Semiparametric Discrete Choice Models for Bundles*

Fu Ouyang† University of Queensland

Thomas Tao Yang‡ Australian National University

June 12, 2020

Abstract

We propose new identification and estimation approaches to semiparametric discrete choice models for bundles in both cross-sectional and panel data settings. The random utility functions of these models take the usual parametric form, while no distributional assumption is imposed on the stochastic disturbances. Our proposed methods permit certain forms of heteroskedasticity and arbitrary correlation in the disturbances across choices. Our identification approach is matching-based; it matches observed covariates across agents for the cross-sectional case, and over time for the panel data case. For the cross-sectional model, we propose a kernel-weighted rank procedure and establish $\sqrt{N}$-asymptotic normality of the resulting estimators. We show the validity of the nonparametric bootstrap for the inference. For the panel data model, we propose localized maximum score type estimators which have a non-standard asymptotic distribution. We show that the numerical bootstrap developed by Hong and Li (2020) is a valid inference method for our panel data estimators. Monte Carlo experiments demonstrate that our proposed estimation and inference procedures perform adequately in finite samples.

JEL classification: C13, C14, C35.

Keywords: Bundle choices; rank estimation; panel data; bootstrap.

1 Introduction

In many circumstances, a consumer may purchase a bundle of goods (e.g., chips, salsa, or both) or services (e.g., mobile phone plan, home internet, cable TV, or any of their combinations). The literature on bundle choices is less well-developed than the literature on the ordinary discrete choice models. An important but less extensively studied empirical question in the industrial organization

*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.

†Email: f.ouyang@uq.edu.au.

‡Email: tao.yang@anu.edu.au.

and marketing research is to investigate the complementary or substitutive effects between goods and explain the bundle choice behavior of consumers.[1] In empirical studies, Gentzkow (2007) estimated a parametric bundle choice model analyzing the demand of print and online newspapers. Similarly, using aggregate data, Fan (2013) examined the ownership consolidation in the newspaper market where households in demand side may purchase two newspapers as a bundle. A substantial literature involves bundle choice behavior, analyzing product demand, pricing strategy, customers subscription, and brand collaboration, among others. Examples include Manski and Sherman (1980), Train, McFadden, and Ben-Akiva (1987), Hendel (1999), Chung and Rao (2003), Dubé (2004), Nevo, Rubinfeld, and McCabe (2005), Augereau, Greenstein, and Rysman (2006), Song and Chintagunta (2006), Foubert and Gijsbrechts (2007), Liu, Chintagunta, and Zhu (2010), Gentzkow, Shapiro, and Sinkinson (2014), Kim, Misra, and Shapiro (2019), and Lewbel and Nesheim (2019). The models in most of these applications are parametric.[2]

Sher and Kim (2014) explored the identification of bundle choice models without stochastic components in the utility from the perspective of microeconomic theory, assuming a finite population of consumers. Dunker, Hoderlein, and Kaido (2017) studied nonparametric identification of market-level demand models in a general framework that nests the demand for bundles. Fox and Lazzati (2017) provided nonparametric identification results for both single-agent discrete choice bundle models and binary games of complete information, and established the mathematically equivalence between these models. Allen and Rehbeck (2019) studied nonparametric identification of a class of latent utility models with additively separable unobserved heterogeneity by exploiting asymmetries in cross-partial derivatives of conditional average demands. Allen and Rehbeck (2020) established partial identification results for complementarity (substitutability) in a latent utility model with binary quantities of two goods.

This work differs from the aforementioned papers in many aspects. We develop a semiparametric point identification of preference coefficients in bundle choice models using individual-level data. This paper adopts a stochastic utility maximization framework similar to Fox and Lazzati (2017), but assumes the deterministic components of the random utility functions to be linear. Our methods are semiparametric in that no parametric restriction is imposed on unobserved error terms. To our knowledge, this is the first work to study bundle choice models in a semiparametric framework. The robustness from the distribution-free specification makes this paper a competitive alternative to existing parametric methods. Fox and Lazzati's (2017) identification heavily relies on the existence of excluded variables. Our approaches, however, are based on a different identification strategy, using a matching insight and the variation in regressors. This work also contributes to the literature in that it is the first to consider bundle choice models in a panel data setting with fixed effects.

Throughout, our approaches allow for arbitrary correlation in the unobserved errors across choices and certain forms of heteroskedasticity. In the cross-sectional setting, we achieve identification by moment inequalities obtained from matching cross-section units in certain way. These

[1]See Berry et al. (2014) for a review of this literature.

[2]See Gentzkow (2007) pp. 722–723 for a brief review on pervasive parametric methods in existing literature.

identification results motivate a two-step, rank-based estimation procedure. We then generalize this matching approach to a panel data bundle choice model with fixed effects, where the moment inequalities used for identification are obtained from variations in the covariates over time for each agent. We propose a maximum score type estimation procedure in the spirit of Manski (1987) for the panel data model. For all these estimators, we derive their asymptotics and justify the validity of using (different) bootstrap methods to conduct inference. All proposed estimation and inference procedures are easy to implement, and can be easily extended to cases with more complicated choice set.

The approaches studied in this paper generalize the work of Khan, Ouyang, and Tamer (2019) to bundle choice models and complement their work by providing asymptotic theory for two-step rank estimators and valid bootstrap inference procedures. More generally, this paper also tightly relates to the growing literature on semiparametric and nonparametric multinomial choice models in both cross-sectional and panel data settings. Examples inlcude Lee (1995), Lewbel (2000), Altonji and Matzkin (2005), Fox (2007), Berry and Haile (2009), Pakes and Porter (2016), Ahn, Ichimura, Powell, and Ruud (2018), Shi, Shum, and Song (2018), Yan and Yoo (2019), Gao and Li (2019), and Chernozhukov, Fernández-Val, and Newey (2019), among others. These approaches, however, do not directly extend to the bundle choices models due to the non-mutually-exclusive choice set and distinctive random utilities associated with bundles.

The remainder of this paper proceeds as follows. Section 2 presents the cross-sectional model and provides sufficient conditions on both observed covariates and unobserved errors that ensure point identification. We then propose a two-step, localized maximum rank correlation (MRC) procedure motivated by the identification strategy, and derive its $\sqrt{N}$-consistency and asymptotic normality. We show that the inference can be conducted by the standard bootstrap. Section 3 extends to a panel data model with unobserved agent, alternative, and bundle specific fixed effects. We propose localized maximum score (MS) estimators for this model, develop its asymptotic properties, and justify the use of the numerical bootstrap for the inference. Section 4 investigates the finite sample performance of our proposed procedures using Monte Carlo experiments. Section 5 concludes. All tables and proofs 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 $\|\cdot\|$, and $\mathbb{R}_+ \equiv \{x \in \mathbb{R} | x \geq 0\}$. $\|\cdot\|_F$ denotes the Frobenius norm, i.e., for any matrix A, $\|A\|_F = \sqrt{\text{trace}\,(AA')}$. We reserve letters i and m for indexing agents, j and l for indexing alternatives, and s and t for indexing time periods. The first element of a vector v is denoted by $v^{(1)}$ and the sub-vector comprising its remaining elements is denoted by $\tilde{v}$. We use $P(\cdot)$ and $\mathbb{E}[\cdot]$ to denote probability and expectation, respectively. $1[\cdot]$ is an indicator function that equals one when the event in the brackets occurs, and zero otherwise. For two random vectors U and V, the notation $U \overset{d}{=} V|\cdot$ means that U and V have identical distribution conditional on $\cdot$, and $U \perp V|\cdot$ means that U and V are independent conditional on $\cdot$. Symbols $\backslash$, $'$, $\Leftrightarrow$, $\propto$, $\xrightarrow{d}$, $\xrightarrow{P}$, and $\rightsquigarrow$ represent set difference, matrix transposition, if and only if, proportionality, convergence in distribution, convergence in probability, and weak convergence in the sense of van der Vaart and Wellner (2000), respectively.

2 Cross-Sectional Model

2.1 Model and Identification

Throughout this paper, we focus on a choice model where the choice set $\mathcal{J}$ consists of 3 mutually exclusive alternatives (numbered from 0 to 2) and a bundle of alternatives 1 and 2, i.e., $\mathcal{J} = \{0, 1, 2, (1, 2)\}$. This simple model is sufficient to illustrate the main intuition that runs through both cross-sectional and panel data models. Our approach can be modified in a straightforward way to be applied to data with more alternatives and their bundles.

For ease of exposition, we re-number alternatives in $\mathcal{J}$ with 2-dimensional vectors of binary indicators $d = (d_1, d_2) \in \{0, 1\}\times\{0, 1\}$, where d_1 and d_2 indicate if alternative 1 and 2 are chosen, respectively. Then the choice set $\mathcal{J}$ can be one-to-one mapped to the set $\mathcal{D} = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. An agent chooses the alternative in $\mathcal{D}$ to maximize the latent utility

$$U_d = \sum_{j=1}^{2}(X_j'\beta + \epsilon_j) \cdot d_j + \eta \cdot (W'\gamma) \cdot d_1 \cdot d_2, \tag{2.1}$$

where $X_j \in \mathbb{R}^{k_1}$ collects covariates affecting the utility associated with stand-alone alternative j,[3] $W \in \mathbb{R}^{k_2}$ is a vector of explanatory variables[4] characterizing the interaction effects of the bundle (e.g., bundle discount), $(\epsilon_1, \epsilon_2, \eta) \in \mathbb{R}^3$ captures unobserved (to the econometrician) heterogeneous effects, and $(\beta', \gamma')' \in \mathbb{R}^{k_1+k_2}$ are unknown preference parameters to estimate.[5]

The specification of model (2.1) follows Fox and Lazzati (2017) but it comes with linear deterministic utilities. The utility of choosing $(0, 0)$ is normalized to zero, the utility of choosing $(1, 0)$ ($(0, 1)$) is $X_1'\beta + \epsilon_1$ ($X_2'\beta + \epsilon_2$), and the utility of choosing the bundle is the sum of the two stand-alone utilities plus the interaction term $\eta \cdot (W'\gamma)$ capturing either complementary ($\eta \cdot (W'\gamma) \geq 0$) or substitution ($\eta \cdot (W'\gamma) \leq 0$) effects.[6] (ϵ_1, ϵ_2) are idiosyncratic shocks associated with each stand-alone alternative as in common multinomial choice models, and η reflects unobserved heterogeneity

[3]Note that by properly re-organizing the covariate vector, expression (2.1) can accommodate both alternative-specific and agent-specific covariates. See Cameron and Trivedi (2005) p.498 for a more detailed discussion. However, the identification of coefficients on agent-specific covariates will need to use different approaches in a second step when all coefficients on alternative-specific covariates are identified. See Remarks 2.4 and 2.5 below.

[4]W and X_j's may have common elements.

[5]Our approaches is flexible enough to accommodate alternative-specific coefficients (e.g., β_j in expression (2.1)). A model with such coefficients is particularly applicable to bundle choice problems involving heterogeneous alternatives. For example, it may be more reasonable to presume that chips and salsa have different price elasticities. The identification of β_j can be achieved using the same method, which we discuss below, and identification inequality only associate with alternative j. In the rest of this paper, we assume $\beta_j = \beta$ for all j to simplify the notations.

[6]Gentzkow (2007) shows that this definition of complements and substitutes is equivalent to the classic definitions based on the sign of cross demand elasticity.

for the bundle. In what follows, we assume $\eta > 0$, i.e., $W'\gamma$ determines the sign of the interaction term and the presence of η allows for agent-specific magnitudes.

Remark 2.1. Our methods can be applied to a more general model with latent utility

$$U_d = \sum_{j=1}^{2} F_j(X_j'\beta, \epsilon_j)d_j + F_b(W'\gamma, \eta) \cdot d_1 \cdot d_2,$$

where $F_j(\cdot,\cdot)$'s and $F_b(\cdot,\cdot)$ are known or unknown (to the econometrician) $\mathbb{R}^2 \mapsto \mathbb{R}$ functions strictly monotonic in each of their arguments. It will be clear that our approaches do not rely on the additive (or multiplicative) separability of the explanatory variables and the disturbances, which is different from Fox and Lazzati (2017) where additive separability plays a key role in their identification strategy. Note that the general framework includes as a special case the model with an additive $W'\gamma + \eta$. This paper focuses on model (2.1) without loss of generality (w.l.o.g.).

Given the latent utility model (2.1), the observed dependent variable Y_d is of the form

$$Y_d = 1[U_d > U_{d'}, \forall d' \in \mathcal{D} \setminus d]. \tag{2.2}$$

Let $Z \equiv (X_1', X_2', W')'$. Then the probabilities of choosing alternatives $d \in \mathcal{D}$ can be expressed as

$$\begin{aligned}
P(Y_{(0,0)} = 1|Z) =& P(\max\{X_1'\beta + \epsilon_1, X_2'\beta + \epsilon_2, (X_1 + X_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (W'\gamma)\} < 0|Z), \\
P(Y_{(1,0)} = 1|Z) =& P(\max\{0, X_2'\beta + \epsilon_2, (X_1 + X_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (W'\gamma)\} < X_1'\beta + \epsilon_1|Z), \\
P(Y_{(0,1)} = 1|Z) =& P(\max\{0, X_1'\beta + \epsilon_1, (X_1 + X_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (W'\gamma)\} < X_2'\beta + \epsilon_2|Z), \\
P(Y_{(1,1)} = 1|Z) =& P(\max\{0, X_1'\beta + \epsilon_1, X_2'\beta + \epsilon_2\} < (X_1 + X_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (W'\gamma)|Z).
\end{aligned} \tag{2.3}$$

When $(\epsilon_1, \epsilon_2, \eta) \perp Z$, expression (2.3) implies that $P(Y_{(1,0)} = 1|Z, X_2 = x_2, W = w)$ and $P(Y_{(1,1)} = 1|Z, X_2 = x_2, W = w)$ are both increasing in $X_1'\beta$ for some constant vectors x_2 and w. Let (Y_i, Z_i) and (Y_m, Z_m) be two independent copies of (Y, Z) with $Y \equiv (Y_{(0,0)}, Y_{(0,1)}, Y_{(1,0)}, Y_{(1,1)})'$. Then we have the following (conditional) moment inequalities: For all d with $d_1 = 1$,

$$X_{i1}'\beta \geq X_{m1}'\beta \Leftrightarrow P(Y_{id} = 1|Z_i, X_{i2} = x_2, W_i = w) \geq P(Y_{md} = 1|Z_m, X_{m2} = x_2, W_m = w). \tag{2.4}$$

Fixing some covariates or their index functions to establish the monotonic relationship between a conditional choice probability and an index of the other covariates motivates all our identification results in this paper.

The moment inequalities above can be repeated for all values of x_2 and w. Furthermore, note that $P(Y_{(0,0)} = 1|Z, X_2 = x_2, W = w)$ and $P(Y_{(0,1)} = 1|Z, X_2 = x_2, W = w)$ are both decreasing in $X_1'\beta$, which means that for all d with $d_1 = 0$,

$$X_{i1}'\beta \geq X_{m1}'\beta \Leftrightarrow P(Y_{id} = 1|Z_i, X_{i2} = x_2, W_i = w) \leq P(Y_{md} = 1|Z_m, X_{m2} = x_2, W_m = w). \tag{2.5}$$

Similar moment inequalities for $X_{i2}'\beta \geq X_{m2}'\beta$ can be obtained by fixing X_1 and W. Collectively, these identification inequalities, along with the regularity conditions stated below, establish the identification of β.

Once β is identified, we can move on to identify γ using the following moment inequalities for $W'\gamma$ constructed by fixing $(X_1'\beta, X_2'\beta)$ at some constant vector (v_1, v_2): For $d = (1,1)$,

$$\begin{aligned}
&W_i'\gamma \geq W_m'\gamma \\
\Leftrightarrow &P(Y_{id} = 1|Z_i, X_{i1}'\beta = v_1, X_{i2}'\beta = v_2) \geq P(Y_{md} = 1|Z_m, X_{m1}'\beta = v_1, X_{m2}'\beta = v_2),
\end{aligned} \tag{2.6}$$

and for all $d \in \mathcal{D} \setminus (1,1)$,

$$\begin{aligned}
&W_i'\gamma \geq W_m'\gamma \\
\Leftrightarrow &P(Y_{id} = 1|Z_i, X_{i1}'\beta = v_1, X_{i2}'\beta = v_2) \leq P(Y_{md} = 1|Z_m, X_{m1}'\beta = v_1, X_{m2}'\beta = v_2).
\end{aligned} \tag{2.7}$$

Remark 2.2. Note that γ can be alternatively identified by "matching" $X \equiv (X_1', X_2')'$ across agents. The idea is the same, i.e., by fixing X at some constant vector x, it can be shown, for example, that $P(Y_{(1,1)} = 1|Z, X = x)$ is increasing in $W'\gamma$. Then we can establish the following moment inequality for γ by matching $X_i = X_m$:

$$W_i'\gamma \geq W_m'\gamma \Leftrightarrow P(Y_{i(1,1)} = 1|Z_i, X_i = x) \geq P(Y_{m(1,1)} = 1|Z_m, X_m = x). \tag{2.8}$$

This approach is particularly useful when X_1 and X_2 have common elements (e.g., agent-specific regressors, whose coefficients generally cannot be point identified using our match and rank approach, see Remarks 2.4 and 2.5 for a further discussion), but X has no covariates in common with W. In contrast, identification based on (2.6) would be best suited for cases where X_1 and X_2 are alternative-specific, but X and W have some covariates in common. In practice, the complementary use of these two approaches could be a useful strategy. As will be clear in Section 2.2, either (2.6) or (2.8) leads to a $\sqrt{N}$-consistent estimator for γ, though the former lowers the dimensionality of the problem.[7]

To establish the identification of β and γ based on (2.4)–(2.7), the following conditions are sufficient.

C1 $\{(Y_i', Z_i')'\}_{i=1}^N$ are i.i.d. across i.

C2 (i) $(\epsilon_1, \epsilon_2, \eta) \perp Z$, and (ii) the joint distribution of $(\epsilon_1, \epsilon_2, \eta)$ is absolutely continuous with respect to (w.r.t.) the Lebesgue measure on $\mathbb{R}^2 \times \mathbb{R}_+$.

C3 For any pair of (i, m) and $j = 1, 2$, denote $X_{imj} = X_{ij} - X_{mj}$ and $W_{im} = W_i - W_m$. Let $\mathcal{X}_{imj}$ and $\mathcal{W}_{im}$ define the supports of X_{imj} and W_{im}, respectively. Then, (i) $X_{im1}^{(1)}$ ($X_{im2}^{(1)}$) has almost everywhere (a.e.) positive Lebesgue density on $\mathbb{R}$ conditional on $\tilde{X}_{im1}$ ($\tilde{X}_{im2}$) and conditional on (X_{im2}, W_{im}) ((X_{im1}, W_{im})) in a neighborhood of (X_{im2}, W_{im}) ((X_{im1}, W_{im})) near zero, (ii) $\mathcal{X}_{im1}$ ($\mathcal{X}_{im2}$), conditional on (X_{im2}, W_{im}) ((X_{im1}, W_{im})) in a neighborhood of (X_{im2}, W_{im}) ((X_{im1}, W_{im})) near zero, is not contained in any proper linear subspace of $\mathbb{R}^{k_1}$, (iii) $W_{im}^{(1)}$ has a.e. positive Lebesgue density on $\mathbb{R}$ conditional on $\tilde{W}_{im}$ and conditional on

[7] We expect that this may improve the finite sample performance of the estimator.

$(X'_{im1}\beta, X'_{im2}\beta)$ in a neighborhood of $(X'_{im1}\beta, X'_{im2}\beta)$ near zero, and (iv) $\mathcal{W}_{im}$, conditional on $(X'_{im1}\beta, X'_{im2}\beta)$ in a neighborhood of $(X'_{im1}\beta, X'_{im2}\beta)$ near zero, is not contained in any proper linear subspace of $\mathbb{R}^{k_2}$.

C4 $(\beta', \gamma')' \in \mathcal{B} \times \mathcal{R}$, where $\mathcal{B} = \{b \in \mathbb{R}^{k_1} | \, \|b\| = 1, b^{(1)} \neq 0\}$ and $\mathcal{R} = \{r \in \mathbb{R}^{k_2} | \, \|r\| = 1, r^{(1)} \neq 0\}$.

As discussed above, Assumptions C1 and C2 suffice to establish the moment inequalities (2.4)–(2.7). Note that Assumption C2 allows arbitrary correlation among $(\epsilon_1, \epsilon_2, \eta)$, and our matching-based approach intrinsically accommodates flexible dependence of $(\epsilon_1, \epsilon_2, \eta)$ on agent-specific characteristics (e.g., interpersonal heteroskedasticity). Assumption C3(i) (C3(iii)) is a standard restriction in MRC and MS types of estimators, which guarantees the point identification, as opposed to a set identification. Assumption C3(ii) (C3(iv)) is the familiar full-rank condition. Assumption C4 is about scale normalization and the parameter space. As usual in discrete choice models, β can be identified only up to scale. In addition, we also need to normalize the scale of γ as the magnitude of the interaction effects is controlled by the unobserved heterogeneity η. Following a substantial literature, we assume that β and γ are both on the unit circle, and their first elements are nonzero.

Our identification result for model (2.1)–(2.2) is stated below, which is proved in Appendix A.

Theorem 2.1. *If Assumptions C1–C4 hold, β and γ are identified.*

Remark 2.3. The identification results established above do not tell the comparative importance of the stand-alone utilities and the interaction effects in determining the choice of the bundle as the scale of the latter is controlled by η. Identifying the scale of η can address this problem but requires additional restriction(s) on the joint distribution of $(\epsilon_1, \epsilon_2, \eta)$. Here we briefly discuss two possible directions for identifying the scale of η: (1) Note that when $W'\gamma < 0$, $P(Y_{(0,0)} = 1|Z) = P(-\epsilon_1 > X'_1\beta, -\epsilon_2 > X'_2\beta)$ and $P(Y_{(1,1)} = 1|Z) = P(X'_1\beta + \epsilon_1 + \eta(W'\gamma) > 0, X'_2\beta + \epsilon_2 + \eta(W'\gamma) > 0)$. With identified $(X'_1\beta, X'_2\beta)$, we can obtain the joint distribution F_ϵ of (ϵ_1, ϵ_2) from $P(Y_{(0,0)} = 1|Z)$. Assuming $(\epsilon_1, \epsilon_2) \perp \eta$, then we can use the knowledge of $(X'_1\beta, X'_2\beta, W'\gamma)$ and F_ϵ to identify the distribution F_η of η from $P(Y_{(1,1)} = 1|Z)$. (2) Let $\mu_\eta \equiv \mathbb{E}[\eta]$ and so $\eta = \mu_\eta + e_\eta$ with $\mathbb{E}[e_\eta] = 0$. Assume that the joint probability density function (PDF) f_ϵ of (ϵ_1, ϵ_2) is symmetric around 0. We can show that $(X_1+X_2)'\beta + \mu_\eta \cdot (W'\gamma) + e_\eta \cdot (W'\gamma) = 0 \Leftrightarrow P(Y_{(0,0)} = 1|Z, X'_1\beta = X'_2\beta) = P(Y_{(1,1)} = 1|Z, X'_1\beta = X'_2\beta)$. Then μ_η can be identified in a local linear regression framework. A detailed and thorough discussion on the identification of F_η (or $\mathbb{E}[\eta]$) with extra distributional restriction(s) is outside the scope of this work. We leave this topic to a separate paper.

Remark 2.4. When the model contains both alternative-specific and agent-specific regressors, for example, in model (2.1), $U_d = \sum_{j=1}^{2}(X'_j\beta + S'\theta_j + \epsilon_j)d_j + \eta \cdot (W'\gamma) \cdot d_1 \cdot d_2$ with S collecting all agent-specific regressors, moment inequalities like (2.4) can only get us identification information about β since with the matching $\{X_{i2} = X_{m2}, S_i = S_m, W_i = W_m\}$, $S'_i\theta_j$ drops out from the moment inequality. But in a special case where $\theta_1 = \theta_2 = \theta$, θ can be point identified in an additional step after β being identified from e.g., (2.4). To see this, note that by conditioning on $\{X'_{i1}\beta = X'_{m1}\beta = v_1, X'_{i2}\beta = X'_{m2}\beta = v_2, W_i = W_m = w\}$, we can establish moment inequalities of the following form for θ:

$$S_i'\theta \geq S_m'\theta \Leftrightarrow \begin{array}{c} P(Y_{i(1,1)} = 1 | Z_i, X_{i1}'\beta = v_1, X_{i2}'\beta = v_2, W_i = w) \\ \geq \\ P(Y_{m(1,1)} = 1 | Z_m, X_{m1}'\beta = v_1, X_{m2}'\beta = v_2, W_m = w), \end{array}$$

based on which θ can be point identified using similar arguments as in establishing Theorem 2.1.

Remark 2.5. This is related to the discussion in Remark 2.4. For general cases where $\theta_1 \neq \theta_2$, once β is "known" from the first step, identifying restrictions for θ_1 and θ_2 can be obtained in a separate step. For example, when conditioning on $\{X_{i1}'\beta = X_{m1}'\beta = v_1, X_{i2}'\beta = X_{m2}'\beta = v_2, W_i = W_m = w\}$, we have

$$\begin{array}{c} P(Y_{i(1,1)} = 1 | Z_i, X_{i1}'\beta = v_1, X_{i2}'\beta = v_2, W_i = w) \\ > \\ P(Y_{m(1,1)} = 1 | Z_m, X_{m1}'\beta = v_1, X_{m2}'\beta = v_2, W_m = w) \end{array} \Rightarrow \neg\{S_i'\theta_1 \leq S_m'\theta_1, S_i'\theta_2 \leq S_m'\theta_2\},$$

where $\Rightarrow$ and $\neg$ denote the logical implication and negation operators, respectively. Gao and Li (2019) studied the set identification of panel data multinomial response models based on identifying restrictions of this type. We think similar approach can be applied here to get bounds for θ.

2.2 Localized MRC Estimator

The local monotonic relations established in (2.4)–(2.7) naturally motivate a two-step localized MRC estimation procedure. Each of the two steps is described in turn below.

In the first step, we consider the localized MRC estimator $\hat{\beta}$ of β, analogous to the MRC estimator proposed in Han (1987), defined as the maximizer over the parameter space $\mathcal{B}$, of the following objective function:

$$\begin{aligned} \mathcal{L}_{N,\beta}(b) = \sum_{i=1}^{N-1} \sum_{m>i} \sum_{d \in \mathcal{D}} \{ & 1[X_{i2} = X_{m2}, W_i = W_m](Y_{md} - Y_{id}) \operatorname{sgn}((X_{i1} - X_{m1})'b) \cdot (-1)^{d_1} \\ & + 1[X_{i1} = X_{m1}, W_i = W_m](Y_{md} - Y_{id}) \operatorname{sgn}((X_{i2} - X_{m2})'b) \cdot (-1)^{d_2} \}. \end{aligned} \quad (2.9)$$

where the notation $\operatorname{sgn}(\cdot)$ denotes the sign function.

In the second step, we use the $\hat{\beta}$ obtained in the first step to define the localized MRC estimator $\hat{\gamma}$ of γ that maximizes the following objective function over the parameter space $\mathcal{R}$:

$$\mathcal{L}_{N,\gamma}(r; \hat{\beta}) = \sum_{i=1}^{N-1} \sum_{m>i} 1[X_{i1}'\hat{\beta} = X_{m1}'\hat{\beta}, X_{i2}'\hat{\beta} = X_{m2}'\hat{\beta}](Y_{i(1,1)} - Y_{m(1,1)}) \operatorname{sgn}((W_i - W_m)'r). \quad (2.10)$$

Note that in the presence of continuous covariates, the probability of getting perfectly matched observations is zero. Thus the values of the objective functions may always be zero. Following the literature, we propose to use kernel weights to approximate the binary weights defined by indicator function $1[\cdot]$ in objective functions (2.9) and (2.10).

