Semiparametric Estimation of Dynamic Binary Choice Panel Data Models

by

Fu Ouyang

and

Thomas Tao Yang

ANU Working Papers in Economics and Econometrics
671

MAY 2020

JEL: C14, C23, C35

ISBN: 0 86831 671 7

Semiparametric Estimation of Dynamic Binary Choice Panel Data Models*

Fu Ouyang[†1] and Thomas Tao Yang[‡2]

[1]School of Economics, University of Queensland
[2]Research School of Economics, Australian National University

April 29, 2020

Abstract

We propose a new approach to the semiparametric analysis of panel data binary choice models with fixed effects and dynamics (lagged dependent variables). The model we consider has the same random utility framework as in Honoré and Kyriazidou (2000). We demonstrate that, with additional serial dependence conditions on the process of deterministic utility and tail restrictions on the error distribution, the (point) identification of the model can proceed in two steps, and only requires matching the value of an index function of explanatory variables over time, as opposed to that of each explanatory variable. Our identification approach motivates an easily implementable, two-step maximum score (2SMS) procedure – producing estimators whose rates of convergence, in contrast to Honoré and Kyriazidou's (2000) methods, are independent of the model dimension. We then derive the asymptotic properties of the 2SMS procedure and propose bootstrap-based distributional approximations for inference. Monte Carlo evidence indicates that our procedure performs adequately in finite samples. We then apply the proposed estimators to study labor market dependence and the effects of health shocks, using data from the Household, Income and Labor Dynamics in Australia (HILDA) survey.

JEL classification Codes: C14, C23, C35.

Keywords: Semiparametric estimation; Binary choice model; Panel data; Fixed effects; Dynamics; Maximum score; Bootstrap

*We are grateful to Yonghong An, Shakeeb Khan, Arthur Lewbel, Takuya Ura, Hanghui Zhang, and Yichong Zhang for useful feedback. We also thank conference and seminar participants at 2019 Asian Meeting of the Econometric Society and 2019 Shanghai Workshop of Econometrics for their helpful comments. Fu Ouyang acknowledges the financial support of the Faculty of Business, Economics, and Law (BEL), University of Queensland, via the 2019 BEL New Staff Research Start-Up Grant. All errors are our responsibility.

†Email: f.ouyang@uq.edu.au.

‡Email: tao.yang@anu.edu.au.

1 Introduction

In this paper, we propose a novel two-step estimation method for panel data binary choice models with fixed effects and dynamics. Specifically, we consider binary choice models of the form:

$$y_{it} = 1\left[x_{it}'\beta + \gamma y_{it-1} + \alpha_i - \epsilon_{it} > 0\right], i = 1, ..., n, t = 1, ..., T^{1}, \tag{1.1}$$

where T is small and n is large, x_{it} is a $K \times 1$ vector of (time-varying) explanatory variables[2], y_{it-1} is the lagged dependent variable, α_i represents a time-invariant, individual-specific (fixed) effect, and ϵ_{it} is an idiosyncratic error term. Both α_i and ϵ_{it} are unobservable to the econometrician. Interest centers on estimating the preference parameter $\theta \equiv (\beta', \gamma)'$. y_{i0} is assumed to be observed, although the model is not specified in the initial period 0. In the literature, lagged terms y_{it-1} and fixed effect α_i are referred to as the "state dependence" (see Heckman (1981a,b)) and the "unobservable heterogeneity", respectively. The co-existence of these two terms complicates the identification and estimation of θ due to the multiple sources of persistence in y_{it}.

This paper resembles other panel data discrete response literature using fixed effects methods, in that there are no restrictions imposed on the distribution of α_i, conditional on the observed explanatory variables. Rasch (1960, 1961) and Andersen (1970) demonstrate that, in the absence of state dependence (y_{it-1}), β can be estimated by conditional maximum likelihood method if ϵ_{it} is assumed to be independent of all the other covariates and i.i.d. across both time periods and individuals with a logistic distribution. The fixed effects approach presented in Manski (1987) enables the identification of β without the parametric and serial independence restrictions placed on ϵ_{it}. Chamberlain (2010) shows that, outside of the logistic case, these "static" binary choice models have zero information bound and the identification requires at least one of the observed covariates having unbounded support.

In the presence of lagged dependent variables, the conditional maximum likelihood method can be used to estimate γ, provided that there are no other explanatory variables x_{it} and that there are at least four observations ($T \geq 3$) per individual[3] (see Chamberlain (1985) and Magnac (2000)). For more general model with x_{it}, Honoré and Kyriazidou (2000) (referred to as HK henceforth) proposed a conditional maximum likelihood estimator (CMLE) for model (1.1) with logistic errors and $T \geq 3$. Hahn (2001) examined the semiparametric efficiency of the CMLE proposed in HK. Bartolucci and Nigro (2010, 2012) demonstrated that the dynamic Logit model for binary panel data may be approximated by a quadratic exponential model. Aguirregabiria, Gu, and Luo (2018) studied dynamic panel data Logit models with forward-looking decision-making process, by deriving the minimal sufficient statistics for the unobserved fixed effect.

HK were the first to consider the semiparametric identification and estimation of model (1.1). They demonstrated that θ could be identified if, in addition to assumptions analogous to the ones

[1] The identification approach and estimation method presented in this paper can be applied to models with unbalanced panels as long as the unbalancedness is not the result of endogenous attrition.

[2] Any time-invariant covariates can be thought of as being part of the fixed effect α_i.

[3] Throughout this paper, this means that the data contains y_{i0} and $(y_{i1}, y_{i2}, y_{i3}, x_{i1}, x_{i2}, x_{i3})$ for each individual i.

in Manski (1987), all explanatory variables are strictly exogenous, ϵ_{it}'s are serially independent, and $T \geq 3$. However, their proposed estimator requires matching all explanatory variables over time, and so rules out time-specific effects. Further, the rate of convergence of their estimator decreases as the number of continuous regressors increases and is slower than the standard maximum score rate derived by Kim and Pollard (1990).

There are several alternative fixed effects approaches to the semi- and nonparametric analysis of dynamic binary choice models. Honoré and Lewbel (2002) proposed an identification strategy for a general (static and dynamic) framework which requires an exclusion restriction (excluded regressor) that one of the explanatory variables is independent of $(\alpha_i, \epsilon_{it})$, conditional on the other regressors including y_{it-1}. More recently, Chen, Khan, and Tang (2018, 2019) showed that the exclusion restriction in Honoré and Lewbel (2002) implicitly required (conditional) serial independence of the excluded regressor in a dynamic setting[4]. Similarly, Williams (2019) studied nonparametric identification of dynamic binary choice models satisfying certain initial conditions, in addition to restrictions on the dynamic process for observed covariates, conditional on α_i. In the absence of excluded regressors, Khan, Ponomareva, and Tamer (2019) established the sharp identified sets of θ under various (weaker) stochastic restrictions on ϵ_{it}, and provided corresponding sufficient conditions for point identifying θ on certain subsets of the support of regressors characterized by a series of moment inequalities.

This paper takes one step in the direction of HK, in the sense that we provide sufficient conditions under which model (1.1) can be identified and estimated without the necessity of matching each of the explanatory variables over time, provided that at least five observations per individual are observed (i.e., $T \geq 4$)[5]. The key insight thereof is that the identification of θ can proceed in two steps. First, β can be identified based on sequences of $\{y_{it}\}$, for which $y_{is-1} = y_{it-1}$ and $y_{is+1} = y_{it+1}$ for some $1 \leq s < t \leq T-1$ with $t \geq s+2$, if the distribution of explanatory variables x_{it} satisfies certain serial dependence and stochastic dominance restrictions. Then, with identified β, the identification of γ can be achieved by simply matching $x_{it}'\beta$ over time. We propose an estimation procedure for β and γ, establish the asymptotics for our estimators, and provide ways of inference using sampling methods. We investigate their small sample properties via Monte Carlo experiments.

As demonstrated by Honoré and Tamer (2006)[6], matching exogenous utilities over time is essential for the point identification of dynamic discrete choice models. However, the approach developed in this paper involves matching an identified linear combination of x_{it}, rather than HK's matching each component of x_{it}. Consequently, in contrast to the results presented by HK, the

[4]Chen, Khan, and Tang (2019) outlined a method, based on adopting the approach in Honoré and Lewbel (2002), to allow for certain serial correlation (e.g., AR(1)) in the excluded regressor.

[5]Namely, at least y_{i0} and $(y_{i1}, y_{i2}, y_{i3}, y_{i4}, x_{i1}, x_{i2}, x_{i3}, x_{i4})$ are observed for each individual i. This is a restriction on the minimum panel length, which is satisfied for many longitudinal panel data sets, such as the HILDA data used for the empirical application in Section 7 of this paper.

[6]More precisely, Honoré and Tamer (2006) provided examples of point identification fails when it is impossible to match x_{it} over time.

rates of convergence of our proposed estimators are independent of the dimension of the regressor space, making the present paper a useful alternative to HK, especially for models with higher dimensional design. Moreover, HK's approach implicitly requires each component of x_{it} to have a time-varying overlap support. This is a restrictive condition, as it rules out explanatory variables, such as time-specific intercepts. The support restriction of this type can be relaxed when the identification only requires matching the index $x_{it}'\beta$, so our procedure enables the econometrician to identify and estimate time-specific effects via including time dummies.

It is known that panel data binary choice models with unobserved heterogeneity and dynamics can be estimated by the random effects or correlated random coefficients approach. Examples include Arellano and Carrasco (2003), Wooldridge (2005), and Honoré and Tamer (2006). In addition to preference parameters, these approaches often allow the econometrician to calculate other quantities of interest, such as choice probabilities and marginal effects. However, these approaches require the specification of the statistical relation between the explanatory variables and α_i. Further, they also require one to specify the distribution of y_{i0}, conditional on the observed explanatory variables and α_i, which raises the so-called initial condition problem. Conversely, the fixed effects approaches attempt to estimate preference parameters without making these subtle specifications. Finally, there is also literature exploring the identification and estimation of various partial effects in panel data models. See, for example, Altonji and Matzkin (2005), Chernozhukov, Fernández-Val, Hahn, and Newey (2013), and Torgovitsky (2019), among others.

Dynamic binary choice models have a wide range of applications. Contemporary empirical literature includes studies of labor force participation (Corcoran and Hill (1985), Hyslop (1999), Lee and Tae (2005), and Damrongplasit, Hsiao, and Zhao (2018)), poverty dynamics (Biewen (2009)), health status (Contoyannis, Jones, and Rice (2004) and Halliday (2008)), educational attainment (Cameron and Heckman (1998, 2001)), stock market participation (Alessie, Hochguertel, and Soest (2004)), product purchase behaviour (Chintagunta, Kyriazidou, and Perktold (2001)), welfare participation (Chay, Hoynes, and Hyslop (1999)), and firm behavior (Roberts and Tybout (1997) and Kerr, Lincoln, and Mishra (2014)). Most applications have typically employed parametric forms of the model (1.1), such as Logit and Probit, or random effects assumptions. The robustness from the distribution-free and fixed effects specification makes the approach proposed in this paper a competitive alternative to existing parametric and random effects methods.

The remainder of this paper is organized as follows. Section 2 establishes the identification of θ under different sets of sufficient conditions, based on which, a 2SMS procedure is proposed in Section 3. Sections 4 and 5 derive asymptotic properties of the 2SMS estimator and proposes bootstrap-based inference methods. We present the results of Monte Carlo experiments in Section 6 investigating the finite-sample performance of the proposed method, and illustrate its empirical application in Section 7 using HILDA data. Section 8 concludes the paper. All proofs and tables are collected in Appendix.

For ease of reference, the notations maintained throughout this paper are listed here.

Notation. All vectors are column vectors. $\mathbb{R}^p$ is a p-dimensional Euclidean space equipped with the Euclidean norm $\|\cdot\|_2$. We reserve letter $i \in \mathcal{N} \equiv \{1, ..., n\}$ for indexing individuals, and letters $s, t \in \mathcal{T} \equiv \{1, ..., T\}$ for indexing time periods. An observation is indexed by (i, t). To simplify the notation, we will suppress the subscript i in the rest of this paper whenever it is clear from the context that all variables are for each individual. Vector x_{its} denotes $x_{it} - x_{is}$. The first element of x_{its} is denoted by $x_{its,1}$ and the sub-vector comprising its remaining elements is denoted by $\tilde{x}_{its}$. Following a substantial panel literature, we use the notation ξ^t to denote $(\xi_1', ..., \xi_t')'$. $F_{\zeta|\cdot}$ and $f_{\zeta|\cdot}$ denote, respectively, the conditional cumulative distribution function (CDF) and probability density function (PDF) of a random vector ζ conditional on $\cdot$. For two random vectors, u and v, the notation $u \overset{d}{=} v|\cdot$ means that u and u have identical distribution, conditional on $\cdot$, and $u \perp v|\cdot$ means that u and u are independent conditional on $\cdot$. We use $P(\cdot)$ and $\mathbb{E}[\cdot]$ to denote probability and expectation, respectively. Function $1[\cdot]$ is an indicator function that equals one when the event in the brackets is true, and zero otherwise. Symbols $\backslash$, $'$, $\propto$, $\Leftrightarrow$, $\overset{d}{\to}$, and $\overset{P}{\to}$ represent set difference, matrix transposition, proportionality, "if and only if", convergence in distribution, and convergence in probability, respectively. For any (random) positive sequences, $\{a_n\}$ and $\{b_n\}$, $a_n = O(b_n)$ ($O_P(b_n)$) means that a_n/b_n is bounded (bounded in probability) and $a_n = o(b_n)$ ($o_P(b_n)$) means that $a_n/b_n \to 0$ ($a_n/b_n \overset{P}{\to} 0$).

2 Identification

This section provides sufficient conditions for identifying the parameter θ with no need of matching observed covariates x_{it} over time. Under these assumptions, we derive a set of identification inequalities that can be taken to data for (point) estimation and inference on the parameter θ.

Note that we use the following notations in this section. $x^T \equiv (x_1', ..., x_T')'$, $\epsilon^T \equiv (\epsilon_1, ..., \epsilon_T)'$, and $x_{ts} \equiv x_t - x_s$. The first element of x_{ts} is denoted by $x_{ts,1}$ and the sub-vector comprising its remaining elements is denoted by $\tilde{x}_{ts}$.

Suppose that a random sample from a population of independent individuals is observed for $T + 1$ ($= |\mathcal{T} \cup \{0\}|$) periods. Recall that, for all $t \in \mathcal{T}$,

$$y_t = 1\left[x_t'\beta + \gamma y_{t-1} + \alpha - \epsilon_t > 0\right]. \tag{2.1}$$

Note that the model is incomplete – in the sense that it does not specify the relationship between y_0 and $(x^T, \alpha, \epsilon^T)$. This is known as the initial condition problem in panel data literature. This paper employs a fixed effects approach, in which we attempt to estimate θ without making any assumptions on the distribution of α, conditional on explanatory variables. This helps us to avoid explicitly specifying the functional form of $p_0(x^T, \alpha) \equiv P(y_0 = 1|x^T, \alpha)$, and so circumvents the initial condition problem.

As mentioned, we impose no restriction on $F_{\alpha|x^T}$, but place the following restrictions on observed covariates x^T and unobserved idiosyncratic errors ϵ^T:

Assumption A. For all α and $s, t \in \mathcal{T}$,

(a) (i) $\epsilon^T \perp (x^T, y_0)|\alpha$, (ii) $\epsilon_s \perp \epsilon_t|\alpha$, and (iii) $\epsilon_s \overset{d}{=} \epsilon_t|\alpha$.

(b) $F_{\epsilon_t|\alpha}$ is absolutely continuous with PDF $f_{\epsilon_t|\alpha}$ and support $\mathbb{R}$.

(c) (i) Without loss of generality (w.l.o.g.), $x_{ts,1}$ has almost everywhere (a.e.) positive probability density on $\mathbb{R}$, conditional on $\tilde{x}_{ts}$ and α, and (ii) the coefficient β_1 on $x_{ts,1}$ is nonzero.

(d) The support $\mathcal{X}_{ts}$ of $F_{x_{ts}|\alpha}$ is not contained in any proper linear subspace of $\mathbb{R}^K$.

(e) $\theta = (\beta', \gamma)' \in \mathcal{B} \times \text{int}(\mathcal{R})$, where $\mathcal{B} \equiv \{b = (b_1, ..., b_K)' \in \mathbb{R}^K | \|b\|_2 = 1, b_1 \neq 0\}$ and $\mathcal{R}$ is a compact subset of $\mathbb{R}$.

Assumption A places the same set of restrictions on the joint distribution of $(x^T, \alpha, \epsilon^T)$ as HK. While not explicitly stated in their Theorem 4, HK used Assumption A(a), the exogeneity of (x^T, y_0) and serial independence of $\{\epsilon_t\}$, conditional on α, to derive the moment inequalities used for the identification. Note that Assumption A(a) implies that the fixed effects α pick up two types of dependence in the model: the dependence over time in the unobservables and the dependence between explanatory variables and unobservables. As a result, in model (2.1), ϵ_t is independent of $\left(x^T, y^{t-1}\right)$, conditional on α. Besides, Assumption A(a) is a special case of the group homogeneity restriction, $\epsilon_s \overset{d}{=} \epsilon_t|(x_s, x_t, \alpha)$, imposed in Manski (1987), Pakes and Porter (2016), and Shi, Shum, and Song (2018) for identifying static discrete choice models (without controlling the lagged term y_{t-1} in the model). This enables us to suppress the time subscript t in $F_{\epsilon_t|\alpha}$ and $f_{\epsilon_t|\alpha}$ in the rest of this paper without ambiguity. Assumption A(b) is a regularity condition to ensure that both $y_s \neq y_t$ and $y_s = y_t$ occur with positive probabilities for all α and $s, t \in \mathcal{T}$.

It is known and documented in the relevant literature (see, e.g., Lemma 1 of Manski (1985)), that to establish the point identification of the parameter θ in a "distribution-free" setting, x_t also needs to satisfy certain regularity conditions. Assumption A(c) requires the existence of a relevant, continuous regressor, with large support, which is a standard restriction imposed in maximum score type estimators. Assumption A(d) is the familiar full-rank condition. Assumptions A(c) and A(d) are identical to Assumption 2 of Manski (1987).

Assumption A(e) is for scale normalization and parameter space. This is a typical practice for discrete choice models because the identification of θ is only up to scale. In the semiparametric framework, where no parametric form of $F_{\epsilon|\alpha}$ is specified, identification is often achieved by normalizing the magnitude of the regression coefficients. Assumption A(e) assumes that β is on the unit circle and has nonzero first element β_1[7].

HK demonstrated that, if $T \geq 3$, θ can be identified under Assumption A[8]. Their proposed approach requires matching all exogenous covariates over time, and results in an estimator with a

[7] It will be made clear that our procedure identifies β and γ sequentially, so it is more convenient to normalize the scale of β rather than that of θ, as in HK.

[8] As is stated in HK, Assumption A is not sufficient for point identifying θ if $T < 3$.

rate that declines as the number of exogenous covariates increases. The main contribution of this paper is the provision of a set of supplementary conditions, under which the identification of θ can escape from the necessity of element-by-element matching. Specifically, our approach is based on the following monotonic relationship between a conditional choice probability and an index of the exogenous covariates: For some $s, t \in \mathcal{T}$ such that $t - s \geq 2$,

$$P(y_t = 1 | x_s, x_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \geq P(y_s = 1 | x_s, x_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\ \Leftrightarrow \\ x_t'\beta \geq x_s'\beta. \tag{2.2}$$

Note that (2.2) implicitly requires that there are at least five ($T \geq 4$) observations per individual observed by the econometrician (i.e., $s = 1$, $t = 3$, and $s + 1 = t - 1 = 2$).

The idea is straightforward. By matching the statuses of the previous ($y_{s-1} = y_{t-1}$) and subsequent ($y_{s+1} = y_{t+1}$) periods, we want the state dependence between y_s and y_t to be cut off in a symmetric manner, so the (conditional) probabilities of choosing 1 in periods s and t to be solely rank ordered by indexes $x_s'\beta$ and $x_t'\beta$. However, to reach this conclusion, two concerns related to conditioning on the states one period ahead (y_{s+1} and y_{t+1}) have to first be addressed. First, x_s (x_t) may affect the value of y_{t+1} (y_{s+1}) via its serial dependence on x_{t+1} (x_{s+1}). Second, the dependence between x_t and y_{t+1} (via x_{t+1}) may change dramatically over time. Both require additional restrictions to be placed on the serial dependence of the stochastic process of x_t.

The following condition, together with Assumption A, is sufficient to ensure (2.2).

Assumption SI. For all $s, t \in \mathcal{T}$, (a) $x_s \perp x_t | \alpha$, and (b) $x_s \stackrel{d}{=} x_t | \alpha$.

Assumption SI imposes strong restriction on the dynamic process of the covariate sequence, which requires the process $\{x_t\}$ to be serially independent and strictly stationary[9], conditional on the individual-specific effects α. In a dynamic fixed effects economic model, α collects all time-invariant covariates, as well as unobserved individual preferences, abilities, or character traits. In such models, if x_t only includes observed individual characteristics naturally correlated with α, it may be reasonable to further assume that the serial dependence in the process $\{x_t\}$ is also derived from α. If x_t contains covariates related to some institutional factors that leads to exogenous variation in, for example, costs of participation, across individuals, Assumption SI may be approximately satisfied by using the differencing, demeaning, or de-trending transformation of these variables. This applies to cases where $\{x_t\}$ exhibits some long-run equilibrium (trend,

[9]In a separate work, we study the identification of θ with Assumption SI(a) being replaced by $x_s \perp x_t | \alpha$ for all $|t-s| > \iota$ with some $\iota > 1$. That is, $x_s \perp x_t | \alpha$ holds if s and t are sufficiently separated from each other. This condition is satisfied for all MA(q) processes with $q \leq \iota$. In the more general case of $\{x_t\}$ being stationary ARMA process, the serial dependence of $\{x_t\}$ decays at an exponential rate. This condition is approximately satisfied for all s, t with $|t - s|$ large enough. The identification approach needs a first step identifying the order of $\{x_t\}$, requires a minimum length of the panel depending on this order, and adopts an objective function different from the one studied in this paper. Due to space constraints, we do not delve into the details of this approach here. For interested readers, a brief note is available upon request.

deterministic or stochastic). The transformed regressor then measures the deviation of x_t from its long-run equilibrium (trend), which, in some cases[10], is assumed to be a white noise process affecting the short-run dynamics of the model.

HK implicitly assume that the support of x_t is overlapping over time, so the differences in regressors across different time periods have a positive density in a neighborhood of 0. If x_t varies across individuals as well as over time[11], Assumption SI requires the distribution of x_t to be time stationary, conditional on α. However, evidence presented in Honoré and Tamer (2006) implies that some additional assumption is needed, to achieve point identification without doing element-by-element match, as in HK.

Further, we try to relax the conditional serial independence assumption and provide the following sufficient, high-level condition, that permits some limited dependence of the covariates, for the identification. We also provide a sufficient (but not necessary) condition in Proposition 2.1 that can imply this high-level condition.

Assumption SD. For all α and $s, t \in \mathcal{T}$,

(a) $f_{\epsilon|\alpha}(\cdot)/F_{\epsilon|\alpha}(\cdot)$ is a non-increasing function, or equivalently, $f_{\epsilon|\alpha}(\cdot)/[1 - F_{\epsilon|\alpha}(\cdot)]$ is a non-decreasing function.

(b) Let $w_t \equiv x_t'\beta$. The following stochastic dominance conditions hold for all $v \in \mathbb{R}$ and $d_0, d_1 \in \{0, 1\}$: If $w_t \geq w_s$, then

$$F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v) \geq F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v),$$

and if $w_t \leq w_s$, then

$$F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v) \leq F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v).$$

Assumption SD(a) says that $F_{\epsilon|\alpha}$ has decreasing inverse Mills ratio, which, together with Assumption SD(b), guarantees the monotonic relation in (2.2). Assumption SD(a) is satisfied by many common continuous distributions, such as Gaussian, logistic, Laplace, uniform, gamma, log-normal, Gumbel, and Weibull[12]. However, this property fails if $F_{\epsilon|\alpha}$ has heavy tails (e.g., student's t-distribution and Cauchy distribution)[13]. Note that Assumption SD(a) is a key condition

[10] As an example, consider a case where $\{x_t\}$ is a random-walk-plus-drift process (i.e., $x_t = x_0 + a_0 t + \sum_{\tau=1}^{t} e_\tau$). Although $\{x_t\}$ violates Assumption SI, its first differencing $\Delta x_t = x_t - x_{t-1} = a_0 + e_t$ is i.i.d. over time.

[11] If x_t only varies over time but is equal for all individuals for given t, Assumption SI is neither weaker nor stronger than HK's overlapping support condition. For example, consider time effects δ_t. HK's approach requires $\delta_s = \delta_t$ for all $s, t \in \mathcal{T}$, while Assumption SI implies i.i.d. δ_t conditional on α, two extremes of the spectrum of serial dependence.

[12] In a mixture model, e.g.,

$$f_{\epsilon|\alpha}(e) = \sum_{m=1}^{M} \pi_m f_{\epsilon|\alpha}(e; \vartheta_m)$$

with mixing proportions π_m, $\sum_{m=1}^{M} \pi_m = 1$, where each component density has a different parameter vector ϑ_m, Assumption SD(a) holds for $F_{\epsilon_t|\alpha}(\cdot)$ if it is satisfied by all component distributions $F_{\epsilon|\alpha}(\cdot; \vartheta_m)$.

[13] More precisely, Assumption SD(a) does not hold globally for these distributions. For example, it is not hard to find that this assumption holds for t and Cauchy on $[-L, \infty)$ for some positive L.

imposed in McFadden (1976) and Silvapulle (1981) for both $-\log F_{\epsilon|\alpha}(\cdot)$ and $-\log(1 - F_{\epsilon|\alpha}(\cdot))$ being convex, which guarantees a unique solution for MLE in cross-sectional models with errors having general distribution.

In model (2.1), the exogenous utility w_t affects the value of y_{t+1} via y_t and its serial dependence with w_{t+1}, conditioning on α. The former is explicitly captured by the coefficient γ. For the latter, Assumption SD(b) restricts the serial dependence of $\{w_t\}$. It says that, conditional on α and the same "initial" and "ending" statuses ($y_{s-1} = y_{t-1}$, $y_{s+1} = y_{t+1}$), if the value of w_t is higher than that of w_s, then w_{t+1} has a better chance of taking a large value than w_{s+1}. This restriction rules out the case in which high utility in one period has negative effects on the utility in the next period. This assumption is more likely to hold in applications where $\{w_t\}$ represents a positively autocorrelated stochastic proccess of the "utility", "benefits", or "profits" of a decision. Assumption SD(b) is high level, for which a sufficient, but not necessary, condition is that the joint distribution of w^T is exchangeable, conditional on α, which is formally stated in Proposition 2.1 below. Similar exchangeability assumptions were imposed in Altonji and Matzkin (2005) and Chen, Khan, and Tang (2018). Assumption SD(b) can be thought of as a conditional "first-order stochastic dominance" condition, which implies that, for any non-decreasing (non-increasing) function $u(\cdot)$,

$$\int u(v)dF_{w_{s+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v) \leq \int u(v)dF_{w_{t+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v)$$

$(\int u(v)dF_{w_{s+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v) \geq \int u(v)dF_{w_{t+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v))$ whenever $w' \geq w$. The property is needed for establishing the monotonic relation in (2.2).

Proposition 2.1. Suppose that Assumption A is satisfied. Then Assumption SD(b) holds with equality, if the joint PDF of w^T conditional on α is exchangeable, i.e.,

$$f_{w^T|\alpha}(\omega_1, ..., \omega_T) = f_{w^T|\alpha}(\omega_{\pi(1)}, ..., \omega_{\pi(T)})$$

for all permutations $\{\pi(1), ..., \pi(T)\}$ defined on the set $\mathcal{T}$.

The proof of Proposition 2.1 can be found in Appendix A.

Remark 2.1. Assumption SD relaxes the serial independence and stationarity restrictions imposed by Assumption SI. To achieve the identification, we do need to restrict the tail behavior of the idiosyncratic error ϵ_t, conditional on α.

Remark 2.2. A few remarks are in order about how our identification conditions are related to the existing literature. First, compared with HK, our approach relies on additional assumptions restricting the serial dependence of strictly exogenous regressors x_t and requires $T \geq 4$. These conditions make identification without element-by-element matching of x_t possible. Second, our identification conditions are non-nested with those in the literature assuming exclusion restrictions, such as Honoré and Lewbel (2002), Chen, Khan, and Tang (2018, 2019), and Williams (2019). Chen, Khan, and Tang (2018, 2019) showed that Honoré and Lewbel (2002) essentially required the serial independence of the excluded regressor. Williams (2019) requires that the other strictly exogenous regressors are conditionally independent of the past values of the excluded regressor. In addition to specific restrictions on the dynamic process for the covariates, the identification results of these papers rely on the existence of at least one "excluded regressor" conditionally independent of the individual fixed effects α. Conversely, our approach allows for arbitrary correlation between x_t and α.

