_t is known (standard logistic) and has a characteristic function with no zeros, thus can then write

$$\mathbb{E}[\exp(-i\zeta C)|V = v] = \frac{\mathbb{E}[\exp(i\zeta(U_t - C))|V = v]}{\mathbb{E}[\exp(i\zeta U_t)]},$$

where the right-hand side is identified from the data.

From the inversion formula for characteristic functions, this implies the conditional distribution of $C|V = v$ is identified for all $v \in \text{supp}(V)$. The distribution of C is also identified by the law of total probability. □

Proof of Proposition A.2. Under $\text{supp}(X_s'\beta_0, X_t'\beta_0, V) = \mathbb{R}^2 \times \text{supp}(V)$, the conditional probability $\mathbb{P}(Y_s = 1, Y_t = 1|X_s'\beta_0 = u_1, X_t'\beta_0 = u_2, V = v)$ is identified for all $(u_1, u_2, v) \in \mathbb{R}^2 \times \text{supp}(V)$. Assumption A2 and basic manipulations show that this conditional probability equals $F_{U_s - C, U_t - C|V}(u_1, u_2|v)$. Therefore, the conditional distribution of $(U_s - C, U_t - C)$ conditional on V is identified. This implies the joint distribution of $(U_s - C, U_t - C)$ is identified by the law of total probability.

Note that (U_s, U_t, C) are assumed jointly independent. Therefore, we can apply Kotlarski's lemma (Kotlarski, 1967) to obtain the distributions of U_s, U_t, and C. □