Specifically, the objective function of $\beta \in \mathcal{B}$ is now of the form

$$\mathcal{L}^K_{N,\beta}(b) = \sum_{i=1}^{N-1} \sum_{m>i} \sum_{d\in\mathcal{D}} \{\mathcal{K}_{h_N}(X_{im2}, W_{im})(Y_{md} - Y_{id})\text{sgn}((X_{i1} - X_{m1})'b) \cdot (-1)^{d_1} + \mathcal{K}_{h_N}(X_{im1}, W_{im})(Y_{md} - Y_{id})\text{sgn}((X_{i2} - X_{m2})'b) \cdot (-1)^{d_2}\}, \tag{2.11}$$

with $\mathcal{K}_{h_N}(X_{imj}, W_{im}) \equiv h_N^{-(k_1+k_2)} \prod_{\iota=1}^{k_1} K(X_{imj,\iota}/h_N) \prod_{\iota=1}^{k_2} K(W_{im,\iota}/h_N)$, where $X_{imj,\iota}$ ($W_{im,\iota}$) is the ιth element of vector X_{imj} (W_{im}), $K(\cdot)$ is a standard kernel density function, and h_N is a bandwidth sequence that converges to 0 as $N \to \infty$. Similarly, we estimate $\gamma \in \mathcal{R}$ with the following objective function

$$\mathcal{L}^K_{N,\gamma}(r; \hat{\beta}) = \sum_{i=1}^{N-1} \sum_{m>i} \mathcal{K}_{\sigma_N}(V_{im}(\hat{\beta}))(Y_{i(1,1)} - Y_{m(1,1)})\text{sgn}((W_i - W_m)'r), \tag{2.12}$$

where $V_{im}(b) \equiv (X'_{im1}b, X'_{im2}b)'$ for all $b \in \mathbb{R}^{k_1}$, $\mathcal{K}_{\sigma_N}(V_{im}(\hat{\beta})) = \sigma_N^{-2} K(X'_{im1}\hat{\beta}/\sigma_N) K(X'_{im2}\hat{\beta}/\sigma_N)$, and σ_N is a bandwidth sequence that converges to 0 as $N \to \infty$.

The rest of this section establishes the asymptotic properties of the estimators computed based on objective functions (2.11) and (2.12). To streamline the exposition, we assume that all regressors are continuous. Before presenting additional regularity conditions and the main results, we introduce some new notations for ease of exposition: For all $d \in \mathcal{D}$ and $j = 1, 2$,

- $V(\beta) \equiv (X'_1\beta, X'_2\beta)'$.
- Let X_{ij} and X_{mj} be two independent copies of X_j and $X_{imj} \equiv X_{ij} - X_{mj}$. Similarly, define $W_i, W_m, W_{im}, Y_i, Y_m, Y_{im}, V_i(\beta), V_m(\beta)$, and $V_{im}(\beta)$.
- $f_{X_{imj}, W_{im}}(\cdot)$ and $f_{V_{im}(\beta)}(\cdot)$ denote the PDF of the random vectors $(X'_{imj}, W'_{im})'$ and $V_{im}(\beta)$, respectively. $f_{X_2,W|X_1}(\cdot)$ ($f_{X_1,W|X_2}(\cdot)$) denotes the PDF of $(X'_2, W')'$ ($(X'_1, W')'$) conditional on X_1 (X_2). $f_{V(\beta)|W}(\cdot)$ denotes the PDF of $V(\beta)$ conditional on W.
- $\varrho_i(b) \equiv -\sum_{d\in\mathcal{D}}\{\varrho_{i21d}(b) \cdot (-1)^{d_1} + \varrho_{i12d}(b) \cdot (-1)^{d_2}\}$, where

  $$\varrho_{ijld}(b) \equiv \mathbb{E}\left(Y_{imd}\left[\text{sgn}(X'_{iml}b) - \text{sgn}(X'_{iml}\beta)\right] | Z_i, X_{mj} = X_{ij}, W_m = W_i\right)$$

  for $j, l = 1, 2$ with $j \neq l$.
- $\tau_i(r) \equiv \mathbb{E}[Y_{im(1,1)}\left[\text{sgn}(W'_{im}r) - \text{sgn}(W'_{im}\gamma)\right] | V_m(\beta) = V_i(\beta), W_i]$.
- For any function g, $\nabla g(v)$ and $\nabla^2 g(v)$ denote the gradient and Hessian matrix for $g(\cdot)$ evaluated at v, respectively.

- Let

$$\mu(v_1, v_2, r)$$
$$\equiv \mathbb{E}\left[Y_{im(1,1)} \left[\text{sgn}(W_{im}' r) - \text{sgn}(W_{im}' \gamma)\right] \left(\begin{array}{c} X_{im1}' \\ X_{im2}' \end{array} \right) \middle| V_i(\beta) = v_1, V_m(\beta) = v_2 \right] f_{V(\beta)}(v_1).$$

- Let $\nabla_1 \mu(v_1, v_2, r)$ denote the partial derivative of $\mu(v_1, v_2, r)$ w.r.t. its first argument. Denote $\nabla_{1k}^2 \mu(v_1, v_2, r)$ as the partial derivative of $\nabla_1 \mu(v_1, v_2, r)$ w.r.t. its k-th argument and $\nabla_{33}^2 \nabla_1 \mu(v_1, v_2, r)$ as the Hessian matrix of $\nabla_1 \mu(v_1, v_2, r)$ w.r.t. its third argument.

We impose the following regularity conditions:

C5 (i) $f_{V_{im}(\beta)}(\cdot)$ and $f_{X_{imj}, W_{im}}(\cdot)$ for $j = 1, 2$ are bounded from above on their supports, strictly positive in a neighborhood of zero, and twice continuously differentiable with bounded second derivatives, (ii) For all $d \in \mathcal{D}$, $b \in \mathcal{B}$, and $r \in \mathcal{R}$, $\mathbb{E}[Y_{imd}[\text{sgn}(X_{im1}' b) - \text{sgn}(X_{im1}' \beta)] | X_{im2} = \cdot, W_{im} = \cdot]$, $\mathbb{E}[Y_{imd}[\text{sgn}(X_{im2}' b) - \text{sgn}(X_{im2}' \beta)] | X_{im1} = \cdot, W_{im} = \cdot]$, and $\mathbb{E}[Y_{im(1,1)}[\text{sgn}(W_{im}' r) - \text{sgn}(W_{im}' \gamma)] | V_{im}(\beta) = \cdot]$ are continuously differentiable with bounded first derivatives, (iii) $\mathbb{E}[Y_{imd} | Z_i = \cdot, Z_m = \cdot]$ is $\kappa_{1\beta}^{\text{th}}$ continuously differentiable with bounded $\kappa_{1\beta}^{\text{th}}$ derivatives. $f_{X_2, W | X_1}(\cdot)$ $(f_{X_1, W | X_2}(\cdot))$ is $\kappa_{2\beta}^{\text{th}}$ continuously differentiable with bounded $\kappa_{2\beta}^{\text{th}}$ derivatives. Denote $\kappa_\beta = \kappa_{1\beta} + \kappa_{2\beta}$. κ_β is an even integer greater than $k_1 + k_2$, (iv) $\mathbb{E}[Y_{im(1,1)} | V_i = \cdot, V_m = \cdot, W_i = \cdot, W_m = \cdot]$ is $\kappa_{1\gamma}^{\text{th}}$ continuously differentiable with bounded $\kappa_{1\gamma}^{\text{th}}$ derivatives, and $f_{V(\beta)|W}(\cdot)$ is $\kappa_{2\gamma}^{\text{th}}$ continuously differentiable with bounded $\kappa_{2\gamma}^{\text{th}}$ derivatives. Denote $\kappa_\gamma = \kappa_{1\gamma} + \kappa_{2\gamma}$. κ_γ is an even integer greater than 2, and (v) $\mathbb{E}[\|X_{imj}\|^2] < \infty$ for all $j = 1, 2$.[8]

C6 Let $\mathcal{N}_\beta$ denote a neighborhood of β. Then,

- All mixed second partial derivatives of $\varrho_i(b)$ exist on $\mathcal{N}_\beta$.
- There is an integrable function $M_1(Z_i)$ such that for all $b \in \mathcal{N}_\beta$, $\|\nabla^2 \varrho_i(b) - \nabla^2 \varrho_i(\beta)\|_F \leq M_1(Z_i) \|b - \beta\|$.
- $\mathbb{E}[\|\nabla \varrho_i(\beta)\|^2] < \infty$, $\mathbb{E}[\|\nabla^2 \varrho_i(\beta)\|_F] < \infty$, and $\mathbb{E}[\nabla^2 \varrho_i(\beta)]$ is negative definite.

C7 Let $\mathcal{N}_\gamma$ denote a neighborhood of γ. Then,

- All mixed second partial derivatives of $\tau_i(r)$ exist on $\mathcal{N}_\gamma$.
- There is an integrable function $M_2(Z_i)$ such that for all $r \in \mathcal{N}_\gamma$ $\|\nabla^2 \tau_i(r) - \nabla^2 \tau_i(\gamma)\|_F \leq M_2(Z_i) \|r - \gamma\|$.
- $\mathbb{E}[\|\nabla \tau_i(\gamma)\|^2] < \infty$, $\mathbb{E}[\|\nabla^2 \tau_i(\gamma)\|_F] < \infty$, and $\mathbb{E}[\nabla^2 \tau_i(\gamma)]$ is negative definite.
- All mixed second partial derivatives of $\mu(v_1, v_2, r)$ exist on $\mathcal{N}_\gamma$.
- There exists an continuous, integrable function $M_3(v_1, v_2)$ such that for all $r \in \mathcal{N}_\gamma$, $\|\nabla_{11}^2 \mu(v_1, v_2, r) - \nabla_{11}^2 \mu(v_1, v_2, \gamma)\|_F \leq M_3(v_1, v_2) \|r - \gamma\|$.

[8] All derivatives in this assumption are with respect to $\cdot$.

- $\mathbb{E}[\|\nabla_{13}^2\mu(V_i(\beta), V_i(\beta), \gamma)\|_F^2] < \infty$ and $\sup_{r\in\mathcal{N}_\gamma} \mathbb{E}[\|\nabla_{33}^2\nabla_1\mu_d(V_i(\beta), V_i(\beta), \gamma)\|_F] < \infty$.

C8 Let $K_\beta(\cdot)$ and $K_\gamma(\cdot)$ be kernel functions such that $K_\beta(\cdot/h_N) = h_N^{k_1+k_2}\mathcal{K}_{h_N}(\cdot)$ and $K_\gamma(\cdot/\sigma_N) = \sigma_N^2\mathcal{K}_{\sigma_N}(\cdot)$, respectively.

- $K_\beta(\cdot)$ is continuously differentiable and assumed to satisfy: (i) $\sup_{\upsilon\in\mathbb{R}^{k_1+k_2}} |K_\beta(\upsilon)| < \infty$, (ii) $\int K_\beta(\upsilon)\mathrm{d}\upsilon = 1$ and $\int |K_\beta(\upsilon)|\mathrm{d}\upsilon < \infty$, (iii) $\int \|\upsilon\||K_\beta(\upsilon)|\mathrm{d}\upsilon < \infty$, and (iv) for positive integers $\iota_1, ..., \iota_{k_1+k_2}$ satisfying $0 < \iota_1 + \cdots + \iota_{k_1+k_2} \le \kappa_\beta$,

$$\int \upsilon_1^{\iota_1}\cdots\upsilon_{k_1+k_2}^{\iota_{k_1+k_2}} K_\beta(\upsilon_1, ..., \upsilon_{k_1+k_2})\mathrm{d}\upsilon_1\cdots\mathrm{d}\upsilon_{k_1+k_2} \begin{cases} = 0 & \text{if } \iota_1 + \cdots + \iota_{k_1+k_2} < \kappa_\beta \\ \neq 0 & \text{if } \iota_1 + \cdots + \iota_{k_1+k_2} = \kappa_\beta \end{cases}$$

  and $\int |\upsilon_1^{\iota_1}\cdots\upsilon_{k_1+k_2}^{\iota_{k_1+k_2}}||K_\beta(\upsilon_1, ..., \upsilon_{k_1+k_2})|\mathrm{d}\upsilon_1\cdots\mathrm{d}\upsilon_{k_1+k_2} < \infty$ for $\iota_1 + \cdots + \iota_{k_1+k_2} = \kappa_\beta$.
- $K_\gamma(\cdot)$ is continuously differentiable and assumed to satisfy: (i) $\sup_{\upsilon\in\mathbb{R}^2} |K_\gamma(\upsilon)| < \infty$, (ii) $\int K_\gamma(\upsilon)\mathrm{d}\upsilon = 1$ and $\int |K_\gamma(\upsilon)|\mathrm{d}\upsilon < \infty$, (iii) $\int \|\upsilon\||K_\gamma(\upsilon)|\mathrm{d}\upsilon < \infty$, and (iv) for positive integers ι_1 and ι_2 satisfying $0 < \iota_1 + \iota_2 \le \kappa_\gamma$,

$$\int \upsilon_1^{\iota_1}\upsilon_2^{\iota_2} K_\gamma(\upsilon_1, \upsilon_2)\mathrm{d}\upsilon_1\mathrm{d}\upsilon_2 \begin{cases} = 0 & \text{if } \iota_1 + \iota_2 < \kappa_\gamma \\ \neq 0 & \text{if } \iota_1 + \iota_2 = \kappa_\gamma \end{cases}$$

  and $\int |\upsilon_1^{\iota_1}\upsilon_2^{\iota_2}||K_\gamma(\upsilon_1, \upsilon_2)|\mathrm{d}\upsilon_1\mathrm{d}\upsilon_2 < \infty$ for $\iota_1 + \iota_2 = \kappa_\gamma$.

C9 h_N and σ_N are sequences of positive numbers such that as $N \to \infty$: (i) $h_N \to 0$, $\sigma_N \to 0$, (ii) $\sqrt{N}h_N^{k_1+k_2} \to \infty$, $\sqrt{N}\sigma_N^2 \to \infty$, and (iii) $\sqrt{N}h_N^{\kappa_\beta} \to 0$, $\sqrt{N}\sigma_N^{\kappa_\gamma} \to 0$.

The boundedness and smoothness restrictions placed in Assumption C5 are needed for proving the uniform convergence of the objective functions to their population analogues. Assumptions C6 and C7 are analogous to the regularity conditions imposed in Sherman (1993). Assumptions C8 and C9 place mild restrictions on kernel functions and tuning parameters, all of which are standard in the literature.

The theorem below establishes the $\sqrt{N}$-consistency and asymptotic normality of the proposed estimators. The proof, as presented in Appendix A, follows from similar arguments to those used in Han (1987), Sherman (1993, 1994a,b), and Abrevaya, Hausman, and Khan (2010).

Theorem 2.2. *If Assumptions C1–C9 hold, then we have*

(i) $\sqrt{N}(\hat\beta-\beta) \xrightarrow{d} N(0, 4\left\{\mathbb{E}[\nabla^2\varrho_i(\beta)]\right\}^{-1}\mathbb{E}[\nabla\varrho_i(\beta)\nabla\varrho_i(\beta)']\left\{\mathbb{E}[\nabla^2\varrho_i(\beta)]\right\}^{-1})$ *or, alternatively,* $\hat\beta-\beta$ *has the linear representation*

$$\hat\beta - \beta = -\frac{2}{N}\left\{\mathbb{E}[\nabla^2\varrho_i(\beta)]\right\}^{-1}\sum_{i=1}^N \varrho_i(\beta) + o_P(N^{-1/2}).$$

(ii) $\sqrt{N}(\hat\gamma - \gamma) \xrightarrow{d} N\left(0, \mathbb{E}[\nabla^2\tau_i(\gamma)]^{-1}\mathbb{E}\left[\varDelta_i\varDelta_i'\right]\mathbb{E}[\nabla^2\tau_i(\gamma)]^{-1}\right)$, *where*

$$\varDelta_i \equiv 2\nabla\tau_i(\gamma) + 2\mathbb{E}\left[\nabla_{13}^2\mu\left(V_i(\beta), V_i(\beta), \gamma\right)\right]\left\{\mathbb{E}[\nabla^2\varrho_i(\beta)]\right\}^{-1}\varrho_i(\beta).$$

2.3 Inference

Theorem 2.2 indicates that our localized MRC estimators are asymptotically normal and have asymptotic variances of usual sandwich structure. The expressions for the asymptotic variances consist of first and second derivatives of the limit of the expectation of the maximands in (2.11) and (2.12). To use these results for statistical inference, Sherman (1993) proposed to apply the numerical derivative method of Pakes and Pollard (1989). Hong, Mahajan, and Nekipelov (2015) investigated the application of the numerical derivative method in extremum estimators, including second-order U-statistics. Alterntively, Cavanagh and Sherman (1998) suggested nonparametrically estimating these quantities. These procedures, however, require selecting additional tuning parameters.

To avoid this complexity, we propose to conduct inference by the classic nonparametric bootstrap for the ease of implementation. Subbotin (2007) proved the consistency of the nonparametric bootstrap for Han's (1987) MRC estimators.[9] The structure of the objective functions of our estimators is similar to that of the standard MRC estimator, but they do differ in two ways. First, our estimators need matching and thus contain kernel functions. Second, the objective function for estimating $\hat{\gamma}$ contains a first-step $\hat{\beta}$ to approximate the true value of β. These two differences usually do not make the bootstrap inconsistent. For example, Horowitz (2001) showed the consistency of the bootstrap for a range of estimators involved with kernel functions; Chen, Linton, and van Keilegom (2003) demonstrated the consistency of the bootstrap for estimators with a first-step estimation component and a well behaved objective function.

The bootstrap estimators $\hat{\beta}^*$ and $\hat{\gamma}^*$ are obtained as follows. Draw $\{Y_i^{*\prime}, Z_i^{*\prime}\}_{i=1}^N$ independently from the collection of the sample values $\{Y_i^\prime, Z_i^\prime\}_{i=1}^N$ with replacement. $\hat{\beta}^*$ is obtained from

$$\begin{aligned}\hat{\beta}^* = \arg\max_{b\in\mathcal{B}} \mathcal{L}_{N,\beta}^{K*}(b) \equiv \arg\max_{b\in\mathcal{B}} \sum_{i=1}^{N-1}\sum_{m>i}\sum_{d\in\mathcal{D}} \{ &\mathcal{K}_{h_N}\left(X_{im2}^*, W_{im}^*\right) Y_{mid}^* \operatorname{sgn}\left(X_{im1}^{*\prime} b\right)\cdot(-1)^{d_1} \\ &+ \mathcal{K}_{h_N}\left(X_{im1}^*, W_{im}^*\right) Y_{mid}^* \operatorname{sgn}\left(X_{im2}^{*\prime} b\right)\cdot(-1)^{d_2} \}.\end{aligned} \tag{2.13}$$

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

$$\hat{\gamma}^* = \arg\max_{r\in\mathcal{R}} \mathcal{L}_{N,\gamma}^{K*}(r, \hat{\beta}^*) \equiv \arg\max_{r\in\mathcal{R}} \sum_{i=1}^{N-1}\sum_{m>i} \mathcal{K}_{\sigma_N}(V_{im}^*(\hat{\beta}^*)) Y_{im(1,1)}^* \operatorname{sgn}\left(W_{im}^{*\prime} r\right). \tag{2.14}$$

We claim that

$$\sqrt{N}(\hat{\beta}^* - \hat{\beta}) \xrightarrow{d} N(0, 4\left\{\mathbb{E}[\nabla^2 \varrho_i(\beta)]\right\}^{-1} \mathbb{E}[\nabla \varrho_i(\beta)\nabla\varrho_i(\beta)^\prime] \left\{\mathbb{E}[\nabla^2 \varrho_i(\beta)]\right\}^{-1}),$$

and

$$\sqrt{N}\left(\hat{\gamma}^* - \hat{\gamma}\right) \xrightarrow{d} N\left(0, \mathbb{E}[\nabla^2\tau_i(\gamma)]^{-1}\mathbb{E}\left[\mathit{\Delta}_i \mathit{\Delta}_i^\prime\right]\mathbb{E}[\nabla^2\tau_i(\gamma)]^{-1}\right).$$

A complete proof of the consistency of the bootstrap procedure is lengthy and tedious. Considering that this process is relatively standard given the literature, we only present an outline of

[9] Jin, Ying, and Wei (2001) proposed an alternative resampling method for U-statistics by perturbing the objective function repeatedly.

the proof in Appendix D.1. We end this section by describing the the algorithm for constructing 95% confidence intervals for β and γ using the nonparametric bootstrap.

1. Draw $(Y_i^{*\prime}, Z_i^{*\prime})'$, $i = 1, ..., N$, independently with replacement from the original sample.
2. Obtain $\hat{\beta}^*$ and $\hat{\gamma}^*$ from equations (2.13) and (2.14).
3. Repeat Steps 1 and 2 for B times independently and arrive at a series of $\hat{\beta}^*$ and $\hat{\gamma}^*$, say, $\{\hat{\beta}^{*(b)}, \hat{\gamma}^{*(b)}\}_{b=1}^B$.
4. Let $Q_{\hat{\beta}^*}(\tau)$ denote the τ-th quantile of $\{\hat{\beta}^{*(b)}\}_{b=1}^B$, $0 \leq \tau \leq 1$. Define $Q_{\hat{\gamma}^*}(\tau)$ analogously. 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

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

3 Panel Data Model

3.1 Identification and Localized MS Estimation

The increasing availability of panel data sets provides new opportunities for the econometrician to control for unobserved heterogeneity across agents. This helps relax the strict exogeneity restriction placed in the cross-sectional setting. Here we apply our matching-based identification strategy presented in Section 2.1 to a panel data bundle choice model. The latent utilities and observed choices are expressed as

$$U_{dt} = \sum_{j=1}^{2} (X_{jt}'\beta + \alpha_j + \epsilon_{jt}) \cdot d_{jt} + \eta_t \cdot (W_t'\gamma + \alpha_b) \cdot d_{1t} \cdot d_{2t}, \tag{3.1}$$

and

$$Y_{dt} = 1[U_{dt} > U_{d't}, \forall d' \in \mathcal{D} \setminus d], \tag{3.2}$$

where we use $t = 1, ..., T$ to denote time periods and suppress the agent subscript i to simplify notations as in this section all variables are for each agent. Note that the random utility specified by expression (3.1) includes a set of unobserved (to the econometrician) agent-specific fixed effects $\alpha \equiv (\alpha_1, \alpha_2, \alpha_b)$ associated with the two stand-alone alternatives and the bundle, respectively. Denote $Z_t = (X_{1t}', X_{2t}', W_t')'$ and $\xi_t = (\epsilon_{1t}, \epsilon_{2t}, \eta_t)'$. In line with the literature on fixed effects methods, we place no restrictions on the distribution of α conditional on $Z^T \equiv (Z_1', ..., Z_T')'$ and $\xi^T \equiv (\xi_1', ..., \xi_T')'$.

Here we consider the identification and estimation of model (3.1)–(3.2) with $T < \infty$ and $N \to \infty$ (i.e., short panel). As will be clear below, our approach works for any $T \geq 2$. To our knowledge, this is the first work in the literature on panel data bundle choice models with fixed effects. Our identification strategy relies on similar group homogeneity condition as the one adopted in Manski (1987), Pakes and Porter (2016), and Shi et al. (2018) for binary and multinomial choice models. Specifically, assuming $\xi_s \overset{d}{=} \xi_t|(\alpha, Z_s, Z_t)$ for any two time periods s and t, we have the following moment inequalities:

For all d with $d_1 = 1$ ($d_1 = 0$) and any fixed $(x_2', w')'$,

$$X_{1t}'\beta \geq X_{1s}'\beta \Leftrightarrow P(Y_{dt} = 1|Z^T, X_{2t} = x_2, W_t = w) \underset{(\leq)}{\geq} P(Y_{ds} = 1|Z^T, X_{2s} = x_2, W_s = w). \quad (3.3)$$

For all d with $d_2 = 1$ ($d_2 = 0$) and any fixed $(x_1', w')'$,

$$X_{2t}'\beta \geq X_{2s}'\beta \Leftrightarrow P(Y_{dt} = 1|Z^T, X_{1t} = x_1, W_t = w) \underset{(\leq)}{\geq} P(Y_{ds} = 1|Z^T, X_{1s} = x_1, W_s = w). \quad (3.4)$$

For $d = (1, 1)$ (all $d \neq (1, 1)$) and any fixed $(x_1', x_2')'$,

$$W_t'\gamma \geq W_s'\gamma \Leftrightarrow P(Y_{dt} = 1|Z^T, X_{1t} = x_1, X_{2t} = x_2) \underset{(\leq)}{\geq} P(Y_{ds} = 1|Z^T, X_{1s} = x_1, X_{2s} = x_2). \quad (3.5)$$

The monotonic relations (3.3)–(3.5)[10] are analogous to (2.4)–(2.8) for the cross-sectional model. In the presence of the fixed effects, we match and make our comparisons within agents over time, as opposed to pairs of agents. These moment inequalities are the foundation of our identification result for model (3.1)–(3.2).

We impose the following conditions for identification. For notational convenience, we let $X_{jts} \equiv X_{jt} - X_{js}$ and $W_{ts} \equiv W_t - W_s$ for all $j = 1, 2$ and (s, t).

- **P1** $\{(Y_i^T, Z_i^T)\}_{i=1}^N$ are i.i.d, where $Y_i^T \equiv \{(Y_{i(0,0)t}, Y_{i(1,0)t}, Y_{i(0,1)t}, Y_{i(1,1)t})\}_{t=1}^T$.
- **P2** For almost all (α, Z^T), (i) $\xi_t \overset{d}{=} \xi_s|\,(\alpha, Z_t, Z_s)$, and (ii) the distribution of ξ_t conditional on α is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^2 \times \mathbb{R}_+$ for all t.
- **P3** $(\beta', \gamma')' \in \mathcal{B} \times \mathcal{R}$, where $\mathcal{B} = \{b \in \mathbb{R}^{k_1}|\,\|b\| = 1, b^{(1)} \neq 0\}$ and $\mathcal{R} = \{r \in \mathbb{R}^{k_2}|\,\|r\| = 1, r^{(1)} \neq 0\}$.
- **P4** (i) $X_{1ts}^{(1)}$ ($X_{2ts}^{(1)}$) has a.e. positive Lebesgue density on $\mathbb{R}$ conditional on $\tilde{X}_{1ts}$ ($\tilde{X}_{2ts}$) and conditional on $(X_{2ts}', W_{ts}')'$ ($(X_{1ts}', W_{ts}')'$) in a neighborhood of $(X_{2ts}', W_{ts}')'$ ($(X_{1ts}', W_{ts}')'$) near zero, (ii) the support of X_{1ts} (X_{2ts}) conditional on $(X_{2ts}', W_{ts}')'$ ($(X_{1ts}', W_{ts}')'$) in a neighborhood of $(X_{2ts}', W_{ts}')'$ ($(X_{1ts}', W_{ts}')'$) near zero is not contained in any proper linear subspace of $\mathbb{R}^{k_1}$, (iii) $W_{ts}^{(1)}$ has a.e. positive Lebesgue density on $\mathbb{R}$ conditional on $\tilde{W}_{ts}$ and conditional on (X_{1ts}, X_{2ts}) in a neighborhood of (X_{1ts}, X_{2ts}) near zero, (iv) the support of W_{ts} conditional on (X_{1ts}, X_{2ts}) in a neighborhood of (X_{1ts}, X_{2ts}) near zero is not contained in any proper linear subspace of $\mathbb{R}^{k_2}$.