Given either of these two sets of sufficient conditions, the identification of θ proceeds in two steps. Proposition 2.2 demonstrates that β can be identified based on moment inequality (2.2), and Proposition 2.3 establishes the identification of γ by matching the value of the index function $x_t'\beta$ in different time periods.

Proposition 2.2 (Identification of β)**.** For all $s, t \in \mathcal{T}$ such that $t \geq s + 2$, define

$$\begin{aligned} Q_1(b) =& \mathbb{E}\left\{[P(y_t = 1|x_s, x_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}) - P(y_s = 1|x_s, x_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1})]\right. \\ & \left. \times sgn(x_{ts}'b)\right\}. \end{aligned}$$

Suppose Assumption A holds. If either Assumption SI or Assumption SD also holds, then $Q_1(\beta) > Q_1(b)$ for all $b \in \mathcal{B} \setminus \{\beta\}$.

The proof of Proposition 2.2 can be found in Appendix A.

Remark 2.3. For the case with $t = s + 2$, the population objective function $Q_1(b)$ reduces to $\mathbb{E}\{[P(y_t = 1|x_s, x_t, y_{t-3} = y_{t-1} = y_{t+1}) - P(y_s = 1|x_s, x_t, y_{t-3} = y_{t-1} = y_{t+1})]sgn(x_{ts}'b)\}$. Particularly, note that Proposition 2.2 indicates that our identification strategy for β requires $T = 4$, as a minimum. In this case, $t = s + 2$ must hold with $s = 1$ and $t = 3$, and thus $Q_1(b)$ is

$$Q_1(b) = \mathbb{E}\{[P(y_3 = 1|x_1, x_3, y_0 = y_2 = y_4) - P(y_1 = 1|x_1, x_3, y_0 = y_2 = y_4)] \cdot sgn(x_{31}'b)\}. \quad (2.3)$$

It is clear that a longer panel allows more population objective functions of similar form, and collectively, all these objective functions (by simply summing them up) can be used to garner identification information on β.

Proposition 2.2 establishes the identification of β, which enables us to move on to identify γ with β being treated as a known, constant vector. Then, the following proposition shows that γ can be identified by matching the deterministic utility w_t in different periods, of which the proof is presented in Appendix A.

Proposition 2.3 (Identification of γ)**.** Consider the event

$$A = \{y_0 = d_0, ..., y_{s-1} = d_{s-1}, y_s = 0, y_{s+1} = d_{s+1}, ..., y_{t-1} = d_{t-1}, y_t = 1, y_{t+1} = d_{t+1}, ..., y_T = d_T\},$$

and its counterpart

$$B = \{y_0 = d_0, ..., y_{s-1} = d_{s-1}, y_s = 1, y_{s+1} = d_{s+1}, ..., y_{t-1} = d_{t-1}, y_t = 0, y_{t+1} = d_{t+1}, ..., y_T = d_T\},$$

where $d_\tau \in \{0, 1\}$ for all $\tau \in \mathcal{T} \cup \{0\}$. Define

$$Q_2(r; \beta) = \mathbb{E}\left\{\left[P(A|x^T, w_t = w_{t+1}) - P(B|x^T, w_t = w_{t+1})\right] sgn\left((w_t - w_{t-1}) + r(d_{t+1} - d_{t-2})\right)\right\}$$

for all $s, t \in \mathcal{T}$ such that $t = s + 1$ and

$$\tilde{Q}_2(r;\beta) = \mathbb{E}\left\{\left[P(A|x^T, w_{s+1} = w_{t+1}, y_{s+1} = y_{t+1}) - P(B|x^T, w_{s+1} = w_{t+1}, y_{s+1} = y_{t+1})\right] \times sgn\left((w_t - w_s) + r(d_{t-1} - d_{s-1})\right)\right\}$$

for all $s, t \in \mathcal{T}$ such that $t > s + 1$. Then, under Assumption A,

(i) $Q_2(\gamma;\beta) > Q_2(r;\beta)$ for all $r \in \mathcal{R} \setminus \{\gamma\}$, and

(ii) $\tilde{Q}_2(\gamma;\beta) > \tilde{Q}_2(r;\beta)$ for all $r \in \mathcal{R} \setminus \{\gamma\}$.

Remark 2.4. Proposition 2.3 indicates that when β is known, the identification of γ can be achieved with $T \geq 3$. For the case $T = 3$, the identification is based on objective function $Q_2(\cdot;\beta)$. When $T \geq 4$, any combination (s,t) of the elements of $\{1, ..., T-1\}$ taken two at a time can be used to construct the population objective function to identify γ. For example, in the simplest case $T = 4$, feasible choices of (s,t) include $(1,2)$, $(1,3)$, and $(2,3)$. One can use any one of these pairs to define population objective function, either $Q_2(\cdot;\beta)$ or $\tilde{Q}_2(\cdot;\beta)$. It is clear that any one, or a combination (again by simply summing them up), of these objective functions can be used to identify γ.

Propositions 2.2 and 2.3 outline a two-step procedure for identifying the preference parameters β and γ, of which Proposition 2.3 uses HK's insight. Note that, as Proposition 2.2 suggests, an additional assumption, SI or SD, enables us to establish the identification of β independently to that of γ in the first step. As a result, it suffices to match the index $x_t'\beta$, rather than each component of x_t over time, as in HK, when identifying γ in the second step. The benefits of doing so are twofold: First, the two-step procedure avoids the curse of dimensionality caused by matching many explanatory variables, which makes it particularly competitive when handling high-dimensional models. As will be demonstrated in Section 4, a two-step estimation method motivated by this identification strategy yields consistent estimators with rates of convergence that are independent of the model dimension, as opposed to those in HK. Second, recall that matching x_t, as in HK, excludes regressors with support that is not overlapping over time (see also the discussion in Honoré and Tamer (2006)). A leading example is time-varying intercept δ_t[14] (or equivalently including a set of time dummies as regressors), which is commonly included in panel data models to control for fixed time effects. Our two-step procedure addresses this limitation by means of matching an index rather than all elements.

The following theorem is an immediate result of Propositions 2.2 and 2.3.

Theorem 2.1 (Identification of θ)**.** *Suppose Assumption A holds. If either Assumption SI or Assumption SD also holds, then β is identified based on population objective function $Q_1(\cdot)$, and γ is identified based on either population objective function $Q_2(\cdot;\beta)$ or $\tilde{Q}_2(\cdot;\beta)$.*

[14] δ_t is assumed to be varying over t but invariant across i with δ_1 normalized to zero.

3 Estimation

Applying the analogy principle, the identification results presented in Section 2 can be translated into a two-step estimation procedure. In the first step, a maximum score (MS) estimator (with binary weights) $\hat{\beta}$ of β is obtained. In the second step, γ is estimated by a localized MS procedure matching the estimated index $x_t'\hat{\beta}$ over time. Each of the two steps is described, in turn, below.

In Sections 3.1 and 3.2, we restrict our discussion to the model with $T = 4$ to streamline exposition. The same method can be applied with straightforward modification to models with longer panels. We provide objective functions for general cases with $T \geq 4$ in Section 3.3.

3.1 Estimation of β with $T = 4$

Assuming a random sample of n individuals, we propose the following weighted MS estimator $\hat{\beta}$ of β, defined as the maximizer over the parameter space $\mathcal{B}$:

$$\hat{\beta} = \arg\max_{b \in \mathcal{B}} Q_{1n}(b), \tag{3.1}$$

where

$$Q_{1n}(b) = \frac{1}{n}\sum_{i=1}^{n} 1[y_{i0} = y_{i2} = y_{i4}](y_{i3} - y_{i1}) \cdot sgn(x_{i31}'b), \tag{3.2}$$

with the notation $sgn(\cdot)$ denoting the sign function.

It is obvious from expression (3.2) that only observations that satisfy $y_{i1} \neq y_{i3}$, $y_{i0} = y_{i2}$, and $y_{i2} = y_{i4}$ are used in the estimation. Namely, the objective function uses only "switchers" whose choice changes in periods 1 and 3, with the same choices in, respectively, their previous and subsequent periods.

3.2 Estimation of γ with $T = 4$

Proposition 2.3 motivates a localized MS estimator $\hat{\gamma}$ of γ, defined here as the maximizer over the parameter space $\mathcal{R}$ of the objective function[15]

$$\begin{aligned} Q_{2n}(r;\beta) =& \frac{1}{n}\sum_{i=1}^{n} \{1[x_{i2}'\beta = x_{i3}'\beta](y_{i2} - y_{i1}) \cdot sgn(x_{i21}'\beta + r(y_{i3} - y_{i0})) \\ &+ 1[x_{i3}'\beta = x_{i4}'\beta](y_{i3} - y_{i2}) \cdot sgn(x_{i32}'\beta + r(y_{i4} - y_{i1}))\} . \end{aligned} \tag{3.3}$$

Expression (3.3) is the sample analogue of $Q_2(r;\beta)$ in Proposition 2.3 after taking the union of events A and B for all possible values of $d_0, d_1, ..., d_4$. Similar to objective function (3.2), (3.3) also

[15] If one magically knew β, the estimation of γ only requires $T = 3$, i.e., using the first line of (3.3) as objective function.

uses only data on "switchers" (i.e., satisfying $A \cup B$), who make different choices in the two periods that are compared. Besides, (3.3) also requires a match in $x_t'\beta$.

Note that this estimator is not feasible because β is unknown and it is of probability zero to have exactly matched indexes ($x_{is}'\beta = x_{it}'\beta$) in the presence of continuous regressors. To resolve the first concern, we propose to replace the unknown parameter β in expression (3.3) with the $\hat{\beta}$ obtained from (3.1), which will be shown to be (cube-root n) consistent in Section 4.

For the second concern, we use kernel weights

$$\mathcal{K}_{h_n}((x_{it} - x_{is})'b), \text{ for all } s, t \in \mathcal{T} \text{ and } b \in \mathcal{B},$$

instead of $1[x_{is}'b = x_{it}'b]$. $\mathcal{K}_{h_n}(\cdot)$ is defined as $h_n^{-1}\mathcal{K}(\cdot/h_n)$, where $\mathcal{K}(\cdot)$ is a kernel density function and h_n is a bandwidth sequence that converges to 0 as $n \to \infty$. The idea is to replace the binary weights for $x_{is}'\hat{\beta} = x_{it}'\hat{\beta}$ in expression (3.3) with weights that depend inversely on the magnitude of $(x_{it} - x_{is})'\hat{\beta}$, giving more weights to observations with $(x_{it} - x_{is})'\hat{\beta}$ being closer to 0.

Then we propose the following kernel weighted MS estimator $\hat{\gamma}$ of γ:

$$\hat{\gamma} = \arg\max_{r \in \mathcal{R}} Q_{2n}^K(r; \hat{\beta}), \tag{3.4}$$

where

$$\begin{aligned} Q_{2n}^K(r; \hat{\beta}) =& \frac{1}{n}\sum_{i=1}^{n}\Big\{\mathcal{K}_{h_n}(x_{i32}'\hat{\beta})(y_{i2} - y_{i1}) \cdot sgn(x_{i21}'\hat{\beta} + r(y_{i3} - y_{i0})) \\ &+ \mathcal{K}_{h_n}(x_{i43}'\hat{\beta})(y_{i3} - y_{i2}) \cdot sgn(x_{i32}'\hat{\beta} + r(y_{i4} - y_{i1}))\Big\}. \end{aligned} \tag{3.5}$$

Remark 3.1. Note that objective function (3.5) is associated with population objective function $Q_2(r; \beta)$ in Proposition 2.3, which only uses observations of adjacent time periods. Applying the same idea to population objective function $\tilde{Q}_2(r; \beta)$ yields the following objective function using observations which are not adjacent.

$$\tilde{Q}_{2n}^K(r; \hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} 1[y_{i2} = y_{i4}]\mathcal{K}_{h_n}(x_{i42}'\hat{\beta})(y_{i3} - y_{i1}) \cdot sgn(x_{i31}'\hat{\beta} + r(y_{i2} - y_{i0})).$$

In practice, to make full use of all observations, one can consider using $Q_{2n}^K(r; \hat{\beta}) + \tilde{Q}_{2n}^K(r; \hat{\beta})$ as objective function for the estimation of γ.

Note that the 2SMS procedure described in (3.1)-(3.2) and (3.4)-(3.5) does not require matching each covariate in x_{it} over time as HK did. As a result, the procedure proposed here allows x_{it} to contain regressors, such as time dummies (time-specific intercepts). Further, as only an index of x_{it} needs to be matched, the rates of convergence of $\hat{\beta}$ and $\hat{\gamma}$ are independent of the number of continuous covariates in x_{it}, which is also in contrast to HK's procedure. In view of existing results on the MS estimators (e.g., Manski (1985, 1987), Kim and Pollard (1990), and Seo and Otsu (2018)), we expect the limiting distributions of $\hat{\beta}$ and $\hat{\gamma}$ to be non-Gaussian and their rates of convergence to be $O_P(n^{-1/3})$ and $O_P((nh_n)^{-1/3})$, respectively. Section 4 states sufficient conditions under which these asymptotic properties can be derived.

3.3 Estimation with $T \geq 4$

For the case with $T \geq 4$, estimators for β and γ that make the best use of the data can be obtained as follows. For β, we find $\hat{\beta}$ via maximizing

$$Q_{1n}(b) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t>s+1} 1[y_{is-1} = y_{it-1}]1[y_{is+1} = y_{it+1}](y_{it} - y_{is})sgn\left((x_{it} - x_{is})'b\right).$$

Once $\hat{\beta}$ is obtained, we then move on to estimate γ via maximizing

$$Q_{2n}^{K}(r;\hat{\beta}) + \tilde{Q}_{2n}^{K}(r;\hat{\beta})$$

with respect to r, where

$$Q_{2n}^{K}(r;\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T-1}\mathcal{K}_{h_n}((x_{it+1} - x_{it})'\hat{\beta})(y_{it} - y_{it-1})sgn((x_{it} - x_{it-1})'\hat{\beta} + r(y_{it+1} - y_{it-2}))$$

is for the case with $t = s + 1$, and

$$\tilde{Q}_{2n}^{K}(r;\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{T-3}\sum_{t=s+2}^{T-1}\left\{1[y_{is+1} = y_{it+1}]\mathcal{K}_{h_n}\left((x_{it+1} - x_{is+1})'\hat{\beta}\right)\right.$$
$$\left.\times(y_{it} - y_{is})sgn\left((x_{it} - x_{is})'\hat{\beta} + r(y_{it-1} - y_{is-1})\right)\right\}$$

is for the case with $t > s + 1$.

4 Asymptotic Properties

The estimators proposed in Section 3 are of the same structure and differ only in that they each use a different fraction of observations in the sample. We expect that they have similar asymptotic properties. Therefore, it suffices to show the asymptotics for the estimators in Sections 3.1 and 3.2, for the case $T = 4$. The asymptotic properties of the estimators in Section 3.3 can be derived in a similar way.

As is standard in the literature, such as Kim and Pollard (1990), we start the analysis from introducing modified objective functions for $\hat{\beta}$ and $\hat{\gamma}$. As will become clear, the new objective functions are monotone (linear) transformations of (3.2) and (3.5), respectively. As a result, working with them does not change the values of $\hat{\beta}$ and $\hat{\gamma}$, but can facilitate the derivation process.

First, note that $1[a > 0] = (sgn(a)+1)/2$ for all $a \in \mathbb{R}$, and hence $\hat{\beta}$ can be obtained equivalently from

$$\hat{\beta} = \arg\max_{b\in\mathcal{B}} n^{-1}\sum_{i=1}^{n} 1\left[y_{i0} = y_{i2} = y_{i4}\right](y_{i3} - y_{i1})\, 1\left[x_{i31}'b > 0\right], \tag{4.1}$$

whose objective function is a monotone transformation of (3.2). Similarly, $\hat{\gamma}$ can be obtained alternatively from

$$\hat{\gamma} = \arg\max_{r \in \mathcal{R}} n^{-1} \sum_{i=1}^{n} \left\{ \mathcal{K}_{h_n} \left(x'_{i32} \hat{\beta} \right) (y_{i2} - y_{i1}) \, 1 \left[x'_{i21} \hat{\beta} + r (y_{i3} - y_{i0}) > 0 \right] \right. \\ \left. + \mathcal{K}_{h_n} \left(x'_{i43} \hat{\beta} \right) (y_{i3} - y_{i2}) \, 1 \left[x'_{i32} \hat{\beta} + r (y_{i4} - y_{i1}) > 0 \right] \right\}. \tag{4.2}$$

To further simplify exposition, we introduce some new notations:

$$\xi_i (b) \equiv 1 \left[y_{i0} = y_{i2} = y_{i4} \right] (y_{i3} - y_{i1}) \left(1 \left[x'_{i31} b > 0 \right] - 1 \left[x'_{i31} \beta > 0 \right] \right), \tag{4.3}$$

$$\varsigma_{ni} (r, b) \equiv \mathcal{K}_{h_n} \left(x'_{i32} b \right) (y_{i2} - y_{i1}) \left(1 \left[x'_{i21} b + r (y_{i3} - y_{i0}) > 0 - 1 \left[x'_{i21} \beta + \gamma (y_{i3} - y_{i0}) > 0 \right] \right] \right. \\ \left. + \mathcal{K}_{h_n} \left(x'_{i43} b \right) (y_{i3} - y_{i2}) \left(1 \left[x'_{i32} b + r (y_{i4} - y_{i1}) > 0 \right] - 1 \left[x'_{i32} \beta + \gamma (y_{i4} - y_{i1}) > 0 \right] \right), \right. \tag{4.4}$$

$$Z_{n,1} (\boldsymbol{s}) \equiv n^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s} n^{-1/3} \right),$$

$$Z_{n,2} (s) \equiv (n h_n)^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \varsigma_{ni} \left(\gamma + s (n h_n)^{-1/3}, \beta \right),$$

and

$$\hat{Z}_{n,2} (s) \equiv (n h_n)^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \varsigma_{ni} \left(\gamma + s (n h_n)^{-1/3}, \hat{\beta} \right).$$

Note that the $\boldsymbol{s}$ in $Z_{n,1}(\boldsymbol{s})$ is a $K \times 1$ vector, and the s in $Z_{n,2}(s)$ and $\hat{Z}_{n,2}(s)$ is a scalar.

As adding terms not related to b will not affect the optimization over b, $\hat{\beta}$ obtained from the following objective function is identical to that from (4.1),

$$\hat{\beta} = \arg\max_{b \in \mathcal{B}} n^{-1} \sum_{i=1}^{n} \xi_i (b).$$

For the same reason, $\hat{\gamma}$ can be equivalently obtained from

$$\hat{\gamma} = \arg\max_{r \in \mathcal{R}} n^{-1} \sum_{i=1}^{n} \varsigma_{ni}(r, \hat{\beta}),$$

where we add terms not related to r. The goal of this section is to derive the asymptotic properties of $\hat{\beta}$ and $\hat{\gamma}$ based on these modified objective functions.

The following technical assumptions are needed for the asymptotics of $\hat{\beta}$ and $\hat{\gamma}$.

Assumption 1. The vectors $\left(x_i^T, y_i^T, y_{i0} \right)'$ with $T \geq 4$ are i.i.d. across individuals[16].

[16] Note that our identification strategy is valid for independent but not necessarily identically distributed observations. While stronger than required, assuming a random sample eases the derivation of the asymptotic properties of our proposed estimators.

Assumption 2. $n^{-1}\sum_{i=1}^{n}\xi_i(\hat{\beta}) \geq \max_{b\in\mathcal{B}} n^{-1}\sum_{i=1}^{n}\xi_i(b) - o_P\left(n^{-2/3}\right)$ and $n^{-1}\sum_{i=1}^{n}\varsigma_{ni}(\hat{\gamma},\hat{\beta}) \geq \max_{r\in\mathcal{R}} n^{-1}\sum_{i=1}^{n}\varsigma_{ni}(r,\hat{\beta}) - o_P\left((nh_n)^{-2/3}\right)$.

Assumption 3. The joint density function for α, covariates x^T, and ϵ^T are continuous differentiable. The density function and its first derivatives are uniformly bounded.

Assumption 4. V_1 in expression (B.1) and V_2 in expression (B.3) are finite and negative definite.

Assumption 5. The kernel function $\mathcal{K}(u)$ is nonnegative, symmetric about 0, continuous differentiable, has compact support, and satisfies $\int_{\mathbb{R}}\mathcal{K}(u)\,du = 1$.

Assumption 6. $h_n \to 0$, $nh_n \to \infty$, and $nh_n^4 \to 0$ as $n \to \infty$.

Assumption 2 is standard in the literature and precisely defines our estimator. Assumption 3 is also standard and is made for technical convenience. V_1 in Assumption 4 will be demonstrated to be the Hessian matrix of the expectation of the objective function in equation (4.1). The assumption on V_1 is needed to ensure that the population mean of the objective function at β is sufficiently larger than it at other b around β. This assumption is in fact quite mild; we provide some details in Remark B.2 in Appendix B. V_2 is the second derivative of the expectation of the objective function in equation (4.2) with respect to r. This condition is also mild. Some discussion on V_2 can be found in Remark B.3 in Appendix B. Note that Assumption 4 implicitly requires the finite second moment of x. Assumption 5 collects some standard restrictions on kernel functions. The symmetry of $\mathcal{K}(u)$ makes the bias term from the nonparametric estimation at the order of h_n^2. In Assumption 6, $nh_n \to \infty$ is standard, and $nh_n^4 \to 0$ is made to ensure the bias term from the kernel estimation asymptotically negligible.

Theorem 4.1. *Suppose Assumptions A, SI (or SD), and 1 - 6 hold. Then*

1. $\hat{\beta} - \beta = O_P\left(n^{-1/3}\right)$, *and*

$$n^{1/3}(\hat{\beta} - \beta) \xrightarrow{d} \arg\max_{\boldsymbol{s}\in\mathbb{R}^K} Z_1(\boldsymbol{s}),$$

where $Z_1(\boldsymbol{s})$ is a Gaussian Process with continuous sample paths, expected value $\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}$ and covariance kernel $H_1(\boldsymbol{s},\boldsymbol{t})$. V_1 and H_1 are defined in expressions (B.1) and (B.2), respectively.

2. $\hat{\gamma} - \gamma = O_P((nh_n)^{-1/3})$, *and*

$$(nh_n)^{1/3}(\hat{\gamma} - \gamma) \xrightarrow{d} \arg\max_{s\in\mathbb{R}} Z_2(s).$$

where $Z_2(s)$ is a Gaussian process with continuous path, expected value $\frac{1}{2}V_2 s^2$, covariance kernel $H_2(s,t)$. V_2 and H_2 are defined in expressions (B.3) and (B.4), respectively.

Kim and Pollard (1990) and Seo and Otsu (2018) derived the cube-root asymptotics for a class of estimators by means of empirical processes. For a comprehensive treatment on this technique, see van der Vaart and Wellner (2000). Our estimators fall into this category. In particular, they are more closely related to Seo and Otsu (2018). The main body of the proof for Theorem 4.1 is to verify the technical conditions in Seo and Otsu (2018), apply their asymptotics results to our estimators, and calculate technical terms needed for the asymptotics such as V_1, H_1, V_2, and H_2.

As will be demonstrated later, the asymptotics of $\hat{\gamma}$ are the same as that in the case where the true value of β is used. The intuition is that $\hat{\beta}$ converges to β faster than $\hat{\gamma}$ does to γ, and the objective function in equation (4.2), after proper normalization, uniformly converges to the limit over a compact set of (r, b) around (γ, β) . The details can be found in the proof of Theorem 4.1, which is presented in Appendix B.

5 Inference

The asymptotic distributions of $\hat{\beta}$ and $\hat{\gamma}$ are complicated and do not have an analytical form. As a result, inference using the asymptotic distribution directly is hard to be implemented. One alternative is to use sampling methods (e.g., bootstrap). Unfortunately, Abrevaya and Huang (2005) have proved the inconsistency of the classic bootstrap for the maximum score estimators. We expect that the classic bootstrap does not work for our estimators, either.

For the ordinary maximum score estimator, valid inference can be conducted using subsampling (Delgado, Rodríguez-Poo, and Wolf (2001)), m-out-of-n bootstrap (Lee and Pun (2006)), the numerical bootstrap (Hong and Li (2020)), and a model-based bootstrap procedure that analytically modifies the criterion function (Cattaneo, Jansson, and Nagasawa (2017)), among other procedures[17]. We think these methods, with certain modifications, can be justified to be valid for our estimators.

Monte Carlo evidence demonstrated in Hong and Li (2020) and Cattaneo, Jansson, and Nagasawa (2017) suggests that their proposed approaches outperform either the subsampling or the m-out-of-n bootstrap in finite samples. Based on these results, we focus on the numerical bootstrap and the bootstrap procedure with a modified criterion function. We provide a brief discussion on the classic bootstrap and the m-out-of-n bootstrap in Appendix E[18].

Some new notations are introduced. Let $(y_j^{T*\prime}, x_j^{T*\prime})'$, $j = 1, ..., n$, be a random sample drawn with replacement from the collection of the sample values $\left(y_1^{T\prime}, x_1^{T\prime}\right)'$, $\left(y_2^{T\prime}, x_2^{T\prime}\right)'$, ..., $\left(y_n^{T\prime}, x_n^{T\prime}\right)'$. Let $\xi_j^*(b)$ denote $\xi(b)$ evaluated at $(y_j^{T*\prime}, x_j^{T*\prime})'$, specifically,

$$\xi_j^*(b) \equiv 1\left[y_{j0}^* = y_{j2}^* = y_{j4}^*\right]\left(y_{j3}^* - y_{j1}^*\right)\left(1\left[x_{j31}^{*\prime} b > 0\right] - 1\left[x_{j31}^{*\prime}\beta > 0\right]\right).$$

[17] The case-specific, smooth bootstrap method proposed by Patra, Seijo, and Sen (2018) is also valid for the maximum score estimator of Manski (1975, 1985). But this method is hard to generalize to our case.

[18] We show in Appendix E that the classic bootstrap is not consistent for our estimators (Appendix E.3), while the m-out-of-n bootstrap is still valid (Appendix E.4).

Similarly, we define $\varsigma^*_{nj}(r,b)$ as

$$\begin{aligned}\varsigma^*_{nj}(r,b) \equiv\ & \mathcal{K}_{h_n}\left(x^{*\prime}_{j32}b\right)\left(y^*_{j2}-y^*_{j1}\right)\left(1\left[x^{*\prime}_{j21}b+r\left(y^*_{j3}-y^*_{j0}\right)>0\right]-1\left[x^{*\prime}_{j21}\beta+\gamma\left(y^*_{j3}-y^*_{j0}\right)>0\right]\right)\\ &+\mathcal{K}_{h_n}\left(x^{*\prime}_{j43}b\right)\left(y^*_{j3}-y^*_{j2}\right)\left(1\left[x^{*\prime}_{j32}b+r\left(y^*_{j4}-y^*_{j1}\right)>0\right]-1\left[x^{*\prime}_{j32}\beta+\gamma\left(y^*_{j4}-y^*_{j1}\right)>0\right]\right).\end{aligned}$$

We may re-use some of these notations in the discussions. To avoid confusion, all notations in each subsection are specific for the procedure discussed in that subsection. This convention also applies to the discussions in Appendix E.

5.1 Numerical Bootstrap

Hong and Li (2020) developed a numerical bootstrap procedure for cases where the classic bootstrap does not work. Hong and Li (2020) demonstrated that their method could work for a class of M-estimators that converge at rate n^a for some $a \in (1/4, 1]$. The estimator $\hat{\beta}$ proposed in Section 3 fits in their framework directly, but $\hat{\gamma}$ does not. With a slight modification of their proof, we show that the numerical bootstrap also works for $\hat{\gamma}$.

