Supplement to "Managing Procurement Auction Failure: Bid Requirements or Reserve Prices?"

Jun Ma* Vadim Marmer† Pai Xu‡

This supplement collects the econometric details omitted from the main text. Section S1 provides a more detailed introduction to our econometric results and an extensive review of the related econometric and statistical literature. Section S2 provides details about the generalized method of moments (GMM) estimation method used in the main text. Section S3 discusses the asymptotic normality property of the GMM estimator (Proposition S2). Section S4 discusses the bootstrap method to estimate the asymptotic variance and shows its consistency (Proposition S3). Appendix A gives the important ancillary results which are of independent interest and their proofs. Appendix B collects the proofs of the main results (Propositions S1, S2 and S3).

Notation. For any $m \in \mathbb{N}$, denote $[m] \coloneqq \{1, ..., m\}$. For a finite set A, let $|A|$ denote the number of elements in A. Let $\mathbf{1}_m$ ($\mathbf{0}_m$) denote an m-dimensional vector whose elements are all one (zero). Let $\mathbf{I}_m$ denote the m-dimensional identity matrix.

S1 Introduction and related literature

In this paper, we develop a complete theory for the identification and estimation of the semiparametric model discussed in the conclusion section of Gentry and Li (2014), which has not been studied in the literature, to the best of our knowledge.[1] We derive testable sufficient conditions that ensure (semiparametric) local and global identification of the copula parameter in the sense of Lewbel (2019) and present them in Proposition 3.3 in the main text. Chen et al. (2025) considers the first-price auction model with both endogenous entry and risk aversion under a parametric assumption on the copula function for the signal and the private value. They show that the utility function and the distribution of private values are nonparametrically identified conditional on the copula parameter, which can be set identified. In our paper, we study the point (global) identification of the copula

This version: February 24, 2025.

*School of Economics, Renmin University of China

†Vancouver School of Economics, University of British Columbia

‡Business School, University of Hong Kong

[1] Gentry and Li (2014) considers high-bid auction with selective entry. The econometric part of our paper considers procurement (low-bid) auctions under the 2-bid requirement. However, our results can be extended straightforwardly to accommodate other scenarios, including high-bid auctions (with or without binding reserve prices) and procurement auctions with binding reserve prices.

parameter, and its limited form (local identification), under the risk neutrality assumption and the parametric assumption on the copula function. Chen et al. (2025) also show that the utility function is nonparametrically identified if there is sufficient variation in the observed instruments and number of potential bidders. They also show parametric identification of the utility function. In either case, as a consequence, the distribution of the private values conditional on entry is nonparametrically identified. Adaptation of our identification results provides testable sufficient conditions that ensure (local or global) identification of the copula parameter, in the semiparametric model with endogenous entry, risk aversion and a parametric assumption on the copula function.

We propose a convenient and practical generalized method of moments (GMM) estimator and develop its first-order asymptotic theory. Several intermediate results are of independent interest. Our semiparametric GMM estimation uses the empirical CDF of pseudo values constructed by using a nonparametrically estimated inverse bidding strategy. Guerre et al. (2000) proposes kernel density estimation using these pseudo values. Ma et al. (2019) derives the first-order asymptotic theory of the pseudo-value-based density estimator.[2] The asymptotic properties of the pseudo-value-based cumulative distribution function (CDF) estimator have not been studied in the literature to the best of our knowledge. The econometric theory derived in this paper complements Ma et al. (2019) and fills the void by providing the first-order asymptotic properties for the pseudo-value-based CDF estimator. Our proof is also different from that of Ma et al. (2019). Recently, Zincenko (2024) studies estimation and inference of seller's expected revenue in high-bid first-price auctions. They derive the asymptotic linearization of a pseudo-value-based CDF estimator. However, Zincenko (2024) uses a kernel-smoothed bid CDF estimator in their estimated inverse bidding strategy for constructing the pseudo values. The pseudo values in Guerre et al. (2000) and Ma et al. (2019) are built from using the empirical CDF of the bids without kernel smoothing. This paper follows Guerre et al. (2000) and Ma et al. (2019) since the empirical CDF incurs no smoothing bias. We show that our estimator admits a desired asymptotic linearization by using a proof completely different from Zincenko (2024)'s.

Estimating the inverse bidding strategy requires plugging in a nonparametric estimator of the compactly supported bid density. The standard kernel density estimator suffers from boundary bias. Either trimming (Guerre et al., 2000) or boundary bias correction (see, e.g., Hickman and Hubbard, 2015 or Ma et al., 2021) has been used to address the issue. Boundary correction for kernel density estimators has received much attention in the statistical literature. In this paper, we follow Ma et al. (2021) to use the boundary adaptive local linear density estimator (Lejeune and Sarda, 1992 and Jones, 1993). One of the advantages of the local linear density estimator is that it does not require selecting additional tuning parameters. The local linear density estimator has received attention in the statistical literature (see, e.g., Cheng et al., 1997 and Chen and Huang, 2007). To the best of our knowledge, the uniform convergence property (over the entire support) of this estimator has not been derived in the literature. In this paper, we derive concentration bounds for the more general

[2] See Marmer and Shneyerov (2012) for another estimation method without using the estimated inverse bidding strategy and calculating the pseudo values.

local polynomial (LP) density estimators (Bickel and Doksum, 2015, Chapter 11.3), and also its first and second derivatives, which extend the classical concentration results for kernel density estimators (Giné and Guillou, 2002). Our new results give the uniform rate of convergence of the LP density estimator over the entire support. These results are also useful in other semiparametric estimation problems which involve compactly supported density functions as nuisance parameters.

The optimal weight matrix of the semiparametric GMM estimator depends on the asymptotic variance of the pseudo-value-based CDF estimator, which takes a complicated form. By using bootstrap, we avoid direct estimation of the asymptotic variance. We show that our bootstrap variance estimator consistently estimates the asymptotic variance of the pseudo-value-based CDF estimator.[3] The proof hinges on the concentration bounds for the local linear density estimator derived in Appendix A.

S2 Generalized method of moments estimation

Let

$$\begin{aligned} q_n &:= \mathrm{E}\left[\frac{N_l^*}{N_l} \mid N_l = n\right] \\ &= \frac{\mathrm{E}\left[\mathbb{1}\left(N_l = n\right)\frac{N_l^*}{N_l}\right]}{\mathrm{E}\left[\mathbb{1}\left(N_l = n\right)\right]}, \end{aligned}$$

where the second equality follows from LIE, and then we have $q_n = \varphi_n\left(p_n\right)$. The sample analogue of q_n is thus given by

$$\widehat{q}_n := \frac{1}{n\left|\left\{l : N_l = n\right\}\right|}\sum_{l:N_l=n} N_l^*. \tag{S1}$$

Let $\widehat{p}_n := \varphi_n^{-1}\left(\widehat{q}_n\right)$. A consistent estimator of $G\left(b \mid n\right)$ is

$$\widehat{G}\left(b \mid n\right) := \frac{\sum_{l:N_l=n}\sum_{i=1}^{N_l^*}\mathbb{1}\left(B_{il} \leq b\right)}{\sum_{l:N_l=n} N_l^*}. \tag{S2}$$

Let $\overline{b}_n := \beta\left(\overline{v} \mid p_n, n\right)$ and $\underline{b}_n := \beta\left(\underline{v} \mid p_n, n\right)$ denote the boundary points. Let

$$\begin{aligned} \widehat{\overline{b}}_n &:= \max\left\{B_{il} : i = 1, ..., N_l^*, N_l = n\right\} \\ \widehat{\underline{b}}_n &:= \min\left\{B_{il} : i = 1, ..., N_l^*, N_l = n\right\} \end{aligned}$$

[3] The bootstrap variance estimator may overestimate the asymptotic variance asymptotically, even when the consistency of the bootstrap distribution can be established. See Hahn and Liao (2021) for examples and general theory. Our consistency result excludes such possibility in our specific context. Ma et al. (2019) shows consistency of the bootstrap distribution for the pseudo-value-based density estimator. Their result does not directly imply the consistency of our bootstrap variance estimator.

be the estimated boundary points. The local-linear-type boundary adaptive kernel density estimator of $g(b \mid n)$ is

$$\widehat{g}(b \mid n) \coloneqq \frac{\sum_{l:N_l=n} \sum_{i=1}^{N_l^*} \frac{1}{h} \mathcal{K}_1 \left(B_{il}, b \mid h, \underline{\widehat{b}}_n, \widehat{\overline{b}}_n \right)}{\sum_{l:N_l=n} N_l^*}, \tag{S3}$$

where the function $\mathcal{K}_1$ is defined in Appendix A. Then we construct the nonparametric estimator $\widehat{\xi}(\cdot \mid n)$ of the inverse bidding function $\xi(\cdot \mid p_n, n)$.

In practical estimation of the copula parameter, we consider finitely many grid points $v_1 < \cdots < v_J$ in $(\underline{v}, \overline{v})$. Then we have a finite set of restrictions

$$F^*(v_j \mid p_n) = C(F(v_j), p_n; \theta_0) / p_n, \text{ for } (j, n) \in \{1, ..., J\} \times \mathcal{N}, \tag{S4}$$

for the $J+1$ parameters $(\theta_0, F(v_1), ..., F(v_J))$. Let $M \coloneqq |\mathcal{N}|$ and let $n_1 < n_2 < \cdots < n_M$ be the elements of $\mathcal{N}$. Write the equations as

$$Q(F^*(v_j \mid p_{n_k}), p_{n_k}; \theta_0) = F(v_j), \text{ for } (j, k) \in [J] \times [M]. \tag{S5}$$

For $\boldsymbol{x} = (x_1, ..., x_M)^\top$ and $\boldsymbol{y} = (y_1, ..., y_M)^\top$, denote

$$\boldsymbol{Q}(\boldsymbol{x}, \boldsymbol{y}; \theta) \coloneqq \begin{pmatrix} Q(x_1, y_1; \theta) \\ \vdots \\ Q(x_M, y_M; \theta) \end{pmatrix}.$$

Denote $\boldsymbol{F}_0 \coloneqq (F(v_1), ..., F(v_J))^\top$, $\boldsymbol{p}_0 \coloneqq (p_{n_1}, ..., p_{n_M})^\top$, $\boldsymbol{F}_j^* \coloneqq (F^*(v_j \mid p_{n_1}), ..., F^*(v_j \mid p_{n_M}))^\top$ for $j \in [J]$, and $\boldsymbol{F}^* \coloneqq \left(\boldsymbol{F}_1^{*\top}, ..., \boldsymbol{F}_J^{*\top}\right)^\top$. For $\boldsymbol{z} \coloneqq \left(\boldsymbol{x}_1^\top, ..., \boldsymbol{x}_J^\top\right)^\top \in \mathbb{R}^{JM}$, denote

$$\boldsymbol{\Psi}(\boldsymbol{z}, \boldsymbol{y}, \theta) \coloneqq \begin{pmatrix} \boldsymbol{Q}(\boldsymbol{x}_1, \boldsymbol{y}; \theta) \\ \vdots \\ \boldsymbol{Q}(\boldsymbol{x}_J, \boldsymbol{y}; \theta) \end{pmatrix}. \tag{S6}$$

(S5) can be written in a vector form as

$$\boldsymbol{\Psi}(\boldsymbol{F}^*, \boldsymbol{p}_0, \theta_0) = (\mathbf{I}_J \otimes \mathbf{1}_M) \boldsymbol{F}_0. \tag{S7}$$

Let $\widehat{\boldsymbol{p}} \coloneqq (\widehat{p}_{n_1}, ..., \widehat{p}_{n_M})^\top$, $\widehat{\boldsymbol{F}}_j^* \coloneqq \left(\widehat{F}^*(v_j \mid p_{n_1}), ..., \widehat{F}^*(v_j \mid p_{n_M})\right)^\top$ for $j \in [J]$, and $\widehat{\boldsymbol{F}}^* \coloneqq \left(\widehat{\boldsymbol{F}}_1^{*\top}, ..., \widehat{\boldsymbol{F}}_J^{*\top}\right)^\top$. Let $\mathbf{W} \in \mathbb{R}^{MJ \times MJ}$ be a weight matrix and $\boldsymbol{F} \coloneqq (F_1, ..., F_J)^\top$. Let

$$\mathbb{F} \coloneqq \left\{ (F_1, ..., F_J)^\top \in [0,1]^J : F_1 \leq F_2 \leq \cdots \leq F_J \right\}.$$

The GMM criterion function is

$$\widehat{D}(\theta, \boldsymbol{F}; \mathbf{W}) := \left(\boldsymbol{\Psi}\left(\widehat{\boldsymbol{F}}^{*}, \widehat{\boldsymbol{p}}, \theta\right) - (\mathbf{I}_J \otimes \mathbf{1}_M)\boldsymbol{F}\right)^{\top} \mathbf{W} \left(\boldsymbol{\Psi}\left(\widehat{\boldsymbol{F}}^{*}, \widehat{\boldsymbol{p}}, \theta\right) - (\mathbf{I}_J \otimes \mathbf{1}_M)\boldsymbol{F}\right).$$

The GMM estimator is defined by

$$\left(\widehat{\theta}(\mathbf{W}), \widehat{\boldsymbol{F}}(\mathbf{W})\right) := \underset{(\theta, \boldsymbol{F}) \in \Theta \times \mathbb{F}}{\arg\min}\, \widehat{D}(\theta, \boldsymbol{F}; \mathbf{W}). \tag{S8}$$

To solve the minimization problem in (S8), we can easily partial out $\boldsymbol{F}$ given fixed θ. This requires solving a quadratic programming under linear inequality constraints. Let

$$\widehat{\boldsymbol{F}}(\theta; \mathbf{W}) := \underset{\boldsymbol{F} \in \mathbb{F}}{\arg\min}\, \widehat{D}(\theta, \boldsymbol{F}; \mathbf{W}). \tag{S9}$$

Then, after partialling out $\boldsymbol{F}$, we can calculate the GMM estimator

$$\widehat{\theta}(\mathbf{W}) = \underset{\theta \in \Theta}{\arg\min} \left(\boldsymbol{\Psi}\left(\widehat{\boldsymbol{F}}^{*}, \widehat{\boldsymbol{p}}, \theta\right) - (\mathbf{I}_J \otimes \mathbf{1}_M)\widehat{\boldsymbol{F}}(\theta; \mathbf{W})\right)^{\top} \mathbf{W} \left(\boldsymbol{\Psi}\left(\widehat{\boldsymbol{F}}^{*}, \widehat{\boldsymbol{p}}, \theta\right) - (\mathbf{I}_J \otimes \mathbf{1}_M)\widehat{\boldsymbol{F}}(\theta; \mathbf{W})\right)$$

by one-dimensional grid search.

We now consider a useful special case. The first-order conditions corresponding to the unconstrained quadratic programming problem $\min_{\boldsymbol{F} \in \mathbb{R}^J} \widehat{D}(\theta, \boldsymbol{F}; \mathbf{W})$ are

$$-2(\mathbf{I}_J \otimes \mathbf{1}_M)^{\top} \mathbf{W}\boldsymbol{\Psi}\left(\widehat{\boldsymbol{F}}^{*}, \widehat{\boldsymbol{p}}, \theta\right) + 2(\mathbf{I}_J \otimes \mathbf{1}_M)^{\top} \mathbf{W}(\mathbf{I}_J \otimes \mathbf{1}_M)\boldsymbol{F} = 0.$$

In case of $\mathbf{W} = \mathbf{I}_{MJ}$, since $(\mathbf{I}_J \otimes \mathbf{1}_M)^{\top}(\mathbf{I}_J \otimes \mathbf{1}_M) = M \cdot \mathbf{I}_J$ and

$$(\mathbf{I}_J \otimes \mathbf{1}_M)^{\top} \boldsymbol{\Psi}\left(\widehat{\boldsymbol{F}}^{*}, \widehat{\boldsymbol{p}}, \theta\right) = \begin{bmatrix} \mathbf{1}_M^{\top}\boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_1^{*}, \widehat{\boldsymbol{p}}; \theta\right) \\ \vdots \\ \mathbf{1}_M^{\top}\boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_J^{*}, \widehat{\boldsymbol{p}}; \theta\right) \end{bmatrix},$$

the minimizer corresponding to the unconstrained problem $\min_{\boldsymbol{F} \in \mathbb{R}^J} \widehat{D}(\theta, \boldsymbol{F}; \mathbf{I}_{MJ})$ is given by

$$\left((\mathbf{I}_J \otimes \mathbf{1}_M)^{\top}(\mathbf{I}_J \otimes \mathbf{1}_M)\right)^{-1}(\mathbf{I}_J \otimes \mathbf{1}_M)^{\top} \boldsymbol{\Psi}\left(\widehat{\boldsymbol{F}}^{*}, \widehat{\boldsymbol{p}}, \theta\right) = \begin{pmatrix} \frac{\mathbf{1}_M^{\top}\boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_1^{*}, \widehat{\boldsymbol{p}}; \theta\right)}{M} \\ \vdots \\ \frac{\mathbf{1}_M^{\top}\boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_J^{*}, \widehat{\boldsymbol{p}}; \theta\right)}{M} \end{pmatrix}. \tag{S10}$$

Since $\mathbf{1}_M^{\top}\boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_1^{*}, \widehat{\boldsymbol{p}}; \theta\right) \le \cdots \le \mathbf{1}_M^{\top}\boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_J^{*}, \widehat{\boldsymbol{p}}; \theta\right)$, the unconstrained minimizer satisfies the linear inequality constraints in (S9). Therefore, $\widehat{\boldsymbol{F}}(\theta;\mathbf{I}_{MJ})$ has a close form solution

$$\widehat{\boldsymbol{F}}(\theta;\mathbf{I}_{MJ})=\begin{pmatrix}\frac{\mathbf{1}_M^\top \boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_1^*,\widehat{\boldsymbol{p}};\theta\right)}{M}\\ \vdots \\ \frac{\mathbf{1}_M^\top \boldsymbol{Q}\left(\widehat{\boldsymbol{F}}_J^*,\widehat{\boldsymbol{p}};\theta\right)}{M}\end{pmatrix},$$

since it must coincide with the unconstrained minimizer.

S3 Asymptotic normality

First, we show the asymptotic normality of the pseudo-value-based CDF estimator. We assume that the following assumption on the data-generating process holds, in addition to Assumptions 2.1 and 3.2 in the main text. Let $f=F'$ be the density of the marginal distribution of the private costs. Let $C_1(x,y):=\partial C(x,y)/\partial x$.