[10]Once β is identified through (3.3) or (3.4), γ can be alternatively identified via matching indexes $(X_{1t}'\beta, X_{2t}'\beta)$, i.e., for $d = (1, 1)$ and any fixed $(v_1, v_2)'$, $W_t'\gamma \geq W_s'\gamma \Leftrightarrow P(Y_{dt} = 1|Z^T, X_{1t}'\beta = v_1, X_{2t}'\beta = v_2) \geq P(Y_{ds} = 1|Z^T, X_{1s}'\beta = v_1, X_{2s}'\beta = v_2)$. But as will be explained in Remark 3.1, these moment inequalities will not be used to construct the MS estimator of γ.

Assumptions P1–P4 are analogous to the identification conditions used in Manski (1987). Note that Assumption P2 allows for arbitrary correlation between the fixed effects and the observed covariates, provided that the correlation is time stationary. This assumption substantially relaxes the strict exogeneity condition in the cross-sectional case. But the price to pay is a slower convergence rate, as will be shown in Theorem 3.2 below. A technical explanation of the slower convergence rate can be found in Remark 3.2.

We summarize our identification results in the following theorem, of which the proof is omitted for brevity as it is very similar to the one for the cross section case.

Theorem 3.1. *Suppose Assumptions P1–P4 hold. β and γ are identified.*

The monotonic relations (3.3)–(3.5) motivate the following two-step localized MS estimation procedure: Given a random sample of N agents $i = 1, ..., N$, the estimator of β is defined as the maximizer, over the parameter space $\mathcal{B}$, of the objective function

$$\mathcal{L}_{N,\beta}^{P}(b) = \sum_{i=1}^{N}\sum_{t>s}\sum_{d\in\mathcal{D}}\{1[X_{i2ts}=0, W_{its}=0](Y_{ids}-Y_{idt})\text{sgn}((X_{i1t}-X_{i1s})'b)\cdot(-1)^{d_1}$$
$$+ 1[X_{i1ts}=0, W_{its}=0](Y_{ids}-Y_{idt})\text{sgn}((X_{i2t}-X_{i2s})'b)\cdot(-1)^{d_2}\}. \quad (3.6)$$

Similarly, the estimator of γ is obtained by maximizing, over the parameter space $\mathcal{R}$, the following objective function

$$\mathcal{L}_{N,\gamma}^{P}(r) = \sum_{i=1}^{N}\sum_{t>s}1[X_{i1ts}=0, X_{i2ts}=0](Y_{i(1,1)t}-Y_{i(1,1)s})\text{sgn}((W_{it}-W_{is})'r). \quad (3.7)$$

Note that objective functions (3.6) and (3.7) take value zero for observations whose choice is time-invariant, i.e., our approach uses only data on "switchers" in a way similar to the estimator in Manski (1987).

For the case with continuous covariates, we replace the binary weights in (3.6) and (3.7) with the corresponding kernel weights analogous to the ones used in Section 2.2. Specifically, we obtain the estimator $\hat{\beta}$ of β with the objective function

$$\mathcal{L}_{N,\beta}^{P,K}(b) = \sum_{i=1}^{N}\sum_{t>s}\sum_{d\in\mathcal{D}}\{\mathcal{K}_{h_N}(X_{i2ts}, W_{its})(Y_{ids}-Y_{idt})\text{sgn}((X_{i1t}-X_{i1s})'b)\cdot(-1)^{d_1}$$
$$+ \mathcal{K}_{h_N}(X_{i1ts}, W_{its})(Y_{ids}-Y_{idt})\text{sgn}((X_{i2t}-X_{i2s})'b)\cdot(-1)^{d_2}\}, \quad (3.8)$$

where $\mathcal{K}_{h_N}(X_{ijts}, W_{its}) \equiv \Pi_{l=1}^{k_1} h_N^{-1} K\left(X_{ijts,l}/h_N\right) \Pi_{l=1}^{k_2} h_N^{-1} K\left(W_{its,l}/h_N\right)$ for $j = 1, 2$, $K(\cdot)$ is a kernel density function, and h_N is a bandwidth sequence that converges to 0 as $N \to \infty$. Similarly, we compute the estimator $\hat{\gamma}$ of γ using the objective function

$$\mathcal{L}_{N,\gamma}^{P,K}(r) = \sum_{i=1}^{N}\sum_{t>s}\mathcal{K}_{\sigma_N}(X_{i1ts}, X_{i2ts})(Y_{i(1,1)t}-Y_{i(1,1)s})\text{sgn}((W_{it}-W_{is})'r), \quad (3.9)$$

where $\mathcal{K}_{\sigma_N}(X_{i1ts}, X_{i2ts}) \equiv \Pi_{l=1}^{k_1}\sigma_N^{-1}K\left(X_{i1ts,l}/\sigma_N\right)\Pi_{l=1}^{k_1}\sigma_N^{-1}K\left(X_{i2ts,l}/\sigma_N\right)$ for $j=1,2$, and σ_N is a bandwidth sequence that converges to 0 as $N \to \infty$.

From a methodology perspective, our panel data estimators fit in the substantial literature on MS approaches stemming from the seminal work of Manski (1975, 1985) analyzing binary choice models. More recently, Fox (2007), Yan and Yoo (2019), and Khan et al. (2019) apply this method to estimation of multinomial choice models in either cross-sectional or panel data settings.

Remark 3.1. Unlike the cross section case, here we choose not to estimate γ by matching $X_{jt}'\hat{\beta}$ and $X_{js}'\hat{\beta}$, $j=1,2$, because of the following observation. As will be shown in Theorem 3.2 below, the convergence rate of $\hat{\beta}$ is $(Nh_N^{k_1+k_2})^{-1/3}$. If we knew the true value of β, $\hat{\gamma}$ could be obtained by matching $X_{jt}'\beta$ and $X_{js}'\beta$. Using the same arguments, we can show that this infeasible estimator has a rate of convergence $(N\sigma_N^2)^{-1/3}$. As a result, it is only possibly beneficial to match $X_{jt}\hat{\beta}$ and $X_{js}\hat{\beta}$ for estimating γ when $\sigma_N^2 = o(h_N^{k_1+k_2})$, because otherwise the asymptotics of $\hat{\beta}$ will dominate the limiting distribution of $\hat{\gamma}$. But this involves careful selection of tuning parameters h_N and σ_N. In more general cases, this selection relies not only on the dimension of the regressor space but also on the cardinality of the choice set, and thus should be made on a case-by-case basis. Due to the lack of a universally applicable treatment, we choose to simply match the covariates to avoid this complexity and the ensuing discussion.

Remark 3.2. The rate of convergence of the estimators for the cross-sectional model is $N^{-1/2}$. The intuition is that each observation can be matched with $Nh_N^{k_1+k_2}$ and $N\sigma_N^{2k_1}$ observations (in probability) for estimating β and γ, respectively. Thus, all observations are useful and the curse of dimensionality is not a concern for the cross-sectional estimators. By contrast, an observation is useful for the panel data estimators if all its covariates are very close to each other across two time periods. For a fixed T, the probability for an observation being useful is proportional to $h_N^{k_1+k_2}$ or $\sigma_N^{2k_1}$. Therefore, the number of the "useful" observations is proportional to $Nh_N^{k_1+k_2}$ and $N\sigma_N^{2k_1}$, and the estimation suffers from the curse of dimensionality. We may remedy this problem if $T \to \infty$ such that $Th_N^{k_1+k_2} \to \infty$ and $T\sigma_N^{2k_1} \to \infty$. Since this paper focuses on models with short panel, we leave the discussion on this issue to future research. Further, the objective functions have a "sharp" edge as in Kim and Pollard (1990). As a result, the convergence rates for $\hat{\beta}$ and $\hat{\gamma}$ are expected to be $(Nh_N^{k_1+k_2})^{-1/3}$ and $(N\sigma_N^{2k_1})^{-1/3}$, respectively. The cross-sectional estimators, however, do not suffer this "sharp" edge effect because the objective functions are U-statistics and the "sharp" edge effect vanishes after we decompose the objective functions (see, e.g., the second line of equation (A.3)) when deriving the asymptotics.

To establish the asymptotic properties of $\hat{\beta}$ and $\hat{\gamma}$, we need the following conditions in addition to Assumptions P1–P4.

P5 Let $f_{X_{jt},W_t}(\cdot)$ denote the density of the random vector $(X_{jt}', W_t')'$ for all $j=1,2$, and $f_{X_{1ts},X_{2ts}}(\cdot)$ denote the density of the random vector (X_{1ts}, X_{2ts}). $f_{X_{jt},W_t}(\cdot)$'s and $f_{X_{1ts},X_{2ts}}(\cdot)$ are absolutely continuous, bounded from above on their supports, strictly positive in a neighborhood of zero, and twice continuous differentiable a.e. with bounded derivatives.

P6 For all $d \in \mathcal{D}$, $\mathbb{E}[Y_{dts}\text{sgn}\,(X_{1ts}'b)\,|X_{2ts}, W_{ts}]$ and $\mathbb{E}[Y_{dts}\text{sgn}\,(X_{2ts}'b)\,|X_{1ts}, W_{ts}]$ are twice continuous differentiable w.r.t. b a.e. with bounded derivatives, and $\mathbb{E}[Y_{dts}\text{sgn}\,(W_{ts}'r)\,|X_{1ts}, X_{2ts}]$ are twice continuous differentiable w.r.t. r a.e. with bounded derivatives.

P7 $K(\cdot)$ is a function of bounded variation and has a compact support. It satisfies: (i) $K(v) \geq 0$ and $\sup_v |K(v)| < \infty$, (ii) $\int K(v)\mathrm{d}v = 1$, (iii) $\int vK(v)\mathrm{d}v = 0$, (iv) $\int v^2K(v)\mathrm{d}v < \infty$ and (v) $K(\cdot)$ is twice continuously differentiable with bounded first and second derivatives.

P8 $(\hat{\beta}, \hat{\gamma})$ satisfies

$$N^{-1}\mathcal{L}_{N,\beta}^{P,K}(\hat{\beta}) \geq \max_{b\in\mathcal{B}} N^{-1}\mathcal{L}_{N,\beta}^{P,K}(b) - o_P((Nh_N^{k_1+k_2})^{-2/3})$$

and

$$N^{-1}\mathcal{L}_{N,\gamma}^{P,K}(\hat{\gamma}) \geq \max_{r\in\mathcal{R}} N^{-1}\mathcal{L}_{N,\gamma}^{P,K}(r) - o_P((N\sigma_N^{2k_1})^{-2/3}).$$

P9 h_N and σ_N are sequences of positive numbers such that as $N \to \infty$: (i) $h_N \to 0$ and $\sigma_N \to 0$, (ii) $Nh_N^{k_1+k_2} \to \infty$ and $N\sigma_N^{2k_1} \to \infty$, and (iii) $(Nh_N^{k_1+k_2})^{2/3}h_N^2 \to 0$ and $(N\sigma_N^{2k_1})^{2/3}\sigma_N^2 \to 0$.

The boundedness and smoothness restrictions placed in Assumptions P5–P7 are regularity conditions needed for proving the uniform convergence of the objective functions to their population analogues. Assumption P8 is a standard technical condition. Assumption P9 is also standard, of which (iii) is placed to ensure the bias term from the kernel estimation asymptotically negligible. Note that Assumption P6 implicitly assumes that the second moments of X_{jts}'s and W_{ts} exist.

For the asymptotic distribution, we focus on the case where all regressors are continuous and introduce the following notations to ease expositions. Let

$$\begin{aligned}\phi_{Ni}\,(b) \equiv \sum_{t>s}\sum_{d\in\mathcal{D}}\{&\mathcal{K}_{h_N}(X_{i2ts}, W_{its})Y_{idst}\,(-1)^{d_1}\left(1[X_{i1ts}'b > 0] - 1[X_{i1ts}'\beta > 0]\right)\\ &+ \mathcal{K}_{h_N}(X_{i1ts}, W_{its})Y_{idst}\,(-1)^{d_2}\left(1[X_{i2st}'b > 0] - 1[X_{i2st}'\beta > 0]\right)\}\end{aligned}$$

and

$$\varphi_{Ni}\,(r) \equiv \sum_{t>s}\mathcal{K}_{\sigma_N}(X_{i1ts}, X_{i2ts})Y_{i(1,1)ts}(1[W_{its}'r > 0] - 1[W_{its}'\gamma > 0]).$$

$\hat{\beta}$ and $\hat{\gamma}$ can be equivalently obtained from

$$\hat{\beta} = \arg\max_{b\in\mathcal{B}} N^{-1}\sum_{i=1}^{N}\phi_{Ni}\,(b) \text{ and } \hat{\gamma} = \arg\max_{r\in\mathcal{R}} N^{-1}\sum_{i=1}^{N}\varphi_{Ni}\,(r)\,,$$

because $\text{sgn}\,(\cdot) = 2\times 1\,[\cdot] - 1$ and adding terms not related to b or r has no effects on the optimization.

Theorem 3.2. *Suppose Assumptions P1–P9 hold. Then*

$$(Nh_N^{k_1+k_2})^{1/3}(\hat{\beta} - \beta) \xrightarrow{d} \arg\max_{\rho\in\mathbb{R}^{k_1}} \mathcal{Z}_1\,(\rho)\,,$$

where $\mathcal{Z}_1(\rho)$ is a Gaussian process with continuous sample path, expected value $\frac{1}{2}\rho'\mathbb{V}\rho$ and covariance kernel $\mathbb{H}_1(\rho_1,\rho_2)$, and

$$(N\sigma_N^{2k_1})^{1/3}(\hat{\gamma}-\gamma) \xrightarrow{d} \arg\max_{\delta\in\mathbb{R}^{k_2}} \mathcal{Z}_2(\delta),$$

where $\mathcal{Z}_2(\delta)$ is a Gaussian process with continuous sample path, expected value $\frac{1}{2}\delta'\mathbb{W}\delta$ and covariance kernel $\mathbb{H}_2(\delta_1,\delta_2)$. $\mathbb{V}$, $\mathbb{W}$, $\mathbb{H}_1$, and $\mathbb{H}_2$ are defined in equations (B.1), (B.2), (B.3), and (B.4), respectively.

In line with existing results on the MS estimators (e.g., Kim and Pollard (1990) and Seo and Otsu (2018)), the limiting distributions of $\hat{\beta}$ and $\hat{\gamma}$ are non-Gaussian and their rates of convergence are slower than $N^{-1/3}$. Inference using the limiting distribution directly is rather hard. We recommend a bootstrap-based procedure for the inference in next section. One alternative is to adopt a smoothed MS approach (e.g., Horowitz (1992)), which may yield a faster and asymptotically normal estimator. We leave this topic for future research.

3.2 Inference

Abrevaya and Huang (2005) proved the inconsistency of the classic bootstrap for the ordinary MS estimator. Our panel data estimators are indeed of the MS type. We expect that the classic bootstrap does not work for our estimators, either. Applying the same arguments in Abrevaya and Huang (2005) with slight modification can prove this result.

Valid inference can be made by the m-out-of-n bootstrap, according to Lee and Pun (2006). Recently, Hong and Li (2020) proposed the numerical bootstrap, and showed the superior performance of this procedure over the m-out-of-n bootstrap. For our panel data estimators, another advantage of the numerical bootstrap is that we do not need to choose another set of bandwidths (h_N and σ_N) for the estimation using bootstrap series, as oppose to the m-out-of-n bootstrap. Based on these considerations, we propose to conduct the inference using the numerical bootstrap.

The numerical bootstrap estimators $\hat{\beta}^*$ and $\hat{\gamma}^*$ are obtained as follows. Draw $\left\{Y_i^{T*\prime}, Z_i^{T*\prime}\right\}_{i=1}^N$ independently from the collection of the sample values $\left\{Y_i^{T\prime}, Z_i^{T\prime}\right\}_{i=1}^N$ with replacement. Then, obtain $\hat{\beta}^*$ from

$$\hat{\beta}^* = \arg\max_{b\in\mathcal{B}} N^{-1}\sum_{i=1}^N \phi_{Ni}(b) + (N\varepsilon_{N1})^{1/2}\cdot[N^{-1}\sum_{i=1}^N \phi_{Ni}^*(b) - N^{-1}\sum_{i=1}^N \phi_{Ni}(b)], \tag{3.10}$$

where ε_{N1} is a tuning parameter to be discussed later and

$$\begin{aligned}\phi_{Ni}^*(b) = \sum_{t>s}\sum_{d\in\mathcal{D}} \{&\mathcal{K}_{h_N}(X_{i2ts}^*, W_{its}^*)Y_{idst}^*(-1)^{d_1}\left[\operatorname{sgn}((X_{i1t}^* - X_{i1s}^*)'b) - \operatorname{sgn}((X_{i1t}^* - X_{i1s}^*)'\beta)\right] \\ &+ \mathcal{K}_{h_N}(X_{i1ts}^*, W_{its}^*)Y_{idst}^*(-1)^{d_2}\left[\operatorname{sgn}((X_{i2t}^* - X_{i2s}^*)'b) - \operatorname{sgn}((X_{i2t}^* - X_{i2s}^*)'\beta)\right]\}.\end{aligned}$$

Similarly, compute $\hat{\gamma}^*$ from

$$\hat{\gamma}^* = \arg\max_{r \in \mathcal{R}} N^{-1} \sum_{i=1}^{N} \varphi_{Ni}(r) + (N\varepsilon_{N2})^{1/2} \cdot [N^{-1} \sum_{i=1}^{N} \varphi^*_{Ni}(r) - N^{-1} \sum_{i=1}^{N} \varphi_{Ni}(r)], \tag{3.11}$$

where ε_{N2} is a tuning parameter and

$$\varphi^*_{Ni}(r) = \sum_{t>s} \mathcal{K}_{\sigma_N}(X^*_{i1ts}, X^*_{i2ts}) Y^*_{i(1,1)ts} \left[\mathrm{sgn}((W^*_{it} - W^*_{is})' r) - \mathrm{sgn}((W^*_{it} - W^*_{is})' \gamma)\right].$$

$\phi^*_{Ni}(b)$ and $\varphi^*_{Ni}(r)$ are simply $\phi_{Ni}(b)$ and $\varphi_{Ni}(b)$, respectively, using bootstrap series.

The following conditions for the ε_N's are needed for the consistency of the numerical bootstrap.

P10 ε_{N1} and ε_{N2} are sequences of positive numbers such that as $N \to \infty$: (i) $\varepsilon_{N1} \to 0$ and $\varepsilon_{N2} \to 0$, (ii) $N\varepsilon_{N1} \to \infty$ and $N\varepsilon_{N2} \to \infty$, (iii) $\varepsilon_{N1}^{-1} h_N^{k_1+k_2} \to \infty$ and $\varepsilon_{N2}^{-1} \sigma_N^{2k_1} \to \infty$, and (iv) $(\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{2/3} h_N^2 \to 0$ and $(\varepsilon_{N2}^{-1} \sigma_N^{2k_1})^{2/3} \sigma_N^2 \to 0$.

When $\varepsilon_{N1}^{-1} = \varepsilon_{N2}^{-1} = N$, the numerical bootstrap reduces to the classic nonparametric bootstrap. The numerical bootstrap excludes this case and requires $N\varepsilon_{N1} \to \infty$ and $N\varepsilon_{N2} \to \infty$ as $N \to 0$. To understand conditions (iii) and (iv) for the ε_N's, we note that ε_{N1}^{-1} and ε_{N2}^{-1} play a similar role as the m in the m-out-of-n bootstrap (use only m observations for the estimation).[11] See Appendix D.2 for more details on how these conditions ensure the consistency of the bootstrap procedure.

Our estimators do not directly satisfy all conditions required by Hong and Li (2020). We illustrate this point for β. First, we need to change their $n^{-\gamma}$ (the inverse of the convergence rate in their notations) to $(Nh_N^{k_1+k_2})^{-1/3}$. That is, we replace n with $Nh_N^{k_1+k_2}$ and set γ as $1/3$. Second, condition (vi) in Theorem 4.1 of Hong and Li (2020) is not satisfied. However, the results in Hong and Li (2020) hold by modifying this condition to (using their notations), for example for β, that

$$\Sigma_{1/2}(\rho_1, \rho_2) = \lim_{N\to\infty} (Nh_N^{k_1+k_2})^{1/3} \mathbb{E}[h_N^{k_1+k_2} \phi_{Ni}(\beta + \rho_1 (Nh_N^{k_1+k_2})^{-1/3}) \phi_{Ni}(\beta + \rho_2 (Nh_N^{k_1+k_2})^{-1/3})]$$

exists for all $\rho_1, \rho_2 \in \mathbb{R}^{k_1}$. This is indeed the case as shown by our Lemma B.2 in Appendix B.

Then, applying similar arguments as in Hong and Li (2020) leads to

$$(\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3} (\hat{\beta}^* - \hat{\beta}) \xrightarrow{d} \arg\max_{\rho \in \mathbb{R}^{k_1}} \mathcal{Z}^*_1(\rho),$$

where $\mathcal{Z}^*_1(\rho)$ is an independent copy of $\mathcal{Z}_1(\rho)$ defined in Theorem 3.2. Similarly, we have

$$(\varepsilon_{N2}^{-1} \sigma_N^{2k_1})^{1/3} (\hat{\gamma}^* - \hat{\gamma}) \xrightarrow{d} \arg\max_{\delta \in \mathbb{R}^{k_2}} \mathcal{Z}^*_2(\delta),$$

where $\mathcal{Z}^*_2(\delta)$ is an independent copies of $\mathcal{Z}_2(\delta)$ defined in Theorem 3.2. We outline the proof of these results in Appendix D.2.

[11] Note that our ε_N was written as $\varepsilon_N^{1/2}$ in Hong and Li (2020).

We now discuss the choice of the tuning parameters. Recall that $\hat{\beta}$ is of the rate $(Nh_N^{k_1+k_2})^{-1/3}$, and additionally we need $(Nh_N^{k_1+k_2})^{2/3}h_N^2 \to 0$. To attain a fast rate of convergence, we tend to set h_N as large as possible. For example, we may set $(Nh_N^{k_1+k_2})^{2/3}h_N^2 \propto \log(N)^{-1}$. For the same reason, we may set $(N\sigma_N^{2k_1})^{2/3}\sigma_N^2 \propto \log(N)^{-1}$. These lead to

$$h_N \propto \log(N)^{-\frac{3}{2k_1+2k_2+6}} N^{-\frac{1}{k_1+k_2+3}} \text{ and } \sigma_N \propto \log(N)^{-\frac{3}{4k_1+6}} N^{-\frac{1}{2k_1+3}}$$

As recommended in Hong and Li (2020), we can choose ε_{N1} and ε_{N2} such that

$$\varepsilon_{N1}^{-1}h_N^{k_1+k_2} \propto (Nh_N^{k_1+k_2})^{2/3} \text{ and } \varepsilon_{N2}^{-1}\sigma_N^{2k_1} \propto (N\sigma_N^{2k_1})^{2/3},$$

which then imply that

$$\varepsilon_{N1} \propto N^{-\frac{k_1+k_2+2}{k_1+k_2+3}} \log(N)^{-\frac{k_1+k_2}{2k_1+2k_2+6}} \text{ and } \varepsilon_{N2} \propto N^{-\frac{2k_1+2}{2k_1+3}} \log(N)^{-\frac{k_1}{2k_1+3}}. \tag{3.12}$$

As the last component of this section, we provide the algorithm for constructing 95% confidence intervals for β and γ using the numerical bootstrap.

1. Draw $(Y_i^{T*\prime}, Z_i^{T*\prime})'$, $i = 1, ..., N$, independently with replacement from the original sample.
2. Obtain $\hat{\beta}^*$ and $\hat{\gamma}^*$ from equations (3.10) and (3.11), respectively.
3. Repeat Steps 1 and 2 for B times independently and arrive at a series of $\hat{\beta}^*$ and $\hat{\gamma}^*$, 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

$$[\hat{\beta} - N^{-1/3} \cdot \varepsilon_{N1}^{-1/3} \cdot (Q_{\hat{\beta}^*}(0.975) - \hat{\beta}),\ \hat{\beta} - N^{-1/3} \cdot \varepsilon_{N1}^{-1/3} \cdot (Q_{\hat{\beta}^*}(0.025) - \hat{\beta})]$$

and

$$[\hat{\gamma} - N^{-1/3} \cdot \varepsilon_{N2}^{-1/3} \cdot (Q_{\hat{\gamma}^*}(0.975) - \hat{\gamma}),\ \hat{\gamma} - N^{-1/3} \cdot \varepsilon_{N2}^{-1/3} \cdot (Q_{\hat{\gamma}^*}(0.025) - \hat{\gamma})].$$

4 Monte Carlo Experiments

In this section, we investigate the finite-sample properties of our proposed estimation and inference procedures, in both cross-sectional and panel data models, by means of Monte Carlo experiments. We set sample sizes $N = 250, 500, 1000$ for cross-sectional designs, while we set $N = 1000, 2500, 5000$ for panel data designs, considering the slower convergence rate. All simulation results are obtained from 1000 independent replications. We report these results in the tables collected in Appendix C.

First, we consider the two-step MRC procedure for the cross-sectional model. We start from a benchmark design (Design 1) specified as follows

$$U_{id} = \sum_{j=1}^{2}(\beta_1 X_{ij,1} + \beta_2 X_{ij,2} + \epsilon_{ij}) \cdot d_j + \eta_i \cdot (\gamma_1 W_{i,1} + \gamma_2 W_{i,2}) \cdot d_1 \cdot d_2,$$
$$Y_{id} = 1[U_{id} > U_{id'}, \forall d' \neq d],$$

where $\beta_1 = \beta_2 = 1$, $\gamma_1 = \gamma_2 = 1$, and $d = (d_1, d_2) \in \{(0,0),(1,0),(0,1),(1,1)\}$. In this design, we let $\{\{X_{ij,\iota}\}_{j=1,2;\iota=1,2}, \{W_{i,\iota}\}_{\iota=1,2}, \{\epsilon_{ij}\}_{j=1,2}, \eta_i\}_{i=1,\ldots,N}$ be independent with each other and across i, j, and ι. $X_{ij,1}$'s, $W_{i,1}$, and ϵ_{ij}'s are $N(0,1)$, $X_{ij,2}$'s and $W_{i,2}$ are uniform in $[-1,1]$, and $\eta_i \sim \text{Beta}(2,2)$. Imposing the scale normalization (Assumption C4), we treat β_2 and γ_2 as free parameters to estimate. After the normalization, true values of β_2 and γ_2 become $\sqrt{2}/2$.

Note that $k_1 = k_2 = 2$, and larger bandwidth means smaller variation. In light of this observation and the requirements in Assumption C9, we employ the sixth-order Gaussian kernel ($\kappa_\beta = 6$) and tuning parameter (bandwidth) $h_N = c_1 \cdot N^{-1/8}\log(N)^{1/6}$ for the first step estimation, and thus $\sqrt{N}h_N^4 \propto \log(N)^{2/3} \to \infty$ and $\sqrt{N}h_N^6 \to 0$. We adpot the fourth-order Gaussian kernel ($\kappa_\gamma = 4$) and bandwidth $\sigma_N = c_2 \cdot N^{-1/4}\log(N)^{1/4}$ for the second step, and thus $\sqrt{N}\sigma_N^2 \propto \log(N)^{1/2} \to \infty$ and $\sqrt{N}\sigma_N^4 \to 0$. To test the sensitivity of our methods to the choice of tuning parameters, we experiment several values of $c_1 \in \{1.0, 1.2, 1.4, 1.6, 1.8\}$ and several values of $c_2 \in \{1.8, 2.0, 2.2, 2.4, 2.6\}$. It turns out that the results are not sensitive to the choice of (c_1, c_2). To save space, we only report the results of $c_1 = 1.4$ and $c_2 = 2.4$, with which the root mean squared errors are the smallest among all choices. 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 MRC estimators and the nonparametric bootstrap in tables labeled "A" and "B", respectively. For example, Table 1A summarizes the performance of the estimators for Design 1, in which we report the mean bias (MBIAS) and root mean squared error (RMSE) of the estimator. Since these statistics are sensitive to outliers, we also present the median bias (MED) and the median absolute deviation (MAD). Table 1B reports the empirical coverage frequencies (COVERAGE) and lengths (LENGTH) of the 95% confidence intervals (CI) constructed using the standard bootstrap with 199 independent draws and estimations.