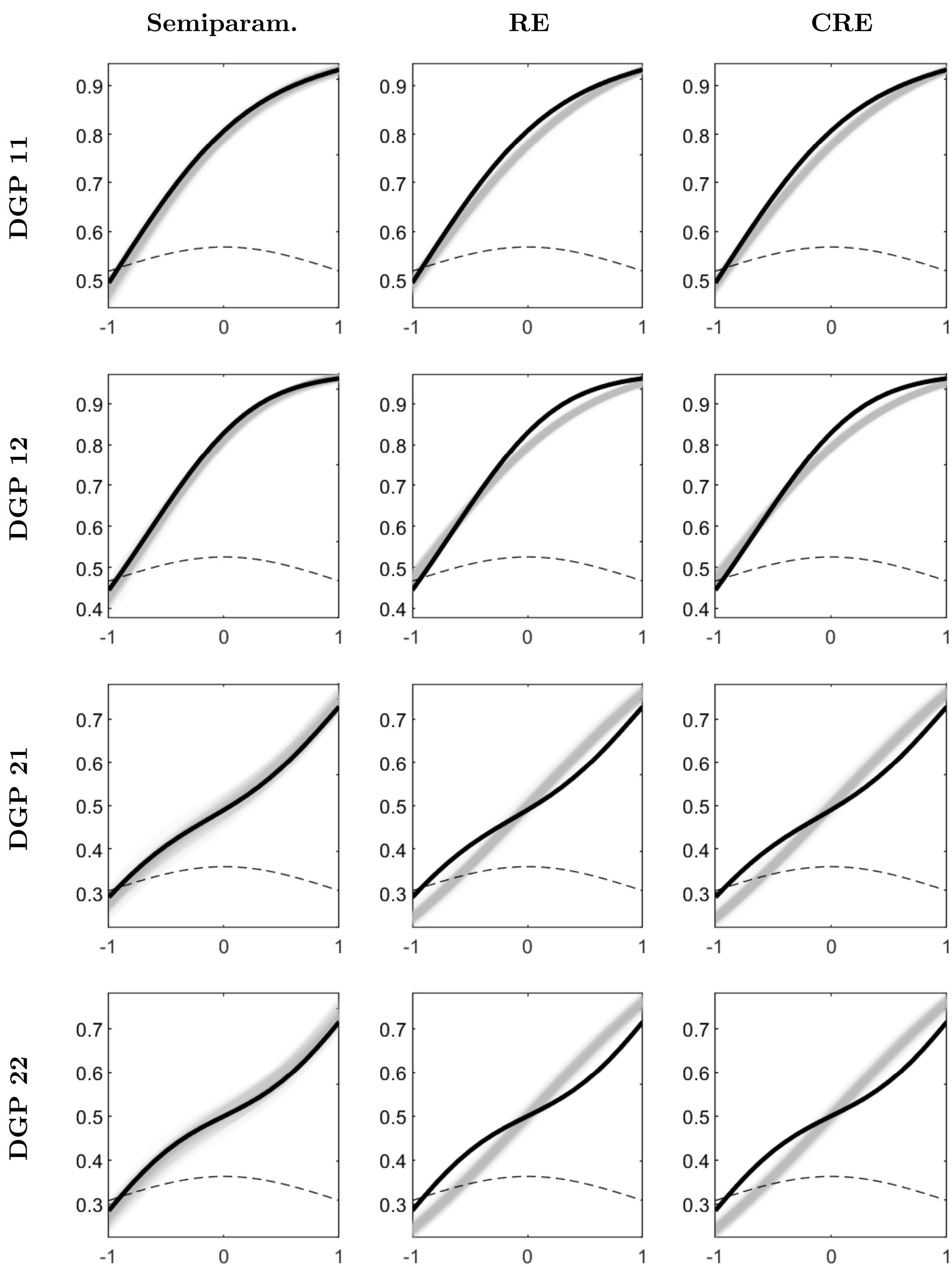
G Additional Figures and Tables

This section presents additional figures and tables that supplement the main results in the text.

G.1 Monte Carlo Simulation Case 1: Conditioning on V

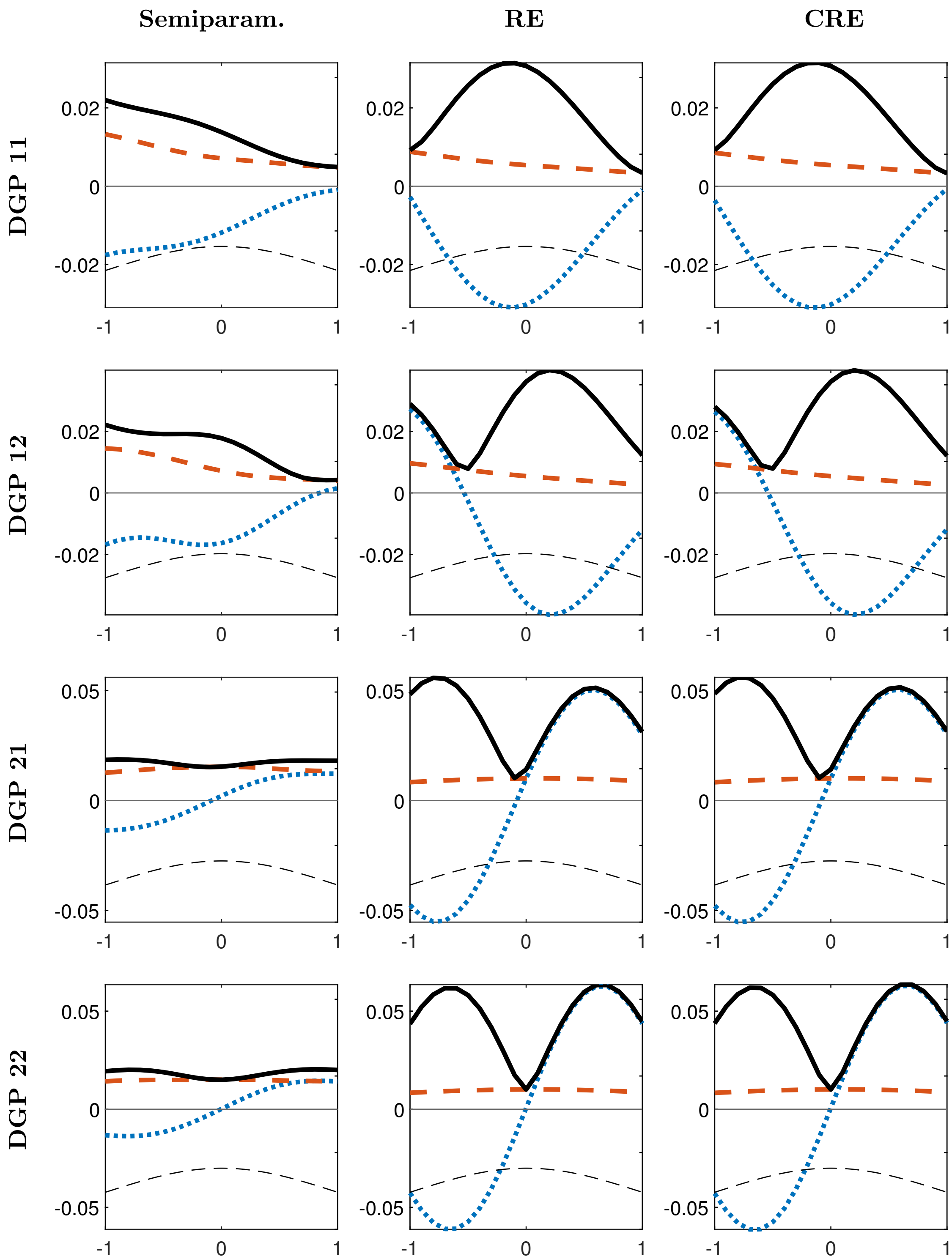
Figures 5 and 6 show the estimated ASFs and their corresponding performance statistics. Table 4 summarizes the estimation performance for the common parameter and ASF.