Assumption S1. *(a) f is twice continuously differentiable and bounded away from zero on $[\underline{v},\overline{v}]$. (b) $C_1(\cdot,y)$ is bounded away from zero for all $y\in(0,1)$.*

Since

$$f^*(v\mid p_n):=\frac{\partial F^*(v\mid p_n)}{\partial v}=\frac{C_1(F(v),p_n)}{p_n}\cdot f(v),$$

Assumption S1(b) guarantees that $f^*(\cdot\mid p_n)$ also satisfies the assumption in Assumption S1(a). If $C(\cdot,\cdot)$ is an Archimedean copula with a twice differentiable strict generator $\varphi:[0,1]\to[0,\infty]$ with $\varphi'(u)<0$ and $\varphi''(u)\geq 0$ for $u\in(0,1)$, $\varphi(1)=0$ and $\varphi(0)=\infty$ (see Nelsen, 2006, Chapter 4.1 for more details about the class of Archimedean copulas). Note that under these requirements, the one-sided derivative $\varphi'(1)$ exists and $\varphi'(1)\leq 0$.[4] Then, it is easy to check that $\partial^2 C(x,y)/\partial x^2\leq 0$ and therefore, $C_1(\cdot,y)$ is non-increasing and

$$\lim_{t\uparrow 1}C_1(t,y)=\frac{\varphi'(1)}{\varphi'(y)}.$$

Then, in this case, Assumption S1(b) is fulfilled if $\varphi'(1)<0$.

We also require that the following mild condition on the kernel function $K(\cdot)$ used in the definition of (S3) holds.

Assumption S2. *The kernel function $K(\cdot)$ is symmetric, compactly supported on $[-1,1]$ and twice continuously differentiable on $\mathbb{R}$.*

We state the asymptotic normality result in the following proposition. Let $\beta'(\cdot\mid p_n,n)$ denote the derivative of $\beta(\cdot\mid p_n,n)$ and let $g''(\cdot\mid n)$ denote the second derivative of $g(\cdot\mid n)$.

[4]Since $\varphi''(u)\geq 0$ for $u\in(0,1)$, φ' is non-decreasing on $(0,1)$. It follows that $\lim_{u\uparrow 1}\varphi'(u)=\sup\{\varphi'(u):u\in(0,1)\}$ and $\lim_{u\uparrow 1}\varphi'(u)\leq 0$. Therefore, by the mean value theorem, $\varphi'(1)$ exists and equals $\lim_{u\uparrow 1}\varphi'(u)$.

Proposition S1. *Assume that Assumptions 2.1 and 3.2 in the main text and Assumptions S1 and S2 hold. Assume that the bandwidth h is chosen to be proportional to $L^{-\gamma}$ with $1/5 \leq \gamma < 1/3$. Then we have the following results. (a) We have*

$$\sqrt{Lh}\left(\widehat{F}^*(v \mid p_n) - F^*(v \mid p_n) - \Xi(v \mid n)\left(\int K(u)\, u^2 du\right) h^2\right) \to_d \mathrm{N}\left(0, \Sigma(v \mid n) \int K^2(u)\, du.\right),$$

where

$$\begin{aligned}\Xi(v \mid n) &:= -\frac{\eta_n(p_n, G(\beta(v \mid p_n, n) \mid n))\, \beta'(v \mid p_n, n)\, g''(\beta(v \mid p_n, n) \mid n)}{2(n-1)\, g(\beta(v \mid p_n, n) \mid n)} \\ \Sigma(v \mid n) &:= \frac{\eta_n^2(p_n, G(\beta(v \mid p_n, n) \mid n))\, (\beta'(v \mid p_n, n))^2}{(n-1)^2\, g(\beta(v \mid p_n, n), n)}.\end{aligned}$$

(b) Let $\boldsymbol{\Sigma}_j := (\Sigma(v_j \mid n_1), ..., \Sigma(v_j \mid n_M))^\top$ and $\boldsymbol{\Xi}_j := (\Xi(v_j \mid n_1), ..., \Xi(v_j \mid n_M))^\top$. We have the following joint asymptotic normality result:

$$\sqrt{Lh}\left(\widehat{\boldsymbol{F}}^* - \boldsymbol{F}^* - \boldsymbol{\Xi}\left(\int K(u)\, u^2 du\right) h^2\right) \to_d \mathrm{N}\left(0, \boldsymbol{\Sigma} \int K^2(u)\, du\right),$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with $\left(\boldsymbol{\Sigma}_1^\top, \boldsymbol{\Sigma}_2^\top, ..., \boldsymbol{\Sigma}_J^\top\right)^\top$ being the diagonal elements and $\boldsymbol{\Xi} := \left(\boldsymbol{\Xi}_2^\top, \boldsymbol{\Xi}_2^\top, ..., \boldsymbol{\Xi}_J^\top\right)^\top$.

Let

$$\begin{aligned}\boldsymbol{\Psi}_0(\boldsymbol{z}, \boldsymbol{y}, \theta) &:= \frac{\partial \boldsymbol{\Psi}(\boldsymbol{z}, \boldsymbol{y}, \theta)}{\partial \theta} \\ \boldsymbol{\Psi}_1(\boldsymbol{z}, \boldsymbol{y}, \theta) &:= \frac{\partial \boldsymbol{\Psi}(\boldsymbol{z}, \boldsymbol{y}, \theta)}{\partial \boldsymbol{z}^\top}\end{aligned}$$

denote the partial derivatives. The asymptotic theory of our GMM estimator is similar to that of the local GMM estimator studied in Lewbel (2007). Denote $\boldsymbol{\vartheta} := \left(\theta, \boldsymbol{F}^\top\right)^\top$, $\boldsymbol{\vartheta}_0 := \left(\theta_0, \boldsymbol{F}_0^\top\right)^\top$, $\widehat{\boldsymbol{\vartheta}}(\mathbf{W}) := \left(\widehat{\theta}(\mathbf{W}), \widehat{\boldsymbol{F}}(\mathbf{W})^\top\right)^\top$ and write

$$\boldsymbol{\Upsilon}(\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\vartheta}) := \boldsymbol{\Psi}(\boldsymbol{z}, \boldsymbol{y}, \theta) - (\mathbf{I}_J \otimes \mathbf{1}_M)\, \boldsymbol{F}.$$

Then, (S7) can be represented more compactly as

$$\boldsymbol{\Upsilon}(\boldsymbol{F}^*, \boldsymbol{p}_0, \boldsymbol{\vartheta}_0) = \mathbf{0}_{JM},$$

and the estimator $\widehat{\boldsymbol{\vartheta}}(\mathbf{W})$ can be represented as

$$\widehat{\boldsymbol{\vartheta}}(\mathbf{W}) = \underset{\boldsymbol{\vartheta} \in \Theta \times \mathbb{F}}{\arg\min}\, \widehat{D}(\boldsymbol{\vartheta}; \mathbf{W}),$$

where

$$\widehat{D}(\boldsymbol{\vartheta};\mathbf{W}) \coloneqq \boldsymbol{\Upsilon}^{\top}\left(\widehat{\boldsymbol{F}}^{*},\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}\right)\mathbf{W}\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^{*},\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}\right).$$

We say that θ_0 is globally identified from the finite set of restrictions if the system of equations

$$\boldsymbol{\Upsilon}(\boldsymbol{F}^{*},\boldsymbol{p}_0,\boldsymbol{\vartheta}) = \mathbf{0}_{JM} \tag{S11}$$

has a unique solution at $\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0$. By similar arguments, θ_0 is globally identified from the finite set of restrictions under a similar but stronger condition: for some $k,l\in[M]$ and $j\in[J]$,

$$\min_{\theta\in\Theta}\frac{\partial}{\partial\theta}\left(Q\left(F^{*}\left(v_j\mid p_{n_k}\right),p_{n_k};\theta\right)-Q\left(F^{*}\left(v_j\mid p_{n_l}\right),p_{n_l};\theta\right)\right)>0. \tag{S12}$$

Similarly, θ_0 is locally identified from the finite set of restrictions if there exists an open neighborhood around θ_0 such that for any $\theta\neq\theta_0$ in the neighborhood and $(F_1,...,F_J)\in[0,1]^J$ that satisfy $F_1\leq F_2\leq\cdots\leq F_J$, $(\theta,F_1,...,F_J)$ cannot be a solution to (S11). The condition in Proposition 3.3(b) guarantees that there must exist some $v\in[\underline{v},\overline{v}]$ so that θ_0 is locally identified from the restrictions

$$F^{*}(v\mid p_n)=C\left(F(v),p_n;\theta_0\right)/p_n,\ n\in\mathcal{N}.$$

We make the following mild assumption on the parametric copula family $\{C(\cdot,\cdot;\theta):\theta\in\Theta\}$. It is satisfied by most commonly used parametric copula families (e.g., Gaussian, Ali-Mikhail-Haq, Clayton, Frank, Gumbel, and Joe). Let $C_1(x,y;\theta)\coloneqq\partial C(x,y;\theta)/\partial x$, $C_2(x,y;\theta)\coloneqq\partial C(x,y;\theta)/\partial y$ and $C_\theta(x,y;\theta)\coloneqq\partial C(x,y;\theta)/\partial\theta$.

Assumption S3. *For all $\epsilon\in(0,1/2)$, $C_1(x,y;\theta)>0$ for all $(x,y,\theta)\in(0,1)\times[\epsilon,1-\epsilon]\times\Theta$.*

Under Assumption S3, $C(\cdot,y;\theta)/y$ is strictly increasing on $[0,1]$ and therefore, $Q(\cdot,y;\theta)$ is also strictly increasing on $[0,1]$. Then we have

$$0<Q(\epsilon,y;\theta)\leq Q(x,y;\theta)\leq Q(1-\epsilon,y;\theta)<1, \tag{S13}$$

for all $(x,y,\theta)\in[\epsilon,1-\epsilon]^2\times\Theta$. Since we can write

$$\frac{C\left(Q(x,y;\theta),y;\theta\right)}{y}-x=0,$$

by (S13) and the implicit function theorem, $Q(\cdot,\cdot;\cdot)$ is continuously differentiable on $[\epsilon,1-\epsilon]^2\times\Theta$.

Let $\boldsymbol{\Psi}_1\coloneqq\boldsymbol{\Psi}_1(\boldsymbol{F}^{*},\boldsymbol{p}_0,\theta_0)$, $\boldsymbol{\Psi}_0\coloneqq\boldsymbol{\Psi}_0(\boldsymbol{F}^{*},\boldsymbol{p}_0,\theta_0)$ and $\boldsymbol{\Omega}_0\coloneqq\boldsymbol{\Psi}_1\boldsymbol{\Sigma}\boldsymbol{\Psi}_1$. Clearly, $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Omega}_0$ are both diagonal matrices. Let

$$\boldsymbol{\Pi}_0\coloneqq\frac{\partial\boldsymbol{\Upsilon}(\boldsymbol{F}^{*},\boldsymbol{p}_0,\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}^{\top}}=\left[\begin{array}{cc}\boldsymbol{\Psi}_0 & -\left(\mathbf{I}_J\otimes\mathbf{1}_M\right)\end{array}\right].$$

Note that the second equality shows that $\mathbf{\Pi}_0$ has full rank under the condition (S12). Denote

$$\begin{aligned}\mathit{\Xi}_{\vartheta}(\mathbf{W}) &:= -\left(\mathbf{\Pi}_0^{\top}\mathbf{W}\mathbf{\Pi}_0\right)^{-1}\mathbf{\Pi}_0^{\top}\mathbf{W}\mathbf{\Psi}_1\mathit{\Xi}\\ \mathbf{\Sigma}_{\vartheta}(\mathbf{W}) &:= \left(\mathbf{\Pi}_0^{\top}\mathbf{W}\mathbf{\Pi}_0\right)^{-1}\mathbf{\Pi}_0^{\top}\mathbf{W}\mathbf{\Omega}_0\mathbf{W}\mathbf{\Pi}_0\left(\mathbf{\Pi}_0^{\top}\mathbf{W}\mathbf{\Pi}_0\right)^{-1}.\end{aligned}$$

By using Proposition S1 and standard arguments in the proof of consistency and asymptotic normality of M-estimators (e.g., Hansen, Chapter 22), we show that the GMM estimator is consistent and asymptotically normal.

Proposition S2. *Assume that the conditions in the statement of Proposition S1 are satisfied. Also assume that Assumption 3.1 in the main text is satisfied. Assume that $\widehat{\mathbf{W}} \to_p \mathbf{W}_0$ for some deterministic positive definite matrix $\mathbf{W}_0$. Assume that $\boldsymbol{\vartheta}_0$ is in the interior of $\Theta \times \mathbb{F}$, $\mathbf{\Pi}_0$ has full rank and (S11) has a unique solution at $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. Also assume that Assumption S3 is satisfied. Then we have the following results. (a) $\widehat{\boldsymbol{\vartheta}}\left(\widehat{\mathbf{W}}\right) \to_p \boldsymbol{\vartheta}_0$. (b)*

$$\sqrt{Lh}\left(\widehat{\boldsymbol{\vartheta}}\left(\widehat{\mathbf{W}}\right) - \boldsymbol{\vartheta}_0 - \mathit{\Xi}_{\vartheta}(\mathbf{W}_0)\left(\int K(u)\,u^2 du\right)h^2\right) \to_d \mathrm{N}\left(0, \mathbf{\Sigma}_{\vartheta}(\mathbf{W}_0)\int K^2(u)\,du\right).$$

It then follows from standard calculation (e.g., Hansen, 2022, Theorem 13.5) that $\mathbf{\Sigma}_{\vartheta}(\mathbf{W}_0) - \mathbf{\Sigma}_{\vartheta}\left(\mathbf{\Omega}_0^{-1}\right)$ is positive semidefinite and the optimal weight matrix is given by $\mathbf{\Omega}_0^{-1}$.

S4 Bootstrap estimation of the optimal weight matrix

Estimation of the optimal weight matrix $\mathbf{\Omega}_0^{-1}$ requires estimating $\Sigma(v \mid n)$, which takes a complicated form and depends on the derivative $\beta'(v \mid p_n, n)$. We propose a convenient nonparametric bootstrap estimator of $\Sigma(v \mid n)$. Let

$$\left\{\left(B_{1l}^{\dagger}, ..., B_{N_l^{*\dagger}l}^{\dagger}, N_l^{*\dagger}, N_l^{\dagger}\right) : l \in [L]\right\}$$

be the nonparametric bootstrap sample drawn with replacement from

$$\left\{\left(B_{1l}, ..., B_{N_l^* l}, N_l^*, N_l\right) : l \in [L]\right\}. \tag{S14}$$

Let $\mathrm{E}_{\dagger}[\cdot]$ and $\mathrm{Var}_{\dagger}[\cdot]$ denote the conditional expectation and variance given (S14).

Let

$$\widehat{q}_n^{\dagger} := \frac{1}{n\left|\left\{l : N_l^{\dagger} = n\right\}\right|}\sum_{l:N_l^{\dagger}=n} N_l^{*\dagger}$$

be the bootstrap analogue of $\widehat{q}_n$ and let

$$\widehat{G}_{\dagger}(b, n) := \frac{1}{L}\sum_{l:N_l^{\dagger}=n}\sum_{i=1}^{N_l^{*\dagger}} \mathbb{1}\left(B_{il}^{\dagger} \leq b\right)$$

$$\widehat{g}_\dagger(b,n) \;\; := \;\; \frac{1}{L}\sum_{l:N_l^\dagger=n}\sum_{i=1}^{N_l^{*\dagger}}\frac{1}{h}\mathcal{K}_1\left(B_{il}^\dagger, b \mid h, \underline{\widehat{b}}_n, \overline{\widehat{b}}_n\right).$$

It is easy to check that $\mathrm{E}_\dagger\left[\widehat{G}_\dagger(b,n)\right]=\widehat{G}(b,n)$. Let

$$\begin{aligned}
\widetilde{r}_n^\dagger &:= \frac{1}{L}\sum_{l:N_l^\dagger=n} N_l^{*\dagger}\\
\widehat{r}_n &:= \frac{1}{L}\sum_{l:N_l=n} N_l^{*}\\
\widehat{\sigma}_{r,n}^2 &:= \frac{1}{L}\sum_{l:N_l=n}^{L}\left(N_l^*\right)^2-\widehat{r}_n^2\\
\widehat{r}_n^\dagger &:= \widetilde{r}_N^\dagger \vee \left(\widehat{r}_n-\widehat{\sigma}_{r,n}\sqrt{\frac{2\cdot\log(L)}{L}}\right),
\end{aligned}$$

$\widehat{G}_\dagger(b\mid n) := \widehat{G}_\dagger(b,n)/\widehat{r}_n^\dagger$ and $\widehat{g}_\dagger(b\mid n) := \widehat{g}_\dagger(b,n)/\widehat{r}_n^\dagger$. Then let $\widehat{p}_n^\dagger := \varphi_n^{-1}\left(\widehat{q}_n^\dagger\right)$ and let

$$\widehat{\xi}_\dagger(b\mid n) = b-\frac{\eta_n\left(\widehat{p}_n^\dagger, \widehat{G}_\dagger(b\mid n)\right)}{(n-1)\,\widehat{g}_\dagger(b\mid n)}$$

be the bootstrap analogue of $\widehat{\xi}(b\mid n)$ and let $\widehat{V}_{il}^\dagger := \widehat{\xi}_\dagger\left(B_{il}^\dagger \mid N_l^\dagger\right)$. Then let

$$\widehat{F}_\dagger^*(v,n) := \frac{1}{L}\sum_{l:N_l^\dagger=n}\sum_{i=1}^{N_l^{*\dagger}}\mathbb{1}\left(\widehat{V}_{il}^\dagger \le v\right) \tag{S15}$$

be the bootstrap analogue of $\widehat{F}^*(v,n)$. Let $\widehat{F}_\dagger^*(v\mid p_n) := \widehat{F}_\dagger^*(v,n)/\widehat{r}_n^\dagger$.

Proposition S3. *Suppose that the assumptions in the statement of Proposition S2 are satisfied. Then, for all $(j,k)\in[J]\times[M]$,*

$$\frac{\mathrm{Var}_\dagger\left[\widehat{F}_\dagger^*\left(v_j\mid p_{n_k}\right)\right]}{\Sigma\left(v_j\mid n_k\right)/(Lh)} \to_p 1.$$

Let $\widetilde{\theta} := \widehat{\theta}(\mathbf{I}_{mJ})$ be the preliminary estimator. Let $\widehat{\mathbf{\Psi}}_1 := \boldsymbol{\Psi}_1\left(\widehat{\boldsymbol{F}}^*, \widehat{\boldsymbol{p}}, \widetilde{\theta}\right)$ and let $\widehat{\mathbf{\Sigma}}$ be the diagonal matrix with $\left(\widehat{\boldsymbol{\Sigma}}_1^\top, \widehat{\boldsymbol{\Sigma}}_2^\top, ..., \widehat{\boldsymbol{\Sigma}}_J^\top\right)^\top \in \mathbb{R}^{JM}$ being the diagonal elements, where

$$\widehat{\boldsymbol{\Sigma}}_j := (Lh)\left(\mathrm{Var}_\dagger\left[\widehat{F}_\dagger^*\left(v_j\mid p_{n_1}\right)\right], ..., \mathrm{Var}_\dagger\left[\widehat{F}_\dagger^*\left(v_j\mid p_{n_M}\right)\right]\right)^\top.$$

The estimated optimal weight matrix is given by $\left(\widehat{\mathbf{\Psi}}_1\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Psi}}_1\right)^{-1}$.