The numerically bootstraped $\hat{\beta}^*$ and $\hat{\gamma}^*$ are constructed from

$$\hat{\beta}^* = \arg\max_{b\in\mathcal{B}}\left\{n^{-1}\sum_{i=1}^{n}\xi_i(b)+(n\varepsilon_n)^{1/2}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi^*_j(b)-n^{-1}\sum_{i=1}^{n}\xi_i(b)\right)\right\} \tag{5.1}$$

and

$$\hat{\gamma}^* = \arg\max_{r\in\mathcal{R}}\left\{n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(r,\hat{\beta}\right)+(n\varepsilon_n)^{1/2}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma^*_{nj}\left(r,\hat{\beta}\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(r,\hat{\beta}\right)\right)\right\}, \tag{5.2}$$

where $\varepsilon_n \to 0$, $n\varepsilon_n \to \infty$, and $(y^{T*\prime}_j, x^{T*\prime}_j)'$, $j = 1, ..., n$, are drawn independently from the collection of the sample values $\left(y^{T\prime}_1, x^{T\prime}_1\right)'$, $\left(y^{T\prime}_2, x^{T\prime}_2\right)'$, ..., $\left(y^{T\prime}_n, x^{T\prime}_n\right)'$ with replacement. ε^{-1}_n plays a similar role as m in the m-out-of-n bootstrap procedure. For $\hat{\gamma}^*$, we additionally require $\varepsilon^{-1}_n h_n \to \infty$ and $\varepsilon^{-1}_n h^4_n \to 0$, similar to the additional restrictions on m.

We claim that

$$\varepsilon_n^{-1/3}\left(\hat{\beta}^*-\hat{\beta}\right)\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})\right)$$

and

$$\left(\varepsilon_n^{-1}h_n\right)^{1/3}\left(\hat{\gamma}^*-\hat{\gamma}\right)\xrightarrow{d}\arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_2 s^2+W_2(s)\right).$$

Some intuition on why the numerical bootstrap works and the way to modify the proof in Hong and Li (2020) to accommodate $\hat{\gamma}$ is provided in Appendix E.1.

5.2 Bootstrap Using a Modified Objective Function

Cattaneo, Jansson, and Nagasawa (2017) proposed a valid bootstrap procedure for the maximum score estimator by means of modifying the objective function. This idea can be applied here analogously. The procedure is as follows. Draw $(y_j^{T*\prime}, x_j^{T*\prime})'$, $j = 1, ..., n$, independently from the collection of the sample values $\left(y_1^{T\prime}, x_1^{T\prime}\right)'$, $\left(y_2^{T\prime}, x_2^{T\prime}\right)'$, ..., $\left(y_n^{T\prime}, x_n^{T\prime}\right)'$ with replacement. The bootstrap estimator $\hat{\beta}^*$ is obtained from

$$\hat{\beta}^* = \arg\max_{b\in\mathcal{B}} \left\{ n^{-1}\sum_{j=1}^{n} \xi_j^*(b) - n^{-1}\sum_{i=1}^{n}\xi_i(b) + \frac{1}{2}\left(b-\hat{\beta}\right)'\hat{V}_{n,1}\left(b-\hat{\beta}\right)\right\}, \tag{5.3}$$

where $\hat{V}_{n,1}$ is a consistent estimate of V_1.

Recall that V_1 is the second derivative of $\left.\frac{\partial\mathbb{E}(\xi_i(b))}{\partial b\partial b'}\right|_{b=\beta}$. $\hat{V}_{n,1}$ can be estimated by numerical derivatives. For instance, the (k,l)-th element of $\hat{V}_{n,1}$ can be obtained by

$$\begin{aligned}\hat{V}_{n,1}^{(k,l)} = \frac{1}{4\omega_n^2} n^{-1}\sum_{i=1}^{n} &\left(\xi_i\left(\hat{\beta}+\omega_n e_k+\omega_n e_l\right) - \xi_i\left(\hat{\beta}+\omega_n e_k-\omega_n e_l\right)\right.\\ &\left.-\xi_i\left(\hat{\beta}-\omega_n e_k+\omega_n e_l\right)+\xi_i\left(\hat{\beta}-\omega_n e_k-\omega_n e_l\right)\right),\end{aligned} \tag{5.4}$$

where e_k is a $K\times 1$ vector with its k-th element being 1 and 0 otherwise, and e_l is similarly defined. The result of Lemma 1 in Cattaneo, Jansson, and Nagasawa (2017) implies $\hat{V}_{n,1} \xrightarrow{P} V_1$ under $\omega_n \to 0$ and $n\omega_n^3 \to \infty$.

Similarly, the bootstrap estimator $\hat{\gamma}^*$ is obtained from

$$\hat{\gamma}^* = \arg\max_{r\in\mathcal{R}} \left\{ n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*\left(r,\hat{\beta}\right) - n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(r,\hat{\beta}\right) + \frac{1}{2}\hat{V}_{n,2}\left(r-\hat{\gamma}\right)^2\right\}, \tag{5.5}$$

where $\hat{V}_{n,2}$ is a consistent estimate of V_2. Recall that V_2 is $\left.\frac{\partial^2\mathbb{E}(\varsigma_{ni}(r,\beta))}{\partial r^2}\right|_{r=\gamma}$. Then an estimate of V_2 can be

$$\hat{V}_{n,2} = \frac{1}{4\omega_n^2}n^{-1}\sum_{i=1}^{n}\left(\varsigma_{ni}\left(\hat{\gamma}+2\omega_n,\hat{\beta}\right) - 2\varsigma_{ni}\left(\hat{\gamma},\hat{\beta}\right) + \varsigma_{ni}\left(\hat{\gamma}-2\omega_n,\hat{\beta}\right)\right). \tag{5.6}$$

It similarly requires $\omega_n \to 0$ and $nh_n\omega_n^3 \to \infty$ to guarantee the consistency of $\hat{V}_{n,2}$.

Then, we can show that

$$n^{1/3}\left(\hat{\beta}^*-\hat{\beta}\right) \xrightarrow{d} \arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1(\boldsymbol{s})\right),$$

and

$$(nh_n)^{1/3}\left(\hat{\gamma}^*-\hat{\gamma}\right) \xrightarrow{d} \arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_2 s^2 + W_2(s)\right).$$

An outline of the proof can be found in Appendix E.2.

5.3 Procedures in Details

We investigate the finite-sample properties of the two bootstrap methods discussed in Sections 5.1 and 5.2 through Monte Carlo experiments in Section 6 and defer the discussion on the choices of their tuning parameters to Section 6.2. Here we provide the algorithms for constructing 95% confidence intervals for β and γ using these two procedures.

The numerical bootstrap proceeds as follows.

1. Draw $(y_j^{T*\prime}, x_j^{T*\prime})'$, $j = 1, ..., n$, independently with replacement from the original sample.
2. Obtain $\hat{\beta}^*$ and $\hat{\gamma}^*$ from equations (5.1) and (5.2).
3. Repeat Steps 1 and 2 for B times independently and arrive at a sequence of $(\hat{\beta}^*, \hat{\gamma}^*)$, say, $\{(\hat{\beta}^{*(b)}, \hat{\gamma}^{*(b)})\}_{b=1}^B$.
4. Let $Q_{\hat{\beta}^*}(\tau)$ denote the τ-th quantile of $\{\hat{\beta}^{*(b)}\}_{b=1}^B$, $0 \le \tau \le 1$. Define $Q_{\hat{\gamma}^*}(\tau)$ analogously. The 95% confidence intervals for β and γ are constructed, respectively, as

$$\left[\hat{\beta} - n^{-1/3} \cdot \varepsilon_n^{-1/3}(Q_{\hat{\beta}^*}(0.975) - \hat{\beta}),\ \hat{\beta} - n^{-1/3} \cdot \varepsilon_n^{-1/3}(Q_{\hat{\beta}^*}(0.025) - \hat{\beta})\right]$$

and

$$\left[\hat{\gamma} - n^{-1/3} \cdot \varepsilon_n^{-1/3}\left(Q_{\hat{\gamma}^*}(0.975) - \hat{\gamma}\right),\ \hat{\gamma} - n^{-1/3} \cdot \varepsilon_n^{-1/3}\left(Q_{\hat{\gamma}^*}(0.025) - \hat{\gamma}\right)\right].$$

The bootstrap procedure with a modified objective function is as follows.

1. Estimate $\hat{V}_{n,1}$ and $\hat{V}_{n,2}$ based on expressions (5.4) and (5.6).
2. Draw $(y_j^{T*\prime}, x_j^{T*\prime})'$, $j = 1, ..., n$, independently with replacement from the original sample.
3. Obtain $\hat{\beta}^*$ and $\hat{\gamma}^*$ from equations (5.3) and (5.5), using the $(\hat{V}_{n,1}, \hat{V}_{n,2})$ obtained from Step 1.
4. Repeat Step 2 and Step 3 B times independently and arrive at a sequence of $(\hat{\beta}^*, \hat{\gamma}^*)$, say, $\{(\hat{\beta}^{*(b)}, \hat{\gamma}^{*(b)})\}_{b=1}^B$.
5. The 95% confidence intervals for β and γ are constructed, respectively, as

$$[\hat{\beta} - (Q_{\hat{\beta}^*}(0.975) - \hat{\beta}),\ \hat{\beta} - (Q_{\hat{\beta}^*}(0.025) - \hat{\beta})]$$

and

$$[\hat{\gamma} - (Q_{\hat{\gamma}^*}(0.975) - \hat{\gamma}),\ \hat{\gamma} - (Q_{\hat{\gamma}^*}(0.025) - \hat{\gamma})].$$

6 Monte Carlo Experiments

6.1 Simulation Setup

In this section, we investigate the finite-sample performance of the proposed estimators by means of Monte Carlo experiments. We start by considering a benchmark design similar to that used in HK. Specifically, this design (referred to as Design 1 hereafter) is specified as follows:

$$y_{i0} = 1\left[\beta_1 x_{i0,1} + \beta_2 x_{i0,2} + \alpha_i - \epsilon_{i0} > 0\right],$$
$$y_{it} = 1\left[\beta_1 x_{it,1} + \beta_2 x_{it,2} + \gamma y_{it-1} + \alpha_i - \epsilon_{it} > 0\right], \quad t \in \{1,2,3,4\},$$

where

- $\beta \equiv (\beta_1, \beta_2)' = (1,1)'$ and $\gamma = -1$,
- $x_{it,1}, x_{it,2} \overset{d}{\sim} N(0,1)$ and are i.i.d. across i and t,
- $\alpha_i = (x_{i0,2} + x_{i1,2} + x_{i2,2} + x_{i3,2} + x_{i4,2})/5$,
- $\epsilon_{it} \overset{d}{\sim} \left(\pi^2/3\right)^{-1/2} \cdot \text{Logistic}(0,1)$ and are i.i.d. across i and t, and
- $x_{\cdot,1}, x_{\cdot,2}$, and $\epsilon_{\cdot}$ are independent of each other.

In the second design (referred to as Design 2 hereafter), the model and the coefficients are the same as in Design 1, but $x_{\cdot,2}$ are autocorrelated over time. Specifically, we have

- $x_{i0,2} \overset{d}{\sim} N(0,1)$ and $x_{it,2} = 0.5x_{it-1,2} + u_{it}$ for all $t \geq 1$, where $u_{it} \overset{d}{\sim} N(0,1)$ and u_{it} are i.i.d. across i and t, and
- $u_{\cdot}$, $x_{\cdot,1}$, $x_{i0,2}$, and $\epsilon_{\cdot}$ are independent of each other.

Note that the setup of Design 2 violates either Assumption SI or the exchangeability condition stated in Proposition 2.1. We conduct this Monte Carlo study to develop more insight into the practical consequences of the failure of these sufficient (but not necessary) conditions. That is, to what extent serial dependence in exogenous covariates may affect the identification.

In the third design (referred to as Design 3 hereafter), the setup is the same as that in Design 1, except we add one more covariate to examine how our estimators perform in a higher dimensional, more complicated design. Specifically,

$$y_{i0} = 1\left[\beta_1 x_{i0,1} + \beta_2 x_{i0,2} + \beta_3 x_{i0,3} + \alpha_i - \epsilon_{i0} > 0\right],$$
$$y_{it} = 1\left[\beta_1 x_{it,1} + \beta_2 x_{it,2} + \beta_3 x_{it,3} + \gamma y_{it-1} + \alpha_i - \epsilon_{it} > 0\right], \quad t \in \{1,2,3,4\},$$

where

- $\beta \equiv (\beta_1, \beta_2, \beta_3)' = (1, 1, 1)'$ and $\gamma = -1$,

- $x_{it,1}, x_{it,2}, x_{it,3} \overset{d}{\sim} N(0, 1)$ are i.i.d. across i and t, and

- $x_{\cdot,1}, x_{\cdot,2}, x_{\cdot,3}$, and $\epsilon_{\cdot}$ are independent of each other.

For estimation of β, we adopt objective function (3.2). To estimate γ, we use objective function (3.5) with the Epanechnikov kernel function. That is,

$$\mathcal{K}(u) = \frac{3}{4}\left(1 - u^2\right) 1\left[|u| \leq 1\right],$$

which satisfies Assumption 5 with a compact support. The choice of bandwidth sequence h_n will be discussed in Section 6.2.

For inference, we investigate the finite-sample performance of the numerical bootstrap (Section 5.1) and the bootstrap with a modified objective function (Section 5.2). The 95% confidence intervals are obtained through $B = 199$ independent draws and estimations for both procedures. See Section 5.3 for the details of the implementation.

Recall that only the observations with $\{y_{i0} = y_{i2} = y_{i4}$ and $y_{i1} \neq y_{i3}\}$ are used for estimating $\hat{\beta}$. In Design 1, the "effective" observations, that are useful for estimating $\hat{\beta}$, are about 14% of the whole sample. Similarly for $\hat{\gamma}$, only observations with either $\{y_{i1} \neq y_{i2}$ and $y_{i0} \neq y_{i3}\}$ or $\{y_{i2} \neq y_{i3}$ and $y_{i1} \neq y_{i4}\}$ are useful. In Design 1, about 39% of the observations are "effective" for $\hat{\gamma}$. In Design 2, the "effective" observations take about 15% and 31% of the original sample, for $\hat{\beta}$ and $\hat{\gamma}$, respectively. In Design 3, the proportions of the "effective" observations for estimating $\hat{\beta}$ and $\hat{\gamma}$ are about 14% and 39%, respectively. For each design, we consider sample sizes of 5000, 10000, and 20000. All the estimation and inference (based on 199 draws and estimation) results presented in this section are based on 1000 replications of each design and each sample size.

6.2 Tuning Parameters

There is only one tuning parameter used for estimation, namely h_n, in objective function (3.5). In Assumption 6, we restrict $nh_n^4 \to 0$, so that the bias term (of order h_n^2) is a small order term of $(nh_n)^{-2/3}$. Since the convergence rate of $\hat{\gamma}$ is $(nh_n)^{-1/3}$, the condition, $nh_n^4 \to 0$, makes the bias term much smaller than the convergence rate. To attain a faster convergence rate, we tend to set h_n as large as possible, and thus we simply set $h_n = n^{-1/4}(\log n)^{-1}$.

For the numerical bootstrap, we have one more tuning parameter ε_n. As recommended in Hong and Li (2020), we set ε_n proportional to $n^{-2/3}\log n$ for the inferences of $\hat{\beta}$ and $\hat{\gamma}$. Apparently, ε_n of this order satisfies the additional requirements for $\hat{\gamma}^*$ that $\varepsilon_n^{-1}h_n \to \infty$ and $\varepsilon_n^{-1}h_n^4 \to 0$. To check how sensitive the procedure to the choice of ε_n is, we conduct the procedure with $\varepsilon_n = c \cdot n^{-2/3}\log n$ and $c = 0.8, 0.9, 1.0, 1.1,$ and 1.2.

The bootstrap procedure with a modified objective function in Section 5.2 also requires one additional tuning parameter, ω_n, to estimate the kernel variance V. As demonstrated in Cattaneo, Jansson, and Nagasawa (2017), the optimal ω_n is proportional to the convergence rate to $3/7$. It leads to $\omega_n \propto n^{-1/7}$ for $\hat{\beta}_n^*$ and $\omega_n \propto n^{-3/28} (\log n)^{1/7}$ for $\hat{\gamma}_n^*$. Our Monte Carlo results suggest that a slight modification by setting $\omega_n = c \cdot n^{-1/7} \log n$ for $\hat{\beta}^*$ leads to a better performance. We adopt this modification for $\hat{\beta}^*$. Meanwhile, we follow the recommendation $\omega_n = c \cdot n^{-3/28} (\log n)^{1/7}$ for $\hat{\gamma}_n^*$. Again, to check the sensitivity of the results to the choice of ω_n, we conduct this procedure with $c = 0.8, 0.9, 1.0, 1.1$, and 1.2.

6.3 Simulation Results

We normalize the preference coefficients β on exogenous covariates to 1 in Euclidean norm. The true values of the parameters, due to this scale normalization, are

$$\beta_1 = \beta_2 = \frac{\sqrt{2}}{2} \approx 0.707 \text{ and } \gamma = -\frac{\sqrt{2}}{2} \approx -0.707$$

in Design 1 and Design 2, and

$$\beta_1 = \beta_2 = \beta_3 = \frac{\sqrt{3}}{3} \approx 0.577 \text{ and } \gamma = -\frac{\sqrt{3}}{3} \approx -0.577,$$

in Design 3. Because of this normalization, we lose one degree of freedom and essentially only estimate one element of β in the first two designs and two elements of β in the last design. As a result, we only report the results for (β_2, γ) in Designs 1 and 2 and the results for $(\beta_2, \beta_3, \gamma)$ in Design 3.

We report the mean (MEAN), the mean bias (BIAS), the median absolute deviation (MAD), and the root mean squared error (RMSE) for $\hat{\beta}$ and $\hat{\gamma}$. For inference, we report the coverage rates (COVERAGE) of the true values and lengths (LENGTH) of the 95% confidence intervals (CI) for both inference procedures.

All results are reported in the tables collected in Appendix C. Results for Design 1 are reported in tables numbered "1" and so on and so forth, for other designs. We report the performance of the estimators, the numerical bootstrap procedure, and the bootstrap procedure with a modified objective function in tables labeled "A", "B", and "C", respectively. For example, Table 1A reports the performance of the estimators for Design 1. We briefly summarize our findings as follows.

The RMSEs of $\hat{\beta}$ and $\hat{\gamma}$ become smaller as the sample size increases in all designs, with RMSE of $\hat{\gamma}$ slightly greater than that of $\hat{\beta}$. This shows the consistency of our estimators, though the rates of convergence are clearly slower than $\sqrt{n}$. Both inference procedures perform reasonably well in all designs. In general, they both yield shrinking CIs with coverage rates approaching 95% as the sample size grows. The coverage rates of CI are greater than 90%, but are slightly lower than 95% in most cases. The coverage rates of CI for γ do not perform as well as those for β, which is not surprising, considering the complication of using two tuning parameters. Neither procedure is very sensitive to the choice of tuning parameters, particularly the numerical bootstrap procedure.

The results for Design 1 (benchmark) and Design 2 are similar. A more careful inspection of the numbers in Tables 2A - 2C reveals that our estimators are slightly biased but in general still work adequately in Design 2, especially for the estimator of γ[19]. In practice, if the robustness and efficiency are considered comprehensively, the complementary use of our approach combined with HK's parametric (conditional Logit) and semiparametric methods could be a useful strategy. The results for Design 3 are also encouraging, demonstrating that our estimators do not suffer in a higher-dimensional design. This points to one of the main advantages of using our methods – the rates of convergence of our estimators are independent of the number of continuous regressors.

Only data on "switchers" provides variation useful for identification, so the effective sample used for estimation is typically a small fraction of the total sample. Intuitively, we would expect that longer panels would improve the finite-sample performance of our estimators. In practice, with a reasonably long panel, our estimators may not need a very large sample size n to produce sufficiently precise estimates, as illustrated in the next section.

7 Empirical Illustration

We apply our estimator to analyze the state dependence and the short-run and long-run response to health shocks on labor market participation, using 15 waves (wave 2 to wave 16) of the Household, Income and labor Dynamics in Australia (HILDA) Survey data. The HILDA Survey follows the lives of more than 17,000 Australians each year. As a rich panel data set, it collects information on many aspects of life in Australia[20].

Damrongplasit, Hsiao, and Zhao (2018) investigated the same problem using eight waves of HILDA data, by means of random effects estimators and HK's fixed effects conditional Logit estimator. We use a more recent data set with more waves. The background to the empirical question has been thoroughly introduced in Damrongplasit, Hsiao, and Zhao (2018), so is omitted here. We simply refer interested readers to Damrongplasit, Hsiao, and Zhao (2018) and references therein for more detailed information.

The model setup for the application is as follows:

$$y_{it} = 1\left[x_{it}'\beta + \gamma y_{it-1} + \alpha_i - \epsilon_{it} > 0\right].$$

The dependent variable y_{it} is labor force participation of individual i in period (wave) t. We are interested in the effects of "Health Shock" (short-run health shocks, HS_{it}), "Activity Limiting Con-

[19]For β, it seems that our estimator generates an upward bias of approximately 0.025, which is small relative to the true value of the coefficient.

[20]More details about the data set can be found at "https://melbourneinstitute.unimelb.edu.au/hilda".

dition" (long-run health shocks, ACL_{it}), and the state persistence (y_{it-1}) on the labor force participation (y_{it}) . Following the example set by Damrongplasit, Hsiao, and Zhao (2018)[21], we also include "Unemployment Rate" (UR_t) and "log(Income)" (log of household income, I_{it}) as control variables. Collectively, we have $x_{it} = (\text{HS}_{it}, \text{ACL}_{it}, \text{UR}_t, I_{it})'$, in which I_{it} is a continuous regressor with rich enough support required for identification. Explanation and summary statistics of all observed variables are provided in Tables 4 and 5 of Appendix C, respectively.

After dropping data with missing information, the sample consists of 6,848 individual males and 7,927 individual females. The panel is unbalanced, with each individual being observed in up to 15 waves (i.e., $T = 14$ using the notation in previous sections). In total, there are 42,416 male observations and 48,121 female observations.

For ease of implementation, we adopt the estimator in Section 3.3 wherein we only use adjacent observations for each individual[22]. The conditions we impose for observations to be useful for the estimation are stringent. Missing observations here and there make it even worse. Observations on only 755 individual males and 1,319 individual females are useful for estimating β. That is about 14% of the original sample. Observations on 1,278 individual males and 2,074 individual females (about 23% of the original sample) are useful for estimating γ.

We conduct estimation on the whole sample, the sample of males, and the sample of females, in order. The estimates are reported in Table 6 of Appendix C. We conduct the inference using the two procedures described in Section 5. The 95% CIs obtained from both procedures are reported, along with the estimates. The 95% CIs obtained from the numerical bootstrap procedure are put on top of the 95% CIs from the bootstrap with a modified objective function for each coefficient in the same table. The tuning parameters are set as in Section 6.2 with $c = 1$.

If a 95% CI does not cover 0, we label "**" next to the CI in the table, indicating that the corresponding estimate is significantly different from 0 at the 5% level, based on the CI. The two inference procedures agree in almost all cases in terms of signs and whether significant or not. The only exception is for the coefficient on "log(Income)" for the whole sample.

We restrict our attention to those estimates where both inference procedures do agree. First, the estimated coefficients on the lagged labor force participation are significantly positive for all three samples. This indicates that the labor force participation decision is sticky over time, so ignoring the state dependence may lead to mis-specification of the model. The estimated coefficients on "Health Shock" are significantly negative for all samples. This implies that temporary deterioration in health status does have a negative effect on labor force participation, which is not surprising at all. The estimated coefficients before "log(Income)" are only significant (and positive) for the male sample. The result is also in line with our economic intuition. Family income is more likely to be a crucial determinant of labor force participation for males, rather than females.

[21] Damrongplasit, Hsiao, and Zhao (2018) included time-invariant regressors, e.g., gender, in the random effects estimation. These covariates are excluded here as their coefficients cannot be identified in a fixed effects framework.

[22] Namely, we only use observations with $t = s + 2$ to estimate β and observations with $t = s + 1$ to estimate γ.

The estimated coefficients before the long-run health shock ("Activity Limiting Condition") are not significant, at 5% level for all three samples. This suggests that permanent changes in health status may not have much effect on the decision to switch from working to not-working, or the other way around. The estimates on "Unemployment Rate" are not significant for all samples. This basically means that an individual's working decision is somewhat independent of the unemployment rate, a macro-level variable indicating general conditions in the labor market.

Our results agree with the fixed effects estimates in Damrongplasit, Hsiao, and Zhao (2018), in terms of the signs of the preference coefficients. However, the magnitudes of these coefficients are very different. For example, their results indicate that the effect of "Health Shock" on an individual's labor force participation is much smaller in magnitude than that of the state dependence (less than 1/3), while our results show that their effects are quite comparable in magnitude. We note that their estimates were obtained using HK's parametric (conditional Logit) estimator, which might suffer from mis-specification. Our results can be thought of as a robust check of their results. Damrongplasit, Hsiao, and Zhao (2018) also obtained the random effects estimates, and demonstrated that the fixed effects estimates are more reasonable. We conjecture that one could arrive at similar random effects estimates using our samples.

As a final note, recall that the preference coefficients are only identified up to scale. The magnitudes of these coefficients are difficult to interpret. Our semiparametric estimates are most useful if several coefficients are included in the regression and coefficient estimates are computed to compare the relative effects of changes in regressors on the choice.

8 Conclusions

In this paper, we provide new identification results for preference parameters in panel data binary choice models that allow for both fixed effects (Heckman's "spurious" state dependence) and lagged dependent variables ("true" state dependence). The same semiparametric random utility framework as in Honoré and Kyriazidou (2000) is considered. A key, novel idea in this paper is that, with additional restrictions on the dynamic process of observed covariates and the tail behavior of the error distribution, the point identification no longer needs element-by-element matching of regressors over time, in contrast to the method proposed in Honoré and Kyriazidou (2000). Our approach assumes a minimum panel length of five ($T \geq 4$), which fits in most empirical settings. Our identification arguments motivate a two-step estimation procedure, adapting Manski's maximum score estimator. The proposed estimators are consistent with rates of convergence independent of the model dimension, as opposed to the estimator proposed in Honoré and Kyriazidou (2000). We further derive limiting distributions of the proposed estimators, which are non-Gaussian, in line with existing literature. We justify the application of several bootstrap procedures for making inference. The results of a small Monte Carlo study suggest that our estimators and inference procedures perform well in finite samples. We apply the proposed approach to study labor force participation using HILDA data.

The work here leaves some open questions for future research. For example, one may consider smoothing the objective functions (in the spirit of Horowitz (1992)) to attain faster rates of convergence and asymptotic normality. One may also consider extending the framework in this paper to study the identification with more than one lag of the dependent variable or the identification in panel data multinomial response models.

References

ABREVAYA, J. AND J. HUANG (2005): "On the bootstrap of the maximum score estimator," *Econometrica*, 73, 1175–1204.

AGUIRREGABIRIA, V., J. GU, AND Y. LUO (2018): "Sufficient Statistics for Unobserved Heterogeneity in Structural Dynamic Logit Models," *arXiv preprint arXiv:1805.04048*.

ALESSIE, R., S. HOCHGUERTEL, AND A. V. SOEST (2004): "Ownership of stocks and mutual funds: A panel data analysis," *Review of Economics and Statistics*, 86, 783–796.

ALTONJI, J. G. AND R. L. MATZKIN (2005): "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica*, 73, 1053–1102.

ANDERSEN, E. B. (1970): "Asymptotic properties of conditional maximum likelihood estimators," *Journal of the Royal Statistical Society. Series B (Methodological)*, 32, 283–301.

ARELLANO, M. AND R. CARRASCO (2003): "Binary choice panel data models with predetermined variables," *Journal of Econometrics*, 115, 125–157.

BARTOLUCCI, F. AND V. NIGRO (2010): "A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator," *Econometrica*, 78, 719–733.

——— (2012): "Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data," *Journal of Econometrics*, 170, 102–116.

BIEWEN, M. (2009): "Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition," *Journal of Applied Econometrics*, 24, 1095–1116.

CAMERON, S. V. AND J. J. HECKMAN (1998): "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males," *Journal of Political economy*, 106, 262–333.

——— (2001): "The dynamics of educational attainment for black, hispanic, and white males," *Journal of Political Economy*, 109, 455–499.

CATTANEO, M. D., M. JANSSON, AND K. NAGASAWA (2017): "Bootstrap-based inference for cube root consistent estimators," *arXiv preprint arXiv:1704.08066*.

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

——— (2010): "Binary response models for panel data: identification and information," *Econometrica*, 78, 159–168.

CHAY, K. Y., H. W. HOYNES, AND D. HYSLOP (1999): "A non-experimental analysis of true state dependence in monthly welfare participation sequences," in *American Statistical Association*, 9–17.

CHEN, S., S. KHAN, AND X. TANG (2018): "Exclusion restrictions in dynamic binary choice panel data models," Tech. rep., Boston College Department of Economics.

——— (2019): "Exclusion Restrictions in Dynamic Binary Choice Panel Data Models: Comment on "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors"," *Econometrica*, 87, 1781–1785.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, J. HAHN, AND W. NEWEY (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81, 535–580.