Figure 5: Estimated ASF vs True ASF - Monte Carlo Case 1



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

Figure 6: Bias, Standard Deviation, and RMSE in ASF Estimation - Monte Carlo Case 1



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

Table 4: 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 allow comparisons across estimators. |Bias| indicates the absolute value of the bias. The |Bias|, SD, and RMSE of the ASF are weighted averages across the collection of evaluation points $\underline{x}$, where the weights are proportional to $f_{X_t}(\underline{x})$. 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 $\text{RMSE}(\underline{x})/\text{ASF}(\underline{x}) \times 100\%$ over $\underline{x}$.

Table 5: 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:	$U_{it} \sim \frac{1}{5}\mathcal{N}(0,4) + \frac{4}{5}\mathcal{N}\left(0, \frac{1}{4}\right)$

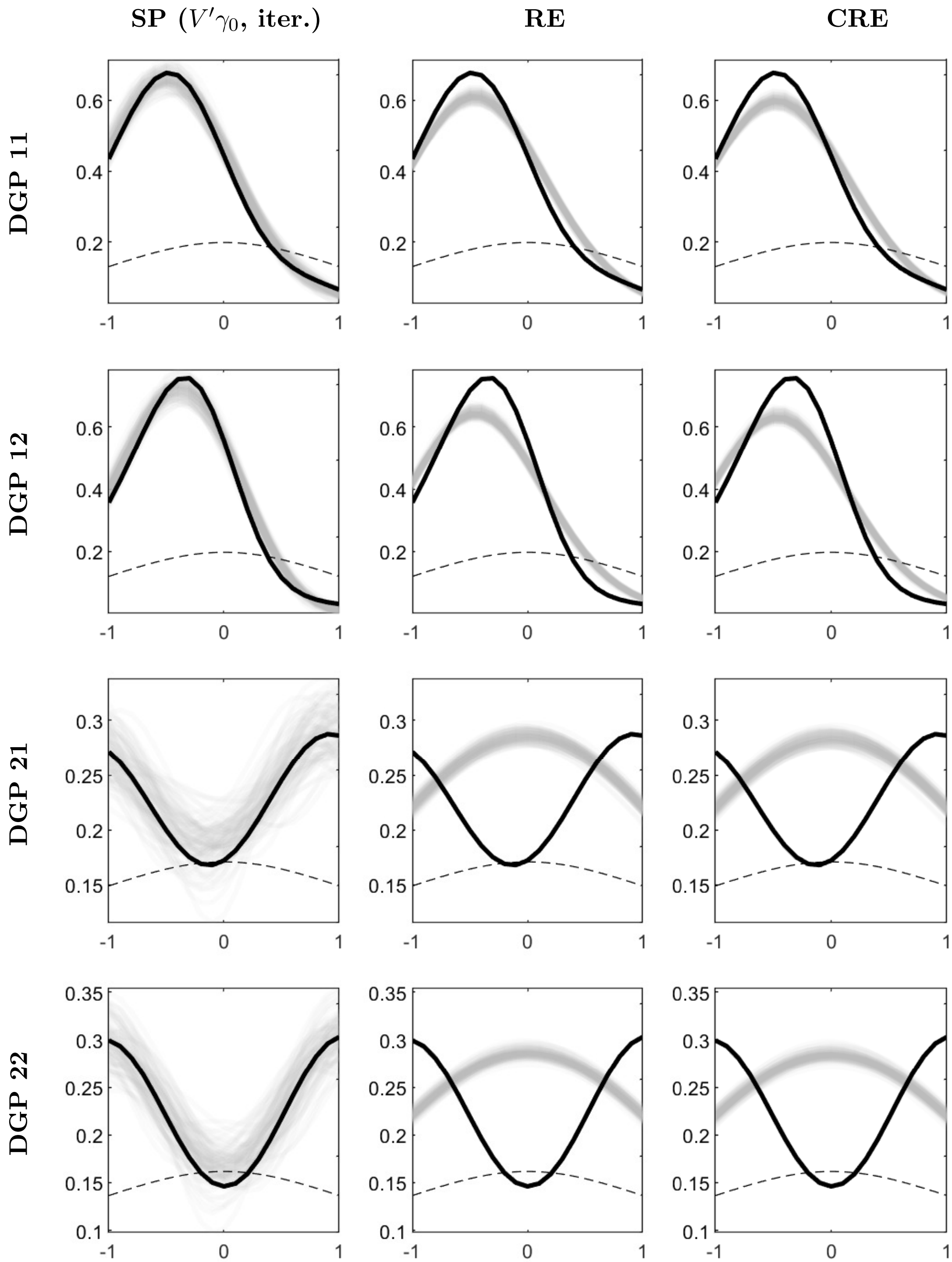
Notes: See the description in Table 1.