Appendix

Notation. For any function $f : A \to \mathbb{R}$, let $\|f\|_A \coloneqq \sup_{x\in A} |f(x)|$. For a univariate function f, denote $f^{(j)}(x) \coloneqq (d/dx)^j f(x)$. For a bivariate function f, denote $D_1 f(x,y) \coloneqq \partial f(x,y)/\partial x$ and $D_2 f(x,y) \coloneqq \partial f(x,y)/\partial y$. For a symmetric matrix $\mathbf{A}$, let $\operatorname{mineig}(\mathbf{A})$ denote its smallest eigenvalue. Let $[a \pm b]$ be shorthand notation for the interval $[a-b, a+b]$. "$a \eqqcolon b$" is understood as "b is defined by a.". We write $a \lesssim b$ if $a \leq C \cdot b$ for some positive constant C that does not depend on the sample size L. Let $\|\boldsymbol{x}\|$ denote the Euclidean norm of a real vector $\boldsymbol{x}$. For a matrix $\mathbf{A}$, $\|\mathbf{A}\|$ is understood as the operator norm of $\mathbf{A}$. "Law of iterated expectations" is abbreviated as "LIE". "First-order conditions" is abbreviated as "FOCs". "With probability approaching one" is abbreviated as "wpa1". "Law of iterated expectations" is abbreviated as "LIE".

Let $\mathfrak{F}$ denote a class of $\mathbb{R}$-valued functions defined on a compact set S in a finite-dimensional Euclidean space. Let $\mathfrak{F}$ be equipped with a norm $\|\cdot\|$. We say that a finite subset $\mathfrak{F}^\circ$ of $\mathfrak{F}$ is an ε-net if the union of the closed $\|\cdot\|$-balls of radius ε centered at points in $\mathfrak{F}^\circ$ covers $\mathfrak{F}$. $N(\varepsilon, \mathfrak{F}, \|\cdot\|) \coloneqq \inf\{|\mathfrak{F}^\circ| : \mathfrak{F}^\circ \text{ is an } \varepsilon\text{-net of } \mathfrak{F}\}$ is called the ε-covering number. A function $F : S \to \mathbb{R}_+$ is an envelope of $\mathfrak{F}$ if $\sup_{f\in\mathfrak{F}} |f| \leq F$. We say that $\mathfrak{F}$ is a (uniform) Vapnik–Chervonenkis-type (VC-type) class with respect to the envelope F (see, e.g., Giné and Nickl 2015, Definition 3.6.10) if there exist some positive constants (A, V) that are independent of the sample size such that for all $\varepsilon \in (0,1]$,

$$\sup_{Q\in\mathcal{Q}} N\left(\varepsilon \|F\|_{Q,2}, \mathfrak{F}, \|\cdot\|_{Q,2}\right) \leq \left(\frac{A}{\varepsilon}\right)^V, \tag{S16}$$

where $\mathcal{Q}$ denotes the collection of all finitely discrete probability measures on S and $\|f\|_{Q,2} \coloneqq \sqrt{\int f^2 dQ}$.

Appendix A Concentration analysis of the local polynomial density estimator

Let $X_1, ..., X_n$ be an i.i.d. sample where X_i has a bounded PDF f supported on $\mathcal{X} \coloneqq [\underline{x}, \overline{x}]$. For some $p \in \mathbb{N}$, let $\boldsymbol{r}_p(t) \coloneqq (1, t, ..., t^p)^\top$, $\boldsymbol{K}_p(t) \coloneqq \boldsymbol{r}_p(t) K(t)$, $\mathbf{R}_p(t) \coloneqq \boldsymbol{r}_p(t) \boldsymbol{r}_p^\top(t)$ and

$$\begin{aligned} \mathbf{W}_p(x \mid h, \underline{x}, \overline{x}) &\coloneqq \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \mathbf{R}_p(t) K(t)\, dt \\ \mathcal{K}_p(X_i, x \mid h, \underline{x}, \overline{x}) &\coloneqq \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{X_i - x}{h}\right), \end{aligned}$$

where $\mathbf{e}_1 \coloneqq (1, 0, ..., 0)^\top \in \mathbb{R}^{p+1}$. It is easy to see that when h is sufficiently small,

$$\sup_{x\in\mathcal{X}} \left\|\mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x})\right\| \leq \left(\operatorname{mineig}\left(\int_0^1 \mathbf{R}_p(t) K(t)\, dt\right)\right)^{-1} \eqqcolon \varpi_p. \tag{S17}$$

The local polynomial density estimator of f is

$$\widehat{f}(x) := \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}_p\left(X_i, x \mid h, \underline{x}, \overline{x}\right). \tag{S18}$$

In practice, the boundary points $\underline{x}$ and $\overline{x}$ are unknown. However, validity of the first-order asymptotic results derived in this section is unaffected if we replace the unknown $\underline{x}$ and $\overline{x}$ with their super-consistent estimators $\min\{X_1, X_2, ..., X_n\}$ and $\max\{X_1, X_2, ..., X_n\}$. Let $\widehat{f}'$ and $\widehat{f}''$ be the first and second derivatives of $\widehat{f}$.

Denote

$$\boldsymbol{K}_p'(t) := \frac{d\boldsymbol{K}_p(t)}{dt} = \frac{d\boldsymbol{r}_p(t)}{dt}K(t) + \boldsymbol{r}_p(t)K'(t)$$

and $\boldsymbol{K}_p''(t) := d^2\boldsymbol{K}_p(t)/dt^2$. It is easy to check that

$$\frac{d}{dx}\mathbf{W}_p\left(x \mid h, \underline{x}, \overline{x}\right) = \frac{1}{h}\mathbf{D}_p\left(x \mid h, \underline{x}, \overline{x}\right), \tag{S19}$$

where

$$\mathbf{D}_p\left(x \mid h, \underline{x}, \overline{x}\right) := -\mathbf{R}_p\left(\frac{\overline{x}-x}{h}\right)K\left(\frac{\overline{x}-x}{h}\right) + \mathbf{R}_p\left(\frac{\underline{x}-x}{h}\right)K\left(\frac{\underline{x}-x}{h}\right)$$

and $\sup_{x\in\mathcal{X}}\left\|\mathbf{D}_p\left(x \mid h, \underline{x}, \overline{x}\right)\right\| < \infty$, since $K(\cdot)$ is compactly supported on $[-1, 1]$. By this result, the product rule and

$$\frac{d}{dx}\mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right) = -\mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right)\frac{d}{dx}\mathbf{W}_p\left(x \mid h, \underline{x}, \overline{x}\right)\mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right), \tag{S20}$$

we have

$$\frac{\partial}{\partial x}\mathcal{K}_p\left(X_i, x \mid h, \underline{x}, \overline{x}\right) = \frac{1}{h}\dot{\mathcal{K}}_p\left(X_i, x \mid h, \underline{x}, \overline{x}\right),$$

where

$$\begin{aligned}\dot{\mathcal{K}}_p\left(X_i, x \mid h, \underline{x}, \overline{x}\right) \quad = \quad & -\mathbf{e}_1^\top\mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right)\boldsymbol{K}_p'\left(\frac{X_i-x}{h}\right)\\ & -\mathbf{e}_1^\top\mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right)\mathbf{D}_p\left(x \mid h, \underline{x}, \overline{x}\right)\mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right)\boldsymbol{K}_p\left(\frac{X_i-x}{h}\right)\end{aligned} \tag{S21}$$

Then we have

$$\widehat{f}'(x) = \frac{1}{nh^2}\sum_{i=1}^{n}\dot{\mathcal{K}}_p\left(X_i, x \mid h, \underline{x}, \overline{x}\right). \tag{S22}$$

It is easy to see that for some matrix $\mathbf{S}_p\left(x \mid h, \underline{x}, \overline{x}\right)$,

$$\frac{d}{dx}\mathbf{D}_p\left(x \mid h, \underline{x}, \overline{x}\right) = \frac{1}{h}\cdot\mathbf{S}_p\left(x \mid h, \underline{x}, \overline{x}\right)$$

and $\sup_{x\in\mathcal{X}} \|\mathbf{S}_p(x \mid h, \underline{x}, \overline{x})\| < \infty$. By (S19), (S20) and the product rule,

$$\frac{\partial^2}{\partial x^2}\mathcal{K}_p(X_i, x \mid h, \underline{x}, \overline{x}) = \frac{1}{h^2}\ddot{\mathcal{K}}_p(X_i, x \mid h, \underline{x}, \overline{x}),$$

where

$$\begin{aligned}
\ddot{\mathcal{K}}_p(X_i, x \mid h, \underline{x}, \overline{x}) &:= \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p''\left(\frac{X_i - x}{h}\right) \\
&+ 2 \cdot \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \mathbf{D}_p(x \mid h, \underline{x}, \overline{x}) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p'\left(\frac{X_i - x}{h}\right) \\
&+ 2 \cdot \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \mathbf{D}_p(x \mid h, \underline{x}, \overline{x}) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \mathbf{D}_p(x \mid h, \underline{x}, \overline{x}) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{X_i - x}{h}\right) \\
&- \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \mathbf{S}_p(x \mid h, \underline{x}, \overline{x}) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{X_i - x}{h}\right).
\end{aligned}$$

Then we have

$$\widehat{f}''(x) = \frac{1}{nh^3}\sum_{i=1}^{n} \ddot{\mathcal{K}}_p(X_i, x \mid h, \underline{x}, \overline{x}).$$

Let $\overline{f}(x) := \mathrm{E}\left[\widehat{f}(x)\right]$, $\overline{f}'(x) := \mathrm{E}\left[\widehat{f}'(x)\right]$ and $\overline{f}''(x) := \mathrm{E}\left[\widehat{f}''(x)\right]$. We have the following results on the bias $\overline{f}(x) - f(x)$ and its first and second derivatives.

Proposition A.1. *Assume that f is $(p+1)$-times continuously differentiable on $[\underline{x}, \overline{x}]$ and $h \downarrow 0$ as $n \uparrow \infty$. Assume that the kernel function $K(\cdot)$ satisfies Assumption S2. Then we have the following results. (a)*

$$\overline{f}(x) = f(x) + \left(\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p(u) u^{p+1} du\right) \frac{f^{(p+1)}(x)}{(p+1)!} h^{p+1} + o\left(h^{p+1}\right),$$

uniformly in $x \in \mathcal{X}$. (b) $\overline{f}'(x) = f'(x) + O(h^p)$, uniformly in $x \in \mathcal{X}$; (c) $\overline{f}''(x) = f''(x) + O\left(h^{p-1}\right)$, uniformly in $x \in \mathcal{X}$.

Proof of Proposition A.1. Denote $\boldsymbol{\phi}(x) := \left(f^{(0)}(x)/0!, f^{(1)}(x)/1!, ..., f^{(p)}(x)/p!\right)^\top$. Let $\mathbf{H}$ be the $(p+1)$-dimensional diagonal matrix with diagonal elements $(1, h, ..., h^p)$. By change of variables and Taylor expansion,

$$\begin{aligned}
\overline{f}(x) &= \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\underline{x}}^{\overline{x}} \frac{1}{h}\boldsymbol{K}_p\left(\frac{y - x}{h}\right) f(y)\, dy \\
&= \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p(u) \left\{\boldsymbol{r}_p(u)^\top \mathbf{H}\boldsymbol{\phi}(x) + \frac{f^{(p+1)}(\dot{x})(hu)^{p+1}}{(p+1)!}\right\} du,
\end{aligned}$$

where $\dot{x}$ denotes the mean value that lies between x and $x + hu$. The conclusion in Part (a) follows from this result.

Now by (S21), (S22), integration by parts and tedious algebra, we have

$$\begin{aligned}\overline{f}'(x) &= \int_{\underline{x}}^{\overline{x}} \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{y-x}{h}\right) f'(y)\,dy \\ &+\frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\underline{x}-x}{h}\right) \underline{I}(x) - \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\overline{x}-x}{h}\right) \overline{I}(x),\end{aligned} \tag{S23}$$

where

$$\underline{I}(x) \coloneqq f(\underline{x}) - \boldsymbol{r}_p^\top\left(\frac{\underline{x}-x}{h}\right) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\underline{x}}^{\overline{x}} \frac{1}{h}\boldsymbol{K}_p\left(\frac{y-x}{h}\right) f(y)\,dy$$

and $\overline{I}(x)$ is defined similarly. By Taylor expansion,

$$\begin{aligned}\underline{I}(x) &= f(\underline{x}) - \boldsymbol{r}_p^\top\left(\frac{\underline{x}-x}{h}\right) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\underline{x}}^{\overline{x}} \frac{1}{h}\boldsymbol{K}_p\left(\frac{y-x}{h}\right) f(y)\,dy \\ &= f(\underline{x}) - \boldsymbol{r}_p^\top(\underline{x}-x)\boldsymbol{\phi}(x) \\ &\quad -\boldsymbol{r}_p^\top\left(\frac{\underline{x}-x}{h}\right) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p(u) \frac{f^{(p+1)}(\dot{x})(hu)^{p+1}}{(p+1)!}du,\end{aligned} \tag{S24}$$

where $\dot{x}$ denotes the mean value that lies between x and $x+hu$. Since

$$\left|K\left(\frac{\underline{x}-x}{h}\right)\right| \lesssim \mathbb{1}(|\underline{x}-x| \leq h), \tag{S25}$$

by this result, (S17) and Taylor expansion,

$$\frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\underline{x}-x}{h}\right)\left\{f(\underline{x}) - \boldsymbol{r}_p^\top(\underline{x}-x)\boldsymbol{\phi}(x)\right\} = O(h^p),$$

uniformly in $x \in \mathcal{X}$. By this result, (S17) and (S24),

$$\frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\underline{x}-x}{h}\right) \underline{I}(x) = O(h^p), \tag{S26}$$

uniformly in $x \in \mathcal{X}$. Similarly,

$$\frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\overline{x}-x}{h}\right) \overline{I}(x) = O(h^p), \tag{S27}$$

uniformly in $x \in \mathcal{X}$. By Taylor expansion and (S17),

$$\begin{aligned}&\int_{\underline{x}}^{\overline{x}} \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{y-x}{h}\right) f'(y)\,dy \\ &= \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p(u)\left\{\boldsymbol{r}_p(u)^\top \mathbf{H}\boldsymbol{\phi}'(x) + \frac{f^{(p+1)}(\dot{x})(hu)^p}{p!}\right\}du\end{aligned}$$

$$= \quad f'(x) + O(h^p),$$

uniformly in $x \in \mathcal{X}$, where $\boldsymbol{\phi}'(x) \coloneqq \left(f^{(1)}(x)/0!, f^{(2)}(x)/1!, ..., f^{(p)}(x)/(p-1)!, 0\right)^\top$ and $\dot{x}$ is the mean value that lies between x and $x + hu$. The conclusion in Part (b) follows from this result, (S23), (S26) and (S27).

For Part (c), first note that we can write

$$\begin{aligned}\overline{f}''(x) &= \frac{d}{dx}\int_{\underline{x}}^{\overline{x}} \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{y-x}{h}\right) f'(y)\,dy \\ &\quad + \frac{1}{h}\frac{d}{dx}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\underline{x}-x}{h}\right) \underline{I}(x) - \frac{1}{h}\frac{d}{dx}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\overline{x}-x}{h}\right) \overline{I}(x).\end{aligned} \tag{S28}$$

By integration by parts and (S20),

$$\begin{aligned}&\frac{d}{dx}\int_{\underline{x}}^{\overline{x}} \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{y-x}{h}\right) f'(y)\,dy \\ &= \int_{\underline{x}}^{\overline{x}} \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{y-x}{h}\right) f''(y)\,dy \\ &\quad + \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\underline{x}-x}{h}\right) \dot{\underline{I}}(x) - \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\overline{x}-x}{h}\right) \dot{\overline{I}}(x),\end{aligned} \tag{S29}$$

where

$$\dot{\underline{I}}(x) \coloneqq f'(\underline{x}) - \boldsymbol{r}_p^\top\left(\frac{\underline{x}-x}{h}\right) \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\underline{x}}^{\overline{x}} \frac{1}{h}\boldsymbol{K}_p\left(\frac{y-x}{h}\right) f'(y)\,dy$$

and $\dot{\overline{I}}(x)$ is defined similarly. By Taylor expansion and (S17),

$$\begin{aligned}&\int_{\underline{x}}^{\overline{x}} \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{y-x}{h}\right) f''(y)\,dy \\ &= \mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p(u) \left\{\boldsymbol{r}_p(u)^\top \mathbf{H}\boldsymbol{\phi}''(x) + \frac{f^{(p+1)}(\dot{x})(hu)^{p-1}}{(p-1)!}\right\} du \\ &= f''(x) + O\left(h^{p-1}\right),\end{aligned}$$

uniformly in $x \in \mathcal{X}$, where $\boldsymbol{\phi}''(x) \coloneqq \left(f^{(2)}(x)/0!, f^{(3)}(x)/1!, ..., f^{(p)}(x)/(p-2)!, 0, 0\right)^\top$ and $\dot{x}$ is the mean value that lies between x and $x + hu$. By using arguments similar to those used in the proof of (S26) (Taylor expansion, (S17) and (S25)), we have

$$\frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}(x \mid h, \underline{x}, \overline{x}) \boldsymbol{K}_p\left(\frac{\underline{x}-x}{h}\right) \dot{\underline{I}}(x) = O\left(h^{p-1}\right),$$

uniformly in $x \in \mathcal{X}$. Similarly,

$$\frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right) \boldsymbol{K}_p\left(\frac{\overline{x}-x}{h}\right) \dot{\overline{I}}\left(x\right) = O\left(h^{p-1}\right),$$

uniformly in $x \in \mathcal{X}$. By these results and (S29),

$$\frac{d}{dx}\int_{\underline{x}}^{\overline{x}} \frac{1}{h}\mathbf{e}_1^\top \mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right) \boldsymbol{K}_p\left(\frac{y-x}{h}\right) f'\left(y\right) dy = f''\left(x\right) + O\left(h^{p-1}\right). \tag{S30}$$

By Taylor expansion with remainder terms written in their integral forms,

$$f\left(\overline{x}\right) - \boldsymbol{r}_p^\top\left(\overline{x}-x\right)\boldsymbol{\phi}\left(x\right) = \int_x^{\overline{x}} \frac{\left(\overline{x}-t\right)^p}{p!} f^{p+1}\left(t\right) dt$$

and

$$\begin{aligned}\overline{I}\left(x\right) \;&=\; \int_x^{\overline{x}} \frac{\left(\overline{x}-t\right)^p}{p!} f^{p+1}\left(t\right) dt \\ &\quad -\boldsymbol{r}_p^\top\left(\frac{\overline{x}-x}{h}\right)\mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right) \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p\left(u\right)\left\{\int_x^{x+hu} f^{(p+1)}\left(t\right)\frac{\left(x+hu-t\right)^p}{p!} dt\right\} du.\end{aligned}$$