CHINTAGUNTA, P., E. KYRIAZIDOU, AND J. PERKTOLD (2001): "Panel data analysis of household brand choices," *Journal of Econometrics*, 103, 111–153.

CONTOYANNIS, P., A. M. JONES, AND N. RICE (2004): "Simulation-based inference in dynamic panel probit models: An application to health," *Empirical Economics*, 29, 49–77.

CORCORAN, M. AND M. S. HILL (1985): "Reoccurrence of unemployment among adult men," *The Journal of Human Resources*, 20, 165–183.

DAMRONGPLASIT, K., C. HSIAO, AND X. ZHAO (2018): "Health status and labour market outcome: Empirical evidence from Australia," *Pacific Economic Review*, 1–24.

DELGADO, M., J. RODRÍGUEZ-POO, AND M. WOLF (2001): "Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator," *Economics Letters*, 73, 241–250.

HAHN, J. (2001): "The information bound of a dynamic panel logit model with fixed effects," *Econometric Theory*, 17, 913–932.

HALLIDAY, T. J. (2008): "Heterogeneity, state dependence and health," *The Econometrics Journal*, 11, 499–516.

HECKMAN, J. J. (1981a): "The incidental parameters problem and the problem of initial condition in estimating a discrete time-discrete data stochastic process," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski and D. L. McFadden, MIT Press.

——— (1981b): "Statistical models for discrete panel data," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski and D. L. McFadden, MIT Press.

HONG, H. AND J. LI (2020): "The numerical bootstrap," *The Annals of Statistics*, 48, 397–412.

HONORÉ, B. E. AND E. KYRIAZIDOU (2000): "Panel data discrete choice models with lagged dependent variables," *Econometrica*, 68, 839–874.

HONORÉ, B. E. AND A. LEWBEL (2002): "Semiparametric binary choice panel data models without strictly exogeneous regressors," *Econometrica*, 70, 2053–2063.

HONORÉ, B. E. AND E. TAMER (2006): "Bounds on parameters in panel dynamic discrete choice models," *Econometrica*, 74, 611–629.

HOROWITZ, J. L. (1992): "A smoothed maximum score estimator for the binary response model," *Econometrica*, 60, 505–531.

HYSLOP, D. R. (1999): "State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women," *Econometrica*, 67, 1255–1294.

KERR, W. R., W. F. LINCOLN, AND P. MISHRA (2014): "The dynamics of firm lobbying," *American Economic Journal: Economic Policy*, 6, 343–79.

KHAN, S., M. PONOMAREVA, AND E. TAMER (2019): "Identification of dynamic panel binary response models," Tech. rep., Boston College Department of Economics.

KIM, J. AND D. POLLARD (1990): "Cube root asymptotics," *The Annals of Statistics*, 18, 191–219.

LEE, M.-J. AND Y.-H. TAE (2005): "Analysis of labour participation behaviour of Korean women with dynamic Probit and conditional Logit," *Oxford Bulletin of Economics and Statistics*, 67, 71–91.

LEE, S. M. S. AND M. C. PUN (2006): "On m out of n bootstrapping for nonstandard m-estimation with nuisance parameters," *Journal of American Statistical Association*, 101, 1185–1197.

MAGNAC, T. (2000): "Subsidised training and youth employment: Distinguishing unobserved heterogeneity from state dependence in labour market histories," *The Economic Journal*, 110, 805–837.

MANSKI, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3, 205–228.

——— (1985): "Semiparametric analysis of discrete response," *Journal of Econometrics*, 27, 313–333.

——— (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica (1986-1998)*, 55, 357–362.

MCFADDEN, D. L. (1976): "Quantal choice analaysis: A survey," in *Annals of Economic and Social Measurement, Volume 5, number 4*, NBER, 363–390.

PAKES, A. AND J. PORTER (2016): "Moment inequalities for multinomial choice with fixed effects," Tech. rep., National Bureau of Economic Research.

PATRA, R. K., E. SEIJO, AND B. SEN (2018): "A consistent bootstrap procedure for the maximum score estimator," *Journal of Econometrics*, 205, 488–507.

RASCH, G. (1960): "Probabilistic models for some intelligence and achievement tests," *Copenhagen: Danish Institute for Educational Research*.

——— (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, 321–333.

ROBERTS, M. J. AND J. R. TYBOUT (1997): "The decision to export in Colombia: An empirical model of entry with sunk costs," *The American Economic Review*, 545–564.

SEO, M. H. AND T. OTSU (2018): "Local M-estimation with discontinuous criterion for dependent and limited observations," *The Annals of Statistics*, 46, 344–369.

SHI, X., M. SHUM, AND W. SONG (2018): "Estimating semiparametric panel multinomial choice models using cyclic monotonicity," *Econometrica*, 86, 737–761.

SILVAPULLE, M. J. (1981): "On the existence of maximum likelihood estimators for the binomial response models," *Journal of the Royal Statistical Society. Series B (Methodological)*, 310–313.

TORGOVITSKY, A. (2019): "Nonparametric inference on state dependence in unemployment," *Econometrica*, 87, 1475–1505.

VAN DER VAART, A. AND J. WELLNER (2000): *Weak convergence and empirical processes*, Springer, New York.

WILLIAMS, B. (2019): "Nonparametric identification of discrete choice models with lagged dependent variables," *Journal of Econometrics*.

WOOLDRIDGE, J. M. (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 20, 39–54.

Appendixes

These appendixes are organized as follows. In Appendix A, we provide proofs for identification, specifically, proofs for Propositions 2.1, 2.2, and 2.3. We also present some technical lemmas needed for the proofs in this section. In Appendix B, we show the asymptotics of our estimators, which is summarized in Theorem 4.1. Technical lemmas needed for this proof are provided in the same section. All tables in this paper are collected in Appendix C. All proofs for technical lemmas used in Appendixes A and B are relegated to Appendix D. In Appendix E, we provide some technical details for Section 5.

A Technical Lemmas and Main Proofs for Identification

Building on the results of Lemmas A.1 and A.2, Lemma A.3 establishes the identification inequality (2.2) under Assumptions A and SD. Lemma A.4 shows that (2.2) also holds under Assumptions A and SI. We present these lemmas below and leave their proofs to Appendix D. Based on these results, we prove Propositions 2.2 and 2.3. We also prove Proposition 2.1 which provides a sufficient condition for Assumption SD(b). Throughout this appendix, we assume $\gamma < 0$. The proofs for the case with $\gamma \geq 0$ are symmetric. We omit them for conciseness.

For each $t \in \mathcal{T}$, define the following partition of the sample space[23]:

$$E_{t,1} = \{\epsilon_t < w_t + \gamma + \alpha\}, E_{t,2} = \{w_t + \gamma + \alpha \leq \epsilon_t < w_t + \alpha\}, E_{t,3} = \{\epsilon_t \geq w_t + \alpha\}.$$

Lemma A.1. *Let $s, t \in \mathcal{T}$ such that $t \geq s + 2$. Under Assumption A, the following equalities hold for both $\tau = s$ and $\tau = t$.*

$$\begin{aligned} &P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\ =&F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha)P(E_{\tau+1,1}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha), \end{aligned} \tag{A.1}$$

and

$$\begin{aligned} &P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\ =&P(E_{\tau+1,2}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\ &+ F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha)P(E_{\tau+1,3}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha). \end{aligned} \tag{A.2}$$

Lemma A.2. *Let $s, t \in \mathcal{T}$ such that $t \geq s + 2$. Under Assumption A, the following equalities hold for both $\tau = s$ and $\tau = t$.*

$$\begin{aligned} &P(E_{\tau+1,1}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\ =&P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\ &+ \frac{F_{\epsilon|\alpha}(w_{\tau+1} + \gamma + \alpha)}{F_{\epsilon|\alpha}(w_{\tau+1} + \alpha)}[1 - P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha)], \end{aligned} \tag{A.3}$$

$$\begin{aligned} &P(E_{\tau+1,2}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\ =&\frac{F_{\epsilon|\alpha}(w_{s+1} + \alpha) - F_{\epsilon|\alpha}(w_{s+1} + \gamma + \alpha)}{1 - F_{\epsilon|\alpha}(w_{s+1} + \gamma + \alpha)}P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha), \end{aligned} \tag{A.4}$$

[23]For the case with $\gamma \geq 0$, the proofs of Lemmas A.1 - A.3 work through with the partition $E_{t,1} = \{\epsilon_t < w_t + \alpha\}, E_{t,2} = \{w_t + \alpha \leq \epsilon_t < w_t + \gamma + \alpha\}$, and $E_{t,3} = \{\epsilon_t \geq w_t + \gamma + \alpha\}$.

and

$$
\begin{aligned}
&P(E_{\tau+1,3}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
=&\frac{1 - F_{\epsilon|\alpha}(w_{\tau+1} + \alpha)}{1 - F_{\epsilon|\alpha}(w_{\tau+1} + \gamma + \alpha)} P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
&+ 1 - P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha). \qquad \text{(A.5)}
\end{aligned}
$$

Lemma A.3. *If Assumptions A and SD hold, then for all $s, t \in \mathcal{T}$,*

$$
P(y_t = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \geq P(y_s = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha)
$$

if and only if $w_t \geq w_s$.

Lemma A.4. *If Assumptions A and SI hold, then for all $s, t \in \mathcal{T}$,*

$$
P(y_t = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \geq P(y_s = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha)
$$

if and only if $w_t \geq w_s$.

We next prove Propositions 2.1 - 2.3 in order.

Proof of Proposition 2.1. For the sake of brevity, we only prove the case $y_{s-1} = y_{s+1} = y_{t-1} = y_{t+1} = 1$. The proofs for the others cases are similar. Denote

$$
C = \{y_0 = d_0, y_1 = d_1, ..., y_{s-1} = 1, y_s = d_s, y_{s+1} = 1, ..., y_{t-1} = 1, y_t = d_t, y_{t+1} = 1, ..., y_T = d_T\}
$$

and $\varpi = (w_1, ...w_{s-1}, w_{s+2}, ..., w_{t-1}, w_{t+2}, ..., w_T)$. Then, by model (2.1) and Assumption A(a)

$$
\begin{aligned}
&P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) \\
=&\int P(C|\varpi, (w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) dF_{\varpi|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha} \\
=&\int p_0(w^T, \alpha)^{d_0}(1 - p_0(w^T, \alpha))^{1-d_0} \times F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha)^{d_1}(1 - F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha))^{1-d_1} \times \cdots \\
&\times F_{\epsilon|\alpha}(\omega_{s-1} + \gamma d_{s-2} + \alpha) F_{\epsilon|\alpha}(\omega_0 + \gamma + \alpha)^{d_s}(1 - F_{\epsilon|\alpha}(\omega_0 + \gamma + \alpha))^{1-d_s} F_{\epsilon|\alpha}(\omega_1 + \gamma d_s + \alpha) \times \cdots \\
&\times F_{\epsilon|\alpha}(\omega_{t-1} + \gamma d_{t-2} + \alpha) F_{\epsilon|\alpha}(\omega_0' + \gamma + \alpha)^{d_t}(1 - F_{\epsilon|\alpha}(\omega_0' + \gamma + \alpha))^{1-d_t} F_{\epsilon|\alpha}(\omega_1' + \gamma d_t + \alpha) \times \cdots \\
&\times F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha)^{d_T}(1 - F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha))^{1-d_T} dF_{\varpi|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha}.
\end{aligned}
$$

Given the exchangeability assumption, If $d_s = d_t$, we have

$$
P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) = P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha),
$$

and if $d_s \neq d_t$, we have

$$
P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) = P(\tilde{C}|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha),
$$

where $\tilde{C} = \{y_0 = d_0, y_1 = d_1, ..., y_{s-1} = 1, y_s = d_t, y_{s+1} = 1, ..., y_{t-1} = 1, y_t = d_s, y_{t+1} = 1, ..., y_T = d_T\}$. Then, adding up $P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha)$ across all possible events C and $\tilde{C}$ yields

$$
\begin{aligned}
&P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) \\
=&P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha). \qquad (A.6)
\end{aligned}
$$

Invoke Bayes' theorem to deduce

$$
\begin{aligned}
&f_{w_s,w_{s+1},w_t,w_{t+1}|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0, \omega_1, \omega_0', \omega_1') \\
=&\frac{P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha)}{P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|\alpha)} \\
&\times f_{w_s,w_{s+1},w_t,w_{t+1}|\alpha}(\omega_0, \omega_1, \omega_0', \omega_1') \\
=&\frac{P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha)}{P(y_{s+1} = y_{t+1} = 1|y_{s-1} = 1, \alpha)} \\
&\times f_{w_s,w_{s+1},w_t,w_{t+1}|\alpha}(\omega_0', \omega_1', \omega_0, \omega_1) \\
=&f_{w_s,w_{s+1},w_t,w_{t+1}|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0', \omega_1', \omega_0, \omega_1), \qquad (A.7)
\end{aligned}
$$

where the second equality follows from (A.6) and the exchangeability assumption.

Applying similar arguments to obtain

$$
f_{w_s,w_t|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0, \omega_0') = f_{w_s,w_t|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0', \omega_0). \qquad (A.8)
$$

Combine (A.7) and (A.8) to deduce

$$
\begin{aligned}
&f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0,\omega_0'),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1, \omega_1') \\
=&f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0',\omega_0),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1', \omega_1).
\end{aligned}
$$

Then, the desired result follows from

$$
\begin{aligned}
&f_{w_{s+1}|(w_s,w_t)=(\omega_0,\omega_0'),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1) \\
=&\int f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0,\omega_0'),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1, \omega_1')d\omega_1' \\
=&\int f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0',\omega_0),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1', \omega_1)d\omega_1' \\
=&f_{w_{t+1}|(w_s,w_t)=(\omega_0',\omega_0),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1).
\end{aligned}
$$

□

Proof of Proposition 2.2. The monotonic relation established in either Lemma A.3 or Lemma A.4 implies that β maximizes $Q_1(\cdot; \alpha)$. The remaining task is to show the uniqueness of β in $\mathcal{B}$, i.e., $Q_1(b; \alpha) = Q_1(\beta; \alpha)$ implies $b = \beta$. Here we assume $\beta_1 > 0$ w.l.o.g. as the case $\beta_1 < 0$ is symmetric.

First note that for any $b \in \mathcal{B}$ such that $Q_1(b;\alpha) = Q_1(\beta;\alpha)$, if

$$P([x_{ts,1}b_1 + \tilde{x}'_{ts}\tilde{b} < 0 < x_{ts,1}\beta_1 + \tilde{x}'_{ts}\tilde{\beta}] \cup [x_{ts,1}\beta_1 + \tilde{x}'_{ts}\tilde{\beta} < 0 < x_{ts,1}b_1 + \tilde{x}'_{ts}\tilde{b}]) > 0,$$

then β and b will yield different realized values of the sign function in $Q_1(\cdot;\alpha)$ with strictly positive probability, and thus $Q_1(\beta;\alpha) > Q_1(b;\alpha)$. It then follows that $b_1 > 0$ must hold, for otherwise by Assumption A(c) we have

$$P(x_{ts,1}b_1 + \tilde{x}'_{ts}\tilde{b} < 0 < x_{ts,1}\beta_1 + \tilde{x}'_{ts}\tilde{\beta}) = P(x_{ts,1} > -\tilde{x}'_{ts}\tilde{b}/b_1, x_{ts,1} > -\tilde{x}'_{ts}\tilde{\beta}/\beta_1) > 0.$$

Then focusing on the case with $b_1 > 0$, we can write

$$\begin{aligned}&P([x_{ts,1}b_1 + \tilde{x}'_{ts}\tilde{b} < 0 < x_{ts,1}\beta_1 + \tilde{x}'_{ts}\tilde{\beta}] \cup [x_{ts,1}\beta_1 + \tilde{x}'_{ts}\tilde{\beta} < 0 < x_{ts,1}b_1 + \tilde{x}'_{ts}\tilde{b}])\\ =&P([-\tilde{x}'_{ts}\tilde{\beta}/\beta_1 < x_{ts,1} < -\tilde{x}'_{ts}\tilde{b}/b_1] \cup [-\tilde{x}'_{ts}\tilde{b}/b_1 < x_{ts,1} < -\tilde{x}'_{ts}\tilde{\beta}/\beta_1]),\end{aligned}$$

which implies that to make $Q_1(b;\alpha) = Q_1(\beta;\alpha)$ hold we must have $P(\tilde{x}'_{ts}\tilde{\beta}/\beta_1 = \tilde{x}'_{ts}\tilde{b}/b_1) = 1$ by Assumption A(c).

However, whenever b is not a scalar multiple of β, $P(\tilde{x}'_{ts}\tilde{\beta}/\beta_1 = \tilde{x}'_{ts}\tilde{b}/b_1) = 1$ implies that $\tilde{\mathcal{X}}_{ts}$ is contained in a proper linear subspace of $\mathbb{R}^{K-1}$ a.e., violating Assumption A(d). As a result, we must have b being a scalar multiple of β, which leads to the desired result $b = \beta$ as $\|b\|_2 = \|\beta\|_2 = 1$ by the construction of the parameter space $\mathcal{B}$ in Assumption A(e). □

Proof of Proposition 2.3. The proof uses the insight of HK. Here we only prove case (ii) of Proposition 2.3 for $t > s+1$ as the same method can be applied to case (i) where s and t are adjacent. Note that it also suffices to prove that γ uniquely maximizes the following population objective function conditional on α:

$$\begin{aligned}Q_{2,2}(\gamma;\beta,\alpha) \equiv&\mathbb{E}\left\{\left[P(A|x^T, w_{s+1} = w_{t+1}, y_{s+1} = y_{t+1}, \alpha) - P(B|x^T, w_{s+1} = w_{t+1}, y_{s+1} = y_{t+1}, \alpha)\right]\right.\\ &\left.\times sgn\left((w_t - w_s) + r(d_{t-1} - d_{s-1})\right)|\alpha\right\}.\end{aligned}$$

First, note that under Assumptions A(a) and A(b), we can write

$$\begin{aligned}&P(A|x^T, w_{s+1} = w_{t+1} = w, y_{s+1} = y_{t+1} = d, \alpha)\\ =&p_0(x^T,\alpha)^{d_0}(1-p_0(x^T,\alpha))^{1-d_0} \times F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha)^{d_1}(1 - F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha))^{1-d_1}\\ &\times \cdots \times (1 - F_{\epsilon|\alpha}(w_s + \gamma d_{s-1} + \alpha)) \times F_{\epsilon|\alpha}(w + \alpha)^d(1 - F_{\epsilon|\alpha}(w + \alpha))^{1-d}\\ &\times \cdots \times F_{\epsilon|\alpha}(w_t + \gamma d_{t-1} + \alpha) \times F_{\epsilon|\alpha}(w + \gamma + \alpha)^d(1 - F_{\epsilon|\alpha}(w + \gamma + \alpha))^{1-d}\\ &\times \cdots \times F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha)^{d_T}(1 - F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha))^{1-d_T}\end{aligned}$$

for all $w \in \mathbb{R}$ and $d \in \{0,1\}$, and similarly,

$$\begin{aligned}&P(B|x^T, w_{s+1} = w_{t+1} = w, y_{s+1} = y_{t+1} = d, \alpha)\\ =&p_0(x^T,\alpha)^{d_0}(1-p_0(x^T,\alpha))^{1-d_0} \times F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha)^{d_1}(1 - F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha))^{1-d_1}\\ &\times \cdots \times F_{\epsilon|\alpha}(w_s + \gamma d_{s-1} + \alpha) \times F_{\epsilon|\alpha}(w + \gamma + \alpha)^d(1 - F_{\epsilon|\alpha}(w + \gamma + \alpha))^{1-d}\\ &\times \cdots \times (1 - F_{\epsilon|\alpha}(w_t + \gamma d_{t-1} + \alpha)) \times F_{\epsilon|\alpha}(w + \alpha)^d(1 - F_{\epsilon|\alpha}(w + \alpha))^{1-d}\\ &\times \cdots \times F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha)^{d_T}(1 - F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha))^{1-d_T}.\end{aligned}$$

Then, we obtain

$$\begin{aligned}
&\frac{P(A|x^T, w_{s+1} = w_{t+1} = w, y_{s+1} = y_{t+1} = d, \alpha)}{P(B|x^T, w_{s+1} = w_{t+1} = w, y_{s+1} = y_{t+1} = d, \alpha)} \\
=&\frac{(1 - F_{\epsilon|\alpha}(w_s + \gamma d_{s-1} + \alpha)) \times F_{\epsilon|\alpha}(w_t + \gamma d_{t-1} + \alpha)}{F_{\epsilon|\alpha}(w_s + \gamma d_{s-1} + \alpha) \times (1 - F_{\epsilon|\alpha}(w_t + \gamma d_{t-1} + \alpha))} \\
&\times \frac{F_{\epsilon|\alpha}(w + \alpha)^d (1 - F_{\epsilon|\alpha}(w + \alpha))^{1-d} \times F_{\epsilon|\alpha}(w + \gamma + \alpha)^d (1 - F_{\epsilon|\alpha}(w + \gamma + \alpha))^{1-d}}{F_{\epsilon|\alpha}(w + \gamma + \alpha)^d (1 - F_{\epsilon|\alpha}(w + \gamma + \alpha))^{1-d} \times F_{\epsilon|\alpha}(w + \alpha)^d (1 - F_{\epsilon|\alpha}(w + \alpha))^{1-d}} \\
=&\frac{(1 - F_{\epsilon|\alpha}(w_s + \gamma d_{s-1} + \alpha)) \times F_{\epsilon|\alpha}(w_t + \gamma d_{t-1} + \alpha)}{F_{\epsilon|\alpha}(w_s + \gamma d_{s-1} + \alpha) \times (1 - F_{\epsilon|\alpha}(w_t + \gamma d_{t-1} + \alpha))}
\end{aligned}$$

and therefore,

$$P(A|x^T, w_{s+1} = w_{t+1} = w, y_{s+1} = y_{t+1} = d, \alpha) \geq P(B|x^T, w_{s+1} = w_{t+1} = w, y_{s+1} = y_{t+1} = d, \alpha)$$

if and only if $w_t + \gamma d_{t-1} \geq w_s + \gamma d_{s-1}$, which implies that γ maximizes $Q_{2,2}(\gamma; \beta, \alpha)$.

The remaining task is to show that γ is unique in $\mathcal{R}$. Suppose that there exists an $r \in \mathcal{R} \setminus \{\gamma\}$ such that $Q_{2,2}(r; \beta, \alpha) = Q_{2,2}(\gamma; \beta, \alpha)$. Note that the value of r (and γ) affects $Q_{2,2}(\cdot; \beta, \alpha)$ only when $d_{s-1} \neq d_{t-1}$. Here we assume that $d_{t-1} = 1$ and $d_{s-1} = 0$ (the case with $d_{t-1} = 0$ and $d_{s-1} = 1$ is symmetric). Then by Assumption A(c), the following probability is non-zero:

$$P\left([-\gamma < w_t - w_s < -r] \cup [-r < w_t - w_s < -\gamma]\right).$$

Consequently, γ and r yield different realized values of the sign function in objective function $Q_{2,2}(\cdot; \beta, \alpha)$ with strictly positive probability, and hence $Q_{2,2}(r; \beta, \alpha) < Q_{2,2}(\gamma; \beta, \alpha)$, a contradiction. Then we can conclude that $Q_{2,2}(r; \beta, \alpha) = Q_{2,2}(\gamma; \beta, \alpha)$ if and only if $r = \gamma$, or equivalently γ uniquely maximizes $Q_{2,2}(\cdot; \beta, \alpha)$ in $\mathcal{R}$. □

B Technical Lemmas and Main Proofs for Asymptotics

In this section, we define a few technical terms and a few more technical notations, present some technical lemmas, and provide the proof of our main asymptotic theory, Theorem 4.1. The proofs for the technical lemmas are relegated to Appendix D.

The outline of the proof of Theorem 4.1 is as follows. Lemmas B.1 and B.2 verify the technical conditions as required in Seo and Otsu (2018). Those conditions can ensure the class of functions is *manageable* as in Kim and Pollard (1990). After that, the maximal inequalities and asymptotics in Seo and Otsu (2018) can be readily applied to our estimator. Lemma B.4 deals with the impact of using $\hat{\beta}$ on estimating $\hat{\gamma}$, using maximal inequalities established in Seo and Otsu (2018). Lemmas B.3 and B.5 obtain the technical terms for the final asymptotics for $\hat{\beta}$ and $\hat{\gamma}$.

Let c and C denote some constants that may vary from line to line. $\mathbb{E}_n$ denotes the expectation conditional on observations being fixed. $\rightsquigarrow$ denotes weakly convergence in the sense of van der

Vaart and Wellner (2000). Let

$$\mathbb{G}_n\left(f_{ni}\right) \equiv n^{1/2} \sum_{i=1}^{n} \left[f_{ni} - \mathbb{E}_n\left(f_n\right)\right],$$

for any f_{ni}. To facilitate calculation, occasionally we may decompose covariate x into $\varpi\beta + x_\beta$ with a scalar ϖ and x_β orthogonal to β.

Lemma B.1. *Suppose Assumptions A, SI (or SD), 3, and 4 hold. Then $\xi_i(b)$ satisfies Assumption M in Seo and Otsu (2018).*

Lemma B.2. *Suppose Assumptions A, SI (or SD) and 3 - 6 hold. Then $\varsigma_{ni}(r, b)$ satisfies Assumption M in Seo and Otsu (2018).*

Lemma B.3. *Suppose Assumptions A and 3 hold. Then*

$$\lim_{n\to\infty} n^{2/3}\mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s}n^{-1/3}\right)\right) = \frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s},$$

and

$$\lim_{n\to\infty} n^{1/3}\mathbb{E}\left[\xi_i\left(\beta + \boldsymbol{s}n^{-1/3}\right)\xi_i\left(\beta + \boldsymbol{t}n^{-1/3}\right)\right] = H_1\left(\boldsymbol{s}, \boldsymbol{t}\right).$$

V_1 is defined as

$$V_1 = -\int 1\left[x_{31}'\beta = 0\right]\left(\frac{\partial\kappa\left(x_{31}\right)'}{\partial x_{31}}\beta\right) f_{x_{31}}\left(x_{31}\right) x_{31}x_{31}' d\sigma_0, \tag{B.1}$$

with σ_0 being the surface measure on $\{x_{31} : x_{31}'\beta = 0\}$ and

$$\kappa(x) = \mathbb{E}\left\{P\left(y_{i0} = y_{i2} = y_{i4}|x_{i1}, x_{i3}\right)\left\{\mathbb{E}\left[y_{i3}|y_{i2} = y_{i4}, x_{i3}\right] - \mathbb{E}\left[y_{i1}|y_{i0} = y_{i2}, x_{i1}\right]\right\}|x_{i31} = x\right\}.$$

$H_1(\boldsymbol{s}, \boldsymbol{t})$ is defined as

$$H_1\left(\boldsymbol{s}, \boldsymbol{t}\right) = \frac{1}{2}\int_{\mathbb{R}^{K-1}} \psi\left(x_\beta\right)\left[\left|x_\beta'\boldsymbol{s}\right| + \left|x_\beta'\boldsymbol{t}\right| - \left|x_\beta'\left(\boldsymbol{s} - \boldsymbol{t}\right)\right|\right] f_{x_{31}}\left(x_\beta\right) dx_\beta, \tag{B.2}$$

where $\boldsymbol{s}, \boldsymbol{t}$ are $K \times 1$ vectors,

$$\psi(x) = \mathbb{E}\left\{P\left(y_{i0} = y_{i2} = y_{i4}|x_{i1}, x_{i3}\right)\left|\mathbb{E}\left[y_{i3}|y_{i2} = y_{i4}, x_{i3}\right] - \mathbb{E}\left[y_{i1}|y_{i0} = y_{i2}, x_{i1}\right]\right|\ |x_{i31} = x\right\},$$

and x_β is orthogonal to β.

Lemma B.4. *Suppose Assumptions A and 3 - 6 hold. Then*

$$\hat{Z}_{n,2}(s) - Z_{n,2}(s) = o_P(1),$$

where the small order term holds uniformly over $|s| \leq C$ for any positive C.