G.2 Monte Carlo Simulation Case 2: Conditioning on $V'\gamma_0$, Estimated Indices

The Monte Carlo design is described in Table 5, 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 semiparametric estimators discussed in Appendix B.3: SP, SP ($V'\gamma_0$), and SP ($V'\gamma_0$, iter.). In the current setup, there is no misspecification for all three semiparametric estimators.

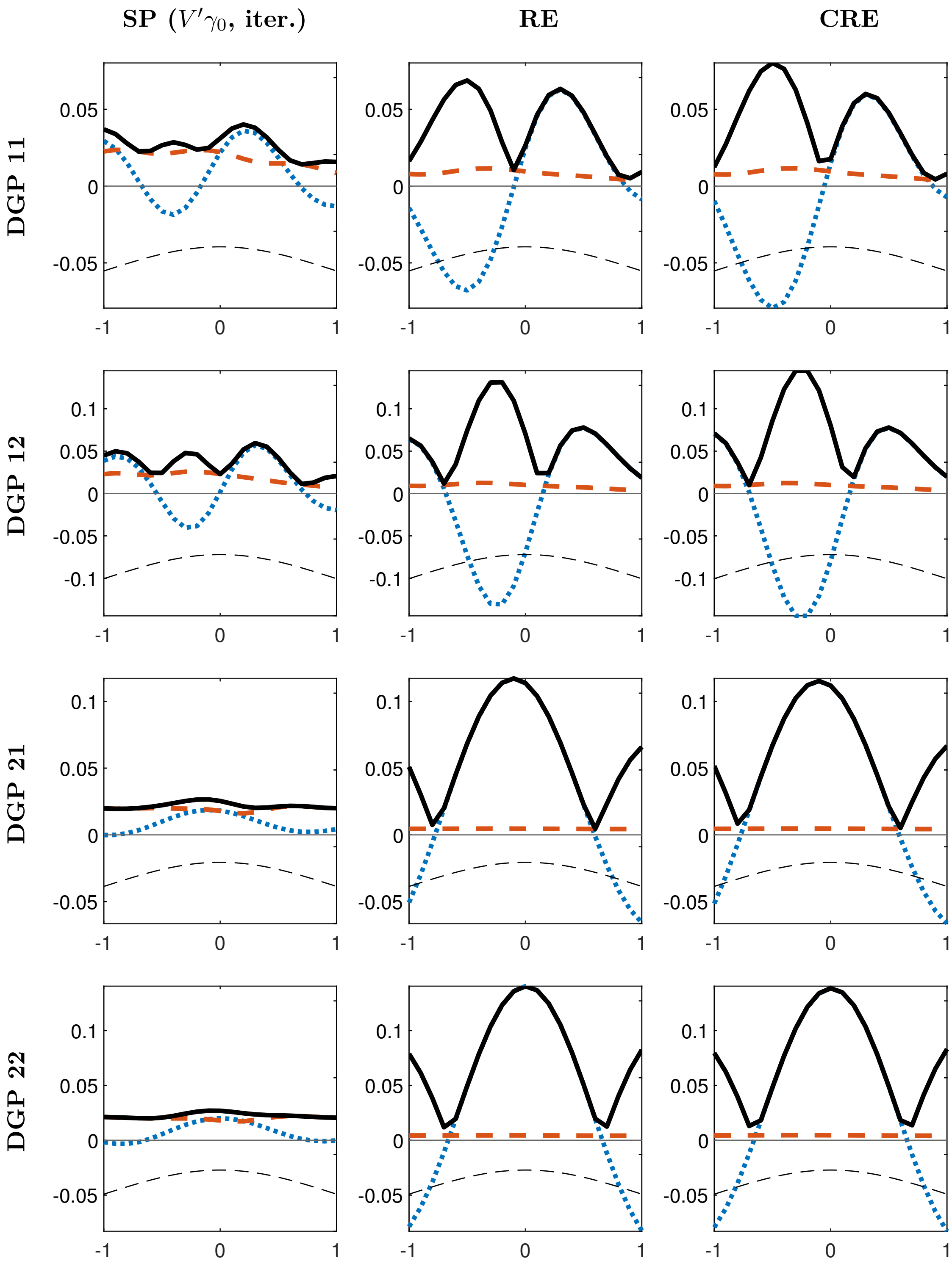
Figure 7 shows the estimated APEs, and Figure 8 depicts the APE performance statistics. Figures 9 and 10 present the estimated ASFs and performance statistics, respectively. Table 6 reports the bias, standard deviation, RMSE, and RMSE ratio statistics for the APE estimators, and Table 7 for the common parameter and ASF. In terms of estimation performance, the differences across the three semiparametric estimators are relatively small, and similar to Case 1, they dominate the RE and CRE.