By calculations,

$$\begin{aligned}&\frac{d}{dx}\int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p\left(u\right)\left\{\int_x^{x+hu} f^{(p+1)}\left(t\right)\frac{\left(x+hu-t\right)^p}{p!} dt\right\} du \\ =\;& -\frac{1}{h}\boldsymbol{K}_p\left(\frac{\overline{x}-x}{h}\right)\int_x^{\overline{x}} f^{(p+1)}\left(t\right)\frac{\left(\overline{x}-t\right)^p}{p!} dt + \frac{1}{h}\boldsymbol{K}_p\left(\frac{\underline{x}-x}{h}\right)\int_x^{\underline{x}} f^{(p+1)}\left(t\right)\frac{\left(\underline{x}-t\right)^p}{p!} dt \\ & -\frac{f^{(p+1)}\left(x\right)}{p!}\int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p\left(u\right)\left(hu\right)^p du + \int_{\frac{\underline{x}-x}{h}}^{\frac{\overline{x}-x}{h}} \boldsymbol{K}_p\left(u\right)\left\{\int_x^{x+hu} f^{(p+1)}\left(t\right)\frac{\left(x+hu-t\right)^{p-1}}{\left(p-1\right)!} dt\right\} du.\end{aligned}$$

By these results, (S17), (S20), and

$$\begin{aligned}\left|K\left(\frac{\overline{x}-x}{h}\right)\right| \;&\lesssim\; \mathbb{1}\left(\left|\overline{x}-x\right| \leq h\right) \\ \left|K'\left(\frac{\overline{x}-x}{h}\right)\right| \;&\lesssim\; \mathbb{1}\left(\left|\overline{x}-x\right| \leq h\right),\end{aligned}$$

we have

$$\frac{1}{h}\frac{d}{dx}\mathbf{e}_1^\top \mathbf{W}_p^{-1}\left(x \mid h, \underline{x}, \overline{x}\right) \boldsymbol{K}_p\left(\frac{\overline{x}-x}{h}\right) \overline{I}\left(x\right) = O\left(h^{p-1}\right),$$

uniformly in $x \in \mathcal{X}$. A similar result holds for the second term on the right hand side of (S29). The conclusion in Part (c) follows from these results, (S28) and (S30). ■

Consider

$$
\begin{aligned}
\mathfrak{K} &:= \{\mathcal{K}_p(\cdot, x \mid h, \underline{x}, \overline{x}) : x \in \mathcal{X}\} \\
\sigma^2_{\mathfrak{K}} &:= \sup_{x \in \mathcal{X}} \mathrm{E}\left[\mathcal{K}_p^2(X_i, x \mid h, \underline{x}, \overline{x})\right].
\end{aligned}
$$

Let $\left(\dot{\mathfrak{K}}, \sigma^2_{\dot{\mathfrak{K}}}\right)$ and $\left(\ddot{\mathfrak{K}}, \sigma^2_{\ddot{\mathfrak{K}}}\right)$ be defined similarly. By change of variables and (S17), $\sigma^2_{\mathfrak{K}} = O(h)$. Similarly, $\sigma^2_{\dot{\mathfrak{K}}} \vee \sigma^2_{\ddot{\mathfrak{K}}} = O(h)$. Let $K_j(t) := t^j K(t)$ and $c_K := \sup_{t \in \mathbb{R}} K(t)$. It follows from Giné and Nickl (2015, Proposition 3.6.12) that for all $j = 0, 1, ..., p$,

$$
\left\{t \mapsto K_j\left(\frac{t-x}{b}\right) : x \in \mathcal{X}, b > 0\right\}
$$

is VC-type with respect to the constant envelope c_K. By (S17), Giné and Guillou (1999, Lemma 3 (b,c)) and Chernozhukov et al. (2014, Corollary A.1(i)), $\mathfrak{K}$ is also VC-type with respect to the constant envelope $F_{\mathfrak{K}} := (p+1)\varpi_p c_K$, where ϖ_p is defined in (S17). By similar arguments, $\dot{\mathfrak{K}}$ and $\ddot{\mathfrak{K}}$ are also VC-type with respect to some constant envelopes $F_{\dot{\mathfrak{K}}}$ and $F_{\ddot{\mathfrak{K}}}$.

Proposition A.2. *Assume that the assumptions in the statement of Proposition A.1 are satisfied. Assume that $\sqrt{|\log(h)|/(nh)} \downarrow 0$ as $n \uparrow \infty$. Then we have the following results. (a) There exist some positive constants c_1, c_2, c_3 which depend only on $K(\cdot)$, such that, for all n sufficiently large,*

$$
\Pr\left[\left\|\widehat{f} - \overline{f}\right\|_{\mathcal{X}} > \epsilon\right] \leq c_1 \cdot \exp\left(-c_2 \cdot \frac{(nh)\epsilon^2}{\left(\sigma^2_{\mathfrak{K}}/h\right) \vee F^2_{\mathfrak{K}}}\right) \tag{S31}
$$

if

$$
c_3\left(\frac{\sigma_{\mathfrak{K}}}{\sqrt{h}} \vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}} \leq \epsilon \leq \frac{c_3}{2}.
$$

(b) There exists positive constants c_1', c_2', c_3' which depend only on $K(\cdot)$, such that, for all n sufficiently large,

$$
\Pr\left[\left\|\widehat{f}' - \overline{f}'\right\|_{\mathcal{X}} > \epsilon\right] \leq c_1' \cdot \exp\left(-c_2' \cdot \frac{(nh^3)\epsilon^2}{\left(\sigma^2_{\dot{\mathfrak{K}}}/h\right) \vee F^2_{\dot{\mathfrak{K}}}}\right),
$$

if

$$
c_3'\left(\frac{\sigma_{\dot{\mathfrak{K}}}}{\sqrt{h}} \vee F_{\dot{\mathfrak{K}}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh^3}} \leq \epsilon \leq \frac{c_3'}{2h}.
$$

(c) There exists positive constants c_1'', c_2'', c_3'' which depend only on $K(\cdot)$, such that, for all n sufficiently large,

$$
\Pr\left[\left\|\widehat{f}'' - \overline{f}''\right\|_{\mathcal{X}} > \epsilon\right] \leq c_1'' \cdot \exp\left(-c_2'' \cdot \frac{(nh^5)\epsilon^2}{\left(\sigma^2_{\ddot{\mathfrak{K}}}/h\right) \vee F^2_{\ddot{\mathfrak{K}}}}\right),
$$

if

$$c_3''\left(\frac{\sigma_{\ddot{\mathfrak{K}}}}{\sqrt{h}}\vee F_{\ddot{\mathfrak{K}}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh^5}}\leq\epsilon\leq\frac{c_3''}{2h^2}.$$

Proof of Proposition A.2. We apply Giné and Guillou (2002, Corollary 2.2) with $\mathcal{F}=\mathfrak{K}$, $\sigma=\sigma_{\mathfrak{K}}\vee\left(F_{\mathfrak{K}}\sqrt{h}\right)$ and $U=2F_{\mathfrak{K}}$. Note that $\log\left(U/\sigma\right)\leq\log\left(2h^{-1/2}\right)$ and $\sqrt{n}\sigma\geq\sqrt{nh}F_{\mathfrak{K}}$ under these definitions. Therefore, we have

$$U\sqrt{\log\left(\frac{U}{\sigma}\right)}\leq 2F_{\mathfrak{K}}\sqrt{\log\left(2h^{-1/2}\right)}\leq\sqrt{nh}F_{\mathfrak{K}}\leq\sqrt{n}\sigma,$$

when n is sufficiently large so that $\sqrt{\log\left(2h^{-1/2}\right)/\left(nh\right)}\leq 1/2$. Therefore, Condition (2.5) in the statement of Giné and Guillou (2002, Corollary 2.2) is satisfied. Note that $\sigma^2/\left(Uh\right)\geq 1/2$ and

$$\frac{\sigma\sqrt{\log\left(U/\sigma\right)}}{\sqrt{n}h}\leq\frac{\left(\sigma_{\mathfrak{K}}\vee\left(F_{\mathfrak{K}}\sqrt{h}\right)\right)\sqrt{\log\left(2h^{-1/2}\right)}}{\sqrt{n}h}=O\left(\sqrt{\frac{\left|\log\left(h\right)\right|}{nh}}\right).$$

Therefore, when n is sufficiently large,

$$\sqrt{n}\sigma\sqrt{\log\left(\frac{U}{\sigma}\right)}<\frac{n\sigma^2}{U}.$$

The conclusion in Part (a) follows from applying Giné and Guillou (2002, Corollary 2.2). The conclusions in Part (b) and Part (c) follow from the same arguments. ■

Corollary A.1. *Assume that the assumptions in the statement of Proposition A.2 are satisfied. Then we have the following results. (a) There exists some $M>0$ such that*

$$\Pr\left[\left\|\hat{f}-\overline{f}\right\|_{\mathcal{X}}>M\left(\frac{\sigma_{\mathfrak{K}}}{\sqrt{h}}\vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right]=O\left(n^{-1}\right).$$

(b) There exists some $M'>0$ such that

$$\Pr\left[\left\|\hat{f}'-\overline{f}'\right\|_{\mathcal{X}}>M'\left(\frac{\sigma_{\dot{\mathfrak{K}}}}{\sqrt{h}}\vee F_{\dot{\mathfrak{K}}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh^3}}\right]=O\left(n^{-1}\right).$$

(c) There exists some $M''>0$ such that

$$\Pr\left[\left\|\hat{f}''-\overline{f}''\right\|_{\mathcal{X}}>M''\left(\frac{\sigma_{\ddot{\mathfrak{K}}}}{\sqrt{h}}\vee F_{\ddot{\mathfrak{K}}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh^5}}\right]=O\left(n^{-1}\right).$$

Proof of Corollary A.1. For all $C>1$, $C\left(\left(\sigma_{\mathfrak{K}}/\sqrt{h}\right)\vee F_{\mathfrak{K}}\right)\sqrt{\log\left(2h^{-1/2}\right)/\left(nh\right)}<1/2$ if n is sufficiently large. Therefore, for $\epsilon = Cc_3\left(\left(\sigma_{\mathfrak{K}}/\sqrt{h}\right) \vee F_{\mathfrak{K}}\right)\sqrt{\log\left(2h^{-1/2}\right)/(nh)}$, (S31) holds if n is sufficiently large. And it is easy to see that the right hand side of (S31) is $O\left(n^{-1}\right)$ if C is taken to be sufficiently large. The conclusions in Part (b) and Part (c) follow from similar arguments. ■

Let $\left\{X_1^\dagger, ..., X_n^\dagger\right\}$ be a nonparametric bootstrap sample from $\{X_1, ..., X_n\}$. Let $\widehat{f}_\dagger(x)$ be defined by the right hand side of (S18) with X_i replaced by $X_i^\dagger$. The following result is a bootstrap analogue of Corollary A.1.

Corollary A.2. *Assume that the assumptions in the statement of Proposition A.2 are satisfied. Then we have the following results. (a) There exists some positive constants* (M_1, M_2) *such that*

$$\Pr_\dagger\left[\left\|\widehat{f}_\dagger - \overline{f}\right\|_{\mathcal{X}} > M_1\left(\sqrt{\frac{\sigma_{\mathfrak{K}}^2}{h} + M_2\left(\frac{\sigma_{\mathfrak{Q}}}{\sqrt{h}} \vee F_{\mathfrak{K}}^2\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}} \vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right] = O_p\left(n^{-1}\right).$$

(b) There exists some positive constants (M_1', M_2') *such that*

$$\Pr_\dagger\left[\left\|\widehat{f}_\dagger' - \overline{f}'\right\|_{\mathcal{X}} > M_1'\left(\sqrt{\frac{\sigma_{\dot{\mathfrak{K}}}^2}{h} + M_2'\left(\frac{\sigma_{\dot{\mathfrak{Q}}}}{\sqrt{h}} \vee F_{\dot{\mathfrak{K}}}^2\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}} \vee F_{\dot{\mathfrak{K}}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh^3}}\right] = O_p\left(n^{-1}\right).$$

(c) There exists some positive constants (M_1'', M_2'') *such that*

$$\Pr_\dagger\left[\left\|\widehat{f}_\dagger'' - \overline{f}''\right\|_{\mathcal{X}} > M_1''\left(\sqrt{\frac{\sigma_{\ddot{\mathfrak{K}}}^2}{h} + M_2''\left(\frac{\sigma_{\ddot{\mathfrak{Q}}}}{\sqrt{h}} \vee F_{\ddot{\mathfrak{K}}}^2\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}} \vee F_{\ddot{\mathfrak{K}}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh^5}}\right] = O_p\left(n^{-1}\right).$$

Proof of Corollary A.2. Let

$$\widehat{q}(x) \coloneqq \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}_p^2\left(X_i, x \mid h, \underline{x}, \overline{x}\right)$$

and $\overline{q}(x) \coloneqq \mathrm{E}\left[\widehat{q}(x)\right]$. Let

$$\begin{aligned}
\mathfrak{Q} &\coloneqq \left\{\mathcal{K}_p^2\left(\cdot, x \mid h, \underline{x}, \overline{x}\right) : x \in \mathcal{X}\right\} \\
\sigma_{\mathfrak{Q}}^2 &\coloneqq \sup_{x\in\mathcal{X}} \mathrm{E}\left[\mathcal{K}_p^4\left(X, x \mid h, \underline{x}, \overline{x}\right)\right].
\end{aligned}$$

By Chernozhukov et al. (2014, Corollary A.1(ii)), $\mathfrak{Q}$ is VC-type with respect to the constant envelope

$F_{\mathfrak{K}}^2$. By change of variables and (S17), $\sigma_{\mathfrak{Q}}^2 = O(h)$. By similar arguments as those used in the proof of Part (a) of Corollary A.1, for some $M_2 > 0$,

$$\Pr\left[\|\widehat{q} - \overline{q}\|_{\mathcal{X}} > M_2 \left(\frac{\sigma_{\mathfrak{Q}}}{\sqrt{h}} \vee F_{\mathfrak{K}}^2\right) \sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right] = O\left(n^{-1}\right). \tag{S32}$$

Let

$$\widehat{\sigma}_{\mathfrak{K}}^2 \coloneqq \sup_{x \in \mathcal{X}} \mathrm{E}_{\dagger}\left[\mathcal{K}_p^2\left(X_i^{\dagger}, x \mid h, \underline{x}, \overline{x}\right)\right].$$

Since $\|\widehat{q}\|_{\mathcal{X}} = \widehat{\sigma}_{\mathfrak{K}}^2/h$ and $\|\overline{q}\|_{\mathcal{X}} = \sigma_{\mathfrak{K}}^2/h$, it follows from (S32) and the triangle inequality that

$$\Pr\left[\left|\frac{\widehat{\sigma}_{\mathfrak{K}}^2}{h} - \frac{\sigma_{\mathfrak{K}}^2}{h}\right| > M_2 \left(\frac{\sigma_{\mathfrak{Q}}}{\sqrt{h}} \vee F_{\mathfrak{K}}^2\right) \sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right] = O\left(n^{-1}\right) \tag{S33}$$

and $\left|\widehat{\sigma}_{\mathfrak{K}}^2/h - \sigma_{\mathfrak{K}}^2/h\right| = O_p\left(\sqrt{\log\left(2h^{-1/2}\right)/(nh)}\right)$. By Proposition A.2(a),

$$\Pr_{\dagger}\left[\left\|\widehat{f}_{\dagger} - \widehat{f}\right\|_{\mathcal{X}} > \epsilon\right] \leq c_1 \cdot \exp\left(-c_2 \cdot \frac{(nh)\,\epsilon^2}{\left(\widehat{\sigma}_{\mathfrak{K}}^2/h\right) \vee F_{\mathfrak{K}}^2}\right)$$

if

$$c_3\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}} \vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}} \leq \epsilon \leq \frac{c_3}{2}.$$

By this result, for all $C > 1$,

$$\begin{aligned}
&\Pr_{\dagger}\left[\left\|\widehat{f}_{\dagger} - \widehat{f}\right\|_{\mathcal{X}} > Cc_3\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}} \vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right] \\
&\qquad\qquad \leq c_1\left(2h^{-1/2}\right)^{-c_2 c_3^2 C^2} + \mathbb{1}\left(\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}} \vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}} > \frac{1}{2C}\right).
\end{aligned}$$

By $\left|\widehat{\sigma}_{\mathfrak{K}}^2/h - \sigma_{\mathfrak{K}}^2/h\right| = O_p\left(\sqrt{\log\left(2h^{-1/2}\right)/(nh)}\right)$ and the fact that $\sigma_{\mathfrak{K}}^2/h = O(1)$, the second term on the right hand side of the above inequality is zero wpa1. Therefore,

$$\Pr_{\dagger}\left[\left\|\widehat{f}_{\dagger} - \widehat{f}\right\|_{\mathcal{X}} > Cc_3\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}} \vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right] = O_p\left(n^{-1}\right), \tag{S34}$$

if C is taken to be sufficiently large. By the triangle inequality,

$$\Pr_{\dagger}\left[\left\|\widehat{f}_{\dagger} - \overline{f}\right\|_{\mathcal{X}} > Cc_3\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}} \vee \frac{\sigma_{\mathfrak{K}}}{\sqrt{h}} \vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right]$$

$$\leq \Pr_\dagger\left[\left\|\widehat{f}_\dagger-\widehat{f}\right\|_{\mathcal{X}}>\frac{Cc_3}{2}\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}}\vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right]$$

$$+\mathbb{1}\left(\left\|\widehat{f}-\overline{f}\right\|_{\mathcal{X}}>\frac{Cc_3}{2}\left(\frac{\sigma_{\mathfrak{K}}}{\sqrt{h}}\vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right).$$

By Corollary A.1(a), the second term on the right hand side of the above inequality is zero wpa1, if C is taken to be sufficiently large. By this result and (S34),

$$\Pr_\dagger\left[\left\|\widehat{f}_\dagger-\overline{f}\right\|_{\mathcal{X}}>Cc_3\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}}\vee\frac{\sigma_{\mathfrak{K}}}{\sqrt{h}}\vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right]=O_p\left(n^{-1}\right), \tag{S35}$$

if C is taken to be sufficiently large. Fix any $C>0$ and take M_2 to be such that (S32) holds, we have

$$\Pr_\dagger\left[\left\|\widehat{f}_\dagger-\overline{f}\right\|_{\mathcal{X}}>Cc_3\left(\sqrt{\frac{\sigma_{\mathfrak{K}}^2}{h}+M_2\left(\frac{\sigma_{\mathfrak{Q}}}{\sqrt{h}}\vee F_{\mathfrak{K}}^2\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}}\vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right]$$

$$\leq \Pr_\dagger\left[\left\|\widehat{f}_\dagger-\overline{f}\right\|_{\mathcal{X}}>Cc_3\left(\frac{\widehat{\sigma}_{\mathfrak{K}}}{\sqrt{h}}\vee\frac{\sigma_{\mathfrak{K}}}{\sqrt{h}}\vee F_{\mathfrak{K}}\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right]$$

$$+\mathbb{1}\left(\left|\frac{\widehat{\sigma}_{\mathfrak{K}}^2}{h}-\frac{\sigma_{\mathfrak{K}}^2}{h}\right|>M_2\left(\frac{\sigma_{\mathfrak{Q}}}{\sqrt{h}}\vee F_{\mathfrak{K}}^2\right)\sqrt{\frac{\log\left(2h^{-1/2}\right)}{nh}}\right).$$

It follows from (S33) that the second term on the right hand side of the above inequality is 0 wpa1. It follows from (S35) that the first term on the right hand side of the above inequality is $O_p\left(n^{-1}\right)$, if C is taken to be sufficiently large. The conclusion in Part (a) can be deduced from the above inequality and these results.