Lemma B.5. *Suppose Assumptions A, 3, 5, and 6 hold. Then*

$$\lim_{n\to\infty} \left(nh_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3}, \beta\right)\right) = \frac{1}{2}V_2 s^2,$$

and

$$\lim_{n\to\infty}(nh_n)^{1/3}\mathbb{E}\left(h_n\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\varsigma_{ni}\left(\gamma+t(nh_n)^{-1/3},\beta\right)\right)=H_2(s,t).$$

V_2 is defined as

$$V_2=-\int_{\mathbb{R}^{K-1}}\int 1\left[x_{21}'\beta+\gamma y_{30}=0\right]\left(\frac{\partial\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)'}{\partial\left(y_{30},x_{21}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right) \tag{B.3}$$
$$f\left(x_{21},y_{30}|x_{32}=x_\beta\right)|y_{30}|\,d\sigma_0 f_{x_{32}}\left(x_\beta\right)dx_\beta$$
$$-\int_{\mathbb{R}^{K-1}}\int 1\left[x_{32}'\beta+\gamma y_{41}=0\right]\left(\frac{\partial\mathbb{E}\left(y_{32}|x_{32},y_{41},x_{43}=x_\beta\right)'}{\partial\left(y_{41},x_{32}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right)$$
$$f\left(x_{32},y_{41}|x_{43}=x_\beta\right)|y_{41}|\,d\sigma_0 f_{x_{43}}\left(x_\beta\right)dx_\beta$$

with σ_0 denoting the surface measure of $\left\{\left(x_{21}',y_{30}\right)'\middle|\,x_{21}'\beta+\gamma y_{30}=0\right\}$ in the first integral and the surface measure of $\left\{\left(x_{32}',y_{41}\right)'\middle|\,x_{32}'\beta+\gamma y_{41}=0\right\}$ in the second integral. $H_2(s,t)$ is defined as

$$\begin{aligned}
&H_2(s,t) \qquad\qquad (\text{B.4})\\
&=\frac{1}{2}\left(|s|+|t|-|s-t|\right)\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{21}|\,|x_{21}'\beta=-\gamma,y_{30}=1,x_{32}=x_\beta\right]f\left(y_{30}=1,x_{21}'\beta=-\gamma|x_{32}=x_\beta\right)\right.\\
&\left.+\,\mathbb{E}\left[|y_{21}|\,|x_{21}'\beta=\gamma,y_{30}=-1,x_{32}=x_\beta\right]f\left(y_{30}=-1,x_{21}'\beta=\gamma|x_{32}=x_\beta\right)\right\}f_{x_{32}}\left(x_\beta\right)dx_\beta\\
&+\frac{1}{2}\left(|s|+|t|-|s-t|\right)\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{32}|\,|x_{32}'\beta=-\gamma,y_{41}=1,x_{43}=x_\beta\right]f\left(y_{41}=1,x_{32}'\beta=-\gamma|x_{43}=x_\beta\right)\right.\\
&\left.+\mathbb{E}\left[|y_{32}|\,|x_{32}'\beta=\gamma,y_{41}=-1,x_{43}=x_\beta\right]f\left(y_{41}=-1,x_{32}'\beta=\gamma|x_{43}=x_\beta\right)\right\}f_{x_{43}}\left(x_\beta\right)dx_\beta,
\end{aligned}$$

where s,t are scalars, $\bar{\mathcal{K}}_2=\int_{\mathbb{R}}\mathcal{K}(u)^2\,du$, and x_β is orthogonal to β.

Remark B.1 (On convergence rate)**.** HK put x_{32} in the kernel $\mathcal{K}_{h_n}(\cdot)$ while we put $x_{32}'b$ and $x_{43}'b$ instead. One implication of this difference is that the convergence rate of the estimator in HK is $(nh_n^K)^{-1/3}$ and the convergence rate of $\hat{\gamma}$ here is expected to be $(nh_n)^{-1/3}$. Thus our estimator $\hat{\gamma}$ does not suffer from the curse of dimensionality. The intuition of this result is that we match a single index $x_{32}'b$ or $x_{43}'b$ while HK had to match the entire vector x_{32}.

Remark B.2 (On V_1)**.** By definition,

$$\left.\frac{\partial\kappa\left(x_{31}\right)'}{\partial x_{31}}\beta\right|_{x_{31}'\beta=0}=\left.\lim_{h\to0}\frac{\kappa\left(x_{31}+h\beta\right)-\kappa\left(x_{31}\right)}{h}\right|_{x_{31}'\beta=0}.$$

Notice that $\left(x_{31}+h\beta\right)'\beta=h\|\beta\|$ if $x_{31}'\beta=0$. Similar to the discussion under equation (D.18), for x_{31} satisfied with $x_{31}'\beta=0$, $\kappa\left(x_{31}+h\beta\right)\geq0=\kappa\left(x_{31}\right)$ if $h>0$ and $\kappa\left(x_{31}+h\beta\right)\leq0=\kappa\left(x_{31}\right)$ if $h<0$. Thus $\left.\frac{\partial\kappa(x_{31})'}{\partial x_{31}}\beta\right|_{x_{31}'\beta=0}\geq0$, and V_1 is negative semidefinite. If $\left(\frac{\partial\kappa(x_{31})'}{\partial x_{31}}\beta\right)f_{x_{31}}$ is strictly positive over a nonzero measure on the surface $x_{31}'\beta=0$, V_1 is negative definite.

Remark B.3 (On V_2)**.** Note that V_2 is a scalar.

$$\frac{\partial\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)'}{\partial\left(y_{30},x_{21}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\geq0\text{ and }\frac{\partial\mathbb{E}\left(y_{32}|x_{32},y_{41},x_{43}=\beta\right)'}{\partial\left(y_{41},x_{32}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\geq0$$

hold for the same reason as in Remark B.2. Thus $V_2\leq0$. $V_2<0$ if either term above is strictly positive over a nonzero measure.

Proof of Theorem 4.1. We prove the first part of this theorem first.

Lemma B.1 verifies the key technical conditions needed for applying Theorem 1 in Seo and Otsu (2018). $\hat{\beta}-\beta=O_P\left(n^{-1/3}\right)$ by Assumption 2 and Lemma 1 in Seo and Otsu (2018).

Notice that $\hat{\beta}$ can be equivalently obtained from

$$\arg\max_{b\in\mathcal{B}} n^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+n^{-1/3}\cdot n^{1/3}\left(b-\beta\right)\right).$$

Intuitively, we get the asymptotics of $n^{1/3}\left(\hat{\beta}-\beta\right)$ if we can get the asymptotics of

$$Z_{n,1}\left(\boldsymbol{s}\right)=n^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right).$$

Lemma B.3 calculates the the mean and covariance kernel of $Z_{n,1}\left(\boldsymbol{s}\right)$. $\xi_i\left(b\right)$ is uniformly bounded, so the Lindeberg condition for $Z_{n,1}\left(\boldsymbol{s}\right)$ is satisfied. Therefore, $Z_{n,1}\left(\boldsymbol{s}\right)$ is pointwise asymptotically normal. With Assumption 2, Theorem 1 in Seo and Otsu (2018) implies the equicontinuity of $Z_{n,1}\left(\boldsymbol{s}\right)$, and it yields $Z_{n,1}\left(\boldsymbol{s}\right)\rightsquigarrow Z_1\left(\boldsymbol{s}\right)$, where $Z_1\left(\boldsymbol{s}\right)$ is a Gaussian Process with continuous sample paths, expected value $-\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}$, and covariance kernel $H_1\left(\boldsymbol{s},\boldsymbol{t}\right)$ that can be calculated as in equation (B.2). As a result,

$$n^{1/3}\left(\hat{\beta}-\beta\right)\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^K} Z_1\left(\boldsymbol{s}\right),$$

by applying Theorem 1 in Seo and Otsu (2018).

We now prove the second part. The calculation of equation (D.29) in the proof of Lemma B.5 shows,

$$\mathbb{E}_n\left(\varsigma_{ni}\left(r,\hat{\beta}\right)-\varsigma_{ni}\left(\gamma,\beta\right)\right)=\frac{1}{2}\left(r-\gamma,\left(\hat{\beta}-\beta\right)'\right)\tilde{V}_2\left(\begin{array}{c}r-\gamma\\ \hat{\beta}-\beta\end{array}\right)+o\left(\left\|\left(r-\gamma,\left(\hat{\beta}-\beta\right)'\right)'\right\|_2\right)+o\left(\left(nh_n\right)^{-2/3}\right), \tag{B.5}$$

where $\tilde{V}_2$ is defined in equation (D.28).

The convergence rate of $\hat{\gamma}$ is $\left(nh_n\right)^{-1/3}$, which can be seen from

$$\begin{aligned}
o_P\left(\left(nh_n\right)^{-2/3}\right)&\leq n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\hat{\gamma},\hat{\beta}\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma,\hat{\beta}\right)\\
&=n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\hat{\gamma},\hat{\beta}\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma,\beta\right)+n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma,\beta\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma,\hat{\beta}\right)\\
&\leq\mathbb{E}_n\left(\varsigma_{ni}\left(\hat{\gamma},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma,\beta\right)\right)+\varepsilon\left(\left(\hat{\gamma}-\gamma\right)^2+\left\|\hat{\beta}-\beta\right\|_2^2\right)+O_P\left(\left(nh_n\right)^{-2/3}\right)\\
&+\mathbb{E}_n\left(\varsigma_{ni}\left(\gamma,\hat{\beta}\right)-\varsigma_{ni}\left(\gamma,\beta\right)\right)+\varepsilon\left\|\hat{\beta}-\beta\right\|_2^2+O_P\left(\left(nh_n\right)^{-2/3}\right)\\
&\leq\left(-c+\varepsilon\right)\left(\left(\hat{\gamma}-\gamma\right)^2+2\left\|\hat{\beta}-\beta\right\|_2^2\right)+o\left(\left(\hat{\gamma}-\gamma\right)^2+\left\|\hat{\beta}-\beta\right\|_2^2\right)+O_P\left(\left(nh_n\right)^{-2/3}\right),
\end{aligned}$$

for each $\varepsilon > 0$, where the first line holds by Assumption 2, the third to fourth lines hold by applying Lemma 1 in Seo and Otsu (2018), the fifth line holds by Assumption 4 and equation (B.5). By noting $\left\|\hat{\beta}-\beta\right\|_2 = O_P\left(n^{-1/3}\right) = o_P\left(\left(nh_n\right)^{-1/3}\right)$, the inequality above implies

$$0 \leq \left(-c+\varepsilon\right)\left(\hat{\gamma}-\gamma\right)^2 + o\left(\left(\hat{\gamma}-\gamma\right)^2\right) + O_P\left(\left(nh_n\right)^{-2/3}\right).$$

Taking an ε to satisfy $\varepsilon << c$ yields that $\hat{\gamma}-\gamma = O_P\left(\left(nh_n\right)^{-1/3}\right)$. So we only need to get the limiting distribution of $\hat{Z}_{n,2}(s)$.

The analysis of $\hat{Z}_{n,2}(s)$ is complicated by including the first-stage estimator $\hat{\beta}$. Lemma B.4 shows that $\hat{\beta}$ has no impact on the asymptotics of $\hat{\gamma}$. More specifically,

$$\begin{aligned}\hat{Z}_{n,2}(s) &= Z_{n,2}(s) + o_P(1)\\ &= n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right) + \left(nh_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right) + o_P(1),\end{aligned} \tag{B.6}$$

where $\mathbb{G}_n\left(\varsigma_{ni}(r,b)\right) = n^{-1/2}\sum_{i=1}^{n}\left(\varsigma_{ni}(r,b) - \mathbb{E}_n\left(\varsigma_{ni}(r,b)\right)\right)$. As a result, the asymptotics is established if the weak convergence of the leading term in equation (B.6) is proved.

Lemma B.5 calculates the the mean of $\left(nh_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right)$ and covariance kernel $n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right)$.

Note

$$\begin{aligned}&\sum_{i=1}^{n}\left(\left(nh_n\right)^{2/3}\cdot n^{-1}\right)^{2+\delta}\mathbb{E}\left[\left|\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right|^{2+\delta}\right]\\ &= \left(nh_n\right)^{-\delta/3}\cdot\left(nh_n\right)^{1/3}\mathbb{E}\left[h_n^{1+\delta}\left|\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right|^{2+\delta}\right] \to 0\end{aligned}$$

for a small $\delta > 0$, because $\left(nh_n\right)^{-\delta/3} \to 0$ and $\left(nh_n\right)^{1/3}\mathbb{E}\left[h_n^{1+\delta}\left|\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right|^{2+\delta}\right] \to c$ for a finite c. This verifies the Lyapunov condition for $n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3},\beta\right)\right)$. Therefore, it converges to normal in distribution for each s. Lemma B.2 verifies the key technical conditions for applying Theorem 1 in Seo and Otsu (2018) to $Z_{n,2}(s)$. Together with Assumption 2, all technical conditions in Theorem 1 of Seo and Otsu (2018) are satisfied for $Z_{n,2}(s)$. That implies the stochastic equicontinuity of $Z_{n,2}(s)$ in s and

$$Z_{n,2}(s) \rightsquigarrow Z_2(s),$$

where $Z_2(s)$ is a Gaussian process with continuous path, expected value $\frac{1}{2}V_2s^2$ and covariance kernel $H_2(s,t)$. By means of the continuous mapping Theorem,

$$\left(nh_n\right)^{1/3}\left(\hat{\gamma}-\gamma\right) \xrightarrow{d} \arg\max_{s\in\mathbb{R}} Z_2(s).$$

□

C Tables

Table 1A: Design 1, Performance of $\hat{\beta}$ and $\hat{\gamma}$

	$n = 5000$		$n = 10000$		$n = 20000$	
	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$
MEAN	0.720	−0.711	0.715	−0.705	0.708	−0.694
BIAS	0.013	−0.004	0.008	0.003	0.001	0.014
MAD	0.104	0.115	0.081	0.089	0.066	0.069
RMSE	0.134	0.146	0.104	0.112	0.083	0.086

Table 1B: Design 1, Numerical Bootstrap

		$n = 5000$		$n = 10000$		$n = 20000$	
		$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$
$c = 0.8$	COVERAGE	0.894	0.903	0.917	0.924	0.926	0.913
	LENGTH	0.585	0.603	0.494	0.466	0.412	0.346
$c = 0.9$	COVERAGE	0.905	0.892	0.914	0.925	0.926	0.925
	LENGTH	0.576	0.603	0.484	0.469	0.405	0.349
$c = 1.0$	COVERAGE	0.900	0.905	0.927	0.916	0.927	0.920
	LENGTH	0.567	0.595	0.478	0.470	0.401	0.352
$c = 1.1$	COVERAGE	0.916	0.902	0.921	0.907	0.930	0.923
	LENGTH	0.560	0.587	0.569	0.472	0.394	0.356
$c = 1.2$	COVERAGE	0.907	0.889	0.918	0.915	0.922	0.929
	LENGTH	0.553	0.581	0.465	0.470	0.390	0.357

Table 1C: Design 1, Bootstrap using a Modified Objective Function

		$n = 5000$		$n = 10000$		$n = 20000$	
		$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$
$c = 0.8$	COVERAGE	0.917	0.888	0.903	0.905	0.915	0.907
	LENGTH	0.468	0.469	0.364	0.374	0.283	0.290
$c = 0.9$	COVERAGE	0.907	0.890	0.930	0.914	0.933	0.918
	LENGTH	0.510	0.485	0.392	0.381	0.304	0.295
$c = 1.0$	COVERAGE	0.905	0.921	0.946	0.911	0.948	0.916
	LENGTH	0.560	0.499	0.429	0.389	0.328	0.300
$c = 1.1$	COVERAGE	0.945	0.907	0.945	0.916	0.958	0.927
	LENGTH	0.614	0.509	0.465	0.397	0.357	0.305
$c = 1.2$	COVERAGE	0.936	0.912	0.957	0.924	0.966	0.930
	LENGTH	0.672	0.525	0.512	0.403	0.388	0.310

Table 2A: Design 2, Performance of $\hat{\beta}$ and $\hat{\gamma}$

	$n = 5000$		$n = 10000$		$n = 20000$	
	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$
MEAN	0.736	-0.714	0.734	-0.716	0.732	-0.700
BIAS	0.029	-0.007	0.027	-0.009	0.024	0.007
MAD	0.109	0.123	0.085	0.097	0.070	0.074
RMSE	0.141	0.156	0.107	0.121	0.089	0.092

Table 2B: Design 2, Numerical Bootstrap

		$n = 5000$		$n = 10000$		$n = 20000$	
		$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$
$c = 0.8$	COVERAGE	0.903	0.898	0.936	0.924	0.941	0.915
	LENGTH	0.587	0.632	0.492	0.493	0.414	0.368
$c = 0.9$	COVERAGE	0.907	0.894	0.920	0.924	0.939	0.916
	LENGTH	0.575	0.624	0.483	0.495	0.408	0.373
$c = 1.0$	COVERAGE	0.905	0.905	0.930	0.913	0.935	0.926
	LENGTH	0.567	0.612	0.477	0.498	0.401	0.375
$c = 1.1$	COVERAGE	0.897	0.910	0.905	0.906	0.933	0.932
	LENGTH	0.559	0.607	0.469	0.494	0.396	0.381
$c = 1.2$	COVERAGE	0.908	0.890	0.930	0.916	0.940	0.931
	LENGTH	0.553	0.598	0.464	0.492	0.390	0.382

Table 2C: Design 2, Bootstrap using a Modified Objective Function

		$n = 5000$		$n = 10000$		$n = 20000$	
		$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\gamma}$
$c = 0.8$	COVERAGE	0.861	0.860	0.891	0.901	0.880	0.890
	LENGTH	0.465	0.489	0.357	0.389	0.281	0.304
$c = 0.9$	COVERAGE	0.886	0.887	0.900	0.905	0.890	0.911
	LENGTH	0.496	0.501	0.380	0.399	0.297	0.311
$c = 1.0$	COVERAGE	0.901	0.897	0.906	0.908	0.900	0.914
	LENGTH	0.541	0.522	0.414	0.409	0.318	0.317
$c = 1.1$	COVERAGE	0.906	0.926	0.917	0.906	0.915	0.920
	LENGTH	0.591	0.535	0.452	0.416	0.344	0.321
$c = 1.2$	COVERAGE	0.923	0.925	0.937	0.916	0.926	0.927
	LENGTH	0.647	0.549	0.489	0.427	0.373	0.328

Table 3A: Design 3, Performance of $\hat{\beta}$ and $\hat{\gamma}$

	$n = 5000$			$n = 10000$			$n = 20000$		
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
MEAN	0.596	0.591	−0.585	0.589	0.589	−0.577	0.588	0.591	−0.575
BIAS	0.019	0.013	−0.008	0.012	0.012	0.001	0.011	0.014	0.003
MAD	0.088	0.089	0.109	0.071	0.070	0.083	0.040	0.042	0.052
RMSE	0.115	0.116	0.141	0.090	0.089	0.106	0.053	0.053	0.063

Table 3B: Design 3, Numerical Bootstrap

		$n = 5000$			$n = 10000$			$n = 20000$		
		$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
$c = 0.8$	COVERAGE	0.933	0.937	0.867	0.957	0.943	0.913	0.955	0.949	0.890
	LENGTH	0.526	0.525	0.540	0.439	0.439	0.433	0.364	0.365	0.330
$c = 0.9$	COVERAGE	0.931	0.933	0.861	0.956	0.947	0.904	0.952	0.942	0.900
	LENGTH	0.518	0.507	0.531	0.432	0.433	0.432	0.359	0.360	0.330
$c = 1.0$	COVERAGE	0.931	0.926	0.861	0.956	0.946	0.907	0.953	0.937	0.890
	LENGTH	0.510	0.508	0.521	0.427	0.428	0.430	0.354	0.356	0.334
$c = 1.1$	COVERAGE	0.928	0.927	0.851	0.957	0.950	0.913	0.953	0.943	0.890
	LENGTH	0.501	0.501	0.512	0.423	0.422	0.427	0.350	0.352	0.353
$c = 1.2$	COVERAGE	0.924	0.926	0.851	0.944	0.944	0.903	0.964	0.945	0.900
	LENGTH	0.493	0.492	0.502	0.418	0.417	0.421	0.347	0.348	0.335

Table 3C: Design 3, Bootstrap using a Modified Objective Function

		$n = 5000$			$n = 10000$			$n = 20000$		
		$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
$c = 0.8$	COVERAGE	0.913	0.913	0.850	0.943	0.936	0.870	0.949	0.934	0.863
	LENGTH	0.407	0.408	0.424	0.317	0.317	0.340	0.247	0.248	0.267
$c = 0.9$	COVERAGE	0.920	0.922	0.877	0.951	0.940	0.877	0.952	0.936	0.900
	LENGTH	0.433	0.433	0.454	0.335	0.335	0.350	0.259	0.259	0.272
$c = 1.0$	COVERAGE	0.934	0.927	0.863	0.959	0.948	0.885	0.959	0.968	0.901
	LENGTH	0.466	0.465	0.440	0.360	0.356	0.356	0.276	0.277	0.274
$c = 1.1$	COVERAGE	0.935	0.944	0.892	0.966	0.963	0.902	0.965	0.948	0.890
	LENGTH	0.508	0.507	0.466	0.390	0.389	0.361	0.299	0.300	0.281
$c = 1.2$	COVERAGE	0.947	0.950	0.904	0.969	0.969	0.911	0.972	0.961	0.890
	LENGTH	0.551	0.553	0.479	0.426	0.424	0.369	0.324	0.326	0.285

Table 4: Definition of Variables

Variable	Definition
labor force participation (y_{it})	1 if a person is in the labor force during the past 7 days, 0 otherwise
Health shock (HS_{it})	1 if there is personal injury or illness that has happened to life over the past 12 months, 0 otherwise
Activity limiting condition (ALC_{it})	1 if there is any long-term health condition, impairment or disability that restricts everyday activity has lasted for 6 months or more, 0 otherwise
Unemployment rate (UR_t)	Unemployment rate in major statistical region
log(Income) (I_{it})	Natural logarithm of household's financial year disposable income

Table 5: Summary Statistics

Variable	Mean	Std.Dev.	25% Quantile	Median	75% Quantile
Male (number of individuals: 6,848, number of observations: 42,416)					
labor force participation	0.736	0.441	0	1	1
Health shock	0.087	0.282	0	0	0
Activity limiting condition	0.276	0.447	0	0	1
Unemployment rate	5.077	1.032	4.400	5.200	5.800
ln(income)	11.119	0.718	10.700	11.120	11.585
Female (number of individuals: 7,927, number of observations: 48,121)					
labor force participation	0.621	0.485	0	1	1
Health shock	0.082	0.274	0	0	0
Activity limiting condition	0.282	0.450	0	0	1
Unemployment rate	5.082	1.030	4.400	5.200	5.800
log(Income)	11.030	0.755	10.576	11.127	11.545

Table 6: Estimates of Preference Coefficients

	Whole Sample		Male Sample		Female Sample	
Variable	Estimate	95% CIs	Estimate	95% CIs	Estimate	95% CIs
y_{it-1}	0.898	$[0.071, 1.062]^{**}$	0.915	$[0.014, 1.000]^{**}$	1.189	$[0.199, 1.367]^{**}$
		$[0.797, 0.929]^{**}$		$[0.891, 0.975]^{**}$		$[0.955, 1.214]^{**}$
HS_{it}	-0.799	$[-0.963, -0.374]^{**}$	-0.831	$[-0.919, -0.331]^{**}$	-0.972	$[-1.073, -0.403]^{**}$
		$[-0.927, -0.728]^{**}$		$[-0.908, -0.757]^{**}$		$[-1.000, -0.866]^{**}$
ALC_{it}	-0.144	$[-0.725, 0.545]$	-0.299	$[-0.613, 0.523]$	0.002	$[-0.712, 0.581]$
		$[-0.395, 0.270]$		$[-0.531, 0.191]$		$[-0.277, 0.222]$
UR_t	-0.096	$[-0.357, 0.119]$	0.066	$[-0.440, 0.137]$	-0.039	$[-0.424, 0.107]$
		$[-0.279, 0.216]$		$[-0.202, 0.278]$		$[-0.155, 0.215]$
I_{it}	0.576	$[-0.255, 0.859]$	0.465	$[0.174, 0.892]^{**}$	0.233	$[-0.626, 0.815]$
		$[0.180, 0.652]^{**}$		$[0.270, 0.544]^{**}$		$[-0.129, 0.433]$

D Proofs for Technical Lemmas

Proof of Lemma A.1. Here we only prove the case $\tau = s$. The derivation for case $\tau = t$ is analogous.

First note that by law of total probability, we can write for all $d_1 \in \{0, 1\}$,

$$\begin{aligned}&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = d_1, \alpha)\\=&\sum_{j=1}^{3}\left\{P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = d_1, \alpha, E_{s+1,j})\right.\\&\left.\times P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = d_1, \alpha)\right\}.\end{aligned} \tag{D.1}$$

When $d_1 = 1$, (D.1) reduces to

$$\begin{aligned}&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha)\\=&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha, E_{s+1,1})\\&\times P(E_{s+1,1}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha)\end{aligned} \tag{D.2}$$

as by definition $E_{s+1,3} \cap \{y_{s+1} = 1\} = \emptyset$ and by Bayes' theorem[24]

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha, E_{s+1,2}) \\
=&\frac{P(y_{s+1} = 1, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha, y_s = 1)P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha)}{P(y_{s+1} = 1, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha)} \\
=&0,
\end{aligned}$$

where the last equality is due to fact that $E_{s+1,2} \cap \{y_{s+1} = 1\} = E_{s+1,2} \cap E_{s+1,1} = \emptyset$ conditional on $\{y_s = 1\}$. Furthermore, under Assumption A(a), we can write

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha, E_{s+1,1}) \\
=&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha, E_{s+1,1}) = P(y_s = 1|w^T, y_{s-1}, \alpha, E_{s+1,1}) \\
=&P(y_s = 1|w^T, y_{s-1}, \alpha) = F_{\epsilon|\alpha}(w_s + \gamma y_{s-1} + \alpha),
\end{aligned} \tag{D.3}$$

where the first equality uses the fact that $E_{s+1,1} \subset \{y_{s+1} = 1\}$, the second equality follows from noticing that y_s (depends only on ϵ_s) is independent of (y_{t-1}, y_{t+1}) (depend only on $(\epsilon_{s+2}, ..., \epsilon_{t+1})$) conditional on (w^T, y_{s-1}, α) and event $E_{s+1,1}$, and the third equality is because $y_s \perp E_{s+1,1}$ conditional on (w^T, y_{s-1}, α). Plugging (D.3) into (D.2) yields (A.1).

When $d_1 = 0$, (D.1) reduces to

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
=&\sum_{j=2}^{3} \{P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,j}) \\
&\times P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha)\}
\end{aligned} \tag{D.4}$$

as by definition $E_{s+1,1} \cap \{y_{s+1} = 0\} = \emptyset$.

Using Bayes' theorem and the fact that $E_{s+1,2} \cap \{y_{s+1} = 0\} = E_{s+1,2} \cap E_{s+1,3} = \emptyset$ conditional on $\{y_s = 0\}$, we have

$$\begin{aligned}
&P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,2}) \\
=&\frac{P(y_{s+1} = 0, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)}{P(y_{s+1} = 0, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)} \\
=&\frac{P(E_{s+1,2} \cap E_{s+1,3}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)}{P(y_{s+1} = 0, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)} \\
=&0,
\end{aligned}$$

[24] The Bayes' theorem is stated mathematically as the following equation

$$P(A|B, C) = P(B|A, C)P(A|C)/P(B|C)$$

where A, B and C are events and $P(B|C) > 0$. Here we apply Bayes' theorem by letting $A = \{y_s = 1\}$, $B = \{y_{s+1} = 1, E_{s+1,2}\}$, and $C = \{w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha\}$.

and thus

$$P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,2}) = 1. \tag{D.5}$$

Applying similar arguments for deriving (D.3) gives

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,3}) \\
=&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha, E_{s+1,3}) = P(y_s = 1|w^T, y_{s-1} = d, \alpha, E_{s+1,3}) \\
=&P(y_s = 1|w^T, y_{s-1}, \alpha) = F_{\epsilon|\alpha}(w_s + \gamma y_{s-1} + \alpha).
\end{aligned} \tag{D.6}$$

Then plugging (D.5) and (D.6) into (D.4) yields (A.2). □

Proof of Lemma A.2. Again we only prove the case $\tau = s$ as the same arguments can be applied to derive the case $\tau = t$. Note that for all $j = 1, 2, 3$, we can use law of total probability to write

$$\begin{aligned}
&P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
=&P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
&+ P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha, y_s = 1)P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
=&P(E_{s+1,j}|w^T, y_{s-1}, y_{s+1}, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
&+ P(E_{s+1,j}|w^T, y_{s-1}, y_{s+1}, \alpha, y_s = 1)P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha),
\end{aligned}$$

where the second equality follows from $E_{s+1,j} \perp \{y_{t-1}, y_{t+1}\}|(w^T, y_{s-1}, y_s, y_{s+1}, \alpha)$ by Assumption A(a). Therefore, to prove (A.3) - (A.5), it suffices to verify the following equalities:

(1) $P(E_{s+1,1}|w^T, y_{s-1}, y_{s+1} = 1, \alpha, y_s = 1) = 1$

(2) $P(E_{s+1,1}|w^T, y_{s-1}, y_{s+1} = 1, \alpha, y_s = 0) = \frac{F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}{F_{\epsilon|\alpha}(w_{s+1}+\alpha)}$

(3) $P(E_{s+1,2}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 1) = \frac{F_{\epsilon|\alpha}(w_{s+1}+\alpha)-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}$

(4) $P(E_{s+1,2}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 0) = 0$

(5) $P(E_{s+1,3}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 1) = \frac{1-F_{\epsilon|\alpha}(w_{s+1}+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}$

(6) $P(E_{s+1,3}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 0) = 1$

Equalities (1), (4), and (6) can be easily verified using the facts that $E_{s+1,1} = \{y_{s+1} = 1\}$ conditional on $\{y_s = 1\}$, $E_{s+1,2} \cap \{y_{s+1} = 0\} = \emptyset$ conditional on $\{y_s = 0\}$, and $E_{s+1,3} = \{y_{s+1} = 0\}$ conditional on $\{y_s = 0\}$, respectively.