The results for Design 1 are in line with the asymptotic theory. The RMSEs of $\hat{\beta}_2$ and $\hat{\gamma}_2$ shrink at the parametric rate as the sample size increases, with RMSE of $\hat{\gamma}_2$ greater than that of $\hat{\beta}_2$. This reflects the fact that for this specific design, the number of moment inequalities that can be used to identify β_2 is twice that of γ_2. The standard bootstrap also performs well for our MRC estimators, which yields shrinking CIs with coverage rates approaching 95% as the sample size grows.

The design of the Monte Carlo experiments may have a large effect on the results. For example, all i.i.d. regressors and errors often make estimators perform better than they are in cases where either the regressors or errors are correlated. To check the performance of our estimators with correlated regressors and errors, we modify Design 1 by setting

$$(X_{i1,1}, X_{i2,1}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right),\ (\epsilon_{i1}, \epsilon_{i2}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right),$$

and $(X_{i1,2}, X_{i2,2}) = (\zeta_{i1} + \zeta_{i3}, \zeta_{i2} + \zeta_{i3})$, where ζ_{i1}, ζ_{i2}, and ζ_{i3} are uniform in $[-1/2, 1/2]$ and mutually independent. All other aspects remain the same. We refer to this modified version of the benchmark design as Design 2. We adopt the same kernels and bandwidths (same c_1 and c_2) as in Design 1, and report its simulation results in Tables 2A and 2B. The results are very similar to those for Design 1, though as expected the performance slightly deteriorates. This confirms our theoretical result that our approach allows for arbitrary correlation in errors.

We then turn to examine the finite-sample performance of the MS estimation and the numerical bootstrap procedure for panel data bundle choice models. We consider two designs (Designs 3 and 4) with the same choice set as the cross-sectional designs (i.e., $\{(0,0),(1,0),(0,1),(1,1)\}$) and a panel of two time periods. For each panel data design, we consider sample sizes of 1000, 2500, and 5000. For inference, we construct 95% CI based on 199 independent draws and estimations.

The first panel data design (Design 3) is specified as follows

$$
\begin{aligned}
U_{idt} &= \sum_{j=1}^{2} (\beta_1 X_{ijt,1} + \beta_2 X_{ijt,2} + \alpha_{ij} + \epsilon_{ijt}) \cdot d_j + \eta_{it} \cdot (\gamma_1 W_{it,1} + \gamma_2 W_{it,2} + \alpha_{ib}) \cdot d_1 \cdot d_2, \\
Y_{idt} &= 1[U_{idt} > U_{id't}, \forall d' \neq d],
\end{aligned}
$$

where $\beta_1 = \beta_2 = 1$ and $\gamma_1 = \gamma_2 = 1$. $\{\{X_{ijt,\iota}\}_{j=1,2;\iota=1,2}, \{W_{it,\iota}\}_{\iota=1,2}, \{\epsilon_{ijt}\}_{j=1,2}, \eta_{it}\}_{i=1,\ldots,N;t=1,2}$ are mutually independent random variables, where $X_{ijt,1}$'s, $W_{it,1}$'s, and ϵ_{ijt}'s are $N(0,1)$, $X_{ijt,2}$'s and $W_{it,2}$'s are uniform in $[-1,1]$, and $\eta_{it} \sim \text{Beta}(2,2)$. $\alpha_{ij} = (X_{ij1,2} + X_{ij2,2})/4$ for $j = 1,2$ and $\alpha_{ib} = (W_{i1,2} + W_{i2,2})/4$. Due to scale normalization (Assumption P3), we set β_2 and γ_2 as free parameters to estimate. True values of β_2 and γ_2 are therefore normalized to $\sqrt{2}/2$. Clearly in this design, the (unobserved) fixed effects are correlated with regressors, making the matching method in the cross-sectional case invalid here.

The second panel data design (Design 4) have the same random utility model and coefficients as Design 3. But analogous to Design 2 for the cross-sectional setting, we set

$$
(X_{i1t,1}, X_{i2t,1}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right), \text{ for } t = 1,2,
$$

and

$$
(\epsilon_{i11}, \epsilon_{i21}, \epsilon_{i12}, \epsilon_{i22}) \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix} \right),
$$

and $(X_{i1t,2}, X_{i2t,2}) = (\zeta_{i1t} + \zeta_{i3t}, \zeta_{i2t} + \zeta_{i3t})$ for $t = 1,2$, where ζ_{i1t}, ζ_{i2t}, and ζ_{i3t} are uniform in $[-1/2, 1/2]$ and mutually independent. All other aspects remain the same.

To implement the MS estimation for these designs, we use the Epanechnikov kernel and $h_N = \sigma_N = c_3 \cdot N^{-1/7} \log(N)^{-1/14}$. For the implementation of the numerical bootstrap, we have two more tuning parameters, ε_{N1} and ε_{N2}. We set $\varepsilon_{N1} = \varepsilon_{N2} = c_4 \cdot N^{-5/7} \log(N)^{-5/14}$.[12] This choice of $(h_N, \sigma_N, \varepsilon_{N1}, \varepsilon_{N2})$ meets Assumptions P9 and P10. We experiment $c_3 \in \{2.2, 2.4, 2.6, 2.8, 3.0\}$ and $c_4 \in \{0.8, 1.0, 1.2, 1.4, 1.6\}$. It turns out that the results are not sensitive to the choice of (c_3, c_4). To save space, we only report the results with $c_3 = 2.8$ and $c_4 = 1$.

Simulation results for Designs 3 and 4 are reported, respectively, in Tables 3A–3B and Tables 4A–4B. As the results indicate, our MS estimators perform reasonably well and give similar results for these two designs. The MBIAS and RMSEs of $\hat{\beta}_2$ and $\hat{\gamma}_2$ decline as the sample size increases. $\hat{\gamma}_2$ has overall larger bias and RMSE than $\hat{\beta}_2$, reflecting the fact that its estimation uses less identification inequalities. These results show the consistency of our estimators, though the rates of convergence are clearly slower than $\sqrt{N}$.[13] The inference results are also in line with with the theory. When the sample size is relatively small, the CI of β_2 obtained by the numerical bootstrap tends to be too wide, while that of γ_2 tends to be too narrow. We conjecture that this is mainly because we implement the estimation and bootstrap inference procedures for β_2 and γ_2 with the same set of tuning parameters, while they have different numbers of identification inequalities. As the sample size grows, we observe that the CIs of both β_2 and γ_2 shrink with their coverage rates approaching 95% (from different directions).

5 Conclusions

In this paper, we propose new estimation and inference procedures for semiparametric discrete choice models for bundles. For the cross-sectional model, we propose a two-step kernel weighted rank procedure and establish its $\sqrt{N}$-consistency and asymptotic normality. Based on these results, we further show the validity of using nonparametric bootstrap to carry out inference. Our matching-based identification strategy for the cross-sectional setting is easily extendable to the panel data model, enabling a consistent two-step MS estimation procedure of a model with alternative, bundle, and individual-specific effects. We further derive limiting distributions of the proposed estimators and justify the application of the numerical bootstrap (Hong and Li (2020)) for making inference. A small-scale Monte Carlo study demonstrates that the proposed estimation and inference procedures perform well in finite samples.

Throughout this paper, we have focused on the settings when the researcher can observe individual-level data. We notice that the methods proposed in this paper can also be adjusted to be applied to models using aggregate data, i.e., the researcher can only observe the aggregated choice probabilities (e.g., market shares) in a number of markets and the market-level covariates for each alternative. Such models are often encountered in empirical industrial organization (see,

[12]Though Hong and Li (2020) recommend $\varepsilon_{N1}, \varepsilon_{N2} \propto N^{-6/7} \log(N)^{-2/7}$ (see (3.12) in our Section 3.2), some tentative Monte Carlo simulations suggest that a slight modification like this leads to a better finite-sample performance.

[13]It is a bit surprising that our MS procedure performs slightly better in Design 4 than in Design 3 in terms of RMSE. One explanation is that when the serial dependence in the errors is high, intertemporal variation in the values of observed covariates has larger impact on the changes of the values of Y_{idt} over time.

e.g., Fan (2013)). In-depth discussion of this extension will involve different estimators and asymptotics.[14] Due to space constraints, we leave this topic to a separate study.

The work here leaves many 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 for the panel data estimators. Besides, as pointed out in Sections 2 and 3, our matching-based approaches provide set identification results for, e.g., preference coefficients on individual-specific regressors, but consistent estimators for these sets are lacking in the literature.

References

Abrevaya, J., J. A. Hausman, and S. Khan (2010): "Testing for causal effects in a generalized regression model with endogenous regressors," *Econometrica*, 78, 2043–2061.

Abrevaya, J. and J. Huang (2005): "On the bootstrap of the maximum score estimator," *Econometrica*, 73, 1175–1204.

Ahn, H., H. Ichimura, J. L. Powell, and P. A. Ruud (2018): "Simple estimators for invertible index models," *Journal of Business & Economic Statistics*, 36, 1–10.

Allen, R. and J. Rehbeck (2019): "Identification with additively separable heterogeneity," *Econometrica*, 87, 1021–1054.

——— (2020): "Latent complementarity in bundles models," *Available at SSRN 3257028.*

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

Augereau, A., S. Greenstein, and M. Rysman (2006): "Coordination versus differentiation in a standards war: 56K modems," *The RAND Journal of Economics*, 37, 887–909.

Berry, S., A. Khwaja, V. Kumar, A. Musalem, K. C. Wilbur, G. Allenby, B. Anand, P. Chintagunta, W. M. Hanemann, and P. Jeziorski (2014): "Structural models of complementary choices," *Marketing Letters*, 25, 245–256.

Berry, S. T. and P. A. Haile (2009): "Nonparametric identification of multinomial choice demand models with heterogeneous consumers," Tech. rep., National Bureau of Economic Research.

[14] Shi et al. (2018) discussed semiparametric estimation of panel data multinomial choice models using aggregate data. Their estimator, defined as an optimizer of a linear programming problem, was shown to be $\sqrt{N}$-consistent. Hsieh, Shi, and Shum (2018) proposed an inference procedure for the estimators of this type. Our approaches, if extended to models using aggregate data, will give estimators of similar structure. We conjecture that their arguments, with certain modification, can be applied to our case.

Cameron, A. C. and P. K. Trivedi (2005): *Microeconometrics: methods and applications*, Cambridge University Press.

Cavanagh, C. and R. P. Sherman (1998): "Rank estimators for monotonic index models," *Journal of Econometrics*, 84, 351–382.

Chen, X., O. Linton, and I. van Keilegom (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71, 1591–1608.

Chernozhukov, V., I. Fernández-Val, and W. K. Newey (2019): "Nonseparable multinomial choice models in cross-section and panel data," *Journal of econometrics*, 211, 104–116.

Chung, J. and V. R. Rao (2003): "A general choice model for bundles with multiple-category products: Application to market segmentation and optimal pricing for bundles," *Journal of Marketing Research*, 40, 115–130.

Dubé, J.-P. (2004): "Multiple discreteness and product differentiation: Demand for carbonated soft drinks," *Marketing Science*, 23, 66–81.

Dunker, F., S. Hoderlein, and H. Kaido (2017): "Nonparametric identification of random coefficients in endogenous and heterogeneous aggregated demand models," Tech. rep., cemmap working paper.

Fan, Y. (2013): "Ownership consolidation and product characteristics: A study of the US daily newspaper market," *American Economic Review*, 103, 1598–1628.

Foubert, B. and E. Gijsbrechts (2007): "Shopper response to bundle promotions for packaged goods," *Journal of Marketing Research*, 44, 647–662.

Fox, J. T. (2007): "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *The RAND Journal of Economics*, 38, 1002–1019.

Fox, J. T. and N. Lazzati (2017): "A note on identification of discrete choice models for bundles and binary games," *Quantitative Economics*, 8, 1021–1036.

Gao, W. Y. and M. Li (2019): "Robust Semiparametric Estimation in Panel Multinomial Choice Models," *Available at SSRN 3282293*.

Gentzkow, M. (2007): "Valuing new goods in a model with complementarity: Online newspapers," *American Economic Review*, 97, 713–744.

Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2014): "Competition and ideological diversity: Historical evidence from us newspapers," *American Economic Review*, 104, 3073–3114.

Giné, E. and J. Zinn (1990): "Bootstrapping general empirical measures," *The Annals of Probability*, 851–869.

Han, A. K. (1987): "Nonparametric analysis of a generalized regression model; the maximum rank correlation estimator," *Journal of Econometrics*, 35, 303–316.

Hendel, I. (1999): "Estimating multiple-discrete choice models: An application to computerization returns," *The Review of Economic Studies*, 66, 423–446.

Hong, H. and J. Li (2020): "The numerical bootstrap," *The Annals of Statistics*, 48, 397–412.

Hong, H., A. Mahajan, and D. Nekipelov (2015): "Extremum estimation and numerical derivatives," *Journal of Econometrics*, 188, 250–263.

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

——— (2001): "The bootstrap," in *Handbook of econometrics*, Elsevier, vol. 5, 3159–3228.

Hsieh, Y.-W., X. Shi, and M. Shum (2018): "Inference on estimators defined by mathematical programming," *Available at SSRN 3041040*.

Jin, Z., Z. Ying, and L. J. Wei (2001): "A simple resampling method by perturbing the minimand," *Biometrika*, 88, 381–390.

Khan, S., F. Ouyang, and E. Tamer (2019): "Inference on semiparametric multinomial 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.

Kim, Y., S. Misra, and B. Shapiro (2019): "Valuing brand collaboration: Evidence from a natural experiment," *Available at SSRN*.

Lee, L.-f. (1995): "Semiparametric maximum likelihood estimation of polychotomous and sequential choice models," *Journal of Econometrics*, 65, 381–428.

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.

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177.

Lewbel, A. and L. Nesheim (2019): "Sparse demand systems: corners and complements," Tech. rep., Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Liu, H., P. K. Chintagunta, and T. Zhu (2010): "Complementarities and the demand for home broadband internet services," *Marketing Science*, 29, 701–720.

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.

Manski, C. F. and L. Sherman (1980): "An empirical analysis of household choice among motor vehicles," *Transportation Research Part A: General*, 14, 349–366.

Nevo, A., D. L. Rubinfeld, and M. McCabe (2005): "Academic journal pricing and the demand of libraries," *American Economic Review*, 95, 447–452.

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

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

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

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

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.

Serfling, R. J. (2009): *Approximation theorems of mathematical statistics*, vol. 162, John Wiley & Sons.

Sher, I. and K. i. Kim (2014): "Identifying combinatorial valuations from aggregate demand," *Journal of Economic Theory*, 153, 428–458.

Sherman, R. P. (1993): "The limiting distribution of the maximum rank correlation estimator," *Econometrica*, 61, 123–137.

——— (1994a): "Maximal inequalities for degenerate U-processes with applications to optimization estimators," *The Annals of Statistics*, 439–459.

——— (1994b): "U-processes in the analysis of a generalized semiparametric regression estimator," *Econometric theory*, 10, 372–395.

Shi, X., M. Shum, and W. Song (2018): "Estimating semiparametric panel multinomial choice models using cyclic monotonicity," *Econometrica*, 86, 737–761.

Song, I. and P. K. Chintagunta (2006): "Measuring cross-category price effects with aggregate store data," *Management Science*, 52, 1594–1609.

Subbotin, V. (2007): "Asymptotic and bootstrap properties of rank regressions," *Available at SSRN 1028548.*

Train, K. E., D. L. McFadden, and M. Ben-Akiva (1987): "The demand for local telephone service: A fully discrete model of residential calling patterns and service choices," *The RAND Journal of Economics*, 109–123.

van der Vaart, A. and J. Wellner (2000): *Weak convergence and empirical processes*, Springer, New York.

Wellner, J. A. and Y. Zhan (1996): "Bootstrapping Z-estimators," *University of Washington Department of Statistics Technical Report*, 308.

Yan, J. and H. I. Yoo (2019): "Semiparametric estimation of the random utility model with rank-ordered choice data," *Journal of econometrics*, 211, 414–438.

Appendix

This appendix is organized as follows. Appendix A provides technical lemmas and the proofs of Theorems 2.1 and 2.2 for our localized MRC estimators of the cross-sectional model. Appendix B proves Theorem 3.2 associated with the MS procedure for the panel data model and provides supporting lemmas for the proof. Appendix D provides technical details for conducting bootstrap inference. Finally, Appendix E collects the proofs of all the lemmas used in Appendixes A and B.

Throughout, we use acronyms, SLLN and CMT, for Strong Law of Large Numbers and Continuous Mapping Theorem, 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 $a_N/b_N \to 0$ $(a_N/b_N \overset{P}{\to} 0)$. For any summation indexed by agents, we suppress the statement that the agent is in the set $\{1, ..., N\}$. For example, $\sum_i$ means $\sum_{i=1}^N$, and $\sum_{i \neq m}$ represents $\sum_{i=1}^N \sum_{m \neq i}$.

A Proofs for the Cross-Sectional Model

We provide the proofs of Theorems 2.1 and 2.2 in order. The first proof is relatively straightforward.

Proof of Theorem 2.1. It suffices to show the identification of β based on (2.4) as the same arguments can be applied to the identification of β and γ based on similar moment inequalities. To ease the notation, we denote $\Omega_{im} = \{X_{i2} = X_{m2}, W_i = W_m\}$. By Assumptions C1 and C2, the monotonic relation (2.4) implies that for all $d \in \mathcal{D}$ with $d_1 = 1$, β maximizes

$$Q_1(b) \equiv \mathbb{E}[(P(Y_{id} = 1|Z_i) - P(Y_{md} = 1|Z_m))\text{sgn}(X_{im1}'b)|\Omega_{im}]$$

for each pair of (i, m). To show that β attains a unique maximum, suppose that there is a $b \in \mathcal{B}$ such that $Q_1(b) = Q_1(\beta)$. In what follows, we assume $\beta^{(1)} > 0$ w.l.o.g. (the case $\beta^{(1)} < 0$ is symmetric). We want to show that $b = \beta$ must hold. First, note that $b^{(1)} > 0$ must hold. Otherwise, we have $P[\text{sgn}(X_{im1}'b) \neq \text{sgn}(X_{im1}'\beta)|\Omega_{im}] = P[(\tilde{X}_{im1}'\tilde{\beta}/\beta^{(1)} < -X_{im1}^{(1)}, \tilde{X}_{im1}'\tilde{b}/b^{(1)} < -X_{im1}^{(1)}) \cup$

$(\tilde{X}_{im1}'\tilde{b}/b^{(1)} > -X_{im1}^{(1)}, \tilde{X}_{im1}'\tilde{\beta}/\beta^{(1)} > -X_{im1}^{(1)})|\Omega_{im}] > 0$ by Assumption C3(i). Then, β and b yield different values of the sgn$(\cdot)$ function in $Q_1(\cdot)$ with strictly positive probability, and thus $Q_1(b) < Q_1(\beta)$. For the case $b^{(1)} > 0$, we write $P[\text{sgn}(X_{im1}'b) \neq \text{sgn}(X_{im1}'\beta)|\Omega_{im}] = P[(\tilde{X}_{im1}'\tilde{\beta}/\beta^{(1)} < -X_{im1}^{(1)} < \tilde{X}_{im1}'\tilde{b}/b^{(1)}) \cup (\tilde{X}_{im1}'\tilde{b}/b^{(1)} < -X_{im1}^{(1)} < \tilde{X}_{im1}'\tilde{\beta}/\beta^{(1)})|\Omega_{im}]$. This implies that for all b satisfying $Q_1(b) = Q_1(\beta)$, $P[(\tilde{X}_{im1}'\tilde{\beta}/\beta^{(1)} < -X_{im1}^{(1)} < \tilde{X}_{im1}'\tilde{b}/b^{(1)}) \cup (\tilde{X}_{im1}'\tilde{b}/b^{(1)} < -X_{im1}^{(1)} < \tilde{X}_{im1}'\tilde{\beta}/\beta^{(1)})|\Omega_{im}] = 0$ must hold, which is equivalent to $P(\tilde{X}_{im1}'\tilde{\beta}/\beta^{(1)} = \tilde{X}_{im1}'\tilde{b}/b^{(1)}|\Omega_{im}) = 1$ under Assumption C3(i). Then it follows by Assumption C3(ii) that $\tilde{\beta}/\beta^{(1)} = \tilde{b}/b^{(1)}$ and hence $b^{(1)}\beta = \beta^{(1)}b$. As $\|b\| = \|\beta\| = 1$, we have $b^{(1)} = \beta^{(1)}$ and thus $b = \beta$. This completes the proof. $\square$

The proof of Theorem 2.2 uses results in a series of technical lemmas (i.e., Lemmas A.1–A.7), which we will present right above the main proof. Before delving into the technical details, we first outline the "roadmap" for the proof process.

In what follows, we will only show the asymptotics of $\hat{\gamma}$. The proof for $\hat{\beta}$ is omitted since one can derive the asymptotics of $\hat{\beta}$ by repeating the proof process for $\hat{\gamma}$ but skipping the step handling the plug-in first step estimates (i.e., Lemma A.7). Throughout this section, we take it as a given that $\hat{\beta}$ is $\sqrt{N}$-consistent and asymptotically normal.

For ease of illustration, with a bit abuse of notation, we will work with objective function

$$\hat{\mathcal{L}}_N^K(r) \equiv \frac{1}{\sigma_N^2 N(N-1)} \sum_{i \neq m} K_{\sigma_N,\gamma}(V_{im}(\hat{\beta}))h_{im}(r),$$

where $K_{\sigma_N,\gamma}(\cdot) \equiv \sigma_N^2 \mathcal{K}_{\sigma_N}(\cdot)$ and $h_{im}(r) \equiv Y_{im(1,1)}[\text{sgn}(W_{im}'r) - \text{sgn}(W_{im}'\gamma)]$.[15] Note that here we subtract the term $\text{sgn}(W_{im}'\gamma)$ from objective function (2.12), analogous to Sherman (1993). Doing this does not affect the value of the estimator, and will facilitate the proofs that follow. Besides, we define

$$\mathcal{L}_N^K(r) = \frac{1}{\sigma_N^2 N(N-1)} \sum_{i \neq m} K_{\sigma_N,\gamma}(V_{im}(\beta))h_{im}(r)$$

and

$$\mathcal{L}(r) \equiv f_{V_{im}(\beta)}(0)\mathbb{E}[h_{im}(r)|V_{im}(\beta) = 0].$$

We establish the consistency of $\hat{\gamma}$ in Lemmas A.1–A.3 by applying Theorem 2.1 in Newey and McFadden (1994). The key step is to show the uniform convergence of $\hat{\mathcal{L}}_N^K(r)$ for $r \in \mathcal{R}$. To this end, we bound the differences among $\hat{\mathcal{L}}_N^K(r)$, $\mathcal{L}_N^K(r)$, and $\mathcal{L}(r)$ in Lemmas A.1 and A.2.

The next step is to show the asymptotic normality of $\hat{\gamma}$ by applying Theorem 2 of Sherman (1994a). Sufficient conditions for this theorem are that $\hat{\gamma} - \gamma = O_P(N^{-1/2})$ and uniformly over a neighborhood of γ with a radius proportional to $N^{-1/2}$,

$$\hat{\mathcal{L}}_N^K(r) = \frac{1}{2}(r-\gamma)'\mathbb{V}_\gamma(r-\gamma) + \frac{1}{\sqrt{N}}(r-\gamma)'\Psi_N + o_P(N^{-1}), \tag{A.1}$$

[15] By definition, $K_{\sigma_N,\gamma}(\cdot) = K_\gamma(\cdot/\sigma_N)$.

where $\mathbb{V}_\gamma$ is a negative definite matrix and Ψ_N is asymptotically normal, with mean zero and variance $\mathbb{V}_\Psi$. To verify equation (A.1), we first show that uniformly over a neighborhood of γ, $\mathcal{R}_N \equiv \{r \in \mathcal{R} | \|r-\gamma\| \leq \delta_N\}$ with $\{\delta_N\} = O(N^{-\delta})$ for some $0 < \delta \leq 1/2$,

$$\hat{\mathcal{L}}_N^K(r) = \frac{1}{2}(r-\gamma)'\mathbb{V}_\gamma(r-\gamma) + \frac{1}{\sqrt{N}}(r-\gamma)'\Psi_N + o_P(\|r-\gamma\|^2) + O_P(\varepsilon_N), \tag{A.2}$$

which is the task of Lemmas A.4–A.7. The $O_P(\varepsilon_N)$ term in equation (A.2) can be shown to be of order $o_P(N^{-1})$ uniformly over $\mathcal{R}_N$. Further, by Theorem 1 of Sherman (1994b), equation (A.2) implies $\hat{\gamma} - \gamma = O_P(\sqrt{\varepsilon_N} \vee N^{-1/2}) = O_P(N^{-1/2})$. This rate result, together with equation (A.2), further verifies equation (A.1).

To obtain equation (A.2), we will work with the following expansion

$$\begin{aligned}\hat{\mathcal{L}}_N^K(r) &= \mathcal{L}_N^K(r) + \Delta\mathcal{L}_N^K(r) + R_N \\ &= \mathbb{E}[\mathcal{L}_N^K(r)] + \frac{2}{N}\sum_i \left\{\mathbb{E}[\mathcal{L}_N^K(r)|Z_i] - \mathbb{E}[\mathcal{L}_N^K(r)]\right\} + \rho_N(r) + \Delta\mathcal{L}_N^K(r) + R_N,\end{aligned} \tag{A.3}$$

where

$$\Delta\mathcal{L}_N^K(r) \equiv \frac{1}{\sigma_N^3 N(N-1)} \sum_{i\neq m} \nabla K_{\sigma_N,\gamma}(V_{im}(\beta))'(V_{im}(\hat{\beta}) - V_{im}(\beta))h_{im}(r), \tag{A.4}$$

$$\rho_N(r) = \mathcal{L}_N^K(r) - \frac{2}{N}\sum_i \mathbb{E}[\mathcal{L}_N^K(r)|Z_i] + \mathbb{E}[\mathcal{L}_N^K(r)],$$

and R_N denotes the remainder term in the expansion of higher order (as $\sqrt{N}\sigma_N \to \infty$). The first three terms in (A.3) are the U-statistic decomposition for $\mathcal{L}_N^K(r)$ (see e.g., Sherman (1993) and Serfling (2009)). Lemmas A.4–A.6 establish asymptotic properties of these three terms, respectively. A linear representation for the fourth term in (A.3) is derived in Lemma A.7.

The following notations will be adopted for notational convenience:

- $V_i \equiv V_i(\beta)$, $\hat{V}_i \equiv V_i(\hat{\beta})$, $V_{im} \equiv V_i - V_m$, and $\hat{V}_{im} \equiv \hat{V}_i - \hat{V}_m$.
- For realized values, $v_i \equiv v_i(\beta)$, $\hat{v}_i \equiv v_i(\hat{\beta})$, $v_{im} \equiv v_i - v_m$, and $\hat{v}_{im} \equiv \hat{v}_i - \hat{v}_m$.
- $B(v_i, v_m, w_i, w_m) \equiv \mathbb{E}[Y_{im(1,1)}|V_i = v_i, V_m = v_m, W_i = w_i, W_m = w_m]$.
- $S_{im}(r) \equiv \text{sgn}(w_{im}'r) - \text{sgn}(w_{im}'\gamma)$.

Further, using the above notations and the definition of $\tau_i(r)$, we write

$$\tau_i(r) = f_{V_{im}}(0)\mathbb{E}[h_{im}(r)|V_i = v_i, W_i = w_i, V_{im} = 0] = \int B(v_i, v_i, w_i, w_m)S_{im}(r)f_{V,W}(v_i, w_m)dw_m.$$

Here we present Lemmas A.1–A.7, whose proofs are relegated to Appendix E.