The conclusions in Part (b) and Part (c) follow from using similar arguments. ∎

Appendix B Proofs of the main results

Denote $\mathcal{B}_n \coloneqq \left[\underline{b}_n, \overline{b}_n\right]$. Smoothness results similar to those in Guerre et al. (2000, Proposition 1 and Lemmas A1 and A2) are summarized in the following lemma. Let

$$R\left(v \mid p_n, n\right) \coloneqq \frac{\int_v^{\overline{v}} H\left(t \mid p_n, n\right) dt}{H^2\left(v \mid p_n, n\right)}.$$

Let $\xi'\left(\cdot \mid p_n, n\right)$ denote the derivative of $\xi\left(\cdot \mid p_n, n\right)$. $R'\left(\cdot \mid p_n, n\right)$, $R''\left(\cdot \mid p_n, n\right)$, $H''\left(\cdot \mid p_n, n\right)$, $g'\left(\cdot \mid n\right)$, $g''\left(\cdot \mid n\right)$, $g'\left(\cdot, n\right)$ and $g''\left(\cdot, n\right)$ are defined similarly.

Lemma S1. *Assume that the assumptions in the statement of Proposition S1 are satisfied. Then we have the following results. (a) $\xi(\cdot \mid p_n, n)$ is thrice continuously differentiable on $\mathcal{B}_n$ and $\xi'(\cdot \mid p_n, n)$ is bounded away from zero on $\mathcal{B}_n$. (b) $g(\cdot \mid n)$ is twice continuously differentiable on $\mathcal{B}_n$ and $g(\cdot \mid n)$ is bounded away from zero on $\mathcal{B}_n$.*

Proof of Lemma S1. For any $v \in (\underline{v}, \overline{v})$, we have

$$\begin{aligned}\beta'(v \mid p_n, n) &= -\frac{\beta'(v \mid p_n, n) - v}{H(v \mid p_n, n)} \cdot H'(v \mid p_n, n) \\ &= -H'(v \mid p_n, n) R(v \mid p_n, n) \\ &= -(n-1) p_n f^*(v \mid p_n) (p_n (1 - F^*(v \mid p_n)) + (1 - p_n))^{n-2} R(v \mid p_n, n), \end{aligned} \tag{S36}$$

where

$$H'(v \mid p_n, n) = (n-1) p_n f^*(v \mid p_n) (p_n (1 - F^*(v \mid p_n)) + (1 - p_n))^{n-2}.$$

It is also easy to see that $\beta'(\cdot \mid p_n, n)$ is continuous on $(\underline{v}, \overline{v})$ and $\beta'(v \mid p_n, n) > 0$ for all $v \in (\underline{v}, \overline{v})$. By L'Hopital rule, $\lim_{v \uparrow \overline{v}} R(v \mid p_n, n) = -(2H'(\overline{v} \mid p_n, n))^{-1}$, where

$$H'(\overline{v} \mid p_n, n) = (n-1) p_n (1 - p_n)^{n-2} f^*(\overline{v} \mid p_n) > 0,$$

and hence, $0 < \lim_{v \uparrow \overline{v}} \beta'(v \mid p_n, n) < \infty$. It is straightforward to check that $0 < \lim_{v \downarrow \underline{v}} \beta'(v \mid p_n, n) < \infty$. Therefore, $\beta(\cdot \mid p_n, n)$ is continuously differentiable on $[\underline{v}, \overline{v}]$ and $\beta'(\cdot \mid p_n, n)$ is bounded away from zero on $[\underline{v}, \overline{v}]$. By the inverse function theorem,

$$\xi'(b \mid p_n, n) = \frac{1}{\beta'(\xi(b \mid p_n, n) \mid p_n, n)}, \tag{S37}$$

for $b \in (\underline{b}_n, \overline{b}_n)$. It follows that $\xi(\cdot \mid p_n, n)$ is continuously differentiable on $\mathcal{B}_n$ and $\xi'(\cdot \mid p_n, n)$ is bounded away from zero on $\mathcal{B}_n$. By the quotient rule, for $v \in (\underline{v}, \overline{v})$,

$$R'(v \mid p_n, n) = -\frac{H^2(v \mid p_n, n) + 2H'(v \mid p_n, n) \left(\int_v^{\overline{v}} H(t \mid p_n, n)\, dt\right)}{H^3(v \mid p_n, n)}$$

and

$$\begin{aligned} R''(v \mid p_n, n) = -H^{-4}(v \mid p_n, n) \Bigg\{ & 2H''(v \mid p_n, n) H(v \mid p_n, n) \left(\int_v^{\overline{v}} H(t \mid p_n, n)\, dt\right) \\ & -3H^2(v \mid p_n, n) H'(v \mid p_n, n) - 6\left(\int_v^{\overline{v}} H(t \mid p_n, n)\, dt\right) (H'(v \mid p_n, n))^2 \Bigg\}. \end{aligned}$$

Then, it is easy to check that the limits of $R'(v \mid p_n, n)$ and $R''(v \mid p_n, n)$ (as $v \downarrow \underline{v}$ or $v \uparrow \overline{v}$) all exist and are finite. It follows from this fact and (S36) that $\beta'(\cdot \mid p_n, n)$ is twice continuously differentiable on $[\underline{v}, \overline{v}]$. It follows from this fact and (S37) that $\xi'(\cdot \mid p_n, n)$ is twice continuously differentiable on $\mathcal{B}_n$.

The conclusion in Part (b) follows from results in Part (a) and

$$g(b \mid n) = f^*(\xi(b \mid p_n, n) \mid p_n)\, \xi'(b \mid p_n, n).$$

$\blacksquare$

Denote

$$\widehat{G}(b, n) := \frac{1}{L} \sum_{l: N_l = n} \sum_{i=1}^{N_l^*} \mathbb{1}(B_{il} \leq b)$$

and $G(b, n) := \mathrm{E}\left[\widehat{G}(b, n)\right] = r_n G(b \mid n)$, where

$$\begin{aligned} r_n &:= \mathrm{E}\left[\mathbb{1}(N_l = n) N_l^*\right] \\ &= \pi_n q_n n, \end{aligned}$$

and the second equality follows from LIE. Let

$$\widehat{g}(b, n) := \frac{1}{L} \sum_{l: N_l = n} \sum_{i=1}^{N_l^*} \frac{1}{h} \mathcal{K}_1\left(B_{il}, b \mid h, \underline{\widehat{b}}_n, \overline{\widehat{b}}_n\right) \tag{S38}$$

and $g(b, n) := r_n g(b \mid n)$. We can now write $\widehat{G}(b \mid n) := \widehat{G}(b, n) / \widehat{r}_n$ and $\widehat{g}(b \mid n) = \widehat{g}(b, n) / \widehat{r}_n$. Denote $\mathbb{G}(b \mid n) := \widehat{G}(b \mid n) - G(b \mid n)$ and $\mathbb{H}(b \mid n) := \widehat{g}(b \mid n) - g(b \mid n)$. Let $\mathbb{G}(b, n) := \widehat{G}(b, n) - G(b, n)$ and $\mathbb{H}(b, n) := \widehat{g}(b, n) - g(b, n)$. Let $\widehat{g}'(\cdot \mid n)$, $\widehat{g}''(\cdot \mid n)$, $\widehat{g}'(\cdot, n)$ and $\widehat{g}''(\cdot, n)$ be the first and second derivatives of $\widehat{g}(\cdot \mid n)$ and $\widehat{g}(\cdot, n)$. Let $\mathbb{H}'(\cdot \mid n)$ and $\mathbb{H}''(\cdot \mid n)$ be the first and second derivatives of $\mathbb{H}(\cdot \mid n)$. Let

$$\widehat{\pi}_n := \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}(N_l = n).$$

Then we can write $\widehat{q}_n = \widehat{r}_n / (n \widehat{\pi}_n)$. The following lemma collects results on the rates of convergence of $\widehat{q}_n$, $\widehat{G}(\cdot \mid n)$ and $\widehat{g}(\cdot \mid n)$ (and its derivatives).

Lemma S2. *Assume that the assumptions in the statement of Proposition S1 are satisfied. Then we have the following results. (a)*

$$\Pr\left[|\widehat{p}_n - p_n| \geq \alpha_L^p\right] = O\left(L^{-1}\right),$$

for some $\alpha_L^p = O\left(\sqrt{\log(L)/L}\right)$. *(b)*

$$\Pr\left[\|\mathbb{G}(\cdot \mid n)\|_{\mathcal{B}_n} \geq \bar{\alpha}_L\right] = O\left(L^{-1}\right),$$

for some $\bar{\alpha}_L = O\left(\sqrt{\log(L)/L}\right)$. *(c)*

$$\Pr\left[\|\mathbb{H}(\cdot \mid n)\|_{\mathcal{B}_n} \geq \alpha_L\right] = O\left(L^{-1}\right),$$

for some $\alpha_L = O\left(\sqrt{\log(L)/(Lh)}\right)$ *and similar results with* $\alpha'_L = O\left(\sqrt{\log(L)/(Lh^3)}\right)$ *and* $\alpha''_L = O\left(\sqrt{\log(L)/(Lh^5)}\right)$ *hold for* $\|\mathbb{H}'(\cdot \mid n)\|_{\mathcal{B}_n}$ *and* $\|\mathbb{H}''(\cdot \mid n)\|_{\mathcal{B}_n}$.

Proof of Lemma S2. By Bernstein's inequality (Giné and Nickl, Theorem 3.1.7),

$$\Pr\left[|\widehat{\pi}_n - \pi_n| \geq \sigma_{\pi,n}\sqrt{2 \cdot \frac{\log(L)}{L}}\right] = O\left(L^{-1}\right), \tag{S39}$$

where $\sigma^2_{\pi,n} := \pi_n - \pi_n^2$. By this result and simple calculation,

$$\Pr\left[\left|\frac{\pi_n}{\widehat{\pi}_n} - 1\right| \geq \frac{\sigma_{\pi,n}\sqrt{2 \cdot \log(L)/L}}{\pi_n - \sigma_{\pi,n}\sqrt{2 \cdot \log(L)/L}}\right] = O\left(L^{-1}\right). \tag{S40}$$

Similarly, by Bernstein's inequality,

$$\Pr\left[|\widehat{r}_n - r_n| \geq \sigma_{r,n}\sqrt{2 \cdot \frac{\log(L)}{L}}\right] = O\left(L^{-1}\right), \tag{S41}$$

where $\sigma_{r,n} := \pi_n \mathrm{E}\left[(N_l^*)^2 \mid N_l = n\right] - r_n^2$. By this result, (S40) and

$$\widehat{q}_n - q_n = \frac{\widehat{r}_n - r_n}{n\pi_n} + \frac{\widehat{r}_n}{n\pi_n}\left(\frac{\pi_n}{\widehat{\pi}_n} - 1\right),$$

we have $\Pr\left[|\widehat{q}_n - q_n| \geq \alpha_L^q\right] = O\left(L^{-1}\right)$ for some $\alpha_L^q = O\left(\sqrt{\log(L)/L}\right)$. Then, by the mean value theorem and taking $\alpha_L^p := \left\|\left(\varphi_n^{-1}\right)'\right\|_{\left[q_n \pm \alpha_L^q\right]\cap[0,1]} \alpha_L^q$, we have

$$\Pr\left[|\widehat{p}_n - p_n| \geq \alpha_L^p\right] \leq \Pr\left[|\widehat{q}_n - q_n| \geq \alpha_L^q\right] = O\left(L^{-1}\right).$$

By (S41) and simple calculation,

$$\Pr\left[\left|\frac{r_n}{\widehat{r}_n} - 1\right| \geq \frac{\sigma_{r,n}\sqrt{2 \cdot \log(L)/L}}{r_n - \sigma_{r,n}\sqrt{2 \cdot \log(L)/L}}\right] = O\left(L^{-1}\right). \tag{S42}$$

Let $(B_{1l}, B_{2l}, ..., B_{nl})$ be i.i.d., $\boldsymbol{B}_l := (B_{1l}, ..., B_{nl}, N_l^*, N_l)^\top$ and

$$\mathcal{G}(\boldsymbol{B}_l; b) := \mathbb{1}(N_l = n)\sum_{i=1}^{N_l^*} \mathbb{1}(B_{il} \leq b).$$

Then we can write $\widehat{G}(b, n) := L^{-1}\sum_{l=1}^{L} \mathcal{G}(\boldsymbol{B}_l; b)$. By Kosorok (2008, Lemma 9.8), Giné and Nickl (2015, Theorem 3.6.9) and Nolan and Pollard (1987, Corollary 17), $\{\mathcal{G}(\cdot; b) : b \in \mathcal{B}_n\}$ is VC-type respect to the constant envelope n. Then we apply Giné and Guillou (2002, Corollary 2.2) with

$U = 2n$, $\sigma^2 = n^2\pi_n$ and t taken to be $C\sqrt{\log(L)\,L}$ for some positive constant C. Note that Equations (2.5) and (2.6) of Giné and Guillou (2002) hold for all L large enough. By Giné and Guillou (2002, Corollary 2.2), taking C to be sufficiently large, we have

$$\Pr\left[\|\mathbb{G}(\cdot,n)\|_{\mathcal{B}_n} \geq \breve{\breve{\alpha}}_L\right] = O\left(L^{-1}\right),$$

for some $\breve{\breve{\alpha}}_L = O\left(\sqrt{\log(L)/L}\right)$. The conclusion in Part (b) follows from this result, (S42) and

$$\mathbb{G}(b \mid n) = \frac{\mathbb{G}(b,n)}{r_n} + \frac{\widehat{G}(b,n)}{r_n}\left(\frac{r_n}{\widehat{r}_n} - 1\right).$$

By (straightforward adaptations of) Proposition A.1(a) and Corollary A.1(a),

$$\Pr\left[\|\mathbb{H}(\cdot,n)\|_{\mathcal{B}_n} \geq \alpha_L^g\right] = O\left(L^{-1}\right),$$

for some $\alpha_L^g = O\left(\sqrt{\log(L)/(Lh)}\right)$. The first conclusion in Part (c) follows from this result, (S42) and

$$\mathbb{H}(b \mid n) = \frac{\mathbb{H}(b,n)}{r_n} + \frac{\widehat{g}(b,n)}{r_n}\left(\frac{r_n}{\widehat{r}_n} - 1\right). \tag{S43}$$

The other results follow from similar arguments, Proposition A.1(b,c) and Corollary A.1(b,c). ■

Proof of Proposition S1. Let

$$\widehat{F}^*(v,n) \quad := \quad \frac{1}{L}\sum_{l:N_l=n}\sum_{i=1}^{N_l^*}\mathbb{1}\left(\widehat{V}_{il} \leq v\right).$$

Now we can write $\widehat{F}^*(v \mid p_n) := \widehat{F}^*(v,n)/\widehat{r}_n$. Denote

$$\widetilde{\xi}(b \mid n) := b - \frac{\eta_n(p_n, G(b \mid n))}{(n-1)\,\widehat{g}(b \mid n)}.$$

Let $\widetilde{V}_{il} := \widetilde{\xi}(B_{il} \mid N_l)$ and $\widetilde{F}^*(v \mid p_n) := \widetilde{F}^*(v,n)/r_n$, where

$$\widetilde{F}^*(v,n) := \frac{1}{L}\sum_{l:N_l=n}\sum_{i=1}^{N_l^*}\mathbb{1}\left(\widetilde{V}_{il} \leq v\right).$$

Then, we decompose

$$\widehat{F}^*(v \mid p_n) - F^*(v \mid p_n) = \left\{\widetilde{F}^*(v \mid p_n) - F^*(v \mid p_n)\right\} + \left\{\widehat{F}^*(v \mid p_n) - \widetilde{F}^*(v \mid p_n)\right\}, \tag{S44}$$

and

$$\widehat{F}^*(v \mid p_n) - \widetilde{F}^*(v \mid p_n) = \frac{\widehat{F}^*(v,n) - \widetilde{F}^*(v,n)}{r_n} + \frac{\widehat{F}^*(v,n)}{r_n}\left(\frac{r_n}{\widehat{r}_n} - 1\right). \tag{S45}$$

Let $\mathbb{X}(b \mid n) \coloneqq \widetilde{\xi}(b \mid n) - \xi(b \mid p_n, n)$. Let $\mathbb{X}'(\cdot \mid n)$ and $\mathbb{X}''(\cdot \mid n)$ denote the first and second derivatives of $\mathbb{X}(\cdot \mid n)$. Then, by straightforward calculation, we have

$$\mathbb{X}(b \mid n) = -\frac{\eta_n(p_n, G(b \mid n))}{n-1}\left\{\frac{1}{\widehat{g}(b \mid n)} - \frac{1}{g(b \mid n)}\right\}, \tag{S46}$$

$$\begin{aligned}\mathbb{X}'(b \mid n) &= -\frac{D_2\eta_n(p_n, G(b \mid n))\, g(b \mid n)}{n-1}\left\{\frac{1}{\widehat{g}(b \mid n)} - \frac{1}{g(b \mid n)}\right\} \\ &\quad + \frac{\eta_n(p_n, G(b \mid n))}{n-1}\left\{\frac{\widehat{g}'(b \mid n)}{\widehat{g}^2(b \mid n)} - \frac{g'(b \mid n)}{g^2(b \mid n)}\right\},\end{aligned} \tag{S47}$$

and

$$\begin{aligned}\mathbb{X}''(b \mid n) &= -\frac{D_2^2\eta_n(p_n, G(b \mid n))\, g^2(b \mid n) + D_2\eta_n(p_n, G(b \mid n))\, g'(b \mid n)}{n-1}\left\{\frac{1}{\widehat{g}(b \mid n)} - \frac{1}{g(b \mid n)}\right\} \\ &\quad + \frac{2D_2\eta_n(p_n, G(b \mid n))\, g(b \mid n)}{n-1}\left\{\frac{\widehat{g}'(b \mid n)}{\widehat{g}^2(b \mid n)} - \frac{g'(b \mid n)}{g^2(b \mid n)}\right\} \\ &\quad + \frac{\eta_n(p_n, G(b \mid n))}{n-1}\left\{\frac{\widehat{g}''(b \mid n)\,\widehat{g}^2(b \mid n) - 2\left(\widehat{g}'(b \mid n)\right)^2 \widehat{g}(b \mid n)}{\widehat{g}^4(b \mid n)}\right. \\ &\quad \left. - \frac{g''(b \mid n)\, g^2(b \mid n) - 2\left(g'(b \mid n)\right)^2 g(b \mid n)}{g^4(b \mid n)}\right\}.\end{aligned} \tag{S48}$$