For equality (2), note that using conditional probability formula, we have

$$
\begin{aligned}
&P(E_{s+1,1}|w^T, y_{s-1}, y_{s+1}=1, \alpha, y_s=0) \\
=&\frac{P(y_{s+1}=1, E_{s+1,1}|w^T, y_{s-1}, \alpha, y_s=0)}{P(y_{s+1}=1|w^T, y_{s-1}, \alpha, y_s=0)} = \frac{P(E_{s+1,1}|w^T, y_{s-1}, \alpha, y_s=0)}{P(E_{s+1,1}\cup E_{s+1,2}|w^T, y_{s-1}, \alpha, y_s=0)} \\
=&\frac{P(E_{s+1,1}|w_{s+1}, \alpha)}{P(E_{s+1,1}\cup E_{s+1,2}|w_{s+1}, \alpha)} = \frac{F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}{F_{\epsilon|\alpha}(w_{s+1}+\alpha)}
\end{aligned}
$$

where the second equality uses the fact that $\{y_{s+1}=1\} = E_{s+1,1}\cup E_{s+1,2}$ conditional on $\{y_s=0\}$, and the third equality follows by Assumption A(a).

Similar arguments, along with the fact that $\{y_{s+1}=0\} = E_{s+1,2}\cup E_{s+1,3}$ conditional on $\{y_s = 1\}$, can be used to verify equalities (3) and (5). Specifically, we can write for equality (3),

$$
\begin{aligned}
&P(E_{s+1,2}|w^T, y_{s-1}, y_{s+1}=0, \alpha, y_s=1) \\
=&\frac{P(y_{s+1}=0, E_{s+1,2}|w^T, y_{s-1}, \alpha, y_s=1)}{P(y_{s+1}=0|w^T, y_{s-1}, \alpha, y_s=1)} = \frac{P(E_{s+1,2}|w^T, y_{s-1}, \alpha, y_s=1)}{P(E_{s+1,2}\cup E_{s+1,3}|w^T, y_{s-1}, \alpha, y_s=1)} \\
=&\frac{P(E_{s+1,2}|w_{s+1}, \alpha)}{P(E_{s+1,2}\cup E_{s+1,3}|w_{s+1}, \alpha)} = \frac{F_{\epsilon|\alpha}(w_{s+1}+\alpha) - F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}
\end{aligned}
$$

and analogously for equality (5),

$$
\begin{aligned}
&P(E_{s+1,3}|w^T, y_{s-1}, y_{s+1}, \alpha, y_s=1) \\
=&\frac{P(y_{s+1}=0, E_{s+1,3}|w^T, y_{s-1}, \alpha, y_s=1)}{P(y_{s+1}=0|w^T, y_{s-1}, \alpha, y_s=1)} = \frac{P(E_{s+1,3}|w^T, y_{s-1}, \alpha, y_s=1)}{P(E_{s+1,2}\cup E_{s+1,3}|w^T, y_{s-1}, \alpha, y_s=1)} \\
=&\frac{P(E_{s+1,3}|w_{s+1}, \alpha)}{P(E_{s+1,2}\cup E_{s+1,3}|w_{s+1}, \alpha)} = \frac{1-F_{\epsilon|\alpha}(w_{s+1}+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}.
\end{aligned}
$$

Then the proof completes. □

Proof of Lemma A.3. Let ϖ denote the sub-vector of w^T comprising all its elements other than w_s and w_t. Note that for all $\tau \in \{s,t\}$,

$$
\begin{aligned}
&P(y_\tau=1|w_s, w_t, y_{s-1}=y_{t-1}, y_{s+1}=y_{t+1}, \alpha) \\
=&\int P(y_\tau=1|w^T, y_{s-1}=y_{t-1}, y_{s+1}=y_{t+1}, \alpha)dF_{\varpi|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}.
\end{aligned}
\tag{D.7}
$$

In what follows, we consider two cases, $y_{s+1}=y_{t+1}=1$ and $y_{s+1}=y_{t+1}=0$, in turn.

Case 1 ($y_{s+1}=y_{t+1}=1$) Plug (A.3) into (A.1) to obtain

$$
\begin{aligned}
&P(y_\tau=1|w^T, y_{s-1}=y_{t-1}, y_{s+1}=y_{t+1}=1, \alpha) \\
=&F_{\epsilon|\alpha}(w_\tau+\gamma y_{\tau-1}+\alpha)\{P(y_\tau=1|w^T, y_{s-1}=y_{t-1}, y_{s+1}=y_{t+1}=1, \alpha) \\
&+\frac{F_{\epsilon|\alpha}(w_{\tau+1}+\gamma+\alpha)}{F_{\epsilon|\alpha}(w_{\tau+1}+\alpha)}[1-P(y_\tau=1|w^T, y_{s-1}=y_{t-1}, y_{s+1}=y_{t+1}=1, \alpha)]\}.
\end{aligned}
\tag{D.8}
$$

Let $\psi(w) \equiv F_{\epsilon|\alpha}(w+\gamma y_{\tau-1}+\alpha)$ and $\phi_1(w) \equiv F_{\epsilon|\alpha}(w+\gamma+\alpha)/F_{\epsilon|\alpha}(w+\alpha)$. Deduce from (D.8) that

$$\begin{aligned}&P(y_\tau=1|w^T, y_{s-1}=y_{t-1}, y_{s+1}=y_{t+1}=1,\alpha)\\ =&\frac{\psi(w_\tau)\phi_1(w_{\tau+1})}{1-\psi(w_\tau)+\psi(w_\tau)\phi_1(w_{\tau+1})} \equiv G_1(w_\tau, w_{\tau+1}).\end{aligned}$$

Then, (D.7) reduces to

$$\int G_1(w_\tau, w) dF_{w_{\tau+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w),$$

and hence

$$\begin{aligned}&P(y_t=1|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha)\\ &-P(y_s=1|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha)\\ =&\int G_1(w_t,w)dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w)\\ &-\int G_1(w_s,w)dF_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w)\\ =&\int \left[G_1(w_t,w)-G_1(w_s,w)\right]dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w) \qquad \text{(D.9)}\\ &+\int G_1(w_s,w)d\left[F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w)-F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w)\right].\end{aligned}$$

It is easy to verify that $\psi'(\cdot)>0$, $\phi_1'(\cdot)>0$ (by Assumption SD(a)). Therefore, $G_1'(\cdot,w)>0$ and $G_1'(w,\cdot)>0$ hold true for all w. The former monotonicity result implies that the first term in (D.9) is positive if and only if $w_t \geq w_s$. The latter, together with Assumption SD(b), implies that the second term in (D.9) is positive if and only if $w_t \geq w_s$. Put these results to establish the desired result.

Case 2 ($y_{s+1}=y_{t+1}=0$) Plug (A.4) and (A.5) into (A.2) to obtain

$$\begin{aligned}&P(y_\tau=1|w^T,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha)\\ =&\frac{F_{\epsilon|\alpha}(w_{\tau+1}+\alpha)-F_{\epsilon|\alpha}(w_{\tau+1}+\gamma+\alpha)}{1-F_{\epsilon|\alpha}(w_{\tau+1}+\gamma+\alpha)}P(y_\tau=1|w^T,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha)\\ &+F_{\epsilon|\alpha}(w_\tau+\gamma y_{\tau-1}+\alpha)[\frac{1-F_{\epsilon|\alpha}(w_{\tau+1}+\alpha)}{1-F_{\epsilon|\alpha}(w_{\tau+1}+\gamma+\alpha)}P(y_\tau=1|w^T,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha)\\ &+1-P(y_\tau=1|w^T,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha)]. \qquad \text{(D.10)}\end{aligned}$$

Let $\phi_0(w)\equiv[1-F_{\epsilon|\alpha}(w+\alpha)]/[1-F_{\epsilon|\alpha}(w+\gamma+\alpha)]$. We deduce from (D.10) that

$$\begin{aligned}&P(y_\tau=1|w^T,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha)\\ =&\frac{\psi(w_\tau)}{\psi(w_\tau)+\phi_0(w_{\tau+1})-\psi(w_\tau)\phi_0(w_{\tau+1})}\equiv G_0(w_\tau,w_{\tau+1}).\end{aligned}$$

Then, (D.7) reduces to

$$\int G_0(w_\tau, w) dF_{w_{\tau+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w),$$

and hence

$$\begin{aligned}
&P(y_t = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
&- P(y_s = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
=& \int G_0(w_t, w) dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) \\
&- \int G_0(w_s, w) dF_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) \\
=& \int \left[G_0(w_t, w) - G_0(w_s, w)\right] dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) \qquad \text{(D.11)} \\
&+ \int G_0(w_s, w) d\left[F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) - F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w)\right].
\end{aligned}$$

By Assumption SD(a), $\phi_0'(\cdot) < 0$. Therefore, $G_0'(\cdot, w) > 0$ and $G_0'(w, \cdot) > 0$ hold true for all w. The former monotonicity result implies that the first term in (D.11) is positive if and only if $w_t \geq w_s$. The latter, together with Assumption SD(b), implies that the second term in (D.11) is positive if and only if $w_t \geq w_s$. The proof is complete. □

Proof of Lemma A.4. The proof adopts similar arguments used in the proofs of Lemmas A.1 - A.3. Here we only outline the proof procedure and omit repetitive technical details for brevity, .

First note that, under Assumptions A and SI, we can write for both $\tau = s$ and $\tau = t$,

$$P(y_\tau = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) = P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1}, \alpha). \qquad \text{(D.12)}$$

To see this, note that for $\tau = s$ and all $d_0, d_1 \in \{0, 1\}$

$$\begin{aligned}
&P(y_s = 1|w_s, w_t, y_{s-1} = y_{t-1} = d_0, y_{s+1} = y_{t+1} = d_1, \alpha) \\
=& \frac{P(y_{t-1} = d_0, y_{t+1} = d_1|w_s, w_t, y_{s-1} = d_0, y_s = 1, y_{s+1} = d_1, \alpha)}{P(y_{t-1} = d_0, y_{t+1} = d_1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha)} \\
&\times P(y_s = 1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha) \\
=& \frac{P(y_{t-1} = d_0, y_{t+1} = d_1|w_t, y_{s+1} = d_1, \alpha) P(y_s = 1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha)}{P(y_{t-1} = d_0, y_{t+1} = d_1|w_t, y_{s+1} = d_1, \alpha)} \\
=& P(y_s = 1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha) \\
=& \frac{P(y_{s+1} = d_1|w_s, w_t, y_{s-1} = d_0, y_s = 1, \alpha) P(y_s = 1|w_s, w_t, y_{s-1} = d_0, \alpha)}{P(y_{s+1} = d_1|w_s, w_t, y_{s-1} = d_0, \alpha)} \\
=& \frac{P(y_{s+1} = d_1|w_s, y_{s-1} = d_0, y_s = 1, \alpha) P(y_s = 1|w_s, y_{s-1} = d_0, \alpha)}{P(y_{s+1} = d_1|w_s, y_{s-1} = d_0, \alpha)} \\
=& P(y_s = 1|w_s, y_{s-1} = d_0, y_{s+1} = d_1, \alpha),
\end{aligned}$$

where the first, third, fourth, and last equalities use Bayes' theorem, and the second and fifth equalities follow by Assumptions SI(a) and A(a)[25]. Using similar arguments yields the same sim-

[25] $(y_{t-1}, y_{t+1}) \perp (w_s, y_{s-1}, y_s)|(w_t, y_{s+1}, \alpha)$ and $(y_s, y_{s+1}) \perp w_t|(w_s, y_{s-1}, \alpha)$.

plification for $\tau = t$.

For the case with $d_1 = 1$, we uses the same arguments for deriving (A.1) to write

$$\begin{aligned} & P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\ =& P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha, E_{\tau+1,1}) P(E_{\tau+1,1} | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\ =& F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha) P(E_{\tau+1,1} | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha), \end{aligned} \tag{D.13}$$

where the last equality follows from $E_{\tau+1,1} \subset \{y_{\tau+1} = 1\}$, Assumption SI(a), and Assumption A(a). Then, we use analogous arguments for proving Lemma A.2 to deduce

$$\begin{aligned} & P(E_{\tau+1,1} | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\ =& P(E_{\tau+1,1} | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha, y_\tau = 1) P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\ & + P(E_{\tau+1,1} | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha, y_\tau = 0)[1 - P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha)] \\ =& P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\ & + \frac{P(E_{\tau+1,1} | w_\tau, y_{\tau-1}, \alpha, y_\tau = 0)}{P(E_{\tau+1,1} \cup E_{\tau+1,2} | w_\tau, y_{\tau-1}, \alpha, y_\tau = 0)}[1 - P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha)] \\ =& P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\ & + \frac{P(E_{\tau+1,1} | \alpha)}{P(E_{\tau+1,1} \cup E_{\tau+1,2} | \alpha)}[1 - P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha)], \end{aligned} \tag{D.14}$$

where the last equality follows from Assumptions SI(a) and A(a).

Combine (D.12), (D.13) and (D.14) to solve

$$\begin{aligned} & P(y_\tau = 1 | w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\ =& P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) = \frac{\phi_{1\alpha} \psi(w_\tau)}{1 - \psi(w_\tau) + \phi_{1\alpha} \psi(w_\tau)} \equiv \mathcal{G}_1(w_\tau), \end{aligned}$$

where $\phi_{1\alpha} \equiv P(E_{\tau+1,1}|\alpha)/P(E_{\tau+1,1} \cup E_{\tau+1,2}|\alpha)$ is a positive constant for any given α. The monotonic relation stated in the lemma is then established by verifying the monotonicity of $\mathcal{G}_1(\cdot)$.

For the case with $d_1 = 0$, using the same arguments for deriving (A.2) yields

$$\begin{aligned} & P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\ =& P(E_{\tau+1,2} | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\ & + P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha, E_{\tau+1,3}) P(E_{\tau+1,3} | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\ =& P(E_{\tau+1,2} | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) + F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha) P(E_{\tau+1,3} | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha), \end{aligned} \tag{D.15}$$

where the last equality follows by $E_{\tau+1,3} \subset \{y_{\tau+1} = 0\}$, Assumption SI(a), and Assumption A(a).

Use analogous arguments for proving Lemma A.2 to obtain

$$\begin{aligned} & P(E_{\tau+1,2} | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\ =& \frac{P(E_{\tau+1,2} | \alpha)}{P(E_{\tau+1,2} \cup E_{\tau+1,3} | \alpha)} P(y_\tau = 1 | w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha), \end{aligned} \tag{D.16}$$

and

$$
\begin{aligned}
&P(E_{\tau+1,3}|w_\tau, y_{\tau-1}, y_{\tau+1}=0,\alpha)\\
=&1-P(y_\tau=1|w_\tau, y_{\tau-1}, y_{\tau+1}=0,\alpha)\\
&+\frac{P(E_{\tau+1,3}|\alpha)}{P(E_{\tau+1,2}\cup E_{\tau+1,3}|\alpha)}P(y_\tau=1|w_\tau, y_{\tau-1}, y_{\tau+1}=0,\alpha). \qquad \text{(D.17)}
\end{aligned}
$$

Combine (D.12), (D.15), (D.16), and (D.17) to obtain

$$
\begin{aligned}
&P(y_\tau=1|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha)\\
=&P(y_\tau=1|w_\tau,y_{\tau-1},y_{\tau+1}=1,\alpha)=\frac{\psi(w_\tau)}{\psi(w_\tau)+\phi_{0\alpha}-\phi_{0\alpha}\psi(w_\tau)}\equiv\mathcal{G}_0(w_\tau),
\end{aligned}
$$

where $\phi_{0\alpha}\equiv P(E_{\tau+1,3}|\alpha)/P(E_{\tau+1,2}\cup E_{\tau+1,3}|\alpha)$ is a positive constant for any given α. Note that $\mathcal{G}_0(w_\tau)$ is an increasing function, from which the monotonic relation stated in the lemma is established. Putting all these results together completes the proof. □

Proof of Lemma B.1. **Preparation.** Relating to the notations in Seo and Otsu (2018), $h_n=1$ (in their notations) for our estimator $\hat{\beta}$. $\xi_i(b)$ only takes value $-1,0,$ and 1, so it is bounded. Proposition 2.2 shows that β it the unique solution to $\max_{b\in\mathcal{B}}\mathbb{E}(\xi_i(b))$. The following calculation can help understand this result.

$$
\begin{aligned}
\mathbb{E}(\xi_i(b)) &= \mathbb{E}\left\{\mathbb{E}\left[1\left[y_{i0}=y_{i2}=y_{i4}\right](y_{i3}-y_{i1})|x_{i1},x_{i3}\right]\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}\\
&=\mathbb{E}\{(\mathbb{E}\left[1\left[y_{i0}=y_{i2}=y_{i4}\right](y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]P\left(y_{i0}=y_{i2}=y_{i4}|x_{i1},x_{i3}\right)\\
&+\mathbb{E}\left[1\left[y_{i0}=y_{i2}=y_{i4}\right](y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]P\left(\overline{y_{i0}=y_{i2}=y_{i4}}|x_{i1},x_{i3}\right))\\
&\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\}\\
&=\mathbb{E}\{\mathbb{E}\left[(y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]P\left(y_{i0}=y_{i2}=y_{i4}|x_{i1},x_{i3}\right)\\
&\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\}\\
&\equiv\mathbb{E}\left\{\mathbb{E}\left[(y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]\varphi(x_{i1},x_{i3})\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}\\
&=\mathbb{E}\{(\mathbb{E}\left[y_{i3}|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]-\mathbb{E}\left[y_{i1}|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right])\\
&\varphi(x_{i1},x_{i3})\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\}\\
&=\mathbb{E}\{(\mathbb{E}\left[y_{i3}|y_{i2}=y_{i4},x_{i3}\right]-\mathbb{E}\left[y_{i1}|y_{i0}=y_{i2},x_{i1}\right])\\
&\varphi(x_{i1},x_{i3})\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\},
\end{aligned}
$$

where in the second equality $\overline{A}$ denotes the complement of the set A,

$$
\varphi(x_{i1},x_{i3})\equiv P\left(y_{i0}=y_{i2}=y_{i4}|x_{i1},x_{i3}\right)
$$

in the fourth equality, and the sixth equality follows the same argument as in the proof of Proposition 2.2.

By the stationary condition, the following is true

$$
\mathbb{E}\left[y_{i3}|y_{i2}=y_{i4},x_{i3}=x\right]=\mathbb{E}\left[y_{i1}|y_{i0}=y_{i2},x_{i1}=x\right].
$$

Let

$$\phi(x) \equiv \mathbb{E}[y_{i3}|y_{i2}=y_{i4}, x_{i3}=x] = \mathbb{E}[y_{i1}|y_{i0}=y_{i2}, x_{i1}=x].$$

With the introduction of the above notation,

$$\mathbb{E}(\xi_i(b)) = \mathbb{E}\left\{\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))\left(1\left[x_{i31}'b > 0\right] - 1\left[x_{i31}'\beta > 0\right]\right)\right\}. \tag{D.18}$$

From the results in the proof of Proposition 2.2, $\phi(x_{i3}) - \phi(x_{i1}) > 0$ if $x_{i31}'\beta > 0$, $\phi(x_{i3}) - \phi(x_{i1}) = 0$ if $x_{i31}'\beta = 0$, and $\phi(x_{i3}) - \phi(x_{i1}) < 0$ if $x_{i31}'\beta < 0$. $\varphi(x_{i1}, x_{i3})$ is a conditional probability, so $\varphi(x_{i1}, x_{i3}) \geq 0$. The above observations imply that $\mathbb{E}(\xi_i(b))$ is nonpositive and is equal to 0 if $b = \beta$. Assumption A ensures that the solution is unique. To simply notations, let

$$\kappa(x_{i31}) \equiv \mathbb{E}[\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))|x_{i31}]. \tag{D.19}$$

Easy to see that κ defined here is equal to the κ in the body of Lemma B.3. The above discussion implies $\kappa(x_{i31})$ has the same sign as $x_{i31}'\beta$.

On Assumption M.i in Seo and Otsu (2018). We now try to get the derivatives of $\mathbb{E}(\xi_i(b))$ with respect to b. $\mathbb{E}(\xi_i(b))$ can be rewritten as

$$\mathbb{E}(\xi_i(b)) = \mathbb{E}\left\{\kappa(x_{i31})\left(1\left[x_{i31}'b > 0\right] - 1\left[x_{i31}'\beta > 0\right]\right)\right\}.$$

Following the same idea in Section 5 and Section 6.4 of Kim and Pollard (1990) and Section B.1 of Seo and Otsu (2018), the above expectation can be calculated by means of the classical differential geometry. Since the results here are obtained using essentially the same argument, we omit similar details. Define the following mapping

$$T_b = \left(I - \|b\|_2^{-2} bb'\right)(I - \beta\beta') + \|b\|_2^{-2} b\beta',$$

where T_b maps the region $\{x_{31} : x_{31}'b > 0\}$ onto $\{x_{31} : x_{31}'\beta > 0\}$, taking the boundary of $\{x_{31} : x_{31}'b > 0\}$ onto the boundary of $\{x_{31} : x_{31}'\beta > 0\}$. Equations (5.2) and (5.3) in Kim and Pollard (1990) imply

$$\frac{\partial}{\partial b}\mathbb{E}(\xi_i(b)) = \|b\|_2^{-2} b'\beta\left(I - \|b\|_2^{-2} bb'\right)\int 1\left[x_{31}'\beta = 0\right]\kappa(T_b x_{31}) x_{31} f_{x_{31}}(T_b x_{31}) d\sigma_0,$$

where $f_{x_{31}}(x_{31})$ is the density function of x_{i31} and σ_0 is the surface measure of the boundary of $\{x_{31} : x_{31}'\beta > 0\}$.

$\frac{\partial}{\partial b}\mathbb{E}(\xi_i(b))\big|_{b=\beta} = 0$, by $T_\beta x_{31} = x_{31}$ and $1[x_{31}'\beta = 0]\kappa(x_{31}) = 0$. Consequently, the nonzero component of the second derivative of $\mathbb{E}(\xi_i(b))$ only comes from the derivative of $\kappa(T_b x_{31})$. Notice that $\frac{\partial}{\partial b}\kappa(T_b x_{31})\big|_{b=\beta} = -\left(\frac{\partial \kappa(x_{31})}{\partial x_{31}}'\beta\right)x_{31}$, we have

$$\left.\frac{\partial^2 \mathbb{E}(\xi_i(b))}{\partial b \partial b'}\right|_{b=\beta} = -\int 1\left[x_{31}'\beta = 0\right]\left(\frac{\partial \kappa(x_{31})}{\partial x_{31}}'\beta\right) f_{x_{31}}(x_{31}) x_{31} x_{31}' d\sigma_0.$$

Combining these results on the derivatives of $\mathbb{E}(\xi_i(b))$ implies that Assumption M.i in Seo and Otsu (2018) is satisfied with the matrix

$$V_1 \equiv -\int 1\left[x_{31}'\beta = 0\right]\left(\frac{\partial \kappa(x_{31})}{\partial x_{31}}'\beta\right) f_{x_{31}}(x_{31}) x_{31} x_{31}' d\sigma_0. \tag{D.20}$$

V_1 is negative definite by Assumption 4.

On Assumption M.ii in Seo and Otsu (2018). Note

$$\xi_i(b_1) - \xi_i(b_2) = 1\left[y_{i0} = y_{i2} = y_{i4}\right](y_{i3} - y_{i1})\left(1\left[x_{i31}' b_1 > 0\right] - 1\left[x_{i31}' b_2 > 0\right]\right)$$

and

$$(\xi_i(b_1) - \xi_i(b_2))^2 = 1\left[y_{i0} = y_{i2} = y_{i4}\right]|y_{i3} - y_{i1}|\left|1\left[x_{i31}' b_1 > 0\right] - 1\left[x_{i31}' b_2 > 0\right]\right|, \tag{D.21}$$

this condition can be verified by the following calculation,

$$\begin{aligned}
&\left[\mathbb{E}(\xi_i(b_1) - \xi_i(b_2))^2\right]^{1/2} \\
&= \left[\mathbb{E}\left\{\mathbb{E}\left[|\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))|\,|x_{i31}\right]\left|1\left[x_{i31}' b_1 > 0\right] - 1\left[x_{i31}' b_2 > 0\right]\right|\right\}\right]^{1/2} \\
&\geq \mathbb{E}\left\{\mathbb{E}\left[|\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))|\,|x_{i31}\right]\left|1\left[x_{i31}' b_1 > 0\right] - 1\left[x_{i31}' b_2 > 0\right]\right|\right\} \\
&\geq c_1 \mathbb{E}\left|1\left[x_{i31}' b_1 > 0\right] - 1\left[x_{i31}' b_2 > 0\right]\right| \\
&\geq c_2 \|b_1 - b_2\|_2,
\end{aligned}$$

where the second line holds because the the value of the term in that line is smaller than 1, and a positive c_1 and c_2 can be guaranteed by Assumption A.

On Assumption M.iii in Seo and Otsu (2018). This condition can be similarly verified by

$$\begin{aligned}
&\mathbb{E}\left[\sup_{b_1, b_2 \in \mathcal{B}: \|b_1 - b_2\| < \varepsilon} |\xi_i(b_1) - \xi_i(b_2)|^2\right] \\
&= \mathbb{E}\left\{\sup_{b_1, b_2 \in \mathcal{B}: \|b_1 - b_2\| < \varepsilon} \mathbb{E}\left[|\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))|\,|x_{i31}\right]\left|1\left[x_{i31}' b_1 > 0\right] - 1\left[x_{i31}' b_2 > 0\right]\right|\right\} \\
&\leq c_3 \mathbb{E}\left\{\sup_{b_1 \in \mathcal{B}: \|b_1 - b_2\| < \varepsilon} \left|1\left[x_{i31}' b_1 > 0\right] - 1\left[x_{i31}' b_2 > 0\right]\right|\right\} \\
&\leq c_4 \varepsilon,
\end{aligned}$$

where third line holds because φ and ϕ are conditional probability and are bounded, and the last line holds since the density of x_{31} is assumed to be bounded in Assumption 3. □

Proof of Lemma B.2. The objective function in this lemma is very similar to the one in HK**.** The only difference is that HK put x_{32} in the kernel $\mathcal{K}_{h_n}(\cdot)$ while we put $x_{32}' b$ and $x_{43}' b$ instead.

Seo and Otsu (2018) verified all the technical conditions needed for the estimator in HK and derived its asymptotics in Section B.1. Assumptions A and 3 - 6 can imply the technical conditions assumed in Section B.1 of Seo and Otsu (2018), and the conclusion follows. □

Proof of Lemma B.3. Note that

$$Z_{n,1}(\boldsymbol{s}) = n^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right) = n^{1/6} \mathbb{G}_n \left(\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right)\right) + n^{2/3} \mathbb{E} \left(\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right)\right),$$

where $\mathbb{G}_n \left(\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right)\right) = n^{-1/2} \sum_{i=1}^{n} \left[\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right) - \mathbb{E}\left(\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right)\right)\right]$.

The mean of $Z_{n,1}(\boldsymbol{s})$ is $n^{2/3}\mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s} n^{-1/3}\right)\right)$. With Assumptions A and 3, some calculation in the proof of Lemma B.1 yields

$$\begin{aligned}
& n^{2/3} \mathbb{E}\left(\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right)\right) \\
&= n^{2/3} \left\{ \mathbb{E}\left(\xi_i(\beta)\right) + n^{-1/3} \left.\frac{\partial \mathbb{E}\left(\xi_i(b)\right)}{\partial b}\right|_{b=\beta}' \boldsymbol{s} + \frac{1}{2} n^{-2/3} \boldsymbol{s}' \left.\frac{\partial^2 \mathbb{E}\left(\xi_i(b)\right)}{\partial b \partial b'}\right|_{b=\beta} \boldsymbol{s} + o\left(n^{-2/3}\right) \right\} \\
&= \frac{1}{2} \boldsymbol{s}' V_1 \boldsymbol{s} + o(1),
\end{aligned}$$

where V_1 is defined in equation (B.1).