Figure 7: Estimated APE vs True APE - Monte Carlo Case 2



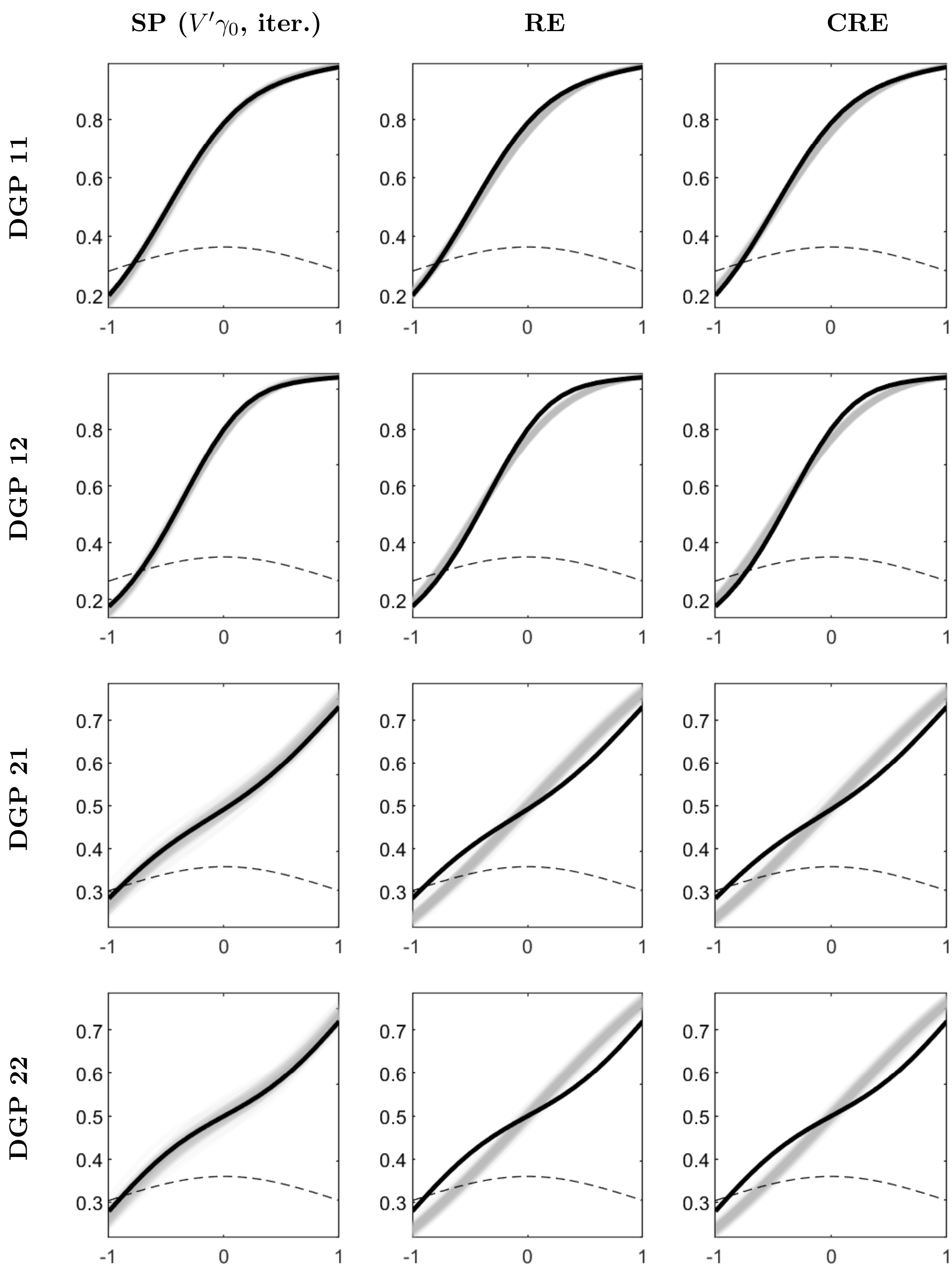
Notes: X-axes are potential values $\underline{x}^{(2)}$. Black solid lines are the true APE. Gray bands are collections of lines where each line corresponds to the estimated APE based on one simulation repetition. Thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}^{(2)}\right)$.