Lemma A.1. *Under Assumptions C1–C6, C8, and C9,* $\sup_{r\in\mathcal{R}} |\hat{\mathcal{L}}_N^K(r) - \mathcal{L}_N^K(r)| = o_P(1)$.

Lemma A.2. *Under Assumptions C1, C5, C8, and C9,* $\sup_{r\in\mathcal{R}} |\mathcal{L}_N^K(r) - \mathcal{L}(r)| = o_P(1)$.

Lemma A.3. *Under Assumptions C1–C6, C8, and C9,* $\hat{\gamma} \xrightarrow{P} \gamma$.

Lemma A.4. *Under Assumptions C1–C5 and C7–C9, uniformly over* $\mathcal{R}_N$*, we have*

$$\mathbb{E}[\mathcal{L}_N^K(r)] = \frac{1}{2}(r-\gamma)'\mathbb{V}_\gamma(r-\gamma) + o(\|r-\gamma\|^2),$$

where $\mathbb{V}_\gamma \equiv \mathbb{E}[\nabla^2\tau_i(\gamma)]$.

Lemma A.5. *Under Assumptions C1–C5 and C7–C9, uniformly over* $\mathcal{R}_N$*, we have*

$$\frac{2}{N}\sum_m \mathbb{E}[\mathcal{L}_N^K(r)|Z_m] - 2\mathbb{E}[\mathcal{L}_N^K(r)] = \frac{1}{\sqrt{N}}(r-\gamma)'\mathit{\Psi}_{N,1} + o_P(\|r-\gamma\|^2),$$

where $\mathit{\Psi}_{N,1} \equiv N^{-1/2}\sum_m 2\nabla\tau_m(\gamma)$.

Lemma A.6. *Under Assumptions C1, C8, and C9, uniformly over* $\mathcal{R}_N$*,* $\rho_N(r) = O_P(N^{-1}\sigma_N^{-2})$.

Lemma A.7. *Suppose*

$$\hat{\beta} - \beta = \frac{1}{N}\sum_i \psi_{i,\beta} + o_P(N^{-1/2}), \tag{A.5}$$

where $\psi_{i,\beta}$ *is the influence function (of* Z_i*).* $\psi_{i,\beta}$ *is i.i.d across* i *with* $\mathbb{E}(\psi_{i,\beta}) = 0$.[16] *Further, suppose Assumptions C1–C9 hold. Then uniformly over* $\mathcal{R}_N$*, we have*

$$\Delta\mathcal{L}_N^K(r) = \frac{1}{\sqrt{N}}(r-\gamma)'\mathit{\Psi}_{N,2} + o_P(\|r-\gamma\|^2) + O_P(N^{-1}\sigma_N^{-2}),$$

with $\mathit{\Psi}_{N,2} = \frac{1}{\sqrt{N}}\sum_i \left(-\int \nabla_{13}^2\mu(v_m, v_m, \gamma) f_V(v_m) dv_m\right)\psi_{i,\beta}$*, where* $\mu(v_i, v_m, r) \equiv G(v_i, v_m, r) f_V(v_i)$*,*

$$G(v_i, v_m, r) \equiv \mathbb{E}\left[B(V_i, V_m, W_i, W_m)S_{im}(r)\begin{pmatrix} x_{im1}' \\ x_{im2}' \end{pmatrix}|V_i = v_i, V_m = v_m\right],$$

$\nabla_1\mu(\cdot,\cdot,\cdot)$ *denotes the partial derivative of* $\mu(\cdot,\cdot,\cdot)$ *w.r.t. its first argument, and* $\nabla_{13}^2\mu(\cdot,\cdot,\cdot)$ *denotes the partial derivative of* $\nabla_1\mu(\cdot,\cdot,\cdot)$ *w.r.t. its third argument.*

Proof of Theorem 2.2. Putting results in Lemmas A.4–A.7 together, we write equation (A.3) as

$$\hat{\mathcal{L}}_N^K(r) = \frac{1}{2}(r-\gamma)'\mathbb{V}_\gamma(r-\gamma) + \frac{1}{\sqrt{N}}(r-\gamma)'\mathit{\Psi}_N + o_P(\|r-\gamma\|^2) + O_P(\varepsilon_N) \tag{A.6}$$

where $\mathit{\Psi}_N = \mathit{\Psi}_{N,1} + \mathit{\Psi}_{N,2}$ and $\varepsilon_N = N^{-1}\sigma_N^{-2}$. Theorem 1 of Sherman (1994b) then implies that $\hat{\gamma} - \gamma = O_P(\sqrt{\varepsilon_N}) = O_P(N^{-1/2}\sigma_N^{-1})$.

Next, take $\delta_N = O(\sqrt{\varepsilon_N})$ and $\mathcal{R}_N = \{r \in \mathcal{R} | \|r-\gamma\| \leq \delta_N\}$. We repeat the proof for Lemma A.6 and deduce from a Taylor expansion around γ that $\sup_{r\in\mathcal{R}_N}\mathbb{E}[\rho_{im}^*(r)^2] = O(\sigma_N^2\delta_N^2)$. Apply

[16] In fact, the linear representation of $\hat{\beta} - \beta$ in (A.5) can be obtained by applying the same arguments in Theorem 2 of Sherman (1993) to a representation for $\mathcal{L}_{N,\beta}^K(b)$ analogous to (A.1).

Theorem 3 of Sherman (1994b) to see that uniformly over $\mathcal{R}_N$, $\rho_N(r) = O_P(N^{-1}\sigma_N^{\lambda-2}\delta_N^{\lambda})$ where $0 < \lambda < 1$. Then we have

$$\rho_N(r) = O_P(N^{-1}\sigma_N^{\lambda-2}\delta_N^{\lambda}) = O_P(N^{-1})O_P(N^{-\lambda/2}\sigma_N^{-2}) = o_P(N^{-1})$$

by invoking Assumption C9 and choosing λ sufficiently close to 1. This result in turn implies that the $O_P(\varepsilon_N)$ term in (A.6) has order $o_P(N^{-1})$, and hence $\hat{\gamma} - \gamma = O_P(N^{-1/2})$ by applying Theorem 1 of Sherman (1994b) once again.

Now (A.6) can be expressed as

$$\hat{\mathcal{L}}_N^K(r) = \frac{1}{2}(r-\gamma)'\mathbb{V}_\gamma(r-\gamma) + \frac{1}{\sqrt{N}}(r-\gamma)'\Psi_N + o_P(N^{-1}).$$

Let $\varDelta_i \equiv 2\nabla\tau_i(\gamma) - \left(\int \nabla_{13}^2\mu(v_m, v_m, \gamma)f_V(v_m)\mathrm{d}v_m\right)\psi_{i,\beta}$. Note that $\mathbb{E}[\varDelta_i] = 0$ since $\mathbb{E}[\nabla\tau_i(\gamma)] = 0$ and $\mathbb{E}[\psi_{i,\beta}] = 0$. We deduce from Assumption C7 and Lindeberg-Lévy CLT that $\Psi_N \xrightarrow{d} N(0, \mathbb{V}_\Psi)$ where $\mathbb{V}_\Psi = \mathbb{E}[\varDelta_i\varDelta_i']$. The asymptotic normality of $\hat{\gamma}$ then follows from Theorem 2 of Sherman (1994a), i.e., $\sqrt{N}(\hat{\gamma}-\gamma) \xrightarrow{d} N(0, \mathbb{V}_\gamma^{-1}\mathbb{V}_\Psi\mathbb{V}_\gamma^{-1})$. □

B Proofs for the Panel Data Model

In this section, we provide the proof of Theorem 3.2, applying the asymptotic theory developed in Seo and Otsu (2018). Before the main proof, we first present two supporting lemmas, Lemmas B.1 and B.2, whose proofs are relegated to Appendix E. The outline of the proof process is as follows. Lemma B.1 verifies the technical conditions in Seo and Otsu (2018). Lemma B.2 obtains technical terms for the final asymptotics. Then we apply the results in Seo and Otsu (2018) and get the asymptotics of our estimators in the proof of Theorem 3.2.

Lemma B.1. *Suppose Assumptions P1–P9 hold. Then $\phi_{Ni}(b)$ and $\varphi_{Ni}(r)$ satisfy Assumption M in Seo and Otsu (2018).*

Lemma B.2. *Suppose Assumptions P1–P9 hold. Then*

$$\lim_{N\to\infty}(Nh_N^{k_1+k_2})^{2/3}\mathbb{E}[\phi_{Ni}(\beta + \rho\left(Nh_N^{k_1+k_2})^{-1/3}\right)] = \frac{1}{2}\rho'\mathbb{V}\rho,$$

$$\lim_{N\to\infty}(N\sigma_N^{2k_1})^{2/3}\mathbb{E}[\varphi_{Ni}(\gamma + \delta(N\sigma_N^{2k_1})^{-1/3})] = \frac{1}{2}\delta'\mathbb{W}\delta,$$

$$\lim_{N\to\infty}(Nh_N^{k_1+k_2})^{1/3}\mathbb{E}[h_N^{k_1+k_2}\phi_{Ni}(\beta + \rho_1(Nh_N^{k_1+k_2})^{-1/3})\phi_{Ni}(\beta + \rho_2(Nh_N^{k_1+k_2})^{-1/3})] = \mathbb{H}_1(\rho_1, \rho_2),$$

and

$$\lim_{N\to\infty}(N\sigma_N^{2k_1})^{1/3}\mathbb{E}[\sigma_N^{2k_1}\varphi_{Ni}(\gamma + \delta_1(N\sigma_N^{2k_1})^{-1/3})\varphi_{Ni}(\gamma + \delta_2(N\sigma_N^{2k_1})^{-1/3})] = \mathbb{H}_2(\delta_1, \delta_2),$$

where ρ, ρ_1, and ρ_2 are $k_1 \times 1$ vectors, δ, δ_1, and δ_2 are $k_2 \times 1$ vectors, $\mathbb{V}$ is a $k_1 \times$ k_1 matrix defined as

$$\mathbb{V} = -\sum_{t>s} \sum_{d \in \mathcal{D}} \left\{ \int 1\left[x'\beta = 0\right] \left(\frac{\partial \kappa_{dts}^{(1)}(x)}{\partial x}' \beta \right) f_{X_{1ts}|\{X_{2ts}=0, W_{ts}=0\}}(x)\, xx' \mathrm{d}\sigma_0^{(1)} \right. \tag{B.1}$$
$$\left. + \int 1\left[x'\beta = 0\right] \left(\frac{\partial \kappa_{dts}^{(2)}(x)}{\partial x}' \beta \right) f_{X_{2ts}|\{X_{1ts}=0, W_{ts}=0\}}(x)\, xx' \mathrm{d}\sigma_0^{(2)} \right\},$$

$\mathbb{W}$ is a $k_2 \times$ k_2 matrix defined as

$$\mathbb{W} = -\sum_{t>s} \int 1\left[w'\gamma = 0\right] \left(\frac{\partial \kappa_{(1,1)ts}^{(3)}(w)}{\partial w}' \gamma \right) f_{W_{ts}|\{X_{1ts}=0, X_{2ts}=0\}}(w)\, ww' \mathrm{d}\sigma_0^{(3)}, \tag{B.2}$$

and $\mathbb{H}_1$ and $\mathbb{H}_2$ are written respectively as

$$\begin{aligned} &\mathbb{H}_1(\rho_1, \rho_2) \\ =&\frac{1}{2} \sum_{t>s} \sum_{d \in \mathcal{D}} \left\{ \int \kappa_{dts}^{(4)}(\bar{x}) \left[\left|\bar{x}'\rho_1\right| + \left|\bar{x}'\rho_2\right| - \left|\bar{x}'(\rho_1 - \rho_2)\right|\right] f_{X_{1ts}|\{X_{2ts}=0, W_{ts}=0\}}(0, \bar{x})\, \mathrm{d}\bar{x} f_{X_{2ts}, W_{ts}}(0,0) \right. \\ &\left. + \int \kappa_{dts}^{(5)}(\bar{x}) \left[\left|\bar{x}'\rho_1\right| + \left|\bar{x}'\rho_2\right| - \left|\bar{x}'(\rho_1 - \rho_2)\right|\right] f_{X_{2ts}|\{X_{1ts}=0, W_{ts}=0\}}(0, \bar{x})\, \mathrm{d}\bar{x} f_{X_{1ts}, W_{ts}}(0,0) \right\} \bar{K}_2^{k_1+k_2}, \end{aligned} \tag{B.3}$$

and

$$\begin{aligned} \mathbb{H}_2(\delta_1, \delta_2) =& \frac{1}{2} \sum_{t>s} \int \kappa_{(1,1)ts}^{(6)}(\bar{w}) \left[\left|\bar{w}'\delta_1\right| + \left|\bar{w}'\delta_2\right| - \left|\bar{w}'(\delta_1 - \delta_2)\right|\right] \\ &\cdot f_{W_{ts}|\{X_{1ts}=0, X_{2ts}=0\}}(0, \bar{w})\, \mathrm{d}\bar{w} f_{X_{1ts}, X_{2ts}}(0,0) \bar{K}_2^{2k_1}. \end{aligned} \tag{B.4}$$

Technical terms that appear in $\mathbb{V}$, $\mathbb{W}$, $\mathbb{H}_1$, $\mathbb{H}_2$ are defined as follows. $\sigma_0^{(1)}$, $\sigma_0^{(2)}$, and $\sigma_0^{(3)}$ are the surface measures of $\{X_{1ts}\beta = 0\}$, $\{X_{2ts}\beta = 0\}$, and $\{W_{ts}\gamma = 0\}$, respectively. $\bar{K}_2 \equiv \int K(u)^2 du$. We decompose X_{1ts}, X_{2ts} and W_{ts} into $X_{jts} = a\beta + \bar{X}_{jts}, j = 1, 2$, and $W_{ts} = a\gamma + \bar{W}_{ts}$, where $\bar{X}_{jts}, j = 1, 2$, are orthogonal to β, and $\bar{W}_{ts}$ is orthogonal to γ. The density for X_{1ts} is written as $f_{X_{1ts}}\left(a, \bar{X}_{1ts}\right)$ under this decomposition. Densities for X_{2ts} and W_{ts} are written similarly. Further,

$$\kappa_{dts}^{(1)}(x) \equiv \mathbb{E}[Y_{idst}(-1)^{d_1} | X_{i1ts} = x, X_{i2ts} = 0, W_{its} = 0],$$

$$\kappa_{dts}^{(2)}(x) \equiv \mathbb{E}[Y_{idst}(-1)^{d_2} | X_{i2ts} = x, X_{i1ts} = 0, W_{its} = 0],$$

$$\kappa_{(1,1)ts}^{(3)}(w) \equiv \mathbb{E}\left[Y_{i(1,1)ts} | W_{its} = w, X_{i1ts} = 0, X_{i2ts} = 0\right],$$

$$\kappa_{dts}^{(4)}(x) \equiv \mathbb{E}\left[|Y_{idst}| \;\; | X_{i1ts} = x, X_{i2ts} = 0, W_{its} = 0\right],$$

$$\kappa_{dts}^{(5)}(x) \equiv \mathbb{E}\left[|Y_{idst}| \;\; | X_{i2ts} = x, X_{i1ts} = 0, W_{its} = 0\right],$$

and

$$\kappa_{(1,1)ts}^{(6)}(w) \equiv \mathbb{E}\left[\left|Y_{i(1,1)ts}\right| \;\; | W_{its} = w, X_{i1ts} = 0, X_{i2ts} = 0\right].$$

Proof of Theorem 3.2. Lemma B.1 verifies the technical conditions for applying the results in Seo and Otsu (2018), and shows that $\phi_{Ni}(b)$ and $\varphi_{Ni}(r)$ are *manageable* in the sense of Kim and Pollard (1990). By Assumption P9, Lemma 1 in Seo and Otsu (2018) and its subsequent analysis, we have

$$\hat{\beta} - \beta = O_P((Nh_N^{k_1+k_2})^{-1/3}) \text{ and } \hat{\gamma} - \gamma = O_P((N\sigma_N^{2k_1})^{-1/3}).$$

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

$$\arg\max_{b\in\mathcal{B}}(Nh_N^{k_1+k_2})^{2/3} \cdot N^{-1}\sum_{i=1}^{N}\phi_{Ni}(\beta + (Nh_N^{k_1+k_2})^{-1/3}\{(Nh_N^{k_1+k_2})^{1/3}(b-\beta)\}).$$

We get the asymptotics of $\hat{\beta}$ if we can get the asymptotics of

$$\arg\max_{\rho\in\mathbb{R}^{k_1}}(Nh_N^{k_1+k_2})^{2/3} \cdot N^{-1}\sum_{i=1}^{N}\phi_{Ni}(\beta + (Nh_N^{k_1+k_2})^{-1/3}\rho).$$

Since $\phi_{Ni}(b)$ is *manageable* in the sense of Kim and Pollard (1990), by Theorem 1 in Seo and Otsu (2018) and Lemma B.2, we have the uniform convergence of the above stochastic process and

$$(Nh_N^{k_1+k_2})^{2/3} \cdot N^{-1}\sum_{i=1}^{N}\phi_{Ni}(\beta + (Nh_N^{k_1+k_2})^{-1/3}\rho) \rightsquigarrow \mathcal{Z}_1(\rho),$$

where $\mathcal{Z}_1(\rho)$ is a Gaussian process with continuous sample path, expected value $\frac{1}{2}\rho'\mathbb{V}\rho$ and covariance kernel $\mathbb{H}_1(\rho_1,\rho_2)$. Apply the CMT to obtain

$$(Nh_N^{k_1+k_2})^{1/3}(\hat{\beta} - \beta) \xrightarrow{d} \arg\max_{\rho}\mathcal{Z}_1(\rho).$$

Apply similar arguments to $\hat{\gamma}$. By Theorem 1 in Seo and Otsu (2018) and Lemma B.2, we get

$$(N\sigma_N^{2k_1})^{1/3}(\hat{\gamma} - \gamma) \xrightarrow{d} \arg\max_{\delta\in\mathbb{R}^{k_2}}\mathcal{Z}_2(\delta),$$

where $\mathcal{Z}_2(\delta)$ is a Gaussian process with continuous sample path, expected value $\frac{1}{2}\delta'\mathbb{W}\delta$ and covariance kernel $\mathbb{H}_2(\delta_1,\delta_2)$. □

C Tables

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

	β_2				γ_2			
	MBIAS	RMSE	MED	MAD	MBIAS	RMSE	MED	MAD
$N = 250$	-0.0167	0.0833	-0.0129	0.0557	-0.0298	0.1691	0.0025	0.0902
$N = 500$	-0.0111	0.0575	-0.0121	0.0385	-0.0204	0.1162	-0.0055	0.0793
$N = 1000$	-0.0078	0.0382	-0.0078	0.0254	-0.0117	0.0828	-0.0047	0.0560

Table 1B: Design 1, Nonparametric Bootstrap

	β_2		γ_2	
	COVERAGE	LENGTH	COVERAGE	LENGTH
$N = 250$	0.9710	0.3579	0.9650	0.7670
$N = 500$	0.9610	0.2613	0.9590	0.4927
$N = 1000$	0.9550	0.1863	0.9570	0.3465

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

	β_2				γ_2			
	MBIAS	RMSE	MED	MAD	MBIAS	RMSE	MED	MAD
$N = 250$	-0.0173	0.1043	-0.0086	0.0797	-0.0357	0.1785	-0.0030	0.1040
$N = 500$	-0.0124	0.0754	-0.0096	0.0492	-0.0243	0.1299	-0.0060	0.0895
$N = 1000$	-0.0087	0.0542	-0.0101	0.0354	-0.0129	0.0825	-0.0065	0.0544

Table 2B: Design 2, Nonparametric Bootstrap

	β_2		γ_2	
	COVERAGE	LENGTH	COVERAGE	LENGTH
$N = 250$	0.9710	0.4597	0.9540	0.8010
$N = 500$	0.9670	0.3374	0.9590	0.5048
$N = 1000$	0.9580	0.2441	0.9510	0.3532

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

	β_2				γ_2			
	MBIAS	RMSE	MED	MAD	MBIAS	RMSE	MED	MAD
$N = 1000$	-0.0302	0.1722	0.0175	0.0856	-0.2099	0.5801	0.0282	0.1985
$N = 2500$	-0.0325	0.1482	-0.0036	0.0903	-0.1345	0.4637	0.0294	0.1541
$N = 5000$	-0.0299	0.1379	-0.0012	0.0876	-0.0800	0.3862	0.0587	0.1224

Table 3B: Design 3, Numerical Bootstrap

	β_2		γ_2	
	COVERAGE	LENGTH	COVERAGE	LENGTH
$N = 1000$	0.962	0.8632	0.833	1.0683
$N = 2500$	0.966	0.7651	0.921	0.9937
$N = 5000$	0.950	0.7014	0.934	0.9397

Table 4A: Design 4, Performance of $\hat{\beta}$ and $\hat{\gamma}$

	β_2				γ_2			
	MBIAS	RMSE	MED	MAD	MBIAS	RMSE	MED	MAD
$N = 1000$	-0.0289	0.1488	0.0137	0.0729	-0.1687	0.5240	0.0318	0.2019
$N = 2500$	-0.0273	0.1342	0.0010	0.0856	-0.1134	0.4254	0.0358	0.1489
$N = 5000$	-0.0218	0.1172	0.0017	0.0826	-0.0741	0.3268	0.0326	0.1022

Table 4B: Design 4, Numerical Bootstrap

	β_2		γ_2	
	COVERAGE	LENGTH	COVERAGE	LENGTH
$N = 1000$	0.981	0.8310	0.785	1.0578
$N = 2500$	0.965	0.7338	0.882	0.9852
$N = 5000$	0.961	0.6501	0.928	0.9301

D Technical Details for Bootstrap Inference

In this section, we provide some technical details for the nonparametric bootstrap for the cross-sectional model (Section 2.3) and the numerical bootstrap for the panel data model (Section 3.2).

D.1 Cross-Sectional Model

We provide an outline to show the validity of the nonparametric bootstrap for our estimators in the cross section case. To fill in the gaps of the outline, one first needs to define a product probability space for both the original random series and the bootstrap series as in Wellner and Zhan (1996) (an application can be found in Abrevaya and Huang (2005)), define a suitable norm for functions, and then establish some uniform convergence results following Sherman (1993, 1994a,b) on this probability space with this norm. To apply Sherman's results, the objective functions need to be *manageable* in the sense of Kim and Pollard (1990). This is indeed the case for our estimators. For example, equation (2.12) is a summation of

$$\mathcal{K}_{\sigma_N}(V_{im}(\hat{\beta}))Y_{im(1,1)}\text{sgn}\left(W_{im}'r\right) = \mathcal{K}_{\sigma_N}(V_{im}(\hat{\beta}))Y_{im(1,1)}\left(2\cdot 1\left[W_{im}'r>0\right]-1\right).$$

The collection of indicator functions $1\left[W_{im}'r>0\right]$ for $r\in\mathcal{R}$ is well known to be a Vapnik–Chervonenkis (VC) class. The kernel function $\mathcal{K}$ is assumed to be uniformly bounded with bounded support, and continuously differentiable in Assumption P7, and thus it is *manageable*. The complication of the bandwidth can be handled in the same way as in Seo and Otsu (2018).

Now suppose we have handled the technical details mentioned above. In what follows, we present the proof outline.

To ease expositions, we assume $\hat{\beta}$ and $\hat{\gamma}$ are obtained from maximizing, respectively, $\mathbb{L}_{N,\beta}^{K}(b)$ and $\mathbb{L}_{N,\gamma}^{K}(r,\hat{\beta})$, which are defined as

$$\mathbb{L}_{N,\beta}^{K}(b) = \sum_{i=1}^{N-1}\sum_{m>i}\mathcal{K}_{h_N}(X_{im})\,\chi_1(Z_i,Z_m,b),$$

and

$$\mathbb{L}_{N,\gamma}^{K}(r,\hat{\beta}) = \sum_{i=1}^{N-1}\sum_{m>i}\mathcal{K}_{\sigma_N}(X_{im}'\hat{\beta})\chi_2(Z_i,Z_m,r).$$

We assume that X_{im} is a sub-vector of Z_i-Z_m. Note that the objective functions here are different from the ones we used. But if we manage to show the consistency of the bootstrap here, it can be easily extended to our original estimators because they share the same structure.

We do normalization such that $\chi_1(Z_m,Z_i,\beta)=0$ and $\chi_2(Z_i,Z_m,\gamma)=0$ for all Z.[17] Assume that $\chi_1(Z_i,Z_m,b)=\chi_1(Z_m,Z_i,b)$, β uniquely maximizes $\mathbb{E}\left[\chi_1(Z_i,Z_m,b)\,|X_{im}=0\right]$, and $\mathcal{K}_{h_N}$ is

[17] Taking equation (2.11) as an example, this can be done by replacing $\text{sgn}(X'b)$ with $\text{sgn}(X'b)-\text{sgn}(X'\beta)$ in the objective function.

a p-th order kernel with h_N that satisfies $Nh_N^p = o(1)$. Similarly, we assume that $\chi_2(Z_i, Z_m, r) = \chi_2(Z_m, Z_i, r)$, γ uniquely maximizes $\mathbb{E}[\chi_2(Z_i, Z_m, r) | X_{im}'\beta = 0]$, and $\mathcal{K}_{\sigma_N}$ is a p-th order kernel with σ_N that satisfies $N\sigma_N^p = o(1)$. We similarly denote $\mathbb{L}_{N,\beta}^{K*}(b)$ and $\mathbb{L}_{N,\gamma}^{K*}(r, \hat{\beta}^*)$ as the corresponding objective functions using the bootstrap series.