By definition, $H_1(\boldsymbol{s}, \boldsymbol{t}) = \lim_{\alpha\to\infty} \alpha \mathbb{E}\left[\xi_i(\beta + \boldsymbol{s}/\alpha)\, \xi_i(\beta + \boldsymbol{t}/\alpha)\right]$ is the covariance kernel for the limiting distribution of $Z_{n,1}(\boldsymbol{s})$. To obtain H_1, define

$$\begin{aligned}
L_1(\boldsymbol{s} - \boldsymbol{t}) &\equiv \lim_{\alpha\to\infty} \alpha \mathbb{E}\left[\left(\xi_i(\beta + \boldsymbol{s}/\alpha) - \xi_i(\beta + \boldsymbol{t}/\alpha)\right)^2\right], \\
L_1(\boldsymbol{s}) &\equiv \lim_{\alpha\to\infty} \alpha \mathbb{E}\left[\left(\xi_i(\beta + \boldsymbol{s}/\alpha) - \xi_i(\beta)\right)^2\right],
\end{aligned}$$

and

$$L_1(\boldsymbol{t}) \equiv \lim_{\alpha\to\infty} \alpha \mathbb{E}\left[\left(\xi_i(\beta + \boldsymbol{t}/\alpha) - \xi_i(\beta)\right)^2\right].$$

Notice that $\xi_i(\beta) = 0$, the relationship between H_1 and L_1 is

$$H_1(\boldsymbol{s}, \boldsymbol{t}) = \frac{1}{2}\left[L_1(\boldsymbol{s}) + L_1(\boldsymbol{t}) - L_1(\boldsymbol{s} - \boldsymbol{t})\right]. \tag{D.22}$$

From equations (D.18) and (D.21),

$$\begin{aligned}
& \alpha \mathbb{E}\left[\left(\xi_i(\beta + \boldsymbol{s}/\alpha) - \xi_i(\beta + \boldsymbol{t}/\alpha)\right)^2\right] \\
&= \alpha \mathbb{E}\left\{\mathbb{E}\left[\left|\varphi(x_{i1}, x_{i3})\left(\phi(x_{i3}) - \phi(x_{i1})\right)\right| |x_{i31}\right] \left|1\left[x_{i31}'(\beta + \boldsymbol{s}/\alpha) > 0\right] - 1\left[x_{i31}'(\beta + \boldsymbol{t}/\alpha) > 0\right]\right|\right\} \\
&\equiv \alpha \mathbb{E}\left\{\psi(x_{i31}) \left|1\left[x_{i31}'(\beta + \boldsymbol{s}/\alpha) > 0\right] - 1\left[x_{i31}'(\beta + \boldsymbol{t}/\alpha) > 0\right]\right|\right\}.
\end{aligned}$$

where in the third line we simply notations by

$$\psi(x_{i31}) \equiv \mathbb{E}\left[\left|\varphi(x_{i1}, x_{i3})\left(\phi(x_{i3}) - \phi(x_{i1})\right)\right| |x_{i31}\right].$$

Not hard to see that ψ defined here is equal to the ψ in the body of this lemma. Following Kim and Pollard (1990), we decompose x_{31} into $\varpi\beta + x_\beta$, with x_β orthogonal to β. The decomposition leads to

$$
\begin{aligned}
&\alpha \mathbb{E}\left[\left(\xi_i\left(\beta+\boldsymbol{s}/\alpha\right)-\xi_i\left(\beta+\boldsymbol{t}/\alpha\right)\right)^2\right] \\
&=\alpha \mathbb{E}\left\{\psi\left(x_{i31}\right)\left|1\left[x_{i31}'\left(\beta+\boldsymbol{s}/\alpha\right)>0\right]-1\left[x_{i31}'\left(\beta+\boldsymbol{t}/\alpha\right)>0\right]\right|\right\} \\
&=\alpha \int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} \psi\left(\varpi\beta+x_\beta\right)\left|1\left[x_\beta'\boldsymbol{s}/\alpha+\varpi+\varpi\beta'\boldsymbol{s}/\alpha>0\right]-1\left[x_\beta'\boldsymbol{t}/\alpha+\varpi+\varpi\beta'\boldsymbol{t}/\alpha>0\right]\right| \\
&f_{x_{31}}\left(\varpi\beta+x_\beta\right) d\varpi dx_\beta \\
&=\alpha \int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} \psi\left(\varpi\beta+x_\beta\right) 1\left[\frac{-x_\beta'\boldsymbol{s}/\alpha}{1+\beta'\boldsymbol{s}/\alpha}<\varpi \leq \frac{-x_\beta'\boldsymbol{t}/\alpha}{1+\beta'\boldsymbol{t}/\alpha}\right] f_{x_{31}}\left(\varpi\beta+x_\beta\right) d\varpi dx_\beta \\
&+\alpha \int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} \psi\left(\varpi\beta+x_\beta\right) 1\left[\frac{-x_\beta'\boldsymbol{t}/\alpha}{1+\beta'\boldsymbol{t}/\alpha}<\varpi \leq \frac{-x_\beta'\boldsymbol{s}/\alpha}{1+\beta'\boldsymbol{s}/\alpha}\right] f_{x_{31}}\left(\varpi\beta+x_\beta\right) d\varpi dx_\beta \\
&=\int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} \psi\left(u/\alpha\beta+x_\beta\right) 1\left[\frac{-x_\beta'\boldsymbol{s}}{1+\beta'\boldsymbol{s}/\alpha}<u \leq \frac{-x_\beta'\boldsymbol{t}}{1+\beta'\boldsymbol{t}/\alpha}\right] f_{x_{31}}\left(\left(u/\alpha\right)\beta+x_\beta\right) du dx_\beta \\
&+\int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} \psi\left(u/\alpha\beta+x_\beta\right) 1\left[\frac{-x_\beta'\boldsymbol{t}}{1+\beta'\boldsymbol{t}/\alpha}<u \leq \frac{-x_\beta'\boldsymbol{s}}{1+\beta'\boldsymbol{s}/\alpha}\right] f_{x_{31}}\left(\left(u/\alpha\right)\beta+x_\beta\right) du dx_\beta,
\end{aligned}
$$

where the fourth equality holds by the change of variables $u=\alpha\varpi$, . As $\alpha\to\infty$,

$$
L_1\left(\boldsymbol{s}-\boldsymbol{t}\right)=\int_{\mathbb{R}^{K-1}} \psi\left(x_\beta\right)\left|x_\beta'\left(\boldsymbol{s}-\boldsymbol{t}\right)\right| f_{x_{31}}\left(x_\beta\right) dx_\beta,
$$

Similarly,

$$
L_1\left(\boldsymbol{s}\right)=\int_{\mathbb{R}^{K-1}} \psi\left(x_\beta\right)\left|x_\beta'\boldsymbol{s}\right| f_{x_{31}}\left(x_\beta\right) dx_\beta
$$

and

$$
L_1\left(\boldsymbol{t}\right)=\int_{\mathbb{R}^{K-1}} \psi\left(x_\beta\right)\left|x_\beta'\boldsymbol{t}\right| f_{x_{31}}\left(x_\beta\right) dx_\beta .
$$

Substituting those L_1 into equation (D.22) yields

$$
H_1\left(\boldsymbol{s},\boldsymbol{t}\right)=\frac{1}{2}\int_{\mathbb{R}^{K-1}} \psi\left(x_\beta\right)\left[\left|x_\beta'\boldsymbol{s}\right|+\left|x_\beta'\boldsymbol{t}\right|-\left|x_\beta'\left(\boldsymbol{s}-\boldsymbol{t}\right)\right|\right] f_{x_{31}}\left(x_\beta\right) dx_\beta .
$$

□

Proof of Lemma B.4. Note

$$
\begin{aligned}
\hat{Z}_{n,2}\left(s\right) &=\left(nh_n\right)^{2/3}\cdot n^{-1}\sum_{i=1}^n \varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right) \\
&=n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)\right)+\left(nh_n\right)^{2/3}\mathbb{E}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)\right) \\
&=n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)+\left(nh_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right) \\
&+n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right) \\
&+\left(nh_n\right)^{2/3}\mathbb{E}_n\left[\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right],
\end{aligned}
\tag{D.23}
$$

where $\mathbb{G}_n\left(\varsigma_{ni}\left(r,b\right)\right)=n^{-1/2}\sum_{i=1}^{n}\left(\varsigma_{ni}\left(r,b\right)-\mathbb{E}_n\left(\varsigma_{ni}\left(r,b\right)\right)\right)$.

We deal with the term in the fourth line of equation (D.23) first. Lemma B.2 verifies the technical conditions in Seo and Otsu (2018). Thus we can applying the result of Lemma M in Seo and Otsu (2018) on ς and it yields[26]

$$\begin{aligned}
&\mathbb{E}\left\{\sup_{|s|\le C,\|b-\beta\|_2\le Mn^{-1/3}} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|\right\}\\
&=n^{1/6}h_n^{1/6}\mathbb{E}\left\{\sup_{|s|\le C,\|b-\beta\|_2\le Mn^{-1/3}}\left|\mathbb{G}_n\left[h_n^{1/2}\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|\right\}\\
&\le cn^{1/6}h_n^{1/6}n^{-1/6}=o\left(1\right),
\end{aligned}$$

for some positive c, any positive constants M and C. By Markov's inequality, the above yields

$$\sup_{|s|\le C,\|b-\beta\|_2\le Mn^{-1/3}} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|=o_P\left(1\right).$$

Since $\hat{\beta}-\beta=O_P\left(n^{-1/3}\right)$, we can take M large enough so that $P\left(\left\|\hat{\beta}-\beta\right\|_2>Mn^{-1/3}\right)<\varepsilon$ for any small $\varepsilon>0$. For any small $\delta>0$,

$$\begin{aligned}
&P\left(\sup_{|s|\le C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|\ge\delta\right)\\
&=P\left(\left\{\sup_{|s|\le C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|\ge\delta\right\}\right.\\
&\left.\cap\left\{\left\|\hat{\beta}-\beta\right\|_2\le Mn^{-1/3}\right\}\right)\\
&+P\left(\left\{\sup_{|s|\le C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|\ge\delta\right\}\right.\\
&\left.\cap\left\{\left\|\hat{\beta}-\beta\right\|_2> Mn^{-1/3}\right\}\right)\\
&\le P\left(\sup_{|s|\le C,\|b-\beta\|_2\le Mn^{-1/3}} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|\ge\delta\right)+\varepsilon.
\end{aligned}$$

In view of the fact that the first term in the above last line can be arbitrary small as $n\to\infty$, after some large n

$$P\left(\sup_{|s|\le C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|\ge\delta\right)\le 2\varepsilon,$$

and it holds for any arbitrary small $\delta>0$. This implies

$$\sup_{|s|\le C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)\right]\right|=o_P\left(1\right). \tag{D.24}$$

[26] It holds by setting the δ in Lemma M of Seo and Otsu (2018) as $n^{-1/3}$.

For the fourth term in equation (D.23), with $\hat{\beta}-\beta=O_P\left(n^{-1/3}\right)$ and $h_n\to 0$, the expansion in equation (D.29) implies

$$(nh_n)^{2/3}\,\mathbb{E}_n\left(\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\hat{\beta}\right)\right)=(nh_n)^{2/3}\,\mathbb{E}\left(\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)\right)+o_P\left(1\right), \quad \text{(D.25)}$$

uniformly over $|s|\le C$. Substituting the results of equations (D.24) and (D.25) into equation (D.23) yields,

$$\begin{aligned}\hat{Z}_{n,2}\left(s\right)&=n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)\right)+(nh_n)^{2/3}\,\mathbb{E}\left(\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)\right)+o_P\left(1\right)\\&=Z_{n,2}\left(s\right)+o_P\left(1\right),\end{aligned}$$

where the small order term holds uniformly over $|s|\le C$ for any positive C. The claim is proved.

□

Proof of Lemma B.5. We could prove the first claim in this lemma by the Taylor expansion of $\mathbb{E}\left(\varsigma_{ni}\left(r,\beta\right)\right)$ with respect to r around γ. We show a more general result instead; we derive the Taylor expansion of $\mathbb{E}\left(\varsigma_{ni}\left(r,b\right)\right)$ with respect to (r,b) around (γ,β). This more general result is useful for understanding Lemma B.5 and part of the derivation in Lemma B.4.

Recall that

$$\begin{aligned}\varsigma_{ni}\left(r,b\right)\equiv{}&\mathcal{K}_{h_n}\left(x_{i32}'b\right)\left(y_{i2}-y_{i1}\right)\left(1\left[x_{i21}'b+r\left(y_{i3}-y_{i0}\right)>0\right]-1\left[x_{i21}'\beta+\gamma\left(y_{i3}-y_{i0}\right)>0\right]\right)\\&+\mathcal{K}_{h_n}\left(x_{i43}'b\right)\left(y_{i3}-y_{i2}\right)\left(1\left[x_{i32}'b+r\left(y_{i4}-y_{i1}\right)>0\right]-1\left[x_{i32}'\beta+\gamma\left(y_{i4}-y_{i1}\right)>0\right]\right).\end{aligned}$$

To ease of notations, let

$$\begin{aligned}\vartheta_1\left(r,b\right)&\equiv\left(y_2-y_1\right)\left(1\left[x_{21}'b+r\left(y_3-y_0\right)>0\right]-1\left[x_{21}'\beta+\gamma\left(y_3-y_0\right)>0\right]\right),\\\vartheta_2\left(r,b\right)&\equiv\left(y_3-y_2\right)\left(1\left[x_{32}'b+r\left(y_4-y_1\right)>0\right]-1\left[x_{32}'\beta+\gamma\left(y_4-y_1\right)>0\right]\right).\end{aligned}$$

We deal with the first component in $\varsigma_{ni}\left(r,b\right)$ first and the second term can be done analogously. First,

$$\begin{aligned}&\mathbb{E}\left[\mathcal{K}_{h_n}\left(x_{32}'b\right)\vartheta_1\left(r,b\right)\right]\\&=\int_{\mathbb{R}^K}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x\right]\mathcal{K}_{h_n}\left(x'b\right)f_{x_{32}}\left(x\right)dx\\&=\int_{\mathbb{R}^K}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x\right]\frac{1}{h_n}\mathcal{K}\left(\frac{x'b}{h_n}\right)f_{x_{32}}\left(x\right)dx.\end{aligned}$$

Decompose x_{32} into $x_{32}=\varpi b+x_b$, where x_b is orthogonal to b. That yields

$$\mathbb{E}\left[\mathcal{K}_{h_n}\left(x_{32}'b\right)\vartheta_1\left(r,b\right)\right]=\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=\varpi b+x_b\right]\frac{1}{h_n}\mathcal{K}\left(\frac{\varpi}{h_n}\right)f_{x_{32}}\left(\varpi b+x_b\right)d\varpi dx_b \quad \text{(D.26)}$$

$$\begin{aligned}&=\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=uh_nb+x_b\right]\mathcal{K}\left(u\right)f_{x_{32}}\left(uh_nb+x_b\right)dudx_b\\&=\int_{\mathbb{R}^{K-1}}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right]f_{x_{32}}\left(x_b\right)dx_b\\&+\frac{h_n^2}{2}\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}u^2\mathcal{K}\left(u\right)\left.\frac{\partial^2\left(\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=tb+x_b\right]f_{x_{32}}\left(tb+x_b\right)\right)}{\partial t^2}\right|_{t=t_u}du\end{aligned}$$

where in the first line we use the fact $\|b\|_2 = 1$, the second line holds by the change of variables $u = \frac{\varpi}{h_n}$, and last two lines hold by the Taylor expansion and t_u is some value between 0 and uh_n. The bias term is of order h_n^2 by Assumption 3 and the symmetry and boundedness conditions of $\mathcal{K}$ in Assumption 5. By $nh_n^4 \to 0$ in Assumption 6, the bias term is $o\left((nh_n)^{-2/3}\right)$ and asymptotically negligible.

Similar results can be obtained for $\mathbb{E}\left[\mathcal{K}_{h_n}\left(x_{43}'b\right)\vartheta_2\left(r,b\right)\right]$.

To summarise,

$$\begin{aligned}\mathbb{E}\left(\varsigma_{ni}\left(r,b\right)\right) = &\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b \\ &+ \int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_2\left(r,b\right)|x_{43}=x_b\right] f_{x_{43}}\left(x_b\right) dx_b + o\left((nh_n)^{-2/3}\right).\end{aligned} \tag{D.27}$$

As a result, to prove the assertion in the lemma, it is enough to derive the first and second derivatives of leading term in the above..

Notice that

$$\vartheta_1|_{(r,b)=(\gamma,\beta)} = 0.$$

Consequently, only the derivative of $E\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right]$ with respect to b in ϑ_1 will appear in

$$\frac{\partial}{\partial b}\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b\bigg|_{r=\gamma,b=\beta}.$$

That leads to

$$\begin{aligned}&\frac{\partial}{\partial b}\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b\bigg|_{r=\gamma,b=\beta} \\ &= \int_{\mathbb{R}^{K-1}} \frac{\partial}{\partial b}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_\beta\right]\bigg|_{(r,b)=(\gamma,\beta)} f_{x_{32}}\left(x_\beta\right) dx_\beta.\end{aligned}$$

By similar derivation as for the derivatives of $\mathbb{E}\left(\xi_i\left(b\right)\right)$, we have

$$\begin{aligned}&\frac{\partial \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_\beta\right]}{\partial\left(r,b'\right)'}\bigg|_{r=\gamma,b=\beta} \\ &= \int 1\left[x_{21}'\beta + \gamma y_{30} = 0\right] \mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)\begin{pmatrix} y_{30} \\ x_{21}\end{pmatrix} f\left(x_{21},y_{30}|x_{32}=x_\beta\right) d\sigma_0,\end{aligned}$$

where σ_0 is the surface measure of $\{(x_{21},y_{30}) : x_{21}'\beta + \gamma y_{30} = 0\}$.

$\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}'\beta=0\right) = 0$ along $x_{21}'\beta + \gamma y_{30} = 0$ by Proposition 2.3. Thus the derivative above is equal to 0 and

$$\frac{\partial}{\partial\left(r,b'\right)'}\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b\bigg|_{r=\gamma,b=\beta} = 0.$$

The fact $\mathbb{E}\left(y_{21}|x_{21}, y_{30}, x_{32}'\beta = 0\right) = 0$ along $x_{21}'\beta + \gamma y_{30} = 0$ implies that only the second derivatives of $\mathbb{E}\left[\vartheta_1\left(r, b\right)|x_{32} = x_\beta\right]$ contribute to the second derivative. By similar derivation as for the second derivative of $\mathbb{E}\left(\xi_i\left(b\right)\right)$,

$$\left.\frac{\partial^2 \mathbb{E}\left[\vartheta_1\left(r, b\right)|x_{32} = x_\beta\right]}{\partial\left(r, b'\right)'\partial\left(r, b'\right)}\right|_{r=\gamma, b=\beta}$$

$$= -\int 1\left[x_{21}'\beta + \gamma y_{30} = 0\right]\left(\frac{\partial \mathbb{E}\left(y_{21}|x_{21}, y_{30}, x_{32} = x_\beta\right)'}{\partial\left(y_{30}, x_{21}'\right)'}\begin{pmatrix}\gamma \\ \beta\end{pmatrix}\right)$$

$$f\left(x_{21}, y_{30}|x_{32} = x_\beta\right)\begin{pmatrix}y_{30} \\ x_{21}\end{pmatrix}\begin{pmatrix}y_{30} & x_{21}'\end{pmatrix}d\sigma_0.$$

Therefore

$$\left.\frac{\partial^2}{\partial\left(r, b'\right)'\partial\left(r, b'\right)}\int_{\mathbb{R}^{K-1}}\mathbb{E}\left[\vartheta_1\left(r, b\right)|x_{32} = x_b\right]f_{x_{32}}\left(x_b\right)dx_b\right|_{r=\gamma, b=\beta}$$

$$= -\int_{\mathbb{R}^{K-1}}\int 1\left[x_{21}'\beta + \gamma y_{30} = 0\right]\left(\frac{\partial \mathbb{E}\left(y_{21}|x_{21}, y_{30}, x_{32} = x_\beta\right)'}{\partial\left(y_{30}, x_{21}'\right)'}\begin{pmatrix}\gamma \\ \beta\end{pmatrix}\right)$$

$$f\left(x_{21}, y_{30}|x_{32} = x_\beta\right)\begin{pmatrix}y_{30} \\ x_{21}\end{pmatrix}\begin{pmatrix}y_{30} & x_{21}'\end{pmatrix}d\sigma_0 f_{x_{32}}\left(x_\beta\right)dx_\beta$$

$$\equiv -\tilde{V}_{21}.$$

Similarly,

$$\left.\frac{\partial^2}{\partial\left(r, b'\right)'\partial\left(r, b'\right)}\int_{\mathbb{R}^{K-1}}\mathbb{E}\left[\vartheta_2\left(r, b\right)|x_{43} = x_b\right]f_{x_{43}}\left(x_b\right)dx_b\right|_{r=\gamma, b=\beta}$$

$$= -\int_{\mathbb{R}^{K-1}}\int 1\left[x_{32}'\beta + \gamma y_{41} = 0\right]\left(\frac{\partial \mathbb{E}\left(y_{32}|x_{32}, y_{41}, x_{43} = x_\beta\right)'}{\partial\left(y_{41}, x_{32}'\right)'}\begin{pmatrix}\gamma \\ \beta\end{pmatrix}\right)$$

$$f\left(x_{32}, y_{41}|x_{43} = x_\beta\right)\begin{pmatrix}y_{41} \\ x_{32}\end{pmatrix}\begin{pmatrix}y_{41} & x_{32}'\end{pmatrix}d\sigma_0 f_{x_{43}}\left(x_\beta\right)dx_\beta$$

$$\equiv -\tilde{V}_{22}.$$

Let

$$\tilde{V}_2 \equiv \tilde{V}_{21} + \tilde{V}_{22}. \tag{D.28}$$

By the Taylor expansion, Assumption 3, and equation (D.27),

$$\mathbb{E}\left(\varsigma_{ni}\left(r, b\right)\right) = -\frac{1}{2}\left(r - \gamma, \left(b - \beta\right)'\right)\tilde{V}_2\begin{pmatrix}r - \gamma \\ b - \beta\end{pmatrix} + o\left(\left\|\begin{pmatrix}r - \gamma \\ b - \beta\end{pmatrix}\right\|_2^2\right) + o\left(\left(nh_n\right)^{-2/3}\right). \tag{D.29}$$

We define V_2 as the first diagonal of $\tilde{V}_2$, that is

$$V_2 \equiv e_1'\tilde{V}_2 e_1,$$

where e_1 is a $(K+1)\times 1$ vector with the first element as 1 and the rest as 0. Not hard to see that

$$V_2 = -\int_{\mathbb{R}^{K-1}}\int 1\left[x_{21}'\beta+\gamma y_{30}=0\right]\left(\frac{\partial\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)'}{\partial\left(y_{30},x_{21}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right) \tag{D.30}$$
$$f\left(x_{21},y_{30}|x_{32}=x_\beta\right)|y_{30}|\,d\sigma_0 f_{x_{32}}\left(x_\beta\right)dx_\beta$$
$$-\int_{\mathbb{R}^{K-1}}\int 1\left[x_{32}'\beta+\gamma y_{41}=0\right]\left(\frac{\partial\mathbb{E}\left(y_{32}|x_{32},y_{41},x_{43}=x_\beta\right)'}{\partial\left(y_{41},x_{32}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right)$$
$$f\left(x_{32},y_{41}|x_{43}=x_\beta\right)|y_{41}|\,d\sigma_0 f_{x_{43}}\left(x_\beta\right)dx_\beta.$$

Using equation (D.29),

$$\lim_{n\to\infty}(nh_n)^{2/3}\,\mathbb{E}_n\left(\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)\right)=\frac{1}{2}V_2 s^2.$$

Now we turn to the covariance kernel. Note

$$H_2(s,t)=\lim_{n\to\infty}(nh_n)^{1/3}\,\mathbb{E}\left(h_n\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)\varsigma_{ni}\left(\gamma+t\,(nh_n)^{-1/3},\beta\right)\right).$$

Similar for the calculation of H_1 in Lemma B.1, define

$$L_2(s-t)\equiv\lim_{n\to\infty}(nh_n)^{1/3}\,\mathbb{E}\left[h_n\left(\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)-\varsigma_{ni}\left(\gamma+t\,(nh_n)^{-1/3},\beta\right)\right)^2\right],$$
$$L_2(s)\equiv\lim_{n\to\infty}(nh_n)^{1/3}\,\mathbb{E}\left[h_n\left(\varsigma_{ni}\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)-\varsigma_{ni}(\gamma,\beta)\right)^2\right],$$
$$L_2(t)\equiv\lim_{n\to\infty}(nh_n)^{1/3}\,\mathbb{E}\left[h_n\left(\varsigma_{ni}\left(\gamma+t\,(nh_n)^{-1/3},\beta\right)-\varsigma_{ni}(\gamma,\beta)\right)^2\right].$$

Since $\varsigma_{ni}(\gamma,\beta)=0$, $H_2(s,t)=\frac{1}{2}\left[L_2(s)+L_2(t)-L_2(s-t)\right].$

The following calculation is useful for $L_2(s-t)$.

$$\begin{aligned}
&\mathbb{E}\left[h_n\left(\varsigma_{ni}(r_1,\beta)-\varsigma_{ni}(r_2,\beta)\right)^2\right]\\
&=\mathbb{E}\left\{h_n\left[\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)\left(\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right)+\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)\left(\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right)\right]^2\right\}\\
&=\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2\left|\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right|+h_n\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)^2\left|\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right|\right.\\
&\left.+2h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)\left(\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right)\left(\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right)\right\}\\
&\equiv\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2\left|\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right|+h_n\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)^2\left|\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right|\right\}+R_n.
\end{aligned}$$

where R_n denotes the term in the fourth line and will be shown to be asymptotic negligible.