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



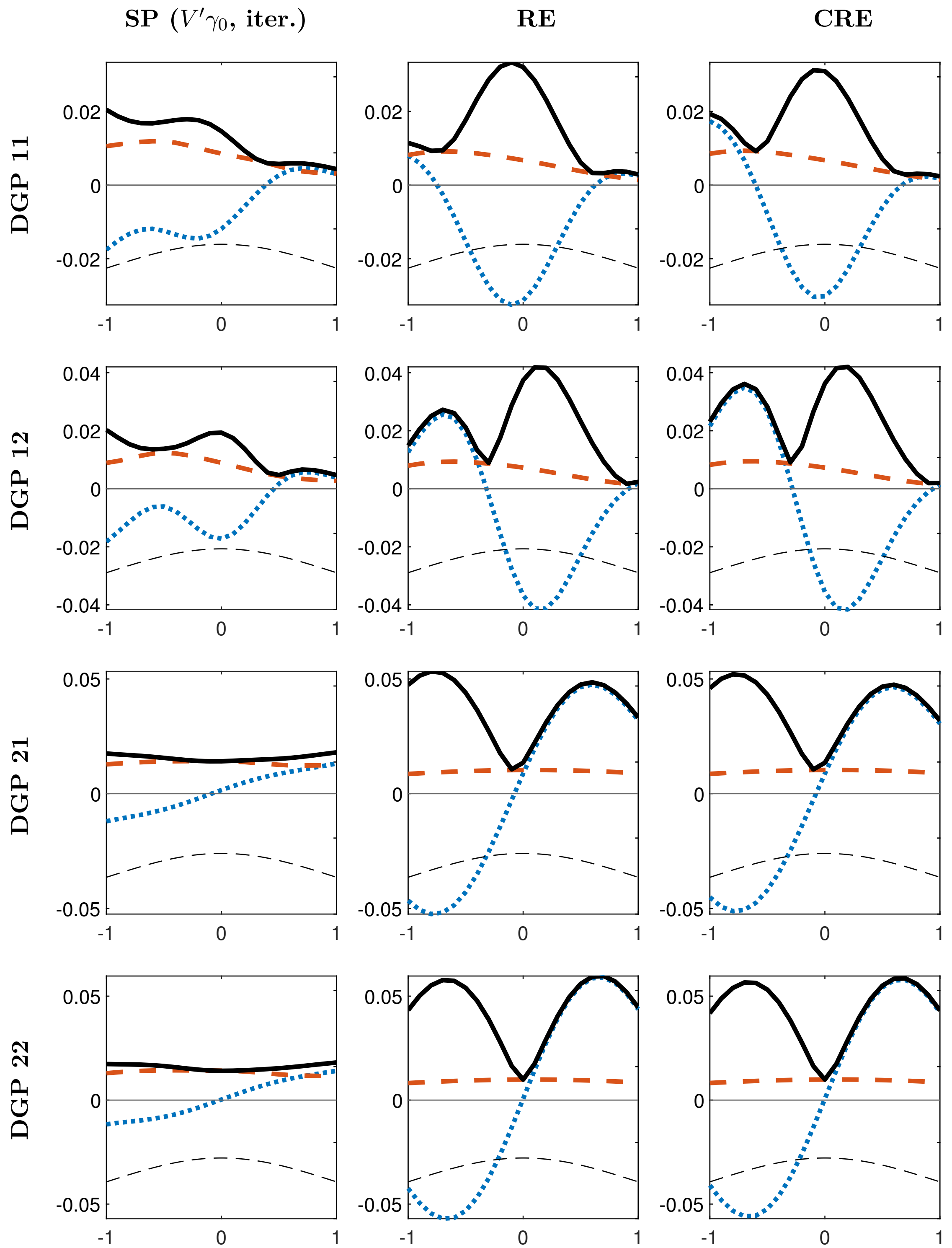
Notes: X-axes are potential values $\underline{x}^{(2)}$. Black solid / blue dotted / red dashed lines represent the RMSEs / biases / standard deviations of the APE estimates. Thin dashed lines at the bottom of all panels show $f_{X_t^{(2)}}\left(\underline{x}^{(2)}\right)$.