Before establishing the consistency of the bootstrap, it is helpful to do a recast of how we derive the asymptotics for $(\hat{\beta}, \hat{\gamma})$. $\mathbb{L}_{N,\beta}^K(b)$ can be decomposed into

$$\begin{aligned} &2[N(N-1)]^{-1}\mathbb{L}_{N,\beta}^K(b) \\ =&\mathbb{E}[\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i, Z_m, b)] \\ &+\frac{2}{N}\sum_{i=1}^N \{\mathbb{E}[\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i, Z_m, b)|Z_i] - \mathbb{E}[\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i, Z_m, b)]\} \\ &+\frac{1}{N(N-1)}\sum_{i=1}^N \sum_{m=1,m\neq i}^N \{\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i, Z_m, b) - \mathbb{E}[\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i, Z_m, b)|Z_i] \\ &-\mathbb{E}[\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i, Z_m, b)|Z_m] + \mathbb{E}[\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i, Z_m, b)]\} \\ \equiv&\mathcal{T}_{N1} + \mathcal{T}_{N2} + \mathcal{T}_{N3}. \end{aligned} \tag{D.1}$$

Since $\chi_1(Z_i, Z_m, \beta) = 0$ and β uniquely maximizes $\mathbb{E}[\chi_1(Z_i, Z_m, b)|X_{im} = 0]$, we have

$$\begin{aligned} \mathcal{T}_{1N} &= \mathbb{E}[\chi_1(Z_i, Z_m, b)|X_{im} = 0] + O_P(h_N^p) \\ &= \frac{1}{2}(b-\beta)'\mathbb{V}_\beta(b-\beta) + O_P\left(\|b-\beta\|^3\right) + O_P(h_N^p), \end{aligned}$$

where the $O_P(h_N^p)$ is the bias term from matching and $\mathbb{V}_\beta$ is a negative definite matrix. For the second term,

$$\begin{aligned} \mathcal{T}_{N2} &= \frac{2}{N}\sum_{i=1}^N \{\mathbb{E}[\chi_1(Z_i, Z_m, b)|Z_i, X_{im} = 0] f_{X_m}(X_i) - \mathbb{E}[\chi_1(Z_i, Z_m, b)|X_{im} = 0]\} + O_P(h_N^p) \\ &= 2(b-\beta)'\frac{W_{N1}}{\sqrt{N}} + o_P\left(\|b-\beta\|^2\right) + O_P(h_N^p), \end{aligned} \tag{D.2}$$

where W_{N1} converges to a normal distribution. For the last term,

$$\mathcal{T}_{N3} = o_P(N^{-1}),$$

by similar arguments in Sherman (1993). Put all these results together to obtain

$$2[N(N-1)]^{-1}\mathbb{L}_{N,\beta}^K(b) = \frac{1}{2}(b-\beta)'\mathbb{V}_\beta(b-\beta) + 2(b-\beta)'\frac{W_{N1}}{\sqrt{N}} + o_P\left(\|b-\beta\|^2\right) + o_P(N^{-1}).$$

$\hat{\beta} - \beta$ can be shown to be $O_P(N^{-1/2})$. Using the rate result and invoking the CMT, we have

$$\sqrt{N}(\hat{\beta} - \beta) = -2\mathbb{V}_\beta^{-1}W_{N1} + o_P(1). \tag{D.3}$$

Let $\mathbb{E}^*$ and P^* denote the expectation and the density for the bootstrap series, respectively. By definition, P^* puts N^{-1} probability on each of $\{Z_i\}_{i=1}^N$. We can do a similar decomposition of $\mathbb{L}_{N,\beta}^{K*}(b)$ as for $\mathbb{L}_{N,\beta}^{K}(b)$:

$$
\begin{aligned}
& 2\left[N(N-1)\right]^{-1}\mathbb{L}_{N,\beta}^{K*}(b) \\
=& \left[N(N-1)\right]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b) \\
=& \mathbb{E}^*\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\right] \\
& +\frac{2}{N}\sum_{i=1}^{N}\left\{\mathbb{E}^*\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\,|Z_i^*\right]-\mathbb{E}^*\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\right]\right\} \\
& +\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\left\{\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)-\mathbb{E}^*\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\,|Z_i^*\right]\right. \\
& \left.-\mathbb{E}^*\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\,|Z_m^*\right]+\mathbb{E}^*\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\right]\right\} \\
\equiv& \mathcal{T}_{N1}^*+\mathcal{T}_{N2}^*+\mathcal{T}_{N3}^*.
\end{aligned} \tag{D.4}
$$

For $\mathcal{T}_{N1}^*$, by the definition of of $\mathbb{E}^*$,

$$
\begin{aligned}
\mathcal{T}_{N1}^* =& \mathbb{E}^*\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\right] \\
=& \frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{N}\mathcal{K}_{h_N}(X_{im})\chi_1(Z_i,Z_m,b) \\
=& \frac{1}{2}(b-\beta)'\mathbb{V}_\beta(b-\beta)+2(b-\beta)'\frac{W_{N1}}{\sqrt{N}}+O_P\left(\|b-\beta\|^3\right)+o_P\left(N^{-1}\right),
\end{aligned} \tag{D.5}
$$

where the last line holds because the second line is the same as $2\left[N(N-1)\right]^{-1}\mathbb{L}_{N,\beta}^{K}(b)$ plus a small order term.

Both $\mathcal{T}_{N2}$ and $\mathcal{T}_{N2}^*$ are sample averages of mean zero series. Moreover, for $\|b-\beta\|=O(N^{-1/2})$, $N\mathcal{T}_{N2}=2\sqrt{N}(b-\beta)'W_{N1}+o_P(1)$ and W_{N1} converges to a normal distribution. Then, we can apply the results in Giné and Zinn (1990) that NT_{N2}^* approximates the distribution of NT_{N2} for $\|b-\beta\|=O(N^{-1/2})$. That is, for $\|b-\beta\|=O(N^{-1/2})$,

$$
\mathcal{T}_{N2}^*=2(b-\beta)'\frac{W_{N1}^*}{\sqrt{N}}+o_P\left(N^{-1}\right), \tag{D.6}
$$

where W_{N1}^* is an independent copy of W_{N1}.

For $\mathcal{T}_{N3}^*$, terms across i and m are uncorrelated with each other under P^*. Therefore, $\left[\mathbb{E}^*\left(\mathcal{T}_{N3}^{*2}\right)\right]^{1/2}$ is of the same order as

$$
N^{-1}\left\{\mathbb{E}^*\left[\left[\mathcal{K}_{h_N}(X_{im}^*)\chi_1(Z_i^*,Z_m^*,b)\right]^2\right]\right\}^{1/2}=N^{-1}\left\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{N}\mathcal{K}_{h_N}(X_{im})^2\chi_1(Z_i,Z_m,b)^2\right\}^{1/2}.
$$

Note that $\chi_1(Z_i, Z_m, \beta) = 0$, so the above term is $o_P\left(N^{-1}\right)$ for $\|b-\beta\| = O\left(N^{-1/2}\right)$. That means $\left[\mathbb{E}^*\left(\mathcal{T}_{N3}^{*2}\right)\right]^{1/2} = o_P\left(N^{-1}\right)$. By Markov inequality,

$$\mathcal{T}_{N3}^* = o_P\left(N^{-1}\right). \tag{D.7}$$

To summarize, equations (D.4)–(D.7) imply that for $\|b-\beta\| = O\left(N^{-1/2}\right)$,

$$2\left[N(N-1)\right]^{-1}\mathbb{L}_{N,\beta}^{K*}(b) = \frac{1}{2}(b-\beta)'\mathbb{V}_\beta(b-\beta) + 2(b-\beta)'\frac{W_{N1}}{\sqrt{N}} + 2(b-\beta)'\frac{W_{N1}^*}{\sqrt{N}} + o_P\left(N^{-1}\right).$$

We can similarly show that $\hat{\beta}^* - \beta = O_P\left(N^{-1/2}\right)$. By the CMT again,

$$\sqrt{N}(\hat{\beta}^* - \beta) = -2\mathbb{V}_\beta^{-1}W_{N1} - 2\mathbb{V}_\beta^{-1}W_{N1}^* + o_P(1).$$

Substitute equation (D.3) into the equation above to get

$$\sqrt{N}(\hat{\beta}^* - \hat{\beta}) = -2\mathbb{V}_\beta^{-1}W_{N1}^* + o_P(1). \tag{D.8}$$

Since W_{N1}^* is an independent copy of W_{N1}, $\sqrt{N}(\hat{\beta}^* - \hat{\beta})$ converges to the same distribution as the one $\sqrt{N}(\hat{\beta} - \beta)$ converges to.

We turn to the inference for $\hat{\gamma}$. Again, we decompose $\mathbb{L}_{N,\gamma}^K(r, \hat{\beta})$ and analyze each term in the decompositions one by one. We start with

$$\begin{aligned}
&2\left[N(N-1)\right]^{-1}\mathbb{L}_{N,\gamma}^K\left(r, \hat{\beta}\right) \\
=&\left[N(N-1)\right]^{-1}\sum_{i=1}^N\sum_{m=1,m\neq i}^N \mathcal{K}_{\sigma_N}\left(X_{im}'\hat{\beta}\right)\chi_2(Z_i, Z_m, r) \\
=&\left[N(N-1)\right]^{-1}\sum_{i=1}^N\sum_{m=1,m\neq i}^N \mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)\chi_2(Z_i, Z_m, r) \\
&+\left[N(N-1)\right]^{-1}\sum_{i=1}^N\sum_{m=1,m\neq i}^N \chi_2(Z_i, Z_m, r)\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'\frac{\left(\hat{\beta}-\beta\right)}{\sigma_N} \\
&+\left[N(N-1)\right]^{-1}\sum_{i=1}^N\sum_{m=1,m\neq i}^N \chi_2(Z_i, Z_m, r)\nabla^2\mathcal{K}_{\sigma_N}\left(X_{im}'\tilde{\beta}\right)\left[X_{im}'\frac{\left(\hat{\beta}-\beta\right)}{\sigma_N}\right]^2 \\
\equiv&\mathcal{T}_{N4} + \mathcal{T}_{N5} + \mathcal{T}_{N6},
\end{aligned} \tag{D.9}$$

where $\tilde{\beta}$ is a vector that lies between β and $\hat{\beta}$ and makes the equality hold.

Note that $\mathcal{T}_{N4}$ shares the same structure as $2\left[N(N-1)\right]^{-1}\mathbb{L}_{N,\beta}^K(b)$, and so

$$\mathcal{T}_{N4} = \frac{1}{2}(r-\gamma)'\mathbb{V}_\gamma(r-\gamma) + 2(r-\gamma)'\frac{W_{N2}}{\sqrt{N}} + O_P\left(\|r-\gamma\|^3\right) + O_P\left(\sigma_N^p\right) + o_P\left(N^{-1}\right), \tag{D.10}$$

where $\mathbb{V}_\gamma$ is a negative definite matrix, and W_{N2} converges to a normal distribution.

For term $\mathcal{T}_{N5}$,

$$\mathcal{T}_{N5}=\left\{[N(N-1)]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'\right\}\left(\hat{\beta}-\beta\right).$$

Note that

$$\begin{aligned}&[N(N-1)]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}' \qquad \text{(D.11)}\\
=&\mathbb{E}\left[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'\right]\\
&+\frac{2}{N}\sum_{i=1}^{N}\left\{\mathbb{E}\left[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'|Z_i\right]-\mathbb{E}\left[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'\right]\right\}\\
&+\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\left\{\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'-\mathbb{E}\left[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'|Z_i\right]\right.\\
&\left.-\mathbb{E}\left[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'|Z_m\right]+\mathbb{E}\left[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}'\right]\right\}\end{aligned}$$

with the lead term $\mathbb{E}[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}(X_{im}'\beta)X_{im}']$. As $\int\nabla\mathcal{K}_{\sigma_N}(u)\,\mathrm{d}u=0$, the lead term cancels one more σ_N^{-1} when calculating the expectation. Since $\chi_2(Z_i,Z_m,\gamma)=0$, we can write

$$\mathbb{E}\left[\chi_2(Z_i,Z_m,r)\,\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}'\beta\right)X_{im}\right]=(r-\gamma)'\,\Gamma+O(\|r-\gamma\|^2)+O\left(\sigma_N^p\right), \qquad \text{(D.12)}$$

for some matrix Γ. Collect all these result to obtain

$$\begin{aligned}\mathcal{T}_{N5}&=(r-\gamma)'\,\Gamma\left(\hat{\beta}-\beta\right)+O_P\left(\|r-\gamma\|^2\left\|\hat{\beta}-\beta\right\|\right)+O\left(\sigma_N^p\right) \qquad \text{(D.13)}\\
&=-2(r-\gamma)'\,\Gamma\mathbb{V}_\beta^{-1}\frac{W_{N1}}{\sqrt{N}}+O_P\left(\|r-\gamma\|^2\left\|\hat{\beta}-\beta\right\|\right)+o_P\left(\frac{\|r-\gamma\|}{\sqrt{N}}\right)+O\left(\sigma_N^p\right),\end{aligned}$$

where the second line follows by substituting in equation (D.3).

For term $\mathcal{T}_{N6}$, it is not hard to see that

$$\mathcal{T}_{N6}=O_P\left(\|r-\gamma\|\left\|\hat{\beta}-\beta\right\|^2\Big/\sigma_N^2\right)=o_P\left(N^{-1}\right), \qquad \text{(D.14)}$$

for $\|r-\gamma\|=O(N^{-1/2})$, and σ_N in Assumption C9.

To summarize, equations (D.9), (D.10), (D.13), and (D.14) imply that for $\|r-\gamma\|=O\left(N^{-1/2}\right)$,

$$\begin{aligned}2[N(N-1)]^{-1}\mathbb{L}_{N,\gamma}^{K}\left(r,\hat{\beta}\right)=&\frac{1}{2}(r-\gamma)'\,\mathbb{V}_\gamma(r-\gamma)\\
&+2(r-\gamma)'\frac{W_{N2}}{\sqrt{N}}-2(r-\gamma)'\,\Gamma\mathbb{V}_\beta^{-1}\frac{W_{N1}}{\sqrt{N}}+o_P\left(N^{-1}\right).\end{aligned}$$

Once $\hat{\gamma}-\gamma=O_P\left(N^{-1/2}\right)$ is established, applying the CMT gives

$$\sqrt{N}(\hat{\gamma}-\gamma)=-2\mathbb{V}_\gamma^{-1}W_{N2}+2\mathbb{V}_\gamma^{-1}\Gamma\mathbb{V}_\beta^{-1}W_{N1}+o_P(1). \qquad \text{(D.15)}$$

For $\mathbb{L}_{N,\gamma}^{K*}(r,b)$, we decompose it similarly:

$$
\begin{aligned}
&2\left[N(N-1)\right]^{-1}\mathbb{L}_{N,\gamma}^{K*}\left(r,\hat{\beta}^{*}\right) \qquad \text{(D.16)}\\
&=\left[N(N-1)\right]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}^{*}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)\\
&=\left[N(N-1)\right]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)\\
&+\left[N(N-1)\right]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\chi_2\left(Z_i^{*},Z_m^{*},r\right)\nabla\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\left(\frac{X_{im}^{*\prime}\left(\hat{\beta}^{*}-\hat{\beta}\right)}{\sigma_N}\right)\\
&+\left[N(N-1)\right]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\chi_2\left(Z_i^{*},Z_m^{*},r\right)\nabla^2\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\tilde{\beta}^{*}\right)\left(\frac{X_{im}^{*\prime}\left(\hat{\beta}^{*}-\hat{\beta}\right)}{\sigma_N}\right)^2\\
&\equiv\mathcal{T}_{N4}^{*}+\mathcal{T}_{N5}^{*}+\mathcal{T}_{N6}^{*},
\end{aligned}
$$

where $\tilde{\beta}^{*}$ is some vector that lies between b and $\hat{\beta}$ and makes the equality hold.

For $\mathcal{T}_{N4}^{*}$, we have

$$
\begin{aligned}
\mathcal{T}_{N4}^{*}=&\left[N(N-1)\right]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)\\
=&\mathbb{E}^{*}\left[\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)\right]\\
&+\frac{2}{N}\sum_{i=1}^{N}\left\{\mathbb{E}^{*}\left[\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)|Z_i^{*}\right]-\mathbb{E}^{*}\left[\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)\right]\right\}\\
&+\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\left\{\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)-\mathbb{E}^{*}\left[\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)|Z_i^{*}\right]\right.\\
&\left.-\mathbb{E}^{*}\left[\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)|Z_m^{*}\right]+\mathbb{E}^{*}\left[\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)\right]\right\}.\\
\equiv&\mathcal{T}_{N4,1}^{*}+\mathcal{T}_{N4,2}^{*}+\mathcal{T}_{N4,3}^{*}. \qquad \text{(D.17)}
\end{aligned}
$$

We analyze the three terms in (D.17) one by one.

For $\mathcal{T}_{N4,1}^{*}$, we have

$$
\begin{aligned}
\mathcal{T}_{N4,1}^{*}&=\mathbb{E}^{*}\left[\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\chi_2\left(Z_i^{*},Z_m^{*},r\right)\right] \qquad \text{(D.18)}\\
&=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{N}\mathcal{K}_{\sigma_N}\left(X_{im}^{\prime}\hat{\beta}\right)\chi_2\left(Z_i,Z_m,r\right)\\
&=\frac{1}{2}(r-\gamma)^{\prime}\mathbb{V}_{\gamma}(r-\gamma)+2(r-\gamma)^{\prime}\frac{W_{N2}}{\sqrt{N}}-2(r-\gamma)^{\prime}\Gamma\mathbb{V}_{\beta}^{-1}\frac{W_{N1}}{\sqrt{N}}+o_P\left(N^{-1}\right),
\end{aligned}
$$

where the last line holds because the term in the second line is $2\left[N(N-1)\right]^{-1}\mathbb{L}_{N,\gamma}^{K}(r,\hat{\beta})$ plus a small order term.

For $\mathcal{T}^*_{N4,2}$, we first note that

$$\begin{aligned}
&\sum_{i=1}^{N}\left\{\mathbb{E}\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\beta\right)\chi_2\left(Z_i,Z_m,r\right)|Z_i\right]-\mathbb{E}\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\beta\right)\chi_2\left(Z_i,Z_m,r\right)\right]\right\} \\
=&\sqrt{N}\left(r-\gamma\right)'W_{N2}+o_P\left(1\right),
\end{aligned} \tag{D.19}$$

for $\|r-\gamma\|=O\left(N^{-1/2}\right)$, which holds for the same reason as for equation (D.2). Since $\hat{\beta}-\beta=O_P\left(N^{-1/2}\right)$, by the equicontinuity we claimed,

$$\begin{aligned}
&\sum_{i=1}^{N}\left\{\mathbb{E}\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\beta\right)\chi_2\left(Z_i,Z_m,r\right)|Z_i\right]-\mathbb{E}\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\beta\right)\chi_2\left(Z_i,Z_m,r\right)\right]\right\} \\
&\quad-\sum_{i=1}^{N}\left\{\mathbb{E}_N\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\hat{\beta}\right)\chi_2\left(Z_i,Z_m,r\right)|Z_i\right]-\mathbb{E}_N\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\hat{\beta}\right)\chi_2\left(Z_i,Z_m,r\right)\right]\right\}=o_P\left(1\right).
\end{aligned} \tag{D.20}$$

Then equations (D.19) and (D.20) imply that

$$\begin{aligned}
&2\sum_{i=1}^{N}\left\{\mathbb{E}_N\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\hat{\beta}\right)\chi_2\left(Z_i,Z_m,r\right)|Z_i\right]-\mathbb{E}_N\left[\mathcal{K}_{\sigma_N}\left(X'_{im}\hat{\beta}\right)\chi_2\left(Z_i,Z_m,r\right)\right]\right\} \\
&=2\sqrt{N}\left(r-\gamma\right)'W_{N2}+o_P\left(1\right),
\end{aligned}$$

which approximates a normal distribution for $\|r-\gamma\|=O\left(N^{-1/2}\right)$. Thus, we are able to apply the results of Giné and Zinn (1990) and get that

$$N\mathcal{T}^*_{N4,2}=2\sum_{i=1}^{N}\left\{\mathbb{E}^*\left[\mathcal{K}_{\sigma_N}\left(X^{*\prime}_{im}\hat{\beta}\right)\chi_2\left(Z^*_i,Z^*_m,r\right)|Z^*_i\right]-\mathbb{E}^*\left[\mathcal{K}_{\sigma_N}\left(X^{*\prime}_{im}\hat{\beta}\right)\chi_2\left(Z^*_i,Z^*_m,r\right)\right]\right\}$$

approximates the distribution of the above term. That is

$$\mathcal{T}^*_{N4,2}=2\left(r-\gamma\right)'\frac{W^*_{N2}}{\sqrt{N}}+o_P\left(N^{-1}\right), \tag{D.21}$$

where W^*_{N2} is an independent copy of W_{N2}.

The same arguments as we used for $\mathcal{T}^*_{N3}$ lead to

$$\mathcal{T}^*_{N4,3}=o_P\left(N^{-1}\right). \tag{D.22}$$

Then, equations (D.17), (D.18), (D.21), and (D.22) together imply that for $\|r-\gamma\|=O(N^{-1/2})$,

$$\begin{aligned}
\mathcal{T}^*_{N4}=&\frac{1}{2}\left(r-\gamma\right)'\mathbb{V}_\gamma\left(r-\gamma\right)+2\left(r-\gamma\right)'\frac{W_{N2}}{\sqrt{N}}-2\left(r-\gamma\right)'\Gamma\mathbb{V}_\beta^{-1}\frac{W_{N1}}{\sqrt{N}} \\
&+2\left(r-\gamma\right)'\frac{W^*_{N2}}{\sqrt{N}}+o_P\left(N^{-1}\right).
\end{aligned} \tag{D.23}$$

For $\mathcal{T}_{N5}^{*}$, we have

$$\begin{aligned}\mathcal{T}_{N5}^{*} &= [N(N-1)]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\chi_2\left(Z_i^*, Z_m^*, r\right)\nabla\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)\left(\frac{X_{im}^{*\prime}\left(\hat{\beta}^*-\hat{\beta}\right)}{\sigma_N}\right)\\ &= \left\{[N(N-1)]^{-1}\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\chi_2\left(Z_i^*, Z_m^*, r\right)\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)X_{im}^{*\prime}\right\}\left(\hat{\beta}^*-\hat{\beta}\right).\end{aligned}$$

Following the same analysis for $\mathcal{T}_{N5}$, the lead term for $\mathcal{T}_{N5}^{*}$ is

$$\mathbb{E}^*\left[\chi_2\left(Z_i^*, Z_m^*, r\right)\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)X_{im}^{*\prime}\right]\left(\hat{\beta}^*-\hat{\beta}\right).$$

Note that

$$\begin{aligned}&\mathbb{E}^*\left[\chi_2\left(Z_i^*, Z_m^*, r\right)\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}^{*\prime}\hat{\beta}\right)X_{im}^{*\prime}\right]\left(\hat{\beta}^*-\hat{\beta}\right)\\ &=\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{N}\chi_2\left(Z_i, Z_m, r\right)\sigma_N^{-1}\nabla\mathcal{K}_{\sigma_N}\left(X_{im}^{\prime}\beta\right)X_{im}\right]\left(\hat{\beta}^*-\hat{\beta}\right)\\ &=(r-\gamma)^{\prime}\Gamma\left(\hat{\beta}^*-\hat{\beta}\right)+O_P\left(\|r-\gamma\|^2\left\|\hat{\beta}^*-\hat{\beta}\right\|\right)+o_P\left(N^{-1}\right)\\ &=-2(r-\gamma)^{\prime}\Gamma\mathbb{V}_{\beta}^{-1}\frac{W_{N1}^{*}}{\sqrt{N}}+o_P\left(\frac{\|r-\gamma\|}{\sqrt{N}}\right)+O_P\left(\|r-\gamma\|^2\left\|\hat{\beta}^*-\hat{\beta}\right\|\right)+o_P\left(N^{-1}\right),\end{aligned}$$

where we use equations (D.11) and (D.12) to get the third line, and we substitute equation (D.8) in the last line. Then we have

$$\mathcal{T}_{N5}^{*}=-2(r-\gamma)^{\prime}\Gamma\mathbb{V}_{\beta}^{-1}\frac{W_{N1}^{*}}{\sqrt{N}}+o_P\left(\frac{\|r-\gamma\|}{\sqrt{N}}\right)+O_P\left(\|r-\gamma\|^2\left\|\hat{\beta}^*-\hat{\beta}\right\|\right)+o_P\left(N^{-1}\right). \quad \text{(D.24)}$$

For $\mathcal{T}_{N6}^{*}$, it is not hard to see that

$$\mathcal{T}_{N6}^{*}=O_P\left(\frac{\|r-\gamma\|\left\|\hat{\beta}^*-\hat{\beta}\right\|^2}{\sigma_N^2}\right). \quad \text{(D.25)}$$

Then, equations (D.16), (D.23), (D.24), and (D.25) together imply that for $\|r-\gamma\|=O\left(N^{-1/2}\right)$,

$$\begin{aligned}&2[N(N-1)]^{-1}\mathbb{L}_{N,\gamma}^{K*}\left(r,\hat{\beta}^*\right)\\ =&\frac{1}{2}(r-\gamma)^{\prime}\mathbb{V}_{\gamma}(r-\gamma)+2(r-\gamma)^{\prime}\frac{W_{N2}}{\sqrt{N}}-2(r-\gamma)^{\prime}\Gamma\mathbb{V}_{\beta}^{-1}\frac{W_{N1}}{\sqrt{N}}+2(r-\gamma)^{\prime}\frac{W_{N2}^{*}}{\sqrt{N}}\\ &-2(r-\gamma)^{\prime}\Gamma\mathbb{V}_{\beta}^{-1}\frac{W_{N1}^{*}}{\sqrt{N}}+o_P\left(N^{-1}\right),\end{aligned}$$

from which we can show that $\hat{\gamma}^*-\gamma=O_P\left(N^{-1/2}\right)$. With this result and by the CMT, we get

$$\sqrt{N}\left(\hat{\gamma}^*-\gamma\right)=-2\mathbb{V}_{\gamma}^{-1}W_{N2}+2\mathbb{V}_{\gamma}^{-1}\Gamma\mathbb{V}_{\beta}^{-1}W_{N1}-2\mathbb{V}_{\gamma}^{-1}W_{N2}^{*}+2\mathbb{V}_{\gamma}^{-1}\Gamma\mathbb{V}_{\beta}^{-1}W_{N1}^{*}+o_P(1),$$

which implies

$$\sqrt{N}\left(\hat{\gamma}^*-\hat{\gamma}\right)=-2\mathbb{V}_\gamma^{-1}W_{N2}^*+2\mathbb{V}_\gamma^{-1}\Gamma\mathbb{V}_\beta^{-1}W_{N1}^*+o_P\left(N^{-1}\right)$$

by substituting equation (D.15) in.

Since W_{N1}^* and W_{N2}^* are independent copies of W_{N1} and W_{N2}, respectively, $\sqrt{N}\left(\hat{\gamma}^*-\hat{\gamma}\right)$ approximates the distribution of $\sqrt{N}\left(\hat{\gamma}-\gamma\right)$.

D.2 Panel Data Model

We illustrate why the numerical bootstrap works for β in this section. The discussion onγ is omitted due to the similarity.

By some standard calculation as Seo and Otsu (2018), we can show that the convergence rate of $\hat{\beta}^*$ is $(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3}$. Thus, if we manage to show the limit of

$$\begin{aligned}&(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}\cdot N^{-1}\sum_{i=1}^N\phi_{Ni}(\beta+\rho(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})+(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}\cdot(N\varepsilon_{N1})^{1/2}\\&\cdot[N^{-1}\sum_{i=1}^N\phi_{Ni}^*(\beta+\rho(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})-N^{-1}\sum_{i=1}^N\phi_{Ni}(\beta+\rho(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})],\end{aligned}\tag{D.26}$$

then the limiting distribution of $\hat{\beta}^*$ can be established.

For the first term in equation (D.26),

$$\begin{aligned}&(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}\cdot N^{-1}\sum_{i=1}^N\phi_{Ni}(\beta+\rho(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})\\=&(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}\mathbb{E}[\phi_{Ni}(\beta+\rho(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})]\\&+(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}\cdot N^{-1}\sum_{i=1}^N[\phi_{Ni}(\beta+\rho(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})-\mathbb{E}[\phi_{Ni}(\beta+\rho(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})]]\\=&\frac{1}{2}\rho'\mathbb{V}\rho+O_P((\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}\cdot N^{-1/2}\cdot h_N^{-(k_1+k_2)/2}(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/6})=\frac{1}{2}\rho'\mathbb{V}\rho+o_P(1),\end{aligned}\tag{D.27}$$

where the second equality follows by Lemma B.2, the assumption on ε_{N1} such that the bias term h_N^2 is a small order term compared to $(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}$, and Markov inequality, and the last equality holds by $N\varepsilon_{N1}\to\infty$.

The second term in equation (D.26) is a sample average of mean 0 series under P^* multiplied by $(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{2/3}(N\varepsilon_{N1})^{1/2}$. The covariance kernel of the second term is

$$\begin{aligned}&(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{4/3}\cdot(N\varepsilon_{N1})\cdot N^{-1}\cdot\mathbb{E}^*[\phi_{Ni}^*(\beta+\rho_1(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})\phi_{Ni}^*(\beta+\rho_2(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})]\\=&(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{1/3}\mathbb{E}^*[h_N^{k_1+k_2}\phi_{Ni}^*(\beta+\rho_1(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})\phi_{Ni}^*(\beta+\rho_2(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})]\\=&(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{1/3}\mathbb{E}[h_N^{k_1+k_2}\phi_{Ni}(\beta+\rho_1(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})\phi_{Ni}(\beta+\rho_2(\varepsilon_{N1}^{-1}h_N^{k_1+k_2})^{-1/3})]+o_P(1)\\=&\mathbb{H}_1(\rho_1,\rho_2)+o_P(1),\end{aligned}\tag{D.28}$$

by the i.i.d. sampling and Lemma B.2.