The first term in the above can be calculated as follows,

$$\begin{aligned}
&\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2\left|\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right|\right\}\\
&=\int_{\mathbb{R}^K}\mathbb{E}\left[\left|\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right|\,|x_{32}=x\right]\frac{1}{h_n}\mathcal{K}\left(\frac{x'\beta}{h_n}\right)^2 f_{x_{32}}(x)\,dx.
\end{aligned}$$

Decompose x_{32} into $x_{32} = \varpi\beta + x_\beta$, where x_β is orthogonal to β. Continue the expression in the above with this decomposition,

$$
\begin{aligned}
&\mathbb{E}\left\{h_n \mathcal{K}_{h_n}\left(x_{32}'\beta\right)^2 \left|\vartheta_1\left(r_1,\beta\right) - \vartheta_1\left(r_2,\beta\right)\right|\right\} \\
&= \int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} \mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right) - \vartheta_1\left(r_2,\beta\right)\right| | x_{32} = \varpi\beta + x_\beta\right] \frac{1}{h_n} \mathcal{K}\left(\frac{\varpi}{h_n}\right)^2 f_{x_{32}}\left(\varpi\beta + x_\beta\right) d\varpi dx_\beta \\
&= \int_{\mathbb{R}^{K-1}} \int_{\mathbb{R}} \mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right) - \vartheta_1\left(r_2,\beta\right)\right| | x_{32} = uh_n\beta + x_\beta\right] \mathcal{K}\left(u\right)^2 f_{x_{32}}\left(uh_n\beta + x_\beta\right) du dx_\beta \\
&= \bar{\mathcal{K}}_2 \int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right) - \vartheta_1\left(r_2,\beta\right)\right| | x_{32} = x_\beta\right] f_{x_{32}}\left(x_\beta\right) dx_\beta + O\left(h_n^2\right)
\end{aligned}
$$

where in the third line we substitute $u = \varpi/h_n$, in the fourth line we do Taylor expansion around $h_n = 0$, the bias term is of order h_n^2 for the same reason as in equation (D.26), and $\bar{\mathcal{K}}_2 = \int_{\mathbb{R}} \mathcal{K}\left(u\right)^2 du$. Using Assumption 6, $\left(nh_n\right)^{2/3} h_n^2 \to 0$, so the bias term is negligible. The rate of the above term can be seen from

$$
\begin{aligned}
&\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right) - \vartheta_1\left(r_2,\beta\right)\right| | x_{32} = x_\beta\right] \\
&= \int_{\mathbb{R}} \mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = \varpi, y_{30} \neq 0, x_{32} = x_\beta\right] \left|1\left[\varpi + r_1\left(y_3 - y_0\right) > 0\right] - 1\left[\varpi + r_2\left(y_3 - y_0\right) > 0\right]\right| \\
&P\left(y_{30} \neq 0 | x_{32} = x_\beta, x_{21}'\beta = \varpi\right) f\left(x_{21}'\beta = \varpi | x_{32} = x_\beta\right) d\varpi \\
&= \left|\int_{-r_1}^{-r_2} \mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = \varpi, y_{30} = 1, x_{32} = x_\beta\right] P\left(y_{30} = 1 | x_{32} = x_\beta, x_{21}'\beta = \varpi\right) f\left(x_{21}'\beta = \varpi | x_{32} = x_\beta\right) d\varpi\right| \\
&+ \left|\int_{r_1}^{r_2} \mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = \varpi, y_{30} = -1, x_{32} = x_\beta\right] P\left(y_{30} = -1 | x_{32} = x_\beta, x_{21}'\beta = \varpi\right) f\left(x_{21}'\beta = \varpi | x_{32} = x_\beta\right) d\varpi\right| \\
&\propto \left|r_2 - r_1\right|.
\end{aligned}
$$

If $r_1 = \gamma + s\left(nh_n\right)^{-1/3}$ and $r_2 = \gamma + t\left(nh_n\right)^{-1/3}$, $\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right) - \vartheta_1\left(r_2,\beta\right)\right| | x_{32} = x_\beta\right] \propto \left(nh_n\right)^{-1/3}$ and

$$
\begin{aligned}
&\lim_{n\to\infty} \left(nh_n\right)^{1/3} \mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right) - \vartheta_1\left(r_2,\beta\right)\right| | x_{32} = x_\beta\right] \\
&= \left\{\mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = -\gamma, y_{30} = 1, x_{32} = x_\beta\right] P\left(y_{30} = 1 | x_{32} = x_\beta, x_{21}'\beta = -\gamma\right) f\left(x_{21}'\beta = -\gamma | x_{32} = x_\beta\right)\right. \\
&+ \left.\mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = \gamma, y_{30} = -1, x_{32} = x_\beta\right] P\left(y_{30} = -1 | x_{32} = x_\beta, x_{21}'\beta = \gamma\right) f\left(x_{21}'\beta = \gamma | x_{32} = x_\beta\right)\right\} \\
&\cdot \left|s - t\right| \\
&= \left\{\mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = -\gamma, y_{30} = 1, x_{32} = x_\beta\right] f\left(y_{30} = 1, x_{21}'\beta = -\gamma | x_{32} = x_\beta\right)\right. \\
&+ \left.\mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = \gamma, y_{30} = -1, x_{32} = x_\beta\right] f\left(y_{30} = -1, x_{21}'\beta = \gamma | x_{32} = x_\beta\right)\right\} \left|s - t\right|
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\lim_{n\to\infty} \left(nh_n\right)^{1/3} \mathbb{E}\left\{h_n \mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2 \left|\vartheta_1\left(\gamma + s\left(nh_n\right)^{-1/3}, \beta\right) - \vartheta_1\left(\gamma + t\left(nh_n\right)^{-1/3}, \beta\right)\right|\right\} \\
&= \left|s - t\right| \bar{\mathcal{K}}_2 \int_{\mathbb{R}^{K-1}} \left\{\mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = -\gamma, y_{30} = 1, x_{32} = x_\beta\right] f\left(y_{30} = 1, x_{21}'\beta = -\gamma | x_{32} = x_\beta\right)\right. \\
&+ \left.\mathbb{E}\left[\left|y_{21}\right| | x_{21}'\beta = \gamma, y_{30} = -1, x_{32} = x_\beta\right] f\left(y_{30} = -1, x_{21}'\beta = \gamma | x_{32} = x_\beta\right)\right\} f_{x_{32}}\left(x_\beta\right) dx_\beta.
\end{aligned}
$$

For the same reason,

$$\lim_{n\to\infty}(nh_n)^{1/3}\,\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x'_{i43}\beta\right)^2\left|\vartheta_2\left(\gamma+s\,(nh_n)^{-1/3},\beta\right)-\vartheta_2\left(\gamma+t\,(nh_n)^{-1/3},\beta\right)\right|\right\}$$
$$=|s-t|\,\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{32}|\,|x'_{32}\beta=-\gamma,y_{41}=1,x_{43}=x_\beta\right]f\left(y_{41}=1,x'_{32}\beta=-\gamma|x_{43}=x_\beta\right)\right.$$
$$\left.+\mathbb{E}\left[|y_{32}|\,|x'_{32}\beta=\gamma,y_{41}=-1,x_{43}=x_\beta\right]f\left(y_{41}=-1,x'_{32}\beta=\gamma|x_{43}=x_\beta\right)\right\}f_{x_{43}}\left(x_\beta\right)dx_\beta.$$

Similar derivation on $R_n=2h_n\mathbb{E}\left[\mathcal{K}_{h_n}\left(x'_{i32}\beta\right)\mathcal{K}_{h_n}\left(x'_{i43}\beta\right)\left(\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right)\left(\vartheta_2\left(r_1,\beta\right)-\vartheta_2\left(r_2,\beta\right)\right)\right]$ can show that $R_n\propto(nh_n)^{-2/3}h_n$ when $r_1=\gamma+s\,(nh_n)^{-1/3}$ and $r_2=\gamma+t\,(nh_n)^{-1/3}$. So $(nh_n)^{1/3}R_n\to0$, as $n\to\infty$.

The results on $L_2(s-t)$ lead to

$$L_2(s-t)$$
$$=|s-t|\,\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{21}|\,|x'_{21}\beta=-\gamma,y_{30}=1,x_{32}=x_\beta\right]f\left(y_{30}=1,x'_{21}\beta=-\gamma|x_{32}=x_\beta\right)\right.$$
$$\left.+\,\mathbb{E}\left[|y_{21}|\,|x'_{21}\beta=\gamma,y_{30}=-1,x_{32}=x_\beta\right]f\left(y_{30}=-1,x'_{21}\beta=\gamma|x_{32}=x_\beta\right)\right\}f_{x_{32}}\left(x_\beta\right)dx_\beta$$
$$+\,|s-t|\,\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{32}|\,|x'_{32}\beta=-\gamma,y_{41}=1,x_{43}=x_\beta\right]f\left(y_{41}=1,x'_{32}\beta=-\gamma|x_{43}=x_\beta\right)\right.$$
$$\left.+\mathbb{E}\left[|y_{32}|\,|x'_{32}\beta=\gamma,y_{41}=-1,x_{43}=x_\beta\right]f\left(y_{41}=-1,x'_{32}\beta=\gamma|x_{43}=x_\beta\right)\right\}f_{x_{43}}\left(x_\beta\right)dx_\beta.$$

$L_2(s)$ and $L_2(t)$ can be obtained by

$$L_2(s)=L_2(s-0)\,,$$
$$L_2(t)=L_2(t-0)\,.$$

As a result

$$H_2(s,t)=\frac{1}{2}\left[L_2(s)+L_2(t)-L_2(s-t)\right],$$

which can be written as in equation (B.4). □

E Some Technical Details for Section 5

E.1 Numerical Bootstrap

If $\varepsilon_n=n^{-1}$, the numerical bootstrap is reduced to the classic bootstrap. Numerical bootstrap excludes the case $\varepsilon_n=n^{-1}$ and requires $n\varepsilon_n\to\infty$. The idea of numerical bootstrap is similar to the m-out-of-n bootstrap; ε_n^{-1} plays a similar role as m. As was shown in Hong and Li (2020), this procedure is less general than the m-out-of-n procedure. However, once it works, it has better finite sample performance than the m-out-of-n bootstrap. We refer to Hong and Li (2020) for the details.

Below is a heuristic illustration for why numerical bootstrap works for $\hat{\beta}$. $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right)$ can be shown to be $O_P(1)$ similarly as in Section E.4. Note that

$$\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \hat{\beta}\right) = \varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right) - \varepsilon_n^{-1/3}\left(\hat{\beta} - \beta\right) = \varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right) + o_P(1) \tag{E.1}$$

by $n\varepsilon_n \to \infty$. Thus the asymptotic distribution $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \hat{\beta}\right)$ is the same as that of $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right)$. Let

$$\mathcal{L}_{n,1}^*(b) \equiv n^{-1}\sum_{i=1}^n \xi_i(b) + (n\varepsilon_n)^{1/2} \cdot n^{-1}\sum_{j=1}^n \left(\xi_j^*(b) - n^{-1}\sum_{i=1}^n \xi_i(b)\right).$$

Then $\hat{\beta}^* = \arg\max_{b\in\mathcal{B}} \mathcal{L}_{n,1}(b)$. By equation (E.1), the asymptotic distribution of $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \hat{\beta}\right)$ can be established if we can show the limiting distribution of $\varepsilon_n^{-2/3}\mathcal{L}_{n,1}^*\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)$.

The previous results suggest that

$$\begin{aligned}
&\varepsilon_n^{-2/3} \cdot n^{-1}\sum_{i=1}^n \xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right) \\
&= \varepsilon_n^{-2/3}\mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)\right) + \varepsilon_n^{-2/3} \cdot n^{-1}\sum_{i=1}^n \left[\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right) - \mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)\right)\right] \\
&= \varepsilon_n^{-2/3}\mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)\right) + o_P(1) \\
&\xrightarrow{P} \frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}
\end{aligned}$$

over a compact set of $\boldsymbol{s}$, where the second equality holds by $n\varepsilon_n \to \infty$. The following holds by the i.i.d. sampling:

$$\varepsilon_n^{-2/3} \cdot (n\varepsilon_n)^{1/2} \cdot n^{-1}\sum_{j=1}^n \left(\xi_j^*\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right) - n^{-1}\sum_{i=1}^n \xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)\right) \rightsquigarrow W_1^*(\boldsymbol{s}),$$

where $W_1^*(\boldsymbol{s})$ is an independent copy of $W_1(\boldsymbol{s})$. As a result,

$$\varepsilon_n^{-2/3}\mathcal{L}_{n,1}^*\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right) \rightsquigarrow \frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1^*(\boldsymbol{s}),$$

as desired.

$\hat{\gamma}$ does not directly fit into the theoretical framework of Hong and Li (2020). More specifically, condition (vi) in Theorem 4.1 in Hong and Li (2020) is not satisfied. The previous results suggest that everything in Hong and Li (2020) can go through by modifying condition (vi) to that

$$\Sigma(s,t) = \lim_{n\to\infty} (nh_n)^{1/3}\,\mathbb{E}\left(h_n\varsigma_{ni}\left(\gamma + s(nh_n)^{-1/3}, \beta\right)\varsigma_{ni}\left(\gamma + t(nh_n)^{-1/3}, \beta\right)\right)$$

exists for each s, t in $\mathbb{R}$. This is true by Lemma B.2. In what follows, we provide an illustration for why numerical bootstrap works for $\hat{\gamma}$.

To concentrate on the key intuition, here we suppose that the effect of the first step estimator $\hat{\beta}$ has been handled and it does not affect the asymptotics of $\hat{\gamma}^*$. Let

$$\mathcal{L}_{n,2}^*(r) \equiv n^{-1}\sum_{i=1}^{n}\varsigma_{ni}(r,\beta) + (n\varepsilon_n)^{1/2}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma_{nj}^*(r,\beta) - n^{-1}\sum_{i=1}^{n}\varsigma_{ni}(r,\beta)\right),$$

where we use the same h_n in $\varsigma_{ni}(r,\beta)$ and $\varsigma_{nj}^*(r,\beta)$. The convergence rate of $\hat{\gamma}_n^*$ to γ can be shown to be $\left(\varepsilon_n^{-1}h_n\right)^{1/3}$. Thus, we only need to show the limit of $\left(\varepsilon_n^{-1}h_n\right)^{2/3}\mathcal{L}_{n,2}^*\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3}\right)$. Previous results suggest that

$$\begin{aligned}
&\left(\varepsilon_n^{-1}h_n\right)^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3},\beta\right)\\
&= \left(\varepsilon_n^{-1}h_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3},\beta\right)\right)\\
&+ \left(\varepsilon_n^{-1}h_n\right)^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\left(\varsigma_{ni}\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3},\beta\right) - \mathbb{E}\left(\varsigma_{ni}\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3},\beta\right)\right)\right)\\
&= \left(\varepsilon_n^{-1}h_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3},\beta\right)\right) + o_P(1)\\
&\xrightarrow{P} \frac{1}{2}V_2 s^2,
\end{aligned}$$

and

$$\begin{aligned}
&\left(\varepsilon_n^{-1}h_n\right)^{2/3}\cdot(n\varepsilon_n)^{1/2}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma_{nj}^*\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3},\beta\right) - n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma + s\left(\varepsilon_n^{-1}h_n\right)^{-1/3},\beta\right)\right)\\
&\rightsquigarrow W_2^*(s)
\end{aligned} \tag{E.2}$$

by i.i.d. and the Central Limit Theorem, where $W_2^*(s)$ is an independent copy of $W_2(s)$. To let equation (E.2) hold, it additionally requires $\varepsilon_n^{-1}h_n \to \infty$ and $\varepsilon_n^{-1}h_n^4 \to 0$, similar to the additional restriction on m.

E.2 Bootstrap Using a Modified Objective Function

In this section, we outline a proof for the consistency of the bootstrap using a modified objective function as in expressions (5.3) and (5.5). $\hat{\beta}^*$ and $\hat{\gamma}^*$ in this section are obtained from (5.3) and (5.5) respectively.

First $\hat{\beta}^* - \hat{\beta} = O_P\left(n^{-1/3}\right)$ can be similarly shown as for $\hat{\beta} - \beta$. Let

$$\tilde{\mathcal{L}}_{n,1}^*(b) \equiv n^{-1}\sum_{j=1}^{n}\xi_j^*(b) - n^{-1}\sum_{i=1}^{n}\xi_i(b) + \frac{1}{2}\left(b - \hat{\beta}\right)'\hat{V}_{n,1}\left(b - \hat{\beta}\right).$$

Thus, $\hat{\beta}^* = \arg\max_{b\in\mathcal{B}}\tilde{\mathcal{L}}_{n,1}^*(b)$. Then the asymptotics of $n^{1/3}\left(\hat{\beta}^* - \hat{\beta}\right)$ can be established if the asymptotics of $n^{2/3}\tilde{\mathcal{L}}^*_{n,1}\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)$ is obtained. $n^{2/3}\tilde{\mathcal{L}}^*_{n,1}\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)$ can be rewritten as:

$$\begin{aligned}
&n^{2/3}\tilde{\mathcal{L}}^*_{n,1}\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)\\
&=n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi_j^*\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)\right)+\frac{1}{2}\boldsymbol{s}'\hat{V}_{n,1}\boldsymbol{s}\\
&=n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi_j^*\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)+\frac{1}{2}\boldsymbol{s}'\hat{V}_{n,1}\boldsymbol{s}\\
&+n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left[\left(\xi_j^*\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)\right)\right.\\
&\left.-\left(\xi_j^*\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)\right].
\end{aligned}$$

Each of the terms in the above can be dealt with as follows. For the first term,

$$n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi_j^*\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)\rightsquigarrow W_1^*\left(\boldsymbol{s}\right),$$

for the same reason as for $n^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\left[\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)-\mathbb{E}\left(\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)\right]$. For the second term,

$$\frac{1}{2}\boldsymbol{s}'\hat{V}_{n,1}\boldsymbol{s}\overset{P}{\to}\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s},$$

if $\hat{V}_{n,1}\overset{P}{\to}V_1$. For the third term,

$$\begin{aligned}
n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}&\left[\left(\xi_j^*\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)\right)\right.\\
&\left.-\left(\xi_j^*\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)\right]=o_P\left(1\right)
\end{aligned}$$

holds uniformly over a compact set of $\boldsymbol{s}$ by the equicontinuity of $n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\xi_j^*\left(b\right)$ over b.

Substituting the above results into $n^{2/3}\tilde{\mathcal{L}}^*_{n,1}\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)$ yields

$$n^{2/3}\tilde{\mathcal{L}}^*_{n,1}\left(\hat{\beta}+\boldsymbol{s}n^{-1/3}\right)\rightsquigarrow W_1^*\left(\boldsymbol{s}\right)+\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s},$$

as desired.

The asymptotic distribution of $\left(nh_n\right)^{1/3}\left(\hat{\gamma}^*-\hat{\gamma}\right)$ can be similarly established.

E.3 Classic Bootstrap

The classic bootstrap estimators for $\hat{\beta}$ and $\hat{\gamma}$, denoted as $\hat{\beta}^*$ and $\hat{\gamma}^*$, are constructed from

$$\hat{\beta}^*=\arg\max_{b\in\mathcal{B}}n^{-1}\sum_{j=1}^{n}\xi_j^*\left(b\right),\ \text{and}\ \hat{\gamma}^*=\arg\max_{r\in\mathcal{R}}n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*\left(r,\hat{\beta}\right).$$

Based on the proof in Abrevaya and Huang (2005), we have

$$n^{1/3}\left(\hat{\beta}^{*}-\beta\right)\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^{K}}\left(\frac{1}{2}\boldsymbol{s}'V_{1}\boldsymbol{s}+W_{1}\left(\boldsymbol{s}\right)+W_{1}^{*}\left(\boldsymbol{s}\right)\right)$$

and

$$\left(nh_{n}\right)^{1/3}\left(\hat{\gamma}^{*}-\gamma\right)\xrightarrow{d}\arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_{2}s^{2}+W_{2}\left(s\right)+W_{2}^{*}\left(s\right)\right),$$

where $W_1(\boldsymbol{s})$ and $W_1^*(\boldsymbol{s})$ are identical and independent Gaussian processes with zero mean and covariance kernel H_1, and $W_2(s)$ and $W_2^*(s)$ are identical and independent Gaussian processes with zero mean and covariance kernel H_2. V_1, V_2, H_1 and H_2 are the same as in Theorem 4.1.

Therefore

$$\begin{aligned}n^{1/3}\left(\hat{\beta}^{*}-\hat{\beta}\right)&=n^{1/3}\left(\hat{\beta}^{*}-\beta\right)-n^{1/3}\left(\hat{\beta}-\beta\right)\\&\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^{K}}\left(\frac{1}{2}\boldsymbol{s}'V_{1}\boldsymbol{s}+W_{1}\left(\boldsymbol{s}\right)+W_{1}^{*}\left(\boldsymbol{s}\right)\right)-\arg\max_{\boldsymbol{s}\in\mathbb{R}^{K}}\left(\frac{1}{2}\boldsymbol{s}'V_{1}\boldsymbol{s}+W_{1}\left(\boldsymbol{s}\right)\right),\end{aligned}$$

and

$$\begin{aligned}\left(nh_{n}\right)^{1/3}\left(\hat{\gamma}^{*}-\hat{\gamma}\right)&=\left(nh_{n}\right)^{1/3}\left(\hat{\gamma}^{*}-\gamma\right)-\left(nh_{n}\right)^{1/3}\left(\hat{\gamma}-\gamma\right)\\&\xrightarrow{d}\arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_{2}s^{2}+W_{2}\left(s\right)+W_{2}^{*}\left(s\right)\right)-\arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_{2}s^{2}+W_{2}\left(s\right)\right).\end{aligned}$$

Here we provide a sketch on showing the inconsistency of the classic bootstrap.

By similar arguments of Lemma 3 in Abrevaya and Huang (2005), the convergence rate of $\hat{\beta}^*$ to β and $\hat{\gamma}^*$ to γ can be shown be at $n^{-1/3}$ and $(nh_n)^{-1/3}$ respectively.

Define

$$Z_{n,1}^{*}\left(\boldsymbol{s}\right)\equiv n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\xi_{j}^{*}\left(\beta+\boldsymbol{s}n^{-1/3}\right).$$

Similar to Theorem 1 in Abrevaya and Huang (2005), one can show

$$Z_{n,1}^{*}\left(\boldsymbol{s}\right)\rightsquigarrow\frac{1}{2}\boldsymbol{s}'V_{1}\boldsymbol{s}+W_{1}\left(\boldsymbol{s}\right)+W_{1}^{*}\left(\boldsymbol{s}\right),\tag{E.3}$$

where $W_1(\boldsymbol{s})$ and $W_1^*(\boldsymbol{s})$ are independent and identical Gaussian processes. The intuition of this result can be seen from the following decomposition of $Z_{n,1}^*(\boldsymbol{s})$:

$$\begin{aligned}Z_{n,1}^{*}\left(\boldsymbol{s}\right)&=n^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\xi_{i}\left(\beta+\boldsymbol{s}n^{-1/3}\right)+n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi_{j}^{*}\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_{i}\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)\\&=Z_{n,1}\left(\boldsymbol{s}\right)+n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi_{j}^{*}\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_{i}\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right),\end{aligned}$$

where the first term weakly converges to $\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})$, and the second term weakly converges to $W_1^*(\boldsymbol{s})$.

Since the convergence rate of $\hat{\beta}^*$ to β is $n^{-1/3}$, (E.3) implies that

$$n^{1/3}\left(\hat{\beta}^*-\beta\right)\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})+W_1^*(\boldsymbol{s})\right),$$

and

$$\begin{aligned}n^{1/3}\left(\hat{\beta}^*-\hat{\beta}\right)&=n^{-1/3}\left(\hat{\beta}^*-\beta\right)-n^{-1/3}\left(\hat{\beta}-\beta\right)\\&\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})+W_1^*(\boldsymbol{s})\right)-\arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})\right).\end{aligned}$$

For $\hat{\gamma}^*$, let

$$\begin{aligned}\hat{Z}_{n,2}^*(s)&\equiv(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right),\text{ and}\\Z_{n,2}^*(s)&\equiv(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\beta\right).\end{aligned}$$

The equicontinuity of $(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*(r,b)$ can be proved using similar arguments as in Theorem 1 of Abrevaya and Huang (2005). By that,

$$\hat{Z}_{n,2}^*(s)=Z_{n,2}^*(s)+o_P(1),$$

holds uniformly over a compact set of s. Thus we only need to establish the asymptotics of $Z_{n,2}^*(s)$. To that end, decompose $Z_{n,2}^*(s)$ as

$$\begin{aligned}Z_{n,2}^*(s)&=Z_{n,2}(s)+Z_{n,2}^*(s)-Z_{n,2}(s)\\&=Z_{n,2}(s)+(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\beta\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\\&=Z_{n,2}(s)+(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\beta\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right).\end{aligned}$$

Using the facts that the re-sampling is i.i.d. and $n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*(r,b)$ is equicontinuous in r, it holds that

$$(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\beta\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\rightsquigarrow W_2^*(s),$$

where $W_2^*(s)$ is identically distributed as $W_2(s)$.

Lemmas B.2 and B.4 imply that

$$Z_{n,2}(s) \rightsquigarrow \frac{1}{2} V_2 s^2 + W_2(s).$$

The independence of $W_2(s)$ and $W_2^*(s)$ can be shown using the same arguments in the proof of Theorem 1 in Abrevaya and Huang (2005).

Combing above results implies

$$\hat{Z}_{n,2}^*(s) \rightsquigarrow \frac{1}{2} V_2 s^2 + W_2(s) + W_2^*(s).$$

Thus,

$$(nh_n)^{1/3}(\hat{\gamma}^* - \gamma) \xrightarrow{d} \arg\max_{s\in\mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s) + W_2^*(s)\right),$$

and

$$\begin{aligned}(nh_n)^{1/3}(\hat{\gamma}^* - \hat{\gamma}) &= (nh_n)^{1/3}(\hat{\gamma}^* - \gamma) - (nh_n)^{1/3}(\hat{\gamma} - \gamma)\\ &\xrightarrow{d} \arg\max_{s\in\mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s) + W_2^*(s)\right) - \arg\max_{s\in\mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s)\right).\end{aligned}$$

E.4 m-out-of-n Bootstrap

Here $m \to \infty$ as $n \to \infty$, but $m/n \to 0$ as $n \to \infty$. This procedure is as follows. Draw $(y_j^{T*}, x_j^{T*\prime})'$, $j = 1, ..., m$, independently from the collection of the sample values $\left(y_1^T, x_1^{T\prime}\right)'$, $\left(y_2^T, x_2^{T\prime}\right)'$, ..., $\left(y_n^T, x_n^{T\prime}\right)'$ with replacement. Let $\hat{\beta}^*$ and $\hat{\gamma}^*$ be the estimator from the sampling observations, that is

$$\hat{\beta}^* = \arg\max_{b\in\mathcal{B}} m^{-1}\sum_{j=1}^{m} \xi_j^*(b) \text{ and } \hat{\gamma}^* = \arg\max_{r\in\mathcal{R}} m^{-1}\sum_{j=1}^{m} \varsigma_{nj}^*\left(r, \hat{\beta}\right), \tag{E.4}$$

where the bandwidth used in ς_{nj}^* is h_n, for simplicity. As the name suggests, this procedure only samples a small portion (m observations) from the data (n observations), with the hope of "correcting" the inconsistency of the classic bootstrap. Lee and Pun (2006) proved the consistency of m-out-of-n bootstrap for nonstandard M-estimators under mild conditions. After proving the general result, they applied it to the maximum score estimator by verifying the required technical conditions. We claim that these technical conditions can be similarly verified for our estimator and

$$m^{1/3}\left(\hat{\beta}^* - \hat{\beta}\right) \xrightarrow{d} \arg\max_{\boldsymbol{s}\in\mathbb{R}^K} \left(\frac{1}{2}\boldsymbol{s}' V_1 \boldsymbol{s} + W_1(\boldsymbol{s})\right)$$

and

$$(mh_n)^{1/3}(\hat{\gamma}^* - \hat{\gamma}) \xrightarrow{d} \arg\max_{s\in\mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s)\right). \tag{E.5}$$

To make equation (E.5) hold, we additionally require $mh_n \to \infty$, $mh_n^4 \to 0$, analogous to the conditions in Assumption 6. Because of the length limitations of the paper, the details are not pursued here. Instead, we have provided a heuristic illustration.

Note $\hat{\beta}^*$ and $\hat{\gamma}^*$ in this section are obtained from expression (E.4). Let

$$Z_{m,1}^*(\boldsymbol{s}) \equiv m^{2/3} \cdot m^{-1} \sum_{j=1}^{m} \xi_j^* \left(\beta + \boldsymbol{s}m^{-1/3}\right).$$

Rewrite $Z_{m,1}^*(\boldsymbol{s})$ as

$$\begin{aligned}
Z_{m,1}^*(\boldsymbol{s}) &= m^{2/3} \cdot m^{-1} \sum_{j=1}^{m} \left(\xi_j^* \left(\beta + \boldsymbol{s}m^{-1/3}\right) - n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s}m^{-1/3}\right) \right) + m^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s}m^{-1/3}\right) \\
&= m^{2/3} \cdot m^{-1} \sum_{j=1}^{m} \left(\xi_j^* \left(\beta + \boldsymbol{s}m^{-1/3}\right) - n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s}m^{-1/3}\right) \right) \\
&\quad + m^{2/3} \mathbb{E}\left(\xi_i \left(\beta + \boldsymbol{s}m^{-1/3}\right)\right) + m^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \left[\xi_i \left(\beta + \boldsymbol{s}m^{-1/3}\right) - \mathbb{E}\left(\xi_i \left(\beta + \boldsymbol{s}m^{-1/3}\right)\right) \right].
\end{aligned}$$

Intuitively, the first term in the above equation weakly converges to $W_1^*(\boldsymbol{s})$, the second term converges to $\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}$, and the last term converges to zero in probability. One can similarly show $\hat{\beta}^* - \beta = O_P\left(m^{-1/3}\right)$.

Therefore,

$$m^{1/3}\left(\hat{\beta}^* - \beta\right) \xrightarrow{d} \arg\max_{\boldsymbol{s} \in \mathbb{R}^K} \frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1^*(\boldsymbol{s}).$$

Finally,

$$\begin{aligned}
m^{1/3}\left(\hat{\beta}^* - \hat{\beta}\right) &= m^{1/3}\left(\hat{\beta}^* - \beta\right) - m^{1/3}\left(\hat{\beta} - \beta\right) \\
&= m^{1/3}\left(\hat{\beta}^* - \beta\right) + o_P(1) \\
&\xrightarrow{d} \arg\max_{\boldsymbol{s} \in \mathbb{R}^K} \frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1^*(\boldsymbol{s}).
\end{aligned}$$

Note that the distribution of $W_1^*(\boldsymbol{s})$ is the same as that of $W_1(\boldsymbol{s})$, and the claim is proved for $\hat{\beta}^*$.

The asymptotic distribution of $(mh_n)^{1/3}(\hat{\gamma}^* - \hat{\gamma})$ can be similarly established.