Denote

$$\mathbb{K}(b \mid n) \coloneqq \eta_n\left(\widehat{p}_n, \widehat{G}(b \mid n)\right) - \eta_n(p_n, G(b \mid n)).$$

By straightforward calculation, we have

$$\|\mathbb{K}(\cdot \mid n)\|_{\mathcal{B}_n} \le \left(\|D_1\eta_n\|_{[p_n \pm \alpha_L^p]\times[0,1]} \vee \|D_2\eta_n\|_{[p_n \pm \alpha_L^p]\times[0,1]}\right)\left(|\widehat{p}_n - p_n| + \|\mathbb{G}(\cdot \mid n)\|_{\mathcal{B}_n}\right),$$

if $|\widehat{p}_n - p_n| < \alpha_L^p$. We have $\|D_1\eta_n\|_{[p_n \pm \alpha_L^p]\times[0,1]} = O(1)$ and $\|D_2\eta_n\|_{[p_n \pm \alpha_L^p]\times[0,1]} = O(1)$ by straightforward calculation. Therefore, by these results and Lemma S2(a,b),

$$\Pr\left[\|\mathbb{K}(\cdot \mid n)\|_{\mathcal{B}_n} \ge \tilde{\alpha}_L\right] = O\left(L^{-1}\right), \tag{S49}$$

for some $\tilde{\alpha}_L = O\left(\sqrt{\log(L)/L}\right)$. Let $\bar{\mathbb{T}} \coloneqq \mathbb{1}\left(\|\mathbb{K}(\cdot \mid n)\|_{\mathcal{B}_n} < \tilde{\alpha}_L\right)$. Let $(\mathbb{T}, \mathbb{T}', \mathbb{T}'')$ be defined by the same formula with $(\mathbb{K}(\cdot \mid n), \tilde{\alpha}_L)$ replaced by $(\mathbb{H}, \alpha_L)$, $(\mathbb{H}', \alpha_L')$ and $(\mathbb{H}'', \alpha_L'')$, where $(\alpha_L, \alpha_L', \alpha_L'')$ are defined in the statement of Lemma S2. Let $\mathbb{I} \coloneqq \bar{\mathbb{T}}\mathbb{T}\mathbb{T}'\mathbb{T}''$. It follows from (S49) and Lemma S2 that $\Pr[\mathbb{I} = 0] = O\left(L^{-1}\right)$.

Note that

$$\widehat{\xi}(b \mid n) - \widetilde{\xi}(b \mid n) = -\frac{\mathbb{K}(b \mid n)}{(n-1)\,\widehat{g}(b \mid n)}.$$

Denote $\delta_L := \left\|\widehat{\xi}\left(\cdot \mid n\right) - \widetilde{\xi}\left(\cdot \mid n\right)\right\|_{\mathcal{B}_n}$. Then, by the triangle inequality,

$$\left|\widehat{F}^*\left(v,n\right) - \widetilde{F}^*\left(v,n\right)\right| \leq \frac{1}{L}\sum_{l:N_l=n}\sum_{i=1}^{N_l^*}\left\{\mathbb{1}\left(\widetilde{V}_{il} \leq v+\delta_L\right) - \mathbb{1}\left(\widetilde{V}_{il} \leq v-\delta_L\right)\right\}.$$

Let $\overline{v}_n := \widetilde{\xi}\left(\overline{b}_n \mid n\right)$ and $\underline{v}_n := \widetilde{\xi}\left(\underline{b}_n \mid n\right)$. It is easy to verify by straightforward calculation that $\left\|\eta_n\left(p_n,\cdot\right)\right\|_{[0,1]} < \infty$ and $\left\|D_2\eta_n\left(p_n,\cdot\right)\right\|_{[0,1]} < \infty$. By these results, (S47) and Lemma S1(b), $\left\|\mathbb{X}'\left(\cdot \mid n\right)\right\|_{\mathcal{B}_n}$ is sufficiently small, when $\mathbb{I} = 1$ and L is large enough. By Lemma S1(a), $\widetilde{\xi}\left(\cdot \mid n\right)$ is strictly increasing on $\mathcal{B}_n$ and its inverse function $\widetilde{\beta}\left(\cdot \mid n\right) := \widetilde{\xi}^{-1}\left(\cdot \mid n\right)$ exists, if $\mathbb{I} = 1$ and L is sufficiently large. $\widetilde{\beta}\left(\cdot \mid n\right)$ is a strictly increasing function on $[\underline{v}_n, \overline{v}_n]$. We can write

$$\mathbb{1}\left(\widetilde{\xi}\left(B_{il} \mid n\right) \leq y\right) = \mathbb{1}\left(y \geq \overline{v}_n\right) + \mathbb{1}\left(y \in \left(\underline{v}_n, \overline{v}_n\right)\right)\mathbb{1}\left(B_{il} \leq \widetilde{\beta}\left(y \mid n\right)\right) \tag{S50}$$

when $\widetilde{\xi}\left(\cdot \mid n\right)$ is strictly increasing. If $\mathbb{I} = 1$ and L is large enough, $\delta_L \lesssim \sqrt{\log\left(L\right)/L}$, $\left|\overline{v}_n - \overline{v}\right| \lesssim \sqrt{\log\left(L\right)/\left(Lh\right)}$ and $\left|\underline{v}_n - \underline{v}\right| \lesssim \sqrt{\log\left(L\right)/\left(Lh\right)}$. We have

$$\begin{aligned}
&\mathbb{I}\cdot\frac{1}{L}\sum_{l:N_l=n}\sum_{i=1}^{N_l^*}\left\{\mathbb{1}\left(\widetilde{V}_{il} \leq v+\delta_L\right) - \mathbb{1}\left(\widetilde{V}_{il} \leq v-\delta_L\right)\right\} \\
&\qquad = \mathbb{I}\left\{\widehat{G}\left(\widetilde{\beta}\left(v+\delta_L \mid n\right),n\right) - \widehat{G}\left(\widetilde{\beta}\left(v-\delta_L \mid n\right),n\right)\right\},
\end{aligned}$$

when L is sufficiently large. Then,

$$\begin{aligned}
\mathbb{I}\left|\widehat{F}^*\left(v,n\right) - \widetilde{F}^*\left(v,n\right)\right| \;&\leq\; \mathbb{I}\left\{\mathbb{G}\left(\widetilde{\beta}\left(v+\delta_L \mid n\right),n\right) - \mathbb{G}\left(\widetilde{\beta}\left(v-\delta_L \mid n\right),n\right)\right\} \\
&\quad + \mathbb{I}\left\{G\left(\widetilde{\beta}\left(v+\delta_L \mid n\right),n\right) - G\left(\widetilde{\beta}\left(v-\delta_L \mid n\right),n\right)\right\},
\end{aligned} \tag{S51}$$

when L is sufficiently large. By Dette et al. (2006, Lemma A.1), if $\mathbb{I} = 1$ and L is large enough, there exists $\lambda \in (0,1)$ such that

$$\widetilde{\beta}\left(v+\delta_L \mid n\right) - \beta\left(v+\delta_L \mid p_n,n\right) = -\frac{\mathbb{X}\left(b_\lambda \mid n\right)}{\xi'\left(b_\lambda \mid p_n,n\right) + \lambda\mathbb{X}'\left(b_\lambda \mid n\right)},$$

where $b_\lambda := \left(\xi\left(\cdot \mid p_n,n\right) + \lambda\mathbb{X}\left(\cdot \mid n\right)\right)^{-1}\left(v+\delta_L\right)$ and by (S46), (S47), Lemmas S1 and S2,

$$\begin{aligned}
\left|\widetilde{\beta}\left(v+\delta_L \mid n\right) - \beta\left(v+\delta_L \mid p_n,n\right)\right| \;&\leq\; \frac{\left\|\mathbb{X}\left(\cdot \mid n\right)\right\|_{\mathcal{B}_n}}{\inf_{b\in\mathcal{B}_n}\xi'\left(b \mid p_n,n\right) - \left\|\mathbb{X}'\left(\cdot \mid n\right)\right\|_{\mathcal{B}_n}} \\
&=\; O\left(\sqrt{\frac{\log\left(L\right)}{Lh}}\right).
\end{aligned} \tag{S52}$$

By mean value expansion, if $\mathbb{I} = 1$ and L is large enough, $\left|\beta\left(v-\delta_L \mid p_n,n\right) - \beta\left(v \mid p_n,n\right)\right| \lesssim \sqrt{\log\left(L\right)/L}$. Therefore, if $\mathbb{I} = 1$ and L is large enough, we have $\left|\widetilde{\beta}\left(v+\delta_L \mid n\right) - \beta\left(v \mid p_n,n\right)\right| \lesssim$

$\sqrt{\log(L)/(Lh)}$ and similarly, $\left|\widetilde{\beta}(v-\delta_L \mid n)-\beta(v \mid p_n, n)\right| \lesssim \sqrt{\log(L)/(Lh)}$. Therefore, when L is sufficiently large, the first term on the right hand side of (S51) can be bounded by

$$\begin{aligned}
&\mathbb{I}\left\{\mathbb{G}\left(\widetilde{\beta}(v+\delta_L \mid n), n\right)-\mathbb{G}\left(\widetilde{\beta}(v-\delta_L \mid n), n\right)\right\} \\
&\qquad\qquad \leq \sup_{(t_1,t_2)\in[-\epsilon_L,\epsilon_L]^2}\left|\mathbb{G}\left(\beta(v \mid p_n, n)+t_1, n\right)-\mathbb{G}\left(\beta(v \mid p_n, n)+t_2, n\right)\right|, \quad \text{(S53)}
\end{aligned}$$

for some $\epsilon_L = O\left(\sqrt{\log(L)/(Lh)}\right)$. By calculation and the mean value theorem,

$$\begin{aligned}
& \sup_{(t_1,t_2)\in[-\epsilon_L,\epsilon_L]^2} \mathrm{E}\left[\left(\mathcal{G}\left(\boldsymbol{B}_l;\beta(v \mid p_n, n)+t_1\right)-\mathcal{G}\left(\boldsymbol{B}_l;\beta(v \mid p_n, n)+t_2\right)\right)^2\right] \\
\leq\ & \sup_{(t_1,t_2)\in[0,\epsilon_L]^2} \mathrm{E}\left[\left(\mathcal{G}\left(\boldsymbol{B}_l;\beta(v \mid p_n, n)+t_1\right)-\mathcal{G}\left(\boldsymbol{B}_l;\beta(v \mid p_n, n)-t_2\right)\right)^2\right] \\
\leq\ & \sup_{(t_1,t_2)\in[0,\epsilon_L]^2} G\left(\beta(v \mid p_n, n)+t_1, n\right)-G\left(\beta(v \mid p_n, n)-t_2, n\right) \qquad \text{(S54)} \\
=\ & O\left(\sqrt{\frac{\log(L)}{Lh}}\right). \qquad \text{(S55)}
\end{aligned}$$

Let $\mathfrak{G} \coloneqq \{\mathcal{G}(\cdot;\beta(v \mid p_n, n)+t) : t \in [-\epsilon_L, \epsilon_L]\}$. By Nolan and Pollard (1987, Corollary 17), the function class $\{f-g : f, g \in \mathfrak{G}\}$ is VC-type respect to a constant envelope. By Chernozhukov et al. (2014, Corollary 5.1) with $\mathcal{F}$ taken to be $\{f-g : f, g \in \mathfrak{G}\}$, F taken to be a constant envelope, and σ^2 taken to be the term on the left hand side of the first inequality in (S55), we have

$$\mathrm{E}\left[\sup_{(t_1,t_2)\in[-\epsilon_L,\epsilon_L]^2}\left|\mathbb{G}\left(\beta(v \mid p_n, n)+t_1, n\right)-\mathbb{G}\left(\beta(v \mid p_n, n)+t_2, n\right)\right|\right] = O\left(L^{-1/2}\left(\frac{\log(L)}{Lh}\right)^{1/4}\right). \tag{S56}$$

By the mean value and inverse function theorems, if $\mathbb{I}=1$ and L is sufficiently large,

$$\begin{aligned}
\left|\widetilde{\beta}(v+\delta_L \mid n)-\widetilde{\beta}(v-\delta_L \mid n)\right| &\leq \frac{2\delta_L}{\inf_{b\in\mathcal{B}_n}\xi'(b \mid p_n, n)-\left\|\mathbb{X}'(\cdot \mid n)\right\|_{\mathcal{B}_n}} \\
&= O\left(\sqrt{\frac{\log(L)}{L}}\right).
\end{aligned}$$

By the above result and Lemma S1(b),

$$\mathbb{I}\left|G\left(\widetilde{\beta}(v+\delta_L \mid n), n\right)-G\left(\widetilde{\beta}(v-\delta_L \mid n), n\right)\right| \lesssim \sqrt{\frac{\log(L)}{L}}, \tag{S57}$$

if L sufficiently large. Now by this result, (S51), (S53) and (S56),

$$\mathbb{I}\left|\widehat{F}^*(v, n)-\widetilde{F}^*(v, n)\right| = O_p\left(\sqrt{\frac{\log(L)}{L}}\right).$$

It follows from this result, (S42), (S45) and $\Pr[\mathbb{I}=0]=O\left(L^{-1}\right)$ that

$$\widehat{F}^{*}(v\mid p_n)-\widetilde{F}^{*}(v\mid p_n)=O_p\left(\sqrt{\frac{\log(L)}{L}}\right).$$

By, this result and (S44), we have

$$\widehat{F}^{*}(v\mid p_n)-F^{*}(v\mid p_n)=\left\{\widetilde{F}^{*}(v\mid p_n)-F^{*}(v\mid p_n)\right\}+O_p\left(\sqrt{\frac{\log(L)}{L}}\right).$$

By using (S50), write

$$\mathbb{I}\left(\widetilde{F}^{*}(v\mid p_n)-F^{*}(v\mid p_n)\right)=\mathbb{I}\cdot\frac{\mathbb{G}\left(\widetilde{\beta}(v\mid n),n\right)}{r_n}+\mathbb{I}\left\{G\left(\widetilde{\beta}(v\mid n)\mid n\right)-G\left(\beta(v\mid p_n,n)\mid n\right)\right\},$$

where the last equality holds when L is sufficiently large. The first term on the right hand side can be bounded by $\|\mathbb{G}(\cdot,n)\|_{\mathcal{B}_n}=O_p\left(\sqrt{\log(L)/L}\right)$. By similar arguments in the proof of (S52), $\left|\widetilde{\beta}(v\mid n)-\beta(v\mid p_n,n)\right|\lesssim\sqrt{\log(L)/(Lh)}$ if $\mathbb{I}=1$ and L is sufficiently large. By using this result and the mean value theorem, we have

$$\begin{aligned}&\mathbb{I}\left\{G\left(\widetilde{\beta}(v\mid n)\mid n\right)-G\left(\beta(v\mid p_n,n)\mid n\right)\right\}\\&\qquad=\mathbb{I}\cdot g\left(\beta(v\mid p_n,n)\mid n\right)\left(\widetilde{\beta}(v\mid n)-\beta(v\mid p_n,n)\right)+O\left(\frac{\log(L)}{Lh}\right).\end{aligned}$$

By Dette et al. (2006, Lemma A.1), if $\mathbb{I}=1$ and L is sufficiently large, there exists $\lambda\in(0,1)$ such that

$$\begin{aligned}\widetilde{\beta}(v\mid n)-\beta(v\mid p_n,n)&=-\frac{\mathbb{X}\left(\beta(v\mid p_n,n)\mid n\right)}{\xi'\left(\beta(v\mid p_n,n)\mid p_n,n\right)}-\frac{2\mathbb{X}\left(\tilde{b}_\lambda\mid n\right)\mathbb{X}'\left(\tilde{b}_\lambda\mid n\right)}{\left(\xi'\left(\tilde{b}_\lambda\mid p_n,n\right)+\lambda\mathbb{X}'\left(\tilde{b}_\lambda\mid n\right)\right)^2}\\&\quad+\frac{\mathbb{X}^2\left(\tilde{b}_\lambda\mid n\right)\left(\xi''\left(\tilde{b}_\lambda\mid p_n,n\right)+\lambda\mathbb{X}''\left(\tilde{b}_\lambda\mid n\right)\right)}{\left(\xi'\left(\tilde{b}_\lambda\mid p_n,n\right)+\lambda\mathbb{X}'\left(\tilde{b}_\lambda\mid n\right)\right)^3},\end{aligned}$$

where $\tilde{b}_\lambda\coloneqq\left(\xi(\cdot\mid p_n,n)+\lambda\mathbb{X}(\cdot\mid n)\right)^{-1}(v)$. It now follows from the above result, (S46), (S47), (S48) and Lemmas S1 and S2 that

$$\widehat{F}^{*}(v\mid p_n)-F^{*}(v\mid p_n)=-\frac{g\left(\beta(v\mid p_n,n)\mid n\right)}{\xi'\left(\beta(v\mid p_n,n)\mid p_n,n\right)}\mathbb{X}\left(\beta(v\mid p_n,n)\mid n\right)+O_p\left(\frac{\log(L)}{Lh}\right).\tag{S58}$$

By (S46),

$$\widehat{g}^{-1}(b\mid n)-g^{-1}(b\mid n)=-\frac{\mathbb{H}(b\mid n)}{g^2(b\mid n)}-\frac{\mathbb{H}(b\mid n)}{g^2(b\mid n)}\left(\frac{g(b\mid n)}{\widehat{g}(b\mid n)}-1\right),$$

Lemma S1(b) and Lemma S2(c), we have

$$\mathbb{X}\left(\beta\left(v \mid p_n, n\right) \mid n\right)=\frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right)}{(n-1)\, g^2\left(\beta\left(v \mid p_n, n\right) \mid n\right)} \mathbb{H}\left(\beta\left(v \mid p_n, n\right) \mid n\right)+O_p\left(\frac{\log (L)}{L h}\right).$$

Then, by this result and (S58),

$$\begin{aligned}\widehat{F}^*\left(v \mid p_n\right)-F^*\left(v \mid p_n\right) &= -\frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right)}{(n-1)\, \xi'\left(\beta\left(v \mid p_n, n\right) \mid p_n, n\right) g\left(\beta\left(v \mid p_n, n\right) \mid n\right)} \mathbb{H}\left(\beta\left(v \mid p_n, n\right) \mid n\right) \\ &\quad +O_p\left(\frac{\log (L)}{L h}\right). \qquad \text{(S59)}\end{aligned}$$