Equations (D.27) and (D.28) imply that the term in equation (D.26) converges to $\mathcal{Z}_1^*(\rho)$ with mean $\frac{1}{2}\rho'\mathbb{V}\rho$ and covariance kernel $\mathbb{H}_1(\rho_1, \rho_2)$. Thus, $\mathcal{Z}_1^*(\rho)$ is an independent copy of $\mathcal{Z}_1(\rho)$ and

$$(\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3}(\hat{\beta}^* - \beta) \xrightarrow{d} \arg\max_{\rho \in \mathbb{R}^{k_1}} \mathcal{Z}_1^*(\rho).$$

Finally, note, by $N\varepsilon_{N1} \to \infty$ and

$$\begin{aligned}(\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3}(\hat{\beta}^* - \hat{\beta}) &= (\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3}(\hat{\beta}^* - \beta) + (\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3}(\hat{\beta} - \beta) \\ &= (\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3}(\hat{\beta}^* - \beta) + O_P((\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3} \cdot (N h_N^{k_1+k_2})^{-1/3}) \\ &= (\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3}(\hat{\beta}^* - \beta) + o_P(1),\end{aligned}$$

we have

$$(\varepsilon_{N1}^{-1} h_N^{k_1+k_2})^{1/3}(\hat{\beta}^* - \hat{\beta}) \xrightarrow{d} \arg\max_{\rho \in \mathbb{R}^{k_1}} \mathcal{Z}_1^*(\rho).$$

E Proofs of Technical Lemmas

We present proofs of Lemmas A.1–A.7 and Lemmas B.1–B.2 in order.

Proof of Lemma A.1. First, note that

$$\frac{1}{2}\sigma_N^2 \sup_{r \in \mathcal{R}} |\hat{\mathcal{L}}_N^K(r) - \mathcal{L}_N^K(r)| \leq \frac{1}{N(N-1)} \sum_{i \neq m} |K_{\sigma_N,\gamma}(\hat{V}_{im}) - K_{\sigma_N,\gamma}(V_{im})|. \tag{E.1}$$

Applying a second-order Taylor expansion to the right-hand side of (E.1) yields the lead term of the form

$$\begin{aligned}&\frac{1}{\sigma_N N(N-1)} \sum_{i \neq m} |\nabla_1 K_{\sigma_N,\gamma}(V_{im}) X_{im1}'(\hat{\beta} - \beta) + \nabla_2 K_{\sigma_N,\gamma}(V_{im}) X_{im2}'(\hat{\beta} - \beta)| \\ \leq &\frac{1}{\sigma_N N(N-1)} \sum_{i \neq m} \sum_{j=1}^{2} |\nabla_j K_\gamma(V_{im}/\sigma_N) X_{imj}'(\hat{\beta} - \beta)|,\end{aligned}$$

where $\nabla_j K_\gamma(\cdot)$ ($\nabla_j K_{\sigma_N,\gamma}(\cdot)$) denotes the partial derivative of $K_\gamma(\cdot)$ ($K_{\sigma_N,\gamma}(\cdot)$) w.r.t. its j-th argument corresponding to alternative j. It suffices to focus on the term with $j = 1$. Then by Assumptions C1–C6, C8, C9, and the Cauchy-Schwarz inequality, we conclude that

$$\begin{aligned}&\frac{1}{N(N-1)} \sum_{i \neq m} \left| \nabla_1 K_\gamma(V_{im}/\sigma_N) \frac{X_{im1}'(\hat{\beta} - \beta)}{\sigma_N} \right| \leq \frac{1}{N(N-1)} \sum_{i \neq m} |\nabla_1 K_\gamma(V_{im}/\sigma_N)| \|X_{im1}\| \frac{\|\hat{\beta} - \beta\|}{\sigma_N} \\ =&[O_P(\sigma_N^2) + o_P(1)] O_P(\|\hat{\beta} - \beta\|/\sigma_N) = o_P(1),\end{aligned}$$

where the penultimate equality follows from the SLLN of U-statistics (see e.g., Serfling (2009)) and $\mathbb{E}[|\nabla_1 K_\gamma(V_{im}/\sigma_N)| \|X_{im1}\|] = O_P(\sigma_N^2)$, and the last equality is due to the $\sqrt{N}$-consistency of $\hat{\beta}$. Then the desired convergence result follows. $\square$

Proof of Lemma A.2. Note that Lemma A.2 follows from $\sup_{r\in\mathcal{R}}|\mathcal{L}_N^K(r)-\mathbb{E}[\mathcal{L}_N^K(r)]|=o_P(1)$ and $\sup_{r\in\mathcal{R}}|\mathbb{E}[\mathcal{L}_N^K(r)]-\mathcal{L}(r)|=o(1)$.

Define $\mathcal{F}_N=\{K_{\sigma_N,\gamma}(v_{im})h_{im}(r)|r\in\mathcal{R}\}$ with $\sigma_N>0$ and $\sigma_N\to 0$. $\mathcal{F}_N$ is a sub-class of the fixed class $\mathcal{F}\equiv\{K_\gamma(v_{im}/\sigma)h_{im}(r)|r\in\mathcal{R},\sigma>0\}=\mathcal{F}_r\mathcal{F}_\sigma$, with $\mathcal{F}_\sigma\equiv\{K_\gamma(v_{im}/\sigma)|\sigma>0\}$, which is Euclidean for the constant envelope $\sup_{v\in\mathbb{R}^2}|K_\gamma(v)|$ by Lemma 22 in Nolan and Pollard (1987) and $\mathcal{F}_r\equiv\{h_{im}(r)|r\in\mathcal{R}\}$. Noticing that $h_{im}(r)$ is uniformly bounded by 2, Example 2.11 and Lemma 2.15 in Pakes and Pollard (1989) then implies that $\mathcal{F}_r$ is Euclidean for the constant envelope 2. Putting all these results together, we conclude using Lemma 2.14 in Pakes and Pollard (1989) that $\mathcal{F}$ is Euclidean for the constant envelope $2\sup_{v\in\mathbb{R}^2}|K_\gamma(v)|$. Applying Corollary 4 in Sherman (1994a), we obtain that $\sup_{r\in\mathcal{R}}|\mathcal{L}_N^K(r)-\mathbb{E}[\mathcal{L}_N^K(r)]|=O_P(N^{-1}/\sigma_N^2)=o_P(1)$ by Assumption C9.

It remains to show that $\sup_{r\in\mathcal{R}}|\mathbb{E}[\mathcal{L}_N^K(r)]-\mathcal{L}(r)|=o(1)$. Let $\eta(\cdot)\equiv f_{V_{im}}(\cdot)\mathbb{E}[h_{im}(r)|V_{im}=\cdot]$. Then $\mathcal{L}(r)=\eta(0)$ by definition. We can write by Assumptions C8 and C9 that

$$
\begin{aligned}
&\sup_{r\in\mathcal{R}}|\mathbb{E}[\mathcal{L}_N^K(r)]-\mathcal{L}(r)|=\sup_{r\in\mathcal{R}}|\sigma_N^{-2}\mathbb{E}[K_\gamma(V_{im}/\sigma_N)h_{im}(r)]-\eta(0)|\\
=&\sup_{r\in\mathcal{R}}|\int\sigma_N^{-2}K_\gamma(v/\sigma_N)\eta(v)\mathrm{d}v-\eta(0)|=\sup_{r\in\mathcal{R}}|\int\sigma_N^{-2}K_\gamma(v/\sigma_N)[\eta(0)+v'\nabla\eta(\bar v)]\mathrm{d}v-\eta(0)|\\
=&\sup_{r\in\mathcal{R}}|\int K_\gamma(u)[\eta(0)+\sigma_N u'\nabla\eta(u_N)]\mathrm{d}u-\eta(0)|=\sup_{r\in\mathcal{R}}|\sigma_N\int K_\gamma(u)u'\nabla\eta(u_N)\mathrm{d}u|\\
\leq&\sigma_N\cdot\sup_{r\in\mathcal{R}}\int|K_\gamma(u)|\|u\|\|\nabla\eta(u_N)\|\mathrm{d}u=O(\sigma_N)=o(1),
\end{aligned}
$$

where the third equality applies a mean-value expansion and the fourth equality uses a change of variables. Then, the desired result follows. □

Proof of Lemma A.3. The proof proceeds by verifying the four sufficient conditions for Theorem 2.1 in Newey and McFadden (1994): (S1) $\mathcal{R}$ is a compact set, (S2) $\sup_{r\in\mathcal{R}}|\hat{\mathcal{L}}_N^K(r)-\mathcal{L}(r)|=o_P(1)$, (S3) $\mathcal{L}(r)$ is continuous in r, and (S4) $\mathcal{L}(r)$ is uniquely maximized at γ.

Condition (S1) is satisfied by Assumption C4. Lemmas A.1 and A.2 together imply that condition (S2) holds. Note that

$$
\begin{aligned}
&\mathbb{E}[h_{im}(r)|V_{im}=0]=\mathbb{E}\left\{\mathbb{E}[Y_{im(1,1)}(\mathrm{sgn}(W_{im}'r)-\mathrm{sgn}(W_{im}'\gamma))|Z_i,Z_m]|V_{im}=0\right\}\\
=&\mathbb{E}\left[(P(Y_{i(1,1)}=1|Z_i)-P(Y_{m(1,1)}=1|Z_m))(\mathrm{sgn}(W_{im}'r)-\mathrm{sgn}(W_{im}'\gamma))|V_{im}=0\right].
\end{aligned}
$$

Then the identification condition (S4) can be verified by similar arguments used in the proof of Theorem 2.1 given that $f_{V_{im}}(0)>0$ by Assumption C5.

It remains to verify the continuity of $\mathcal{L}(r)$ in r. Assuming $r^{(1)}>0$ w.l.o.g., $\mathbb{E}[h_{im}(r)|V_{im}=0]$ can be expressed as the sum of terms like

$$
\begin{aligned}
&P(Y_{im(1,1)}=d,W_{im}^{(1)}>-\tilde W_{im}'\tilde r/r^{(1)}|V_{im}=0)\\
=&\int\int_{-\tilde W_{im}'\tilde r/r^{(1)}}^{\infty}P(Y_{im(1,1)}=d|W_{im}^{(1)}=w,\tilde W_{im},V_{im}=0)f_{W_{im}^{(1)}|\tilde W_{im},V_{im}=0}(w)\mathrm{d}w\mathrm{d}F_{\tilde W_{im}|V_{im}=0}.
\end{aligned}
$$

for some $d \in \{-1, 0, 1\}$. Then $\mathcal{L}(r)$ is continuous in r if $f_{W_{im}^{(1)}|\tilde{W}_{im}, V_{im}=0}(\cdot)$ does not have any mass points, which is guaranteed by Assumption C3. This completes the proof. $\square$

Proof of Lemma A.4. First, we express $\mathbb{E}[\mathcal{L}_N^K(r)]$ as the integral

$$\int \sigma_N^{-2} K_{\sigma_N,\gamma}(v_{im}) B(v_i, v_m, w_i, w_m) S_{im}(r) \mathrm{d}F_{V,W}(v_i, w_i) \mathrm{d}F_{V,W}(v_m, w_m) \tag{E.2}$$
$$= \int K_\gamma(u_{im}) B(v_m + u_{im}\sigma_N, v_m, w_i, w_m) S_{im}(r) f_{V,W}(v_m + u_{im}\sigma_N, w_i) \mathrm{d}w_i \mathrm{d}u_{im} \mathrm{d}F_{V,W}(v_m, w_m),$$

where we apply the change of variables $u_{im} = v_{im}/\sigma_N$ to obtain the equality.

Take the $\kappa_\gamma^{\text{th}}$-order Taylor expansion inside the integral around v_m to obtain the lead term

$$\int K_\gamma(u_{im}) B(v_m, v_m, w_i, w_m) S_{im}(r) f_{V,W}(v_m, w_i) \mathrm{d}w_i \mathrm{d}u_{im} \mathrm{d}F_{V,W}(\nu_m, w_m)$$
$$= \int B(v_m, v_m, w_i, w_m) S_{im}(r) f_{V,W}(v_m, w_i) \mathrm{d}w_i \mathrm{d}F_{V,W}(v_m, w_m) = \mathbb{E}[\tau_m(r)], \tag{E.3}$$

where the first equality follows by $\int K_\gamma(u) du = 1$. All remaining therms are zero except the last one which is of order $O(\sigma_N^\kappa \delta_N) = o(\|r - \gamma\|^2)$ by Assumptions C7–C9.

Note that a second-order Taylor expansion around γ gives

$$\tau_m(r) - \tau_m(\gamma) = (r - \gamma)' \nabla \tau_m(\gamma) + \frac{1}{2}(r - \gamma)' \nabla^2 \tau_m(\bar{r})(r - \gamma)$$
$$= (r - \gamma)' \nabla \tau_m(\gamma) + \frac{1}{2}(r - \gamma)' \nabla^2 \tau_m(\gamma)(r - \gamma) + \frac{1}{2}(r - \gamma)' \left[\nabla^2 \tau_m(\bar{r}) - \nabla^2 \tau_m(\gamma)\right](r - \gamma),$$

and hence by Assumption C7

$$\mathbb{E}[\tau_m(r)] = (r - \gamma)' \mathbb{E}[\nabla \tau_m(\gamma)] + \frac{1}{2}(r - \gamma)' \mathbb{E}[\nabla^2 \tau_m(\gamma)](r - \gamma) + o(\|r - \gamma\|^2)$$
$$= \frac{1}{2}(r - \gamma)' \mathbb{V}_\gamma (r - \gamma) + o(\|r - \gamma\|^2), \tag{E.4}$$

where the second equality uses the fact that $\mathbb{E}[\nabla \tau_m(\gamma)] = 0$ since $\mathbb{E}[\tau_m(r)]$ is maximized at γ. Then, applying (E.2), (E.3), and (E.4) proves the lemma. $\square$

Proof of Lemma A.5. We can establish a representation for $\mathbb{E}[\mathcal{L}_N^K(r)|Z_m]$ using the same arguments as in proving Lemma A.4, but no longer integrating over Z_m. Specifically, a change of variables $u_{im} = v_{im}/\sigma_N$ gives

$$\mathbb{E}[\mathcal{L}_N^K(r)|Z_m = z_m] = \int \sigma_N^{-2} K_{\sigma_N,\gamma}(v_{im}) B(v_i, v_m, w_i, w_m) S_{im}(r) \mathrm{d}F_{V,W}(v_i, w_i)$$
$$= \int K_\gamma(u_{im}) B(v_m + u_{im}\sigma_N, v_m, w_i, w_m) S_{im}(r) f_{V,W}(v_m + u_{im}\sigma_N, w_i) \mathrm{d}w_i \mathrm{d}u_{im}.$$

The lead term of the $\kappa_\gamma^{\text{th}}$-order Taylor expansion inside the integral around v_m is

$$\begin{aligned}
&\int K_\gamma(u_{im})B(v_m, v_m, w_i, w_m)S_{im}(r)f_{V,W}(v_m, w_i)\mathrm{d}w_i\mathrm{d}u_{im} \\
=&\int B(v_m, v_m, w_i, w_m)S_{im}(r)f_{V,W}(v_m, w_i)\mathrm{d}w_i = \tau_m(r)
\end{aligned}$$

by $\int K_\gamma(u)du = 1$, and the sample average of the bias term is of order $o(\|r-\gamma\|^2)$.

Then, applying Lemma A.4 and Assumption C7, we can write

$$\begin{aligned}
&\frac{2}{N}\sum_m \mathbb{E}[\mathcal{L}_N^K(r)|Z_m] - 2\mathbb{E}[\mathcal{L}_N^K(r)] \\
=&\frac{2}{N}\sum_m \tau_m(r) - (r-\gamma)'\mathbb{V}_\gamma(r-\gamma) + o_P(\|r-\gamma\|^2) \\
=&\frac{2}{N}\sum_m (r-\gamma)'\nabla\tau_m(\gamma) + \frac{1}{N}\sum_m (r-\gamma)'\left(\nabla^2\tau_m(\gamma) - \mathbb{V}_\gamma\right)(r-\gamma) + o_P(\|r-\gamma\|^2).
\end{aligned}$$

Then the desired result follows by noticing that $N^{-1}\sum_m \nabla^2\tau_m(\gamma) - \mathbb{V}_\gamma = o_P(1)$ by the SLLN. $\square$

Proof of Lemma A.6. By construction, we can write

$$\rho_N(r) = \frac{1}{N(N-1)}\sum_{i\neq m}\rho_{im}(r),$$

where $\rho_{im}(r) \equiv (K_{\sigma_N,\gamma}(V_{im})h_{im}(r) - 2N^{-1}\sum_i \mathbb{E}[K_{\sigma_N,\gamma}(V_{im})h_{im}(r)|Z_i] + \mathbb{E}[K_{\sigma_N,\gamma}(V_{im})h_{im}(r)])/\sigma_N^2$.

Note that $\rho_{im}(\gamma) = 0$ and $|\rho_{im}(r)|$ is bounded by a multiple of M/σ_N^2 where M is a positive constant. Define $\mathcal{F}_N^* = \{\rho_{im}^*(r)|r\in\mathcal{R}\}$ where $\rho_{im}^*(r) \equiv \sigma_N^2\rho_{im}(r)/M$. Then the Euclidean properties of the (P-degenerate) class of functions $\mathcal{F}_N^*$ are deduced using similar arguments for proving Lemma A.2 in combination with Corollary 17 and Corollary 21 in Nolan and Pollard (1987). As $\int K_{\sigma_N,\gamma}(v)^p\mathrm{d}v = O(\sigma_N^2)$ for $p = 1, 2$, we have $\sup_{r\in\mathcal{R}_N}\mathbb{E}[\rho_{im}^*(r)^2] = O(\sigma_N^2)$. Then applying Theorem 3 of Sherman (1994b) gives

$$\frac{1}{N(N-1)}\sum_{i\neq m}\rho_{im}^*(r) = O_P(N^{-1}\sigma_N^\alpha),$$

where $0 < \alpha < 1$ and hence $q_N(r) = O_P(N^{-1}\sigma_N^{\alpha-2}) = O_P(N^{-1}\sigma_N^{-2})$. $\square$

Proof of Lemma A.7. The first step is to plug (A.5) into (A.4) so that $\Delta\mathcal{L}_N^K(r)$ can be expressed as

$$\frac{1}{\sigma_N^3 N(N-1)(N-2)}\sum_{i\neq m\neq l}[\nabla_1 K_{\sigma_N,\gamma}(V_{im})X_{im1}'\psi_{l,\beta} + \nabla_2 K_{\sigma_N,\gamma}(V_{im})X_{im2}'\psi_{l,\beta}]h_{im}(r) \tag{E.5}$$

plus a remainder term of higher order since $\sqrt{N}\sigma_N \to \infty$. (E.5) is a third order U-process. We use the U-statistic decomposition found in Sherman (1994b) (see also Serfling (2009)) to derive a linear representation. Note that (E.5) has unconditional mean zero, so as is its mean conditional on each of its first two arguments. Furthermore, by Theorem 3 of Sherman (1994b) and similar arguments for proving Lemma A.6, the remainder term of the decomposition (projection) is of order $O_P(N^{-1}\sigma_N^{-2})$. Hence it suffices to derive a linear representation for its mean conditional on its third argument:

$$\frac{1}{\sigma_N^3 N}\sum_l \int B(v_i, v_m, w_i, w_m) S_{im}(r) \nabla K_{\sigma_N,\gamma}(v_{im})' \begin{pmatrix} x'_{im1}\psi_{l,\beta} \\ x'_{im2}\psi_{l,\beta} \end{pmatrix} \mathrm{d}F_{X,W}(x_i, w_i)\mathrm{d}F_{X,W}(x_m, w_m).$$

where $F_{X,W}(\cdot,\cdot)$ denotes the joint distribution function of (X_i, W_i). While the above integral is expressed w.r.t. (x_i, w_i) and (x_m, w_m), it will prove convenient to express the integral in terms of ν_i and ν_m. We do so as follows:

$$\frac{1}{\sigma_N^3 N}\sum_l \left(\int \nabla K_{\sigma_N,\gamma}(v_{im})' G(v_i, v_m, r) f_V(v_i) f_V(v_m)\mathrm{d}v_i \mathrm{d}v_m\right)\psi_{l,\beta}. \tag{E.6}$$

Apply a change of variables in (E.6) with $u_{im} = v_{im}/\sigma_N$ to obtain

$$\begin{aligned}&\frac{1}{\sigma_N^3 N}\sum_l \left(\int \nabla K_{\sigma_N,\gamma}(v_{im})' G(v_i, v_m, r) f_V(v_i) f_V(v_m)\mathrm{d}v_i \mathrm{d}v_m\right)\psi_{l,\beta} \\ =&\frac{1}{\sigma_N N}\sum_l \left(\int \nabla K_{\gamma}(u_{im})' G(v_m + u_{im}\sigma_N, v_m, r) f_V(v_m + u_{im}\sigma_N) f_V(v_m)\mathrm{d}u_{im} \mathrm{d}v_m\right)\psi_{l,\beta}.\end{aligned}$$

As $\int \nabla K_\gamma(u)'\mathrm{d}u = 0$ and $\int u_j \nabla_j K_\gamma(u)\mathrm{d}u = -1$ for $j = 1, 2$, a second-order expansion inside the integral around $u_{im}\sigma_N = 0$ yields the lead term[18]

$$-\frac{1}{N}\sum_l \left(\int \nabla_1 \mu(v_m, v_m, r) f_V(v_m)\mathrm{d}v_m\right)\psi_{l,\beta}.$$

We next apply a second-order Taylor expansion of $\nabla_1\mu(v_m, v_m, r)$ around γ to yield

$$(r-\gamma)'\frac{1}{N}\sum_l \left(-\int \nabla_{13}^2 \mu(v_m, v_m, \gamma) f_V(v_m)\mathrm{d}v_m\right)\psi_{l,\beta} + o_P(\|r-\gamma\|^2),$$

which concludes the proof of this lemma. □

Proof of Lemma B.1. We verify the conditions in Assumption M, specifically, M.i, M.ii, and M.iii, in Seo and Otsu (2018) one by one. Recall that

$$\begin{aligned}\phi_{Ni}(b) \equiv \sum_{t>s}\sum_{d\in\mathcal{D}}\{&\mathcal{K}_{h_N}(X_{i2ts}, W_{its}) Y_{idst}(-1)^{d_1}\left(1[X'_{i1ts}b > 0] - 1[X'_{i1ts}\beta > 0]\right) \\ &+ \mathcal{K}_{h_N}(X_{i1ts}, W_{its}) Y_{idst}(-1)^{d_2}\left(1[X'_{i2st}b > 0] - 1[X'_{i2st}\beta > 0]\right)\},\end{aligned}$$

[18]The second-order term is of order $O_P(\delta_N\sigma_N/\sqrt{N})$, which will be $o_P(\|r-\gamma\|^2)$.

The "h_n" in Seo and Otsu (2018) needs to be set as $h_N^{k_1+k_2}$ for the term $\phi_{Ni}(b)$. Similarly, the "h_n" in Seo and Otsu (2018) needs to be set as $\sigma_N^{2k_1}$ for the term $\varphi_{Ni}(r)$. We conduct the analysis for $\phi_{Ni}(b)$ first, and the results for $\varphi_{Ni}(r)$ follow similarly. For $\phi_{Ni}(b)$, we analyze the term

$$\mathcal{K}_{h_N}(X_{i2ts}, W_{its})Y_{idst}(-1)^{d_1}\left(1[X_{i1ts}'b > 0] - 1[X_{i1ts}'\beta > 0]\right),$$

and the remaining terms can be analyzed similarly, thanks to the similar structure of those terms.

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

$$\begin{aligned}
&\mathbb{E}[\mathcal{K}_{h_N}(X_{i2ts}, W_{its})Y_{idst}(-1)^{d_1}\left(1\left[X_{i1ts}'b > 0\right] - 1\left[X_{i1ts}'\beta > 0\right]\right)] \\
=&\int_{\mathbb{R}^{k_2}}\int_{\mathbb{R}^{k_1}}\mathbb{E}[Y_{idst}(-1)^{d_1}\left(1\left[X_{i1ts}'b > 0\right] - 1\left[X_{i1ts}'\beta > 0\right]\right)|X_{i2ts} = x_2, W_{its} = w] \\
&\cdot\mathcal{K}_{h_N}(x_2, w)f_{X_{2ts},W_{ts}}(x_2, w)\,\mathrm{d}x_2\mathrm{d}w \\
=&\mathbb{E}[Y_{idst}(-1)^{d_1}\left(1\left[X_{i1ts}'b > 0\right] - 1\left[X_{i1ts}'\beta > 0\right]\right)|X_{i2ts} = 0, W_{its} = 0]f_{X_{2ts},W_{ts}}(0,0) + O(h_N^2),
\end{aligned} \tag{E.7}$$

where the second equality holds by change of variables and Taylor expansion, and the small order term holds by the properties of the kernel function and the boundedness conditions imposed in Assumptions P5–P7. Further, by Assumption P8, the small order term satisfies $O(h_N^2) = o((Nh_N^{k_1+k_2})^{-2/3})$. As a result, we only need to focus on the lead term

$$\begin{aligned}
&\mathbb{E}[Y_{idst}(-1)^{d_1}\left(1\left[X_{i1ts}'b > 0\right] - 1\left[X_{i1ts}'\beta > 0\right]\right)|X_{i2ts} = 0, W_{its} = 0] \\
=&\mathbb{E}[\mathbb{E}[Y_{idst}(-1)^{d_1}|X_{i1ts}, X_{i2ts} = 0, W_{its} = 0]\left(1\left[X_{i1ts}'b > 0\right] - 1\left[X_{i1ts}'\beta > 0\right]\right)|X_{i2ts} = 0, W_{its} = 0] \\
\equiv&\mathbb{E}[\kappa_{dts}^{(1)}(X_{i1ts})\left(1\left[X_{i1ts}'b > 0\right] - 1\left[X_{i1ts}'\beta > 0\right]\right)|X_{i2ts} = 0, W_{its} = 0] \\
=&\int_{\mathbb{R}^{k_1}}\kappa_{dts}^{(1)}(x)\left(1\left[x'b > 0\right] - 1\left[x'\beta > 0\right]\right)f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}(x)\,\mathrm{d}x \\
\equiv&\int_{\mathbb{R}^{k_1}}\kappa_{dts}^{(1)}(x)\left(1\left[x'b > 0\right] - 1\left[x'\beta > 0\right]\right)f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}(x)\,\mathrm{d}x,
\end{aligned} \tag{E.8}$$

where to ease notations we denote

$$\kappa_{dts}^{(1)}(x) = \mathbb{E}[Y_{idst}(-1)^{d_1}|X_{i1ts} = x, X_{i2ts} = 0, W_{its} = 0]$$

in the third line.