Figure 9: Estimated ASF vs True ASF - Monte Carlo Case 2



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

Figure 10: Bias, Standard Deviation, and RMSE in ASF Estimation - Monte Carlo Case 2



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

Table 6: APE Estimation - Monte Carlo Case 2

		$\|\text{Bias}\|$	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%

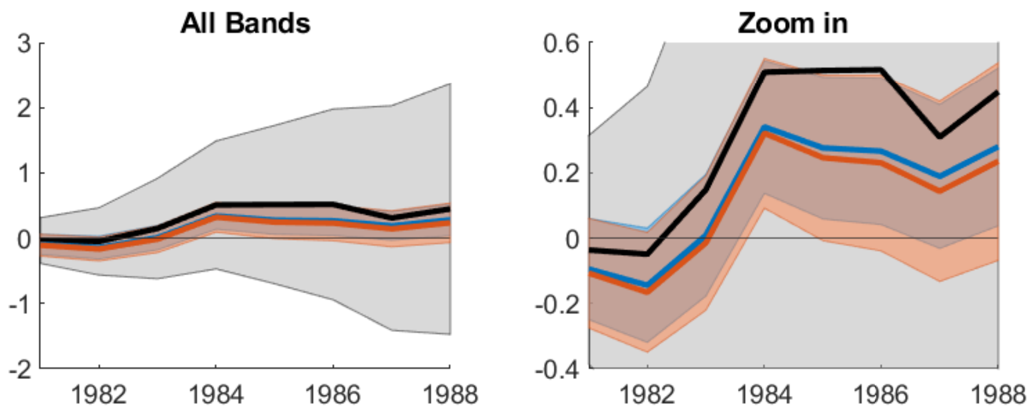
Notes: |Bias| indicates the absolute value of the bias. The reported |Bias|, SD, and RMSE are weighted averages across the collection of evaluation points $\underline{x}$, where the weights are proportional to $f_{X_t}(\underline{x})$. Bold entries indicate the best estimator (i.e., with the smallest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

Table 7: 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 allow comparisons across estimators. |Bias| indicates the absolute value of the bias. The |Bias|, SD, and RMSE of the ASF are weighted averages across the collection of evaluation points $\underline{x}$, where the weights are proportional to $f_{X_t}(\underline{x})$. 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($\underline{x}$)/ASF($\underline{x}$) $\times$ 100% over $\underline{x}$.

Figure 11: Estimated Coefficients on Time Dummies - Female Labor Force Participation



Notes: Black/blue/orange solid lines represent point estimates of the coefficients on time dummies using the smoothed maximum score/RE/CRE. 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.

G.3 Empirical Illustration

Table 8 summarizes descriptive statistics of the observables and supplements Figure 3 in the main text.

Figure 11 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 show 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 12 plots the estimated ASF and APE based on alternative specifications. In the benchmark specification, 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 Section B.3 for detailed comparison across estimators). To explore the effects of these choices on the empirical findings, we examine a range of 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 $(V'\gamma_0)$.[24] Comparing with the benchmark specification in Figure 4, we see that, in general, the estimates do not change much as we vary the timing of V or consider estimated indices.