For $b \in\left[\underline{\widehat{b}}_n+h, \overline{\widehat{b}}_n-h\right]$, we have

$$\widehat{g}(b, n)=\frac{1}{L} \sum_{l: N_l=n} \sum_{i=1}^{N_l^*} \frac{1}{h} K\left(\frac{B_{i l}-b}{h}\right).$$

By Taylor expansion,

$$\mathrm{E}\left[\widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right)\right]=g\left(\beta\left(v \mid p_n, n\right), n\right)+\frac{1}{2} g''\left(\beta\left(v \mid p_n, n\right) \mid n\right) r_n\left(\int u^2 K(u)\, d u\right) h^2+o\left(h^2\right). \qquad \text{(S60)}$$

Let

$$J_l(v, n) \coloneqq \mathbb{1}\left(N_l=n\right) \sum_{i=1}^{N_l^*} \frac{1}{\sqrt{h}} K\left(\frac{B_{i l}-\beta\left(v \mid p_n, n\right)}{h}\right)-\mathrm{E}\left[\mathbb{1}\left(N_l=n\right) \sum_{i=1}^{N_l^*} \frac{1}{\sqrt{h}} K\left(\frac{B_{i l}-\beta\left(v \mid p_n, n\right)}{h}\right)\right]$$

and $\sigma_g^2(v, n) \coloneqq \mathrm{E}\left[\left(\sum_{l=1}^L\left(J_l(v, n) / \sqrt{L}\right)\right)^2\right]=\mathrm{E}\left[J_l^2(v, n)\right]$. Then we can write

$$\sqrt{L h}\left(\widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right)-\mathrm{E}\left[\widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right)\right]\right)=\sum_{l=1}^L \frac{J_l(v, n)}{\sqrt{L}}.$$

By LIE and Taylor expansion, we have

$$\sigma_g^2(v, n)=g\left(\beta\left(v \mid p_n, n\right), n\right) \int K^2(u)\, d u+o(1) \qquad \text{(S61)}$$

and

$$\sum_{l=1}^L \mathrm{E}\left[\left|\frac{J_l(v, n)}{\sigma_g(v, n)}\right|^3\right]=O\left((L h)^{-1 / 2}\right),$$

which shows that Lyapunov's condition holds. By Lyapunov's central limit theorem (Severini, 2005, Theorem 12.2), $\sum_{l=1}^{L} J_l(v,n)/\sigma_g(v,n) \to_d \mathrm{N}(0,1)$. By this result, (S61) and Slutsky's theorem,

$$\sqrt{Lh}\left(\widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right) - \mathrm{E}\left[\widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right)\right]\right) \to_d \mathrm{N}\left(0, g\left(\beta\left(v \mid p_n, n\right), n\right) \int K^2(u)\, du\right). \tag{S62}$$

By (S42), (S43) and Lemma S2(c),

$$\mathbb{H}\left(\beta\left(v \mid p_n, n\right) \mid n\right) = \frac{\widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right) - g\left(\beta\left(v \mid p_n, n\right), n\right)}{r_n} + O_p\left(\frac{\log(L)}{L}\right). \tag{S63}$$

By this result, (S60) and (S62),

$$\begin{aligned}\sqrt{Lh}\left(\mathbb{H}\left(\beta\left(v \mid p_n, n\right) \mid n\right) - \frac{1}{2} g''\left(\beta\left(v \mid p_n, n\right) \mid n\right)\left(\int u^2 K(u)\, du\right) h^2\right)& \\ \to_d \mathrm{N}\left(0, \frac{g\left(\beta\left(v \mid p_n, n\right) \mid n\right)}{r_n} \int K^2(u)\, du\right)&.\end{aligned}$$

The conclusion in Part (a) follows from this result and (S59).

It follows from straightforward calculations that $\mathrm{E}\left[J_l(v,n) J_l(v',n')\right] = o(1)$ for all $(v,n) \neq (v',n')$. The Lyapunov condition for the the multi-dimensional Lyapunov central limit theorem can also be easily verified. Then by these results,

$$\begin{aligned}&\begin{pmatrix} \sqrt{Lh}\left(\widehat{g}\left(\beta\left(v_1 \mid p_{n_1}, n_1\right), n_1\right) - \mathrm{E}\left[\widehat{g}\left(\beta\left(v_1 \mid p_{n_1}, n_1\right), n_1\right)\right]\right) \\ \vdots \\ \sqrt{Lh}\left(\widehat{g}\left(\beta\left(v_J \mid p_{n_M}, n_M\right), n_M\right) - \mathrm{E}\left[\widehat{g}\left(\beta\left(v_J \mid p_{n_M}, n_M\right), n_M\right)\right]\right) \end{pmatrix} \\ &\quad \to_d \mathrm{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} g\left(\beta\left(v_1 \mid p_{n_1}, n_1\right), n_1\right) & & \\ & \ddots & \\ & & g\left(\beta\left(v_J \mid p_{n_M}, n_M\right), n_M\right) \end{pmatrix} \int K^2(u)\, du\right).\end{aligned}$$

The conclusion in Part (b) follows from this result, (S59), (S60) and (S63). ∎

Proof of Proposition S2. For notational simplicity, write $\widehat{\boldsymbol{\vartheta}} \coloneqq \widehat{\boldsymbol{\vartheta}}\left(\widehat{\mathbf{W}}\right)$. Since $Q(\cdot,\cdot;\cdot)$ is continuously differentiable on $[\epsilon, 1-\epsilon]^2 \times \Theta$, under Assumption S3, the uniform convergence

$$\sup_{\boldsymbol{\vartheta} \in \Theta \times \mathbb{F}} \left\| \boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*, \widehat{\boldsymbol{p}}, \boldsymbol{\vartheta}\right) - \boldsymbol{\Upsilon}\left(\boldsymbol{F}^*, \boldsymbol{p}_0, \boldsymbol{\vartheta}\right) \right\| \to_p 0 \tag{S64}$$

follows from the consistency of $\widehat{\boldsymbol{F}}^*$ and $\widehat{\boldsymbol{p}}$. Let

$$D_0\left(\boldsymbol{\vartheta}; \mathbf{W}\right) \coloneqq \boldsymbol{\Upsilon}^\top\left(\boldsymbol{F}^*, \boldsymbol{p}_0, \boldsymbol{\vartheta}\right) \mathbf{W} \boldsymbol{\Upsilon}\left(\boldsymbol{F}^*, \boldsymbol{p}_0, \boldsymbol{\vartheta}\right)$$

and it follows that

$$\boldsymbol{\vartheta}_0 = \underset{\boldsymbol{\vartheta} \in \Theta \times \mathbb{F}}{\arg\min}\, D_0\left(\boldsymbol{\vartheta}; \mathbf{W}_0\right).$$

By the reverse triangle inequality,

$$\left|\sqrt{\widehat{D}\left(\boldsymbol{\vartheta};\mathbf{W}_0\right)}-\sqrt{D_0\left(\boldsymbol{\vartheta};\mathbf{W}_0\right)}\right|$$
$$\leq\left(\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}\right)-\boldsymbol{\Upsilon}\left(\boldsymbol{F}^*,\boldsymbol{p}_0,\boldsymbol{\vartheta}\right)\right)^\top\mathbf{W}_0\left(\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}\right)-\boldsymbol{\Upsilon}\left(\boldsymbol{F}^*,\boldsymbol{p}_0,\boldsymbol{\vartheta}\right)\right).$$

It follows from this result and (S64) that

$$\sup_{\boldsymbol{\vartheta}\in\Theta\times\mathbb{F}}\left|\sqrt{\widehat{D}\left(\boldsymbol{\vartheta};\mathbf{W}_0\right)}-\sqrt{D_0\left(\boldsymbol{\vartheta};\mathbf{W}_0\right)}\right|\to_p 0.$$

It follows from the reverse triangle inequality and (S64) that

$$\sup_{\boldsymbol{\vartheta}\in\Theta\times\mathbb{F}}\left|\widehat{D}\left(\boldsymbol{\vartheta};\widehat{\mathbf{W}}\right)-\widehat{D}\left(\boldsymbol{\vartheta};\mathbf{W}_0\right)\right|\to_p 0.$$

It follows from these results and the triangle inequality that

$$\sup_{\boldsymbol{\vartheta}\in\Theta\times\mathbb{F}}\left|\widehat{D}\left(\boldsymbol{\vartheta};\widehat{\mathbf{W}}\right)-D_0\left(\boldsymbol{\vartheta};\mathbf{W}_0\right)\right|\to_p 0.$$

Consistency of $\widehat{\boldsymbol{\vartheta}}$ follows from this result and the standard arguments used in the proof of the consistency of M-estimators (see, e.g., Hansen, 2022, Theorem 22.1). Compactness of $\Theta\times\mathbb{F}$ and continuity of $D_0\left(\cdot;\mathbf{W}_0\right)$ ensure that the second requirement in the statement of Hansen (2022, Theorem 22.1) is satisfied.

It follows from consistency of $\widehat{\boldsymbol{\vartheta}}$ and the assumption that $\boldsymbol{\vartheta}_0$ is an interior point that $\widehat{\boldsymbol{\vartheta}}$ satisfies the FOCs wpa1. Then, we have

$$\begin{aligned}
o_p\left((Lh)^{-1/2}\right) &= \frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\Upsilon}^\top\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\vartheta}}\right)\widehat{\mathbf{W}}\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\vartheta}}\right)\\
&= \frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\Upsilon}^\top\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\vartheta}}\right)\widehat{\mathbf{W}}\left\{\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\vartheta}}\right)-\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}_0\right)+\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}_0\right)\right\},
\end{aligned}$$

and therefore,

$$\frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\Upsilon}^\top\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\vartheta}}\right)\widehat{\mathbf{W}}\left(\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\vartheta}}\right)-\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}_0\right)\right)$$
$$=-\frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\Upsilon}^\top\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\vartheta}}\right)\widehat{\mathbf{W}}\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}_0\right)+o_p\left((Lh)^{-1/2}\right).$$

Therefore, by the mean value theorem,

$$\left(\frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\Upsilon}^\top\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}_0\right)\widehat{\mathbf{W}}\frac{\partial}{\partial\boldsymbol{\vartheta}^\top}\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\dot{\boldsymbol{\vartheta}}\right)\right)\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)$$
$$=-\frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\Upsilon}^\top\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}_0\right)\widehat{\mathbf{W}}\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*,\widehat{\boldsymbol{p}},\boldsymbol{\vartheta}_0\right)+o_p\left((Lh)^{-1/2}\right),\quad\text{(S65)}$$

where $\dot{\boldsymbol{\vartheta}}$ denotes the mean value. By Proposition S1, the fact that $\widehat{\boldsymbol{p}} - \boldsymbol{p}_0 = O_p\left(L^{-1/2}\right)$ and the mean value theorem,

$$\begin{aligned}\boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*, \widehat{\boldsymbol{p}}, \boldsymbol{\vartheta}_0\right) &= \boldsymbol{\Upsilon}\left(\widehat{\boldsymbol{F}}^*, \widehat{\boldsymbol{p}}, \boldsymbol{\vartheta}_0\right) - \boldsymbol{\Upsilon}\left(\boldsymbol{F}^*, \boldsymbol{p}_0, \boldsymbol{\vartheta}_0\right) \\ &= \boldsymbol{\Psi}_1\left(\widehat{\boldsymbol{F}}^* - \boldsymbol{F}^*\right) + o_p\left((Lh)^{-1/2}\right).\end{aligned}$$

Then by this result, (S65), $\widehat{\mathbf{W}} \to_p \mathbf{W}_0$, $\widehat{\boldsymbol{F}}^* \to_p \boldsymbol{F}^*$ and $\widehat{\boldsymbol{p}} \to_p \boldsymbol{p}_0$, we have

$$\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = -\left(\boldsymbol{\Pi}_0^\top \mathbf{W}_0 \boldsymbol{\Pi}_0\right)^{-1} \boldsymbol{\Pi}_0^\top \mathbf{W}_0 \boldsymbol{\Psi}_1 \left(\widehat{\boldsymbol{F}}^* - \boldsymbol{F}^*\right) + o_p\left((Lh)^{-1/2}\right). \tag{S66}$$

The second conclusion follows from this result and Proposition S1. ■

Denote $\mathbb{G}_\dagger(\cdot \mid n) \coloneqq \widehat{G}_\dagger(\cdot \mid n) - G(\cdot \mid n)$, $\mathbb{G}_\dagger(\cdot, n) \coloneqq \widehat{G}_\dagger(\cdot, n) - G(\cdot, n)$, $\mathbb{H}_\dagger(\cdot \mid n) \coloneqq \widehat{g}_\dagger(\cdot \mid n) - g(\cdot \mid n)$ and $\mathbb{H}_\dagger(\cdot, n) \coloneqq \widehat{g}_\dagger(\cdot, n) - g(\cdot, n)$. Let $\widehat{g}'_\dagger(\cdot \mid n)$, $\widehat{g}''_\dagger(\cdot \mid n)$, $\widehat{g}'_\dagger(\cdot, n)$ and $\widehat{g}''_\dagger(\cdot, n)$ be the first and second derivatives of $\widehat{g}_\dagger(\cdot \mid n)$ and $\widehat{g}_\dagger(\cdot, n)$. Let $\mathbb{H}'_\dagger(\cdot \mid n)$ and $\mathbb{H}''_\dagger(\cdot \mid n)$ denote the first and second derivatives of $\mathbb{H}_\dagger(\cdot \mid n)$. The following lemma is a bootstrap analogue of Lemma S2.

Lemma S3. *Assume that the assumptions in the statement of Proposition S1 are satisfied. Then we have the following results. (a)*

$$\Pr_\dagger\left[\left|\widehat{p}_n^\dagger - p_n\right| \geq \alpha_{\dagger,L}^p\right] = O_p\left(L^{-1}\right),$$

for some deterministic sequence $\alpha_{\dagger,L}^p = O\left(\sqrt{\log(L)/L}\right)$. *(b)*

$$\Pr_\dagger\left[\|\mathbb{G}_\dagger(\cdot \mid n)\|_{\mathcal{B}_n} \geq \bar{\alpha}_{\dagger,L}\right] = O_p\left(L^{-1}\right),$$

for some deterministic sequence $\bar{\alpha}_{\dagger,L} = O\left(\sqrt{\log(L)/L}\right)$. *(c)*

$$\Pr_\dagger\left[\|\mathbb{H}_\dagger(\cdot \mid n)\|_{\mathcal{B}_n} \geq \alpha_{\dagger,L}\right] = O_p\left(L^{-1}\right),$$

for some deterministic sequence $\alpha_{\dagger,L} = O\left(\sqrt{\log(L)/(Lh)}\right)$ *and similar results with deterministic sequences* $\alpha'_{\dagger,L} = O\left(\sqrt{\log(L)/(Lh^3)}\right)$ *and* $\alpha''_{\dagger,L} = O\left(\sqrt{\log(L)/(Lh^5)}\right)$ *hold for* $\left\|\mathbb{H}'_\dagger(\cdot \mid n)\right\|_{\mathcal{B}_n}$ *and* $\left\|\mathbb{H}''_\dagger(\cdot \mid n)\right\|_{\mathcal{B}_n}$.

Proof of Lemma S3. Let

$$\widehat{\pi}_n^\dagger \coloneqq \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}\left(N_l^\dagger = n\right).$$

Then, we can write $\widehat{q}_n^\dagger = \widetilde{r}_n^\dagger / \left(n\widehat{\pi}_n^\dagger\right)$. By Bernstein's inequality, we have

$$\Pr_\dagger\left[\left|\widetilde{r}_n^\dagger - \widehat{r}_n\right| \geq \widehat{\sigma}_{r,n}\sqrt{\frac{2\cdot\log\left(L\right)}{L}}\right] = O_p\left(L^{-1}\right). \tag{S67}$$

By Bernstein's inequality and (S41),

$$\Pr\left[\left|\widehat{\sigma}_{r,n}^2 - \sigma_{r,n}^2\right| \geq C\sqrt{\frac{\log\left(L\right)}{L}}\right] = O\left(L^{-1}\right)$$

for some $C > 0$. Then by this result, (S41), (S67) and the triangle inequality, we have

$$\Pr_\dagger\left[\left|\widetilde{r}_n^\dagger - r_n\right| \geq 2\sqrt{\sigma_{r,n}^2 + C\sqrt{\frac{\log\left(L\right)}{L}}}\sqrt{\frac{2\cdot\log\left(L\right)}{L}}\right] = O_p\left(L^{-1}\right). \tag{S68}$$

It follows from similar arguments that

$$\Pr_\dagger\left[\left|\widehat{\pi}_n^\dagger - \pi_n\right| \geq \alpha_{\dagger,L}^\pi\right] = O_p\left(L^{-1}\right)$$

for some deterministic sequence $\alpha_{\dagger,L}^\pi = O\left(\sqrt{\log\left(L\right)/L}\right)$. By simple calculation,

$$\Pr_\dagger\left[\left|\frac{\pi_n}{\widehat{\pi}_n^\dagger} - 1\right| \geq \frac{\alpha_{\dagger,L}^\pi}{\pi_n - \alpha_{\dagger,L}^\pi}\right] = O_p\left(L^{-1}\right).$$

It follows from this result, (S68) and

$$\widehat{q}_n^\dagger - q_n = \frac{\widetilde{r}_n^\dagger - r_n}{n\pi_n} + \frac{\widetilde{r}_n^\dagger}{n\pi_n}\left(\frac{\pi_n}{\widehat{\pi}_n^\dagger} - 1\right)$$

that $\Pr_\dagger\left[\left|\widehat{q}_n^\dagger - q_n\right| \geq \alpha_{\dagger,L}^q\right] = O_p\left(L^{-1}\right)$ for some deterministic sequence $\alpha_{\dagger,L}^q = O\left(\sqrt{\log\left(L\right)/L}\right)$. The conclusion in Part (a) follows from this result.