We now get the first and second derivatives of the above term w.r.t. b around β. Since the calculation is the classical differential geometry and very similar to that in Sections 5 and 6.4 of Kim and Pollard (1990) and Section B.1 of Seo and Otsu (2018), we only present key steps and omit some standard similar details. First, define the mapping

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

Then T_b maps $\{X_{i1ts}'b > 0\}$ onto $\{X_{i1ts}'\beta > 0\}$ and $\{X_{i1ts}'b = 0\}$ onto $\{X_{i1ts}'\beta = 0\}$. With equation

(E.8), equations (5.2) and (5.3) in Kim and Pollard (1990) imply

$$\begin{aligned}
&\frac{\partial}{\partial b}\mathbb{E}\left[Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'b>0\right]-1\left[X_{i1ts}'\beta>0\right]\right)\middle| X_{i2ts}=0,W_{its}=0\right]\\
&=\frac{\partial}{\partial b}\int_{\mathbb{R}^{k_1}}\kappa_{dts}^{(1)}\left(x\right)\left(1\left[x'b>0\right]-1\left[x'\beta>0\right]\right)f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(x\right)\mathrm{d}x,\\
&=\|b\|_2^{-2}\,b'\beta\left(I-\|b\|_2^{-2}\,bb'\right)\int 1\left[x'\beta=0\right]\kappa_{dts}^{(1)}\left(T_bx\right)xf_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(T_bx\right)\mathrm{d}\sigma_0^{(1)},
\end{aligned}$$

where $\sigma_0^{(1)}$ is the surface measure of $\{X_{i1ts}'\beta=0\}$. Note that $T_\beta x=x$, then

$$1\left[x'\beta=0\right]\kappa_{dts}^{(1)}\left(T_\beta x\right)=1\left[x'\beta=0\right]\kappa_{dts}^{(1)}\left(x\right)=0$$

because

$$\begin{aligned}
\kappa_{dts}^{(1)}\left(X_{i1t}\right)\Big|_{x'\beta=0}&=\mathbb{E}\left[Y_{idst}\left(-1\right)^{d_1}\middle| X_{i1ts}'\beta=0,X_{i2ts}=0,W_{its}=0\right]\\
&=\mathbb{E}\left[\left(Y_{idt}-Y_{ids}\right)\left(-1\right)^{d_1}\middle| X_{i1t}'\beta=X_{i1s}'\beta,X_{i2t}=X_{i2s},W_{it}=W_{is}\right]\\
&=0,
\end{aligned}$$

where the last line holds by Assumption P2 that $\xi_s\overset{d}{=}\xi_t|(\alpha,Z^T)$. This implies that

$$\frac{\partial}{\partial b}\mathbb{E}\left[Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'b>0\right]-1\left[X_{i1ts}'\beta>0\right]\right)\middle| X_{i2ts}=0,W_{its}=0\right]\Bigg|_{b=\beta}=0.$$

For the same reason, the nonzero component of the second order derivative at $b=\beta$ only comes from the derivative of $\kappa_{dts}^{(1)}\left(X_{i1ts}\right)$. Notice that $\frac{\partial\kappa_{dts}^{(1)}(T_bx)}{\partial b}\Big|_{b=\beta}=-\left(\frac{\partial\kappa_{dts}^{(1)}(x)}{\partial x}'\beta\right)x$, we have

$$\begin{aligned}
&\frac{\partial^2}{\partial b\partial b'}\mathbb{E}\left[Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'b>0\right]-1\left[X_{i1ts}'\beta>0\right]\right)\middle| X_{i2ts}=0,W_{its}=0\right]\Bigg|_{b=\beta}\\
&=-\int 1\left[x'\beta=0\right]\left(\frac{\partial\kappa_{dts}^{(1)}\left(x\right)}{\partial x}'\beta\right)f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(x\right)xx'\mathrm{d}\sigma_0^{(1)}\equiv\mathbb{V}_{dts}^{(1)}.
\end{aligned}$$

With the first and second derivatives obtained, we can move on to apply the Taylor expansion to the expectation term in equation (E.8) around $b=\beta$ as

$$\begin{aligned}
&\mathbb{E}\left[Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'b>0\right]-1\left[X_{i1ts}'\beta>0\right]\right)\middle| X_{i2ts}=0,W_{its}=0\right]\\
=&\frac{1}{2}\left(b-\beta\right)'\mathbb{V}_{dts}^{(1)}\left(b-\beta\right)+o\left(\|b-\beta\|^2\right).
\end{aligned}$$

Substitute this back to equation (E.7), we have

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{K}_{h_N}(X_{i2ts},W_{its})Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'b>0\right]-1\left[X_{i1ts}'\beta>0\right]\right)\right]\\
=&\frac{1}{2}\left(b-\beta\right)'\mathbb{V}_{dts}^{(1)}\left(b-\beta\right)+o\left(\|b-\beta\|^2\right)+o\left(\left(Nh_N^{k_1+k_2}\right)^{-2/3}\right).
\end{aligned}$$

We can similarly define

$$\mathbb{V}_{dts}^{(2)} \equiv -\int 1\left[x'\beta=0\right]\left(\frac{\partial \kappa_{dts}^{(2)}\left(x\right)}{\partial x}'\beta\right) f_{X_{2ts}|\{X_{1ts}=0,W_{ts}=0\}}\left(x\right)xx'\mathrm{d}\sigma_0^{(2)},$$

where

$$\kappa_{dts}^{(2)}\left(x\right) \equiv \mathbb{E}\left[Y_{idst}\left(-1\right)^{d_2}\middle| X_{i2ts}=x, X_{i1ts}=0, W_{its}=0\right],$$

and $\sigma_0^{(2)}$ is the surface measure of $\{X_{i2ts}'\beta=0\}$. A similar derivation yields

$$\begin{aligned}&\mathbb{E}\left[\mathcal{K}_{h_N}(X_{i1ts},W_{its})Y_{idst}\left(-1\right)^{d_2}\left(1\left[X_{i2st}'b>0\right]-1\left[X_{i2st}'\beta>0\right]\right)\right]\\=&\frac{1}{2}\left(b-\beta\right)'\mathbb{V}_{dts}^{(2)}\left(b-\beta\right)+o\left(\|b-\beta\|^2\right)+o\left(\left(Nh_N^{k_1+k_2}\right)^{-2/3}\right).\end{aligned}$$

Put everything so far together, we have

$$\mathbb{E}\left[\phi_{Ni}\left(b\right)\right]=\frac{1}{2}\left(b-\beta\right)'\left[\sum_{t>s}\sum_{d\in\mathcal{D}}\left(\mathbb{V}_{dts}^{(1)}+\mathbb{V}_{dts}^{(2)}\right)\right]\left(b-\beta\right)+o(\|b-\beta\|^2)+o((Nh_N^{k_1+k_2})^{-2/3}). \quad \text{(E.9)}$$

Assumption M.i in Seo and Otsu (2018) is then verified by

$$\begin{aligned}\mathbb{V}\equiv&\sum_{t>s}\sum_{d\in\mathcal{D}}\left(\mathbb{V}_{dts}^{(1)}+\mathbb{V}_{dts}^{(2)}\right).\\=&-\sum_{t>s}\sum_{d\in\mathcal{D}}\left\{\int 1\left[x'\beta=0\right]\left(\frac{\partial\kappa_{dts}^{(1)}\left(x\right)}{\partial x}'\beta\right)f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(x\right)xx'\mathrm{d}\sigma_0^{(1)}\right.\\&\left.+\int 1\left[x'\beta=0\right]\left(\frac{\partial\kappa_{dts}^{(2)}\left(x\right)}{\partial x}'\beta\right)f_{X_{2ts}|\{X_{1ts}=0,W_{ts}=0\}}\left(x\right)xx'\mathrm{d}\sigma_0^{(2)}\right\}.\end{aligned}$$

We can obtain similar results for $\varphi_{Ni}\left(r\right)$ using exactly the same line of analysis, i.e.,

$$\mathbb{E}\left[\varphi_{Ni}\left(r\right)\right]=\frac{1}{2}\left(r-\gamma\right)'\mathbb{W}\left(r-\gamma\right)+o(\|r-\gamma\|^2)+o((N\sigma_N^{2k_1})^{-2/3}), \quad \text{(E.10)}$$

where

$$\mathbb{W}\equiv-\sum_{t>s}\int 1\left[w'\gamma=0\right]\left(\frac{\partial\kappa_{(1,1)ts}^{(3)}\left(w\right)}{\partial w}'\gamma\right)f_{W_{ts}|\{X_{1ts}=0,X_{2ts}=0\}}\left(w\right)ww'\mathrm{d}\sigma_0$$

with

$$\kappa_{(1,1)ts}^{(3)}\left(w\right)\equiv\mathbb{E}\left[Y_{i(1,1)ts}|W_{its}=w,X_{i1ts}=0,X_{i2ts}=0\right]$$

and $\sigma_0^{(3)}$ being the surface measure of $\{W_{its}'\gamma=0\}$.

On Assumption M.ii in Seo and Otsu (2018). Again, we first verify this condition for

$$\mathcal{K}_{h_N}(X_{i2ts},W_{its})Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'b>0\right]-1\left[X_{i1ts}'\beta>0\right]\right),$$

and other components in $\phi_{Ni}(b)$ follow similarly. We evaluate norm of the difference of the above term at $b = b_1$ and $b = b_2$ multiplied by $h_N^{(k_1+k_2)/2}$:

$$
\begin{aligned}
& h_N^{(k_1+k_2)/2} \left\| \mathcal{K}_{h_N}(X_{i2ts}, W_{its}) Y_{idst} (-1)^{d_1} \left(1\left[X_{i1ts}' b_1 > 0\right] - 1\left[X_{i1ts}' b_2 > 0\right]\right) \right\| \\
= & \left\{ \mathbb{E}\left[\mathbb{E}\left[h_N^{k_1+k_2} \mathcal{K}_{h_N}^2(X_{i2ts}, W_{its}) Y_{idst}^2 \middle| X_{i1ts} \right] \left(1\left[X_{i1ts}' b_1 > 0\right] - 1\left[X_{i1ts}' b_2 > 0\right]\right)^2 \right] \right\}^{1/2} \\
= & \left\{ \mathbb{E}\left[\mathbb{E}\left[h_N^{k_1+k_2} \mathcal{K}_{h_N}^2(X_{i2ts}, W_{its}) Y_{idst}^2 \middle| X_{i1ts} \right] \left|1\left[X_{i1ts}' b_1 > 0\right] - 1\left[X_{i1ts}' b_2 > 0\right]\right| \right] \right\}^{1/2} \qquad \text{(E.11)} \\
\geq & C_1 \mathbb{E}\left[\left|1\left[X_{i1ts}' b_1 > 0\right] - 1\left[X_{i1ts}' b_2 > 0\right]\right| \right] \geq C_2 \left\| b - \beta \right\|_2,
\end{aligned}
$$

where the second equality holds by the fact that the difference of two indicator functions can only take values $-1, 0,$ or 1. C_1 and C_2 are two positive constants. Applying the same analysis to all other terms in $\phi_{Ni}(b)$, we have

$$
h_N^{(k_1+k_2)/2} \left\| \phi_{Ni}(b_1) - \phi_{Ni}(b_2) \right\| \geq C_3 \left\| b_1 - b_2 \right\|
$$

for some positive constant C_3.

Applying similar analysis to $\varphi_{Ni}(r)$ leads to

$$
\sigma_N^{k_1} \left\| \varphi_{Ni}(r_1) - \varphi_{Ni}(r_2) \right\| \geq C_4 \left\| r_1 - r_2 \right\|
$$

for some positive constant C_4.

On Assumption M.iii in Seo and Otsu (2018). We still begin with the analysis on

$$
\mathcal{K}_{h_N}(X_{i2ts}, W_{its}) Y_{idst} (-1)^{d_1} \left(1\left[X_{i1ts}' b > 0\right] - 1\left[X_{i1ts}' \beta > 0\right]\right),
$$

and other components in $\phi_{Ni}(b)$ follow similarly. We evaluate the square of difference of the above term at $b = b_1$ and $b = b_2$, such that $\|b_1 - b_2\| < \varepsilon$, multiplied by $h_N^{k_1+k_2}$:

$$
\begin{aligned}
& h_N^{k_1+k_2} \mathbb{E}\left[\sup_{b_1, b_2 \in \mathcal{B}, \|b_1 - b_2\| < \varepsilon} \left[\mathcal{K}_{h_N}(X_{i2ts}, W_{its}) Y_{idst} (-1)^{d_1} \left(1\left[X_{i1ts}' b_1 > 0\right] - 1\left[X_{i1ts}' b_2 > 0\right]\right) \right]^2 \right] \\
\leq & \mathbb{E}\left[\mathbb{E}\left[h_N^{k_1+k_2} \mathcal{K}_{h_N}^2(X_{i2ts}, W_{its}) Y_{idst}^2 \middle| X_{i1ts} \right] \sup_{b_1, b_2 \in \mathcal{B}, \|b_1 - b_2\| < \varepsilon} \left|1\left[X_{i1ts}' b_1 > 0\right] - 1\left[X_{i1ts}' b_2 > 0\right]\right| \right] \\
\leq & C_5 \mathbb{E}\left[\sup_{b_1, b_2 \in \mathcal{B}, \|b_1 - b_2\| < \varepsilon} \left|1\left[X_{i1ts}' b_1 > 0\right] - 1\left[X_{i1ts}' b_2 > 0\right]\right| \right] \leq C_6 \varepsilon,
\end{aligned}
$$

where the second line follows by the fact that the maximum of absolute value of the difference of two indicator functions is 1, and the last line holds by Assumptions P4 and P5. Apply the same analysis to all other terms in $\phi_{Ni}(b)$ to obtain

$$
h_N^{k_1+k_2} \mathbb{E}\left[\sup_{b_1, b_2 \in \mathcal{B}, \|b_1 - b_2\| < \varepsilon} \left[\phi_{Ni}(b_1) - \phi_{Ni}(b_2)\right]^2 \right] \leq C_7 \varepsilon
$$

for some positive constant C_7.

Similar analysis on $\varphi_{Ni}(r)$ leads to

$$h_N^{2k_1}\mathbb{E}\left[\sup_{r_1,r_2\in\mathcal{R},\|r_1-r_2\|<\varepsilon}\left[\varphi_{Ni}(r_1)-\varphi_{Ni}(r_2)\right]^2\right]\le C_8\varepsilon$$

for some positive constant C_8. □

Proof of Lemma B.2. The first two equalities are direct results from Lemma B.1 by Taylor expansions. Taking $b=\beta+\rho(Nh_N^{k_1+k_2})^{-1/3}$ in equation (E.9) yields

$$\left(Nh_N^{k_1+k_2}\right)^{2/3}\mathbb{E}\left[\phi_{Ni}\left(\beta+\rho\left(Nh_N^{k_1+k_2}\right)^{-1/3}\right)\right]=\frac{1}{2}\rho'\mathbb{V}\rho+o(1)\to\frac{1}{2}\rho'\mathbb{V}\rho.$$

Similarly, setting $r=\gamma+\delta(N\sigma_N^{2k_1})^{-1/3}$ in equation (E.10) gives

$$\left(N\sigma_N^{2k_1}\right)^{2/3}\mathbb{E}\left[\varphi_{Ni}\left(\gamma+\delta\left(N\sigma_N^{2k_1}\right)^{-1/3}\right)\right]=\frac{1}{2}\delta'\mathbb{W}\delta+o(1)\to\frac{1}{2}\delta'\mathbb{W}\delta.$$

$\mathbb{H}_1$ and $\mathbb{H}_2$ can be obtained in the same way as in Kim and Pollard (1990). We omit some similar yet tedious details and refer interested readers to Section 6.4 in Kim and Pollard (1990). To get $\mathbb{H}_1$, we let $\upsilon_N\equiv(Nh_N^{k_1+k_2})^{1/3}$ and define

$$\begin{aligned}\mathbb{L}(\rho_1-\rho_2)&\equiv\lim_{N\to\infty}\upsilon_N\mathbb{E}\left\{h_N^{k_1+k_2}\left[\phi_{Ni}\left(\beta+\rho_1\upsilon_N^{-1}\right)-\phi_{Ni}\left(\beta+\rho_2\upsilon_N^{-1}\right)\right]^2\right\},\\ \mathbb{L}(\rho_1)&\equiv\lim_{N\to\infty}\upsilon_N\mathbb{E}\left\{h_N^{k_1+k_2}\left[\phi_{Ni}\left(\beta+\rho_1\upsilon_N^{-1}\right)-\phi_{Ni}(\beta)\right]^2\right\},\ \text{and}\\ \mathbb{L}(\rho_2)&\equiv\lim_{N\to\infty}\upsilon_N\mathbb{E}\left\{h_N^{k_1+k_2}\left[\phi_{Ni}\left(\beta+\rho_2\upsilon_N^{-1}\right)-\phi_{Ni}(\beta)\right]^2\right\}.\end{aligned}$$

Since $\phi_{Ni}(\beta)=0$,

$$\mathbb{H}_1(\rho_1,\rho_2)=\frac{1}{2}\left[\mathbb{L}(\rho_1)+\mathbb{L}(\rho_2)-\mathbb{L}(\rho_1-\rho_2)\right].\tag{E.12}$$

We calculate $\mathbb{L}$ as follows. Notice that

$$\begin{aligned}&\mathbb{E}\left[h_N^{k_1+k_2}\mathcal{K}_{h_N}(X_{ijts},W_{its})^2\right]\\ =&\int\frac{1}{h_N^{k_1+k_2}}\left[\Pi_{\iota=1}^{k_1}K\left(\frac{X_{ijts,\iota}}{h_N}\right)\Pi_{\iota=1}^{k_2}K\left(\frac{W_{its,\iota}}{h_N}\right)\right]^2\mathrm{d}F_{X_{ijts},W_{its}}=O(1),\\ &\mathbb{E}\left[h_N^{k_1+k_2}\mathcal{K}_{h_N}(X_{ijts},W_{its})\mathcal{K}_{h_N}(X_{ij't's'},W_{it's'})\right]\\ =&\int\int\frac{1}{h_N^{k_1+k_2}}\Pi_{\iota=1}^{k_1}K\left(\frac{X_{ijts,\iota}}{h_N}\right)\Pi_{\iota=1}^{k_2}K\left(\frac{W_{its,\iota}}{h_N}\right)\\ &\cdot\Pi_{\iota=1}^{k_1}K\left(\frac{X_{ij't's',\iota}}{h_N}\right)\Pi_{\iota=1}^{k_2}K\left(\frac{W_{it's',\iota}}{h_N}\right)\mathrm{d}F_{X_{ijts},W_{its}}\mathrm{d}F_{X_{ij't's'},W_{it's'}}\\ =&o(1),\ \text{for any }\ (j,t,s)\neq\left(j',t',s'\right),\end{aligned}$$

and

$$
\begin{aligned}
&\phi_{Ni}\left(\beta+\rho_1\upsilon_N^{-1}\right)-\phi_{Ni}\left(\beta+\rho_2\upsilon_N^{-1}\right)\\
=&\sum_{t>s}\sum_{d\in\mathcal{D}}\Big\{\mathcal{K}_{h_N}(X_{i2ts},W_{its})Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i1ts}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right)\\
&+\mathcal{K}_{h_N}(X_{i1ts},W_{its})Y_{idst}\left(-1\right)^{d_2}\left(1\left[X_{i2st}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i2st}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right)\Big\}.
\end{aligned}
$$

Thus, the interaction terms after expanding the square in $\mathbb{L}$'s are asymptotically negligible, and we only need to focus on the squares of each term above.

We now derive the square of the first term multiplied by $h_N^{k_1+k_2}$ and the rest follow similarly.

$$
\begin{aligned}
&\mathbb{E}\left\{h_N^{k_1+k_2}\left[\mathcal{K}_{h_N}(X_{i2ts},W_{its})Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i1ts}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right)\right]^2\right\}\\
=&\mathbb{E}\left\{\mathbb{E}\left[\left|Y_{idst}\right|\left|1\left[X_{i1ts}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i1ts}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right|\;\middle|\;X_{i2ts},W_{its}\right]h_N^{k_1+k_2}\mathcal{K}_{h_N}^2(X_{i2ts},W_{its})\right\}\\
=&\mathbb{E}\left[\left|Y_{idst}\right|\left|1\left[X_{i1ts}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i1ts}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right|\;\middle|\;X_{i2ts}=0,W_{its}=0\right]\\
&\cdot f_{X_{2ts},W_{ts}}\left(0,0\right)\left[\int K^2\left(u\right)\mathrm{d}u\right]^{k_1+k_2}+O\left(h_N\right)\\
=&\mathbb{E}\left\{\mathbb{E}\left[\left|Y_{idst}\right|\;\left|X_{i1ts},X_{i2ts}=0,W_{its}=0\right.\right]\right.\\
&\cdot\left|1\left[X_{i1ts}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i1ts}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right|\;\left|\;X_{i2ts}=0,W_{its}=0\right\}\\
&\cdot f_{X_{2ts},W_{ts}}\left(0,0\right)\left[\int K^2\left(u\right)\mathrm{d}u\right]^{k_1+k_2}+O\left(h_N\right)\\
=&\int\kappa_{dts}^{(4)}\left(x\right)\left|1\left[x'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[x'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right|f_{X_{1ts}|\{X_{2ts}=0,W_{its}=0\}}\left(x\right)\mathrm{d}x\\
&\cdot f_{X_{2ts},W_{ts}}\left(0,0\right)\bar{K}_2^{k_1+k_2}+O\left(h_N\right),
\end{aligned}
$$

where $\kappa_{dts}^{(4)}\left(x\right)\equiv\mathbb{E}\left[\left|Y_{idst}\right|\;\left|X_{i1ts}=x,X_{i2ts}=0,W_{its}=0\right.\right]$ and $\bar{K}_2\equiv\int K^2\left(u\right)\mathrm{d}u$. By Assumption P9, $\upsilon_N h_N=(Nh_N^{k_1+k_2})^{1/3}h_N\to 0$, so the bias term is asymptotic negligible. To calculate the above integral, we follow Kim and Pollard (1990) and decompose $X_{1ts}=a\beta+\bar{X}_{1ts}$, with $\bar{X}_{1ts}$ orthogonal to β. We use $f_{X_{1ts}}\left(a,\bar{x}\right)$ to denote the density of X_{1ts} at $X_{1ts}=a\beta+\bar{x}$. Using the results in Kim and Pollard (1990),

$$
\begin{aligned}
&\lim_{N\to\infty}\upsilon_N\cdot\int\kappa_{dts}^{(4)}\left(x\right)\left|1\left[x'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[x'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right|f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(x\right)\mathrm{d}x\\
=&\int\kappa_{dts}^{(4)}\left(\bar{x}\right)\left|\bar{x}'\left(\rho_1-\rho_2\right)\right|f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(0,\bar{x}\right)\mathrm{d}\bar{x}.
\end{aligned}
$$

To summarize, we have

$$
\begin{aligned}
&\lim_{N\to\infty}\upsilon_N\mathbb{E}\left\{h_N^{k_1+k_2}\left[\mathcal{K}_{h_N}(X_{i2ts},W_{its})Y_{idst}\left(-1\right)^{d_1}\left(1\left[X_{i1ts}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i1ts}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right)\right]^2\right\}\\
=&\int\kappa_{dts}^{(4)}\left(\bar{x}\right)\left|\bar{x}'\left(\rho_1-\rho_2\right)\right|f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(0,\bar{x}\right)\mathrm{d}\bar{x}f_{X_{2ts},W_{ts}}\left(0,0\right)\bar{K}_2^{k_1+k_2}.
\end{aligned}
$$

Similarly, we have for the second term

$$\lim_{N\to\infty}\upsilon_N\mathbb{E}\left\{h_N^{k_1+k_2}\left[\mathcal{K}_{h_N}(X_{i1ts},W_{its})Y_{idst}\left(-1\right)^{d_2}\left(1\left[X_{i2st}'\left(\beta+\rho_1\upsilon_N^{-1}\right)>0\right]-1\left[X_{i2st}'\left(\beta+\rho_2\upsilon_N^{-1}\right)>0\right]\right)\right]^2\right\}$$
$$=\int\kappa_{dts}^{(5)}\left(\bar{x}\right)\left|\bar{x}'\left(\rho_1-\rho_2\right)\right|f_{X_{2ts}|\{X_{1ts}=0,W_{ts}=0\}}\left(0,\bar{x}\right)\mathrm{d}\bar{x}f_{X_{1ts},W_{ts}}\left(0,0\right)\bar{K}_2^{k_1+k_2},$$

where

$$\kappa_{dts}^{(5)}\left(x\right)\equiv\mathbb{E}\left[\left|Y_{idst}\right|\ \left|X_{i2ts}=x,X_{i1ts}=0,W_{its}=0\right]\right..$$

Therefore,

$$\begin{aligned}\mathbb{L}\left(\rho_1-\rho_2\right)&=\lim_{N\to\infty}\upsilon_N\mathbb{E}\left\{h_N^{k_1+k_2}\left[\phi_{Ni}\left(\beta+\rho_1\upsilon_N^{-1}\right)-\phi_{Ni}\left(\beta+\rho_2\upsilon_N^{-1}\right)\right]^2\right\}\\&=\sum_{t>s}\sum_{d\in\mathcal{D}}\left\{\int\kappa_{dts}^{(4)}\left(\bar{x}\right)\left|\bar{x}'\left(\rho_1-\rho_2\right)\right|f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(0,\bar{x}\right)\mathrm{d}\bar{x}f_{X_{2ts},W_{ts}}\left(0,0\right)\right.\\&\quad\left.+\int\kappa_{dts}^{(5)}\left(\bar{x}\right)\left|\bar{x}'\left(\rho_1-\rho_2\right)\right|f_{X_{2ts}|\{X_{1ts}=0,W_{ts}=0\}}\left(0,\bar{x}\right)\mathrm{d}\bar{x}f_{X_{1ts},W_{ts}}\left(0,0\right)\right\}\bar{K}_2^{k_1+k_2}.\end{aligned}$$

Finally, by equation (E.12),

$$\begin{aligned}&\mathbb{H}_1\left(\rho_1,\rho_2\right)\\=&\frac{1}{2}\sum_{t>s}\sum_{d\in\mathcal{D}}\left\{\int\kappa_{dts}^{(4)}\left(\bar{x}\right)\left[\left|\bar{x}'\rho_1\right|+\left|\bar{x}'\rho_2\right|-\left|\bar{x}'\left(\rho_1-\rho_2\right)\right|\right]f_{X_{1ts}|\{X_{2ts}=0,W_{ts}=0\}}\left(0,\bar{x}\right)\mathrm{d}\bar{x}f_{X_{2ts},W_{ts}}\left(0,0\right)\right.\\&\left.+\int\kappa_{dts}^{(5)}\left(\bar{x}\right)\left[\left|\bar{x}'\rho_1\right|+\left|\bar{x}'\rho_2\right|-\left|\bar{x}'\left(\rho_1-\rho_2\right)\right|\right]f_{X_{2ts}|\{X_{1ts}=0,W_{ts}=0\}}\left(0,\bar{x}\right)\mathrm{d}\bar{x}f_{X_{1ts},W_{ts}}\left(0,0\right)\right\}\bar{K}_2^{k_1+k_2}.\end{aligned}$$

The same arguments can be applied to obtain $\mathbb{H}_2$ as

$$\begin{aligned}\mathbb{H}_2\left(\delta_1,\delta_2\right)=&\frac{1}{2}\sum_{t>s}\int\kappa_{(1,1)ts}^{(6)}\left(\bar{w}\right)\left[\left|\bar{w}'\delta_1\right|+\left|\bar{w}'\delta_2\right|-\left|\bar{w}'\left(\delta_1-\delta_2\right)\right|\right]f_{W_{ts}|\{X_{1ts}=0,X_{2ts}=0\}}\left(0,\bar{w}\right)\mathrm{d}\bar{w}\\&\cdot f_{X_{1ts},X_{2ts}}\left(0,0\right)\bar{K}_2^{2k_1},\end{aligned}$$

where $W_{ts}=a\gamma+\bar{W}_{ts}$ with $\bar{W}_{ts}$ orthogonal to γ, $f_{W_{ts}}\left(a,\bar{w}\right)$ denotes the density of W_{ts} at $W_{ts}=a\gamma+\bar{W}_{ts}$, and

$$\kappa_{(1,1)ts}^{(6)}\left(w\right)\equiv\mathbb{E}\left[\left|Y_{i(1,1)ts}\right|\ \left|W_{its}=w,X_{i1ts}=0,X_{i2ts}=0\right]\right..$$

□