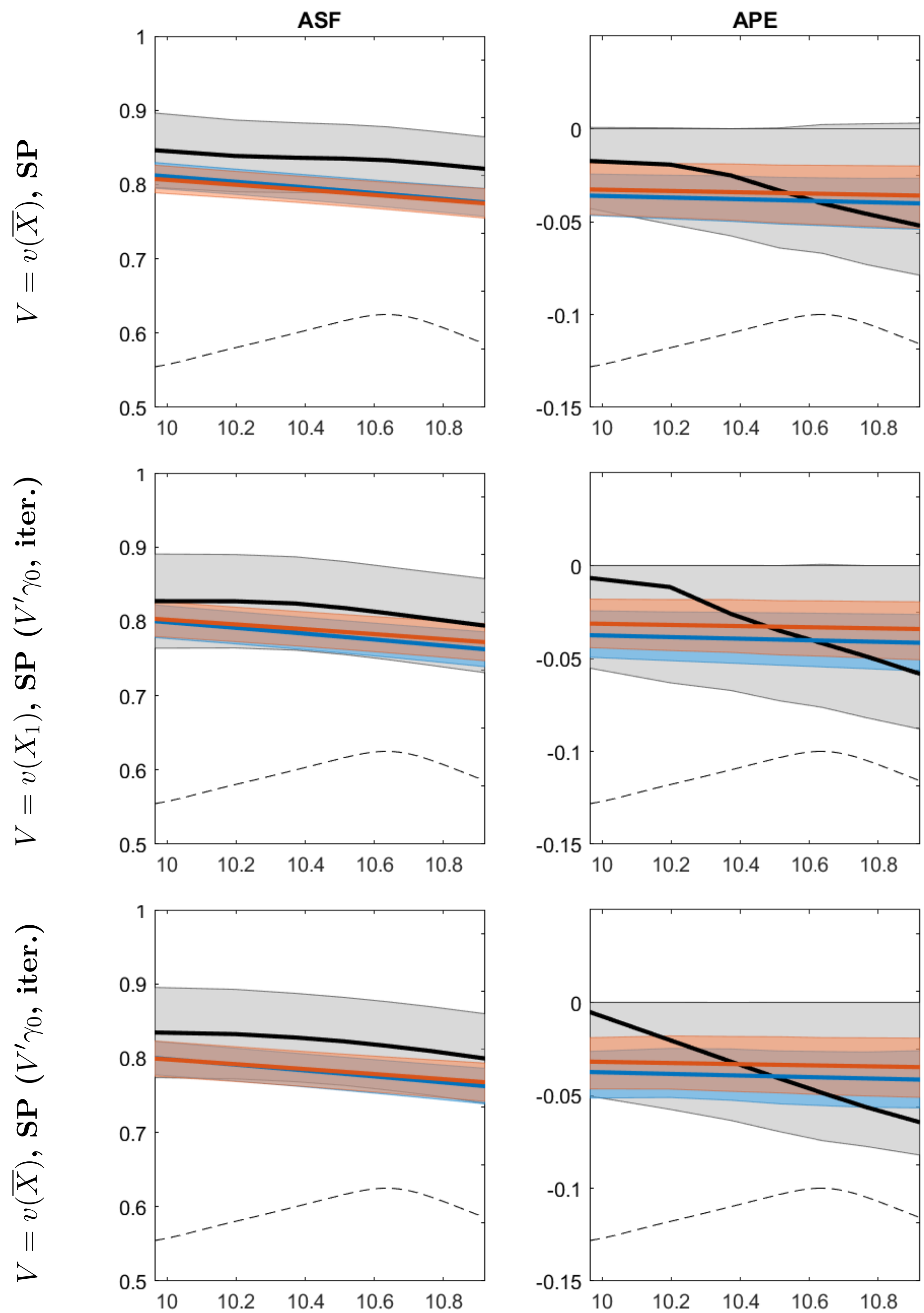
[24]For robustness checks with alternative coarsening schemes and the local logit estimator, see the previous version of this paper (Liu, Poirier, and Shiu, 2021).

Table 8: 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 participate in the labor market in some years but not all. See Fernández-Val (2009) for details.

Figure 12: Estimated ASF and APE under Alternative Specifications - Female Labor Force Participation



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