By (S67) and the fact that for all $x \in \mathbb{R}$ and $c > 0$, $|x \vee (-c)| \geq c$ if and only if $|x| \geq c$, we have

$$\begin{aligned}\Pr_\dagger\left[\left|\widehat{r}_n^\dagger - \widehat{r}_n\right| \geq \widehat{\sigma}_{r,n}\sqrt{\frac{2\cdot\log\left(L\right)}{L}}\right] &= \Pr_\dagger\left[\left|\widetilde{r}_n^\dagger - \widehat{r}_n\right| \geq \widehat{\sigma}_{r,n}\sqrt{\frac{2\cdot\log\left(L\right)}{L}}\right] \\ &= O_p\left(L^{-1}\right).\end{aligned} \tag{S69}$$

Then it follows from the same arguments as those used to prove (S68) that

$$\Pr_\dagger\left[\left|\widehat{r}_n^\dagger - r_n\right| \geq \alpha_{\dagger,L}^r\right] = O_p\left(L^{-1}\right),$$

for some deterministic sequence $\alpha^r_{\dagger,L} = O\left(\sqrt{\log(L)/L}\right)$, and by simple calculation,

$$\Pr_\dagger\left[\left|\frac{r_n}{\widehat{r}^\dagger_n} - 1\right| \geq \frac{\alpha^r_{\dagger,L}}{r_n - \alpha^r_{\dagger,L}}\right] = O_p\left(L^{-1}\right). \tag{S70}$$

Then we apply Giné and Guillou (2002, Corollary 2.2) with $U = 2n$, $\sigma^2 = n^2\widehat{\pi}_n$ and t taken to be $C\sqrt{\log(L)L}$ for some positive constant C. Let $\alpha^\pi_L := \sigma_{\pi,n}\sqrt{2(\log(L)/L)}$. Then, by Giné and Guillou (2002, Corollary 2.2), taking C to be sufficiently large, we have

$$\Pr_\dagger\left[\left\|\widehat{G}_\dagger(\cdot,n) - \widehat{G}(\cdot,n)\right\|_{\mathcal{B}_n} > C\sqrt{\frac{\log(L)}{L}}\right]\mathbb{1}\left(|\widehat{\pi}_n - \pi_n| < \alpha^\pi_L\right) = O_p\left(L^{-1}\right).$$

Note that by (S39), we have $\mathbb{1}\left(|\widehat{\pi}_n - \pi_n| < \alpha^\pi_L\right) = 1$ wpa1. Therefore,

$$\Pr_\dagger\left[\left\|\widehat{G}_\dagger(\cdot,n) - \widehat{G}(\cdot,n)\right\|_{\mathcal{B}_n} > C\sqrt{\frac{\log(L)}{L}}\right] = O_p\left(L^{-1}\right),$$

if C is sufficiently large. The conclusion in Part (a) follows from this result, (S70) and

$$\mathbb{G}_\dagger(b \mid n) = \frac{\mathbb{G}_\dagger(b,n)}{r_n} + \frac{\widehat{G}_\dagger(b,n)}{r_n}\left(\frac{r_n}{\widehat{r}^\dagger_n} - 1\right).$$

The other conclusions follow from using (S70) and Corollary A.2. ■

Proof of Proposition S3. Denote

$$\widetilde{\xi}_\dagger(b \mid n) := b - \frac{\eta_n(p_n, G(b \mid n))}{(n-1)\,\widehat{g}_\dagger(b \mid n)}.$$

Let $\widetilde{V}^\dagger_{il} := \widetilde{\xi}_\dagger\left(B^\dagger_{il} \mid N^\dagger_l\right)$. Let $\widetilde{F}^*_\dagger(v,n)$ be defined by the right hand side of (S15) with $\widehat{V}^\dagger_{il}$ replaced by $\widetilde{V}^\dagger_{il}$ and let $\widetilde{F}^*_\dagger(v \mid p_n) := \widetilde{F}^*_\dagger(v,n)/r_n$. Let $\mathbb{X}_\dagger(b \mid n) := \widetilde{\xi}_\dagger(b \mid n) - \xi(b \mid p_n, n)$. Let $\mathbb{X}'_\dagger(\cdot \mid n)$ and $\mathbb{X}''_\dagger(\cdot \mid n)$ denote the first and second derivatives of $\mathbb{X}_\dagger(\cdot \mid n)$. Then, we have

$$\mathbb{X}_\dagger(b \mid n) = -\frac{\eta_n(p_n, G(b \mid n))}{n-1}\left\{\frac{1}{\widehat{g}_\dagger(b \mid n)} - \frac{1}{g(b \mid n)}\right\}, \tag{S71}$$

$$\begin{aligned}\mathbb{X}'_\dagger(b \mid n) &= -\frac{D_2\eta_n(p_n, G(b \mid n))\,g(b \mid n)}{n-1}\left\{\frac{1}{\widehat{g}_\dagger(b \mid n)} - \frac{1}{g(b \mid n)}\right\} \\ &\quad + \frac{\eta_n(p_n, G(b \mid n))}{n-1}\left\{\frac{\widehat{g}'_\dagger(b \mid n)}{\widehat{g}^2_\dagger(b \mid n)} - \frac{g'(b \mid n)}{g^2(b \mid n)}\right\},\end{aligned} \tag{S72}$$

and

$$
\begin{aligned}
\mathbb{X}_{\dagger}''(b \mid n) \quad = \quad & -\frac{D_2^2\eta_n\left(p_n, G(b \mid n)\right)g^2(b \mid n) + D_2\eta_n\left(p_n, G(b \mid n)\right)g'(b \mid n)}{n-1}\left\{\frac{1}{\widehat{g}_{\dagger}(b \mid n)} - \frac{1}{g(b \mid n)}\right\} \\
& + \frac{2D_2\eta_n\left(p_n, G(b \mid n)\right)g(b \mid n)}{n-1}\left\{\frac{\widehat{g}_{\dagger}'(b \mid n)}{\widehat{g}_{\dagger}^2(b \mid n)} - \frac{g'(b \mid n)}{g^2(b \mid n)}\right\} \\
& + \frac{\eta_n\left(p_n, G(b \mid n)\right)}{n-1}\left\{\frac{\widehat{g}_{\dagger}''(b \mid n)\widehat{g}_{\dagger}^2(b \mid n) - 2\left(\widehat{g}_{\dagger}'(b \mid n)\right)^2\widehat{g}_{\dagger}(b \mid n)}{\widehat{g}_{\dagger}^4(b \mid n)}\right. \\
& \left. - \frac{g''(b \mid n)g^2(b \mid n) - 2\left(g'(b \mid n)\right)^2 g(b \mid n)}{g^4(b \mid n)}\right\}. \qquad \text{(S73)}
\end{aligned}
$$

Denote

$$
\mathbb{K}_{\dagger}(b \mid n) := \eta_n\left(\widehat{p}_n^{\dagger}, \widehat{G}_{\dagger}(b \mid n)\right) - \eta_n\left(p_n, G(b \mid n)\right).
$$

By Lemma S3(a,b) and similar arguments as those used to prove (S49),

$$
\Pr_{\dagger}\left[\left\|\mathbb{K}_{\dagger}(\cdot \mid n)\right\|_{\mathcal{B}_n} > \tilde{\alpha}_{\dagger,L}\right] = O_p\left(L^{-1}\right),
$$

for some deterministic sequence $\tilde{\alpha}_{\dagger,L} = O\left(\sqrt{\log(L)/L}\right)$. Let $\bar{\mathbb{T}}_{\dagger} := \mathbb{1}\left(\left\|\mathbb{K}_{\dagger}(\cdot \mid n)\right\|_{\mathcal{B}_n} < \tilde{\alpha}_{\dagger,L}\right)$. Let $\left(\mathbb{T}_{\dagger}, \mathbb{T}_{\dagger}', \mathbb{T}_{\dagger}''\right)$ be defined by the same formula with $(\mathbb{K}_{\dagger}, \tilde{\alpha}_{\dagger,L})$ replaced by $(\mathbb{H}_{\dagger}, \alpha_{\dagger,L})$, $\left(\mathbb{H}_{\dagger}', \alpha_{\dagger,L}'\right)$ and $\left(\mathbb{H}_{\dagger}'', \alpha_{\dagger,L}''\right)$. Let $\mathbb{I}_{\dagger} := \bar{\mathbb{T}}_{\dagger}\mathbb{T}_{\dagger}\dot{\mathbb{T}}_{\dagger}\ddot{\mathbb{T}}_{\dagger}$. Then, we have $\Pr_{\dagger}\left[\mathbb{I}_{\dagger} = 0\right] = O_p\left(L^{-1}\right)$.

Decompose

$$
\widehat{F}_{\dagger}^*(v \mid p_n) - \widetilde{F}_{\dagger}^*(v \mid p_n) = \frac{\widehat{F}_{\dagger}^*(v, n) - \widetilde{F}_{\dagger}^*(v, n)}{r_n} + \frac{\widehat{F}_{\dagger}^*(v, n)}{r_n}\left(\frac{r_n}{\widehat{r}_n^{\dagger}} - 1\right). \qquad \text{(S74)}
$$

It now follows that

$$
\begin{aligned}
\mathrm{E}_{\dagger}\left[\left(\frac{r_n}{\widehat{r}_n^{\dagger}} - 1\right)^2\right] \quad \leq \quad & \frac{\mathrm{E}_{\dagger}\left[\left(\left(\widetilde{r}_n^{\dagger} - r_n\right) \vee \left(\widehat{r}_n - r_n - \widehat{\sigma}_{r,n}\sqrt{2 \cdot \log(L)/L}\right)\right)^2\right]}{\left(\widehat{r}_n - \widehat{\sigma}_{r,n}\sqrt{2 \cdot \log(L)/L}\right)^2} \\
& + \mathbb{1}\left(\widehat{r}_n \leq \widehat{\sigma}_{r,n}\sqrt{\frac{2 \cdot \log(L)}{L}}\right)\mathrm{E}_{\dagger}\left[\left(\frac{r_n}{\widehat{r}_n^{\dagger}} - 1\right)^2\right],
\end{aligned}
$$

where the second term on the right hand side equals zero wpa1. For the first term, we have

$$
\mathrm{E}_{\dagger}\left[\left(\left(\widetilde{r}_n^{\dagger} - r_n\right) \vee \left(\widehat{r}_n - r_n - \widehat{\sigma}_{r,n}\sqrt{\frac{2 \cdot \log(L)}{L}}\right)\right)^2\right] \leq \mathrm{E}_{\dagger}\left[\left(\widetilde{r}_n^{\dagger} - r_n\right)^2\right]
$$

$$+ \Pr_{\dagger}\left[\left|\widetilde{r}_n^{\dagger} - \widehat{r}_n\right| > \widehat{\sigma}_{r,n}\sqrt{\frac{2\cdot\log(L)}{L}}\right]\left(\widehat{r}_n - r_n - \widehat{\sigma}_{r,n}\sqrt{\frac{2\cdot\log(L)}{L}}\right)^2.$$

It is easy to check that $\mathrm{E}_{\dagger}\left[\left(\widetilde{r}_n^{\dagger} - r_n\right)^2\right] = O_p\left(L^{-1}\right)$. It now follows from this result, (S69) and the above inequalities that $\mathrm{E}_{\dagger}\left[\left(r_n/\widehat{r}_n^{\dagger} - 1\right)^2\right] = O_p\left(L^{-1}\right)$.

Denote $\delta_L^{\dagger} := \left\|\widehat{\xi}_{\dagger}(\cdot \mid n) - \widetilde{\xi}_{\dagger}(\cdot \mid n)\right\|_{\mathcal{B}_n}$. Then,

$$\left|\widehat{F}_{\dagger}^{*}(v,n) - \widetilde{F}_{\dagger}^{*}(v,n)\right| \leq \frac{1}{L}\sum_{l:N_l^{\dagger}=n}\sum_{i=1}^{N_l^{*\dagger}}\left\{\mathbb{1}\left(\widetilde{V}_{il}^{\dagger} \leq v + \delta_L^{\dagger}\right) - \mathbb{1}\left(\widetilde{V}_{il}^{\dagger} \leq v - \delta_L^{\dagger}\right)\right\}.$$

Let $\overline{v}_n^{\dagger} := \widetilde{\xi}_{\dagger}\left(\overline{b}_n \mid n\right)$ and $\underline{v}_n^{\dagger} := \widetilde{\xi}_{\dagger}\left(\underline{b}_n \mid n\right)$. If $\mathbb{I}_{\dagger} = 1$ and L is large enough, $\delta_L^{\dagger} \lesssim \sqrt{\log(L)/L}$, $\left|\overline{v}_n^{\dagger} - \overline{v}\right| \lesssim \sqrt{\log(L)/(Lh)}$ and $\left|\underline{v}_n^{\dagger} - \underline{v}\right| \lesssim \sqrt{\log(L)/(Lh)}$. By (S72), if $\mathbb{I}_{\dagger} = 1$ and L is large enough, $\left\|\mathbb{X}_{\dagger}'(\cdot \mid n)\right\|_{\mathcal{B}_n}$ is sufficiently small, the inverse $\widetilde{\beta}_{\dagger}(\cdot \mid n) := \widetilde{\xi}_{\dagger}^{-1}(\cdot \mid n)$ exists and $\widetilde{\beta}_{\dagger}(\cdot \mid n)$ is a strictly increasing function on $\left[\underline{v}_n^{\dagger}, \overline{v}_n^{\dagger}\right]$ with $\widetilde{\beta}_{\dagger}\left(\underline{v}_n^{\dagger}\right) = \underline{b}_n$ and $\widetilde{\beta}_{\dagger}\left(\overline{v}_n^{\dagger} \mid n\right) = \overline{b}_n$. Then, when L is sufficiently large,

$$\begin{aligned}\mathbb{I}_{\dagger}\left|\widehat{F}_{\dagger}^{*}(v,n) - \widetilde{F}_{\dagger}^{*}(v,n)\right| &\leq \mathbb{I}_{\dagger}\left\{\mathbb{G}_{\dagger}\left(\widetilde{\beta}_{\dagger}\left(v+\delta_L^{\dagger} \mid n\right), n\right) - \mathbb{G}_{\dagger}\left(\widetilde{\beta}_{\dagger}\left(v-\delta_L^{\dagger} \mid n\right), n\right)\right\} \\ &\quad + \mathbb{I}_{\dagger}\left\{G\left(\widetilde{\beta}_{\dagger}\left(v+\delta_L^{\dagger} \mid n\right), n\right) - G\left(\widetilde{\beta}_{\dagger}\left(v-\delta_L^{\dagger} \mid n\right), n\right)\right\}. \quad \text{(S75)}\end{aligned}$$

By arguments similar to those in the proof of (S57),

$$\mathbb{I}_{\dagger}\left|G\left(\widetilde{\beta}_{\dagger}\left(v+\delta_L^{\dagger} \mid n\right), n\right) - G\left(\widetilde{\beta}_{\dagger}\left(v-\delta_L^{\dagger} \mid n\right), n\right)\right| \lesssim \sqrt{\frac{\log(L)}{L}},$$

if L sufficiently large. For the first term on the right hand side of (S75),

$$\begin{aligned}&\mathbb{I}_{\dagger}\left\{\mathbb{G}_{\dagger}\left(\widetilde{\beta}_{\dagger}\left(v+\delta_L^{\dagger} \mid n\right), n\right) - \mathbb{G}_{\dagger}\left(\widetilde{\beta}_{\dagger}\left(v-\delta_L^{\dagger} \mid n\right), n\right)\right\}^2 \\ &\qquad \leq \left(\sup_{(t_1,t_2)\in[-\epsilon_L,\epsilon_L]^2}\left|\mathbb{G}_{\dagger}\left(\beta(v \mid p_n, n) + t_1, n\right) - \mathbb{G}_{\dagger}\left(\beta(v \mid p_n, n) + t_2, n\right)\right|\right)^2,\end{aligned}$$

for some deterministic sequence $\epsilon_L = O\left(\sqrt{\log(L)/(Lh)}\right)$. By arguments similar to those in the proof of (S56),

$$\mathrm{E}_{\dagger}\left[\sup_{(t_1,t_2)\in[-\epsilon_L,\epsilon_L]^2}\left|\mathbb{G}_{\dagger}\left(\beta(v \mid p_n, n) + t_1, n\right) - \mathbb{G}_{\dagger}\left(\beta(v \mid p_n, n) + t_2, n\right)\right|\right] = O_p\left(L^{-1/2}\left(\frac{\log(L)}{Lh}\right)^{1/4}\right).$$

By this result and Ledoux and Talagrand (1991, Theorem 6.20),

$$\mathrm{E}_{\dagger}\left[\left(\sup_{(t_1,t_2)\in[-\epsilon_L,\epsilon_L]^2}\left|\mathbb{G}_{\dagger}\left(\beta\left(v\mid p_n,n\right)+t_1,n\right)-\mathbb{G}_{\dagger}\left(\beta\left(v\mid p_n,n\right)+t_2,n\right)\right|\right)^2\right]=O_p\left(L^{-1}\left(\frac{\log\left(L\right)}{Lh}\right)^{1/2}\right).$$

It now follows that

$$\begin{aligned}\mathrm{E}_{\dagger}\left[\left(\widehat{F}_{\dagger}^{*}\left(v,n\right)-\widetilde{F}_{\dagger}^{*}\left(v,n\right)\right)^2\right] &\lesssim \mathrm{E}_{\dagger}\left[\mathbb{I}_{\dagger}\left(\widehat{F}_{\dagger}^{*}\left(v,n\right)-\widetilde{F}_{\dagger}^{*}\left(v,n\right)\right)^2\right]+\Pr_{\dagger}\left[\mathbb{I}_{\dagger}=0\right]\\ &= O_p\left(\frac{\log\left(L\right)}{L}\right). \qquad \text{(S76)}\end{aligned}$$

Then it follows from this result, (S74) and $\mathrm{E}_{\dagger}\left[\left(r_n/\widehat{r}_n^{\dagger}-1\right)^2\right]=O_p\left(L^{-1}\right)$ that

$$\mathrm{E}_{\dagger}\left[\left(\widehat{F}_{\dagger}^{*}\left(v\mid p_n\right)-\widetilde{F}_{\dagger}^{*}\left(v\mid p_n\right)\right)^2\right]=O_p\left(\frac{\log\left(L\right)}{L}\right). \qquad \text{(S77)}$$

Write

$$\begin{aligned}\widetilde{F}_{\dagger}^{*}\left(v\mid p_n\right) &= \mathbb{I}_{\dagger}\cdot\widetilde{F}_{\dagger}^{*}\left(v\mid p_n\right)+\left(1-\mathbb{I}_{\dagger}\right)\widetilde{F}_{\dagger}^{*}\left(v\mid p_n\right)\\ &= \mathbb{I}_{\dagger}\left(G\left(\widetilde{\beta}_{\dagger}\left(v\mid n\right)\mid n\right)-G\left(\beta\left(v\mid p_n,n\right)\mid n\right)\right)\\ &\quad +\mathbb{I}_{\dagger}\cdot G\left(\beta\left(v\mid p_n,n\right)\mid n\right)+\mathbb{I}_{\dagger}\cdot\frac{\mathbb{G}_{\dagger}\left(\widetilde{\beta}_{\dagger}\left(v\mid n\right),n\right)}{r_n}+\left(1-\mathbb{I}_{\dagger}\right)\widetilde{F}_{\dagger}^{*}\left(v\mid p_n\right). \qquad \text{(S78)}\end{aligned}$$

By Dette et al. (2006, Lemma A.1), if $\mathbb{I}_{\dagger}=1$ and L is large enough, there exists $\lambda\in(0,1)$ such that

$$\begin{aligned}\widetilde{\beta}_{\dagger}\left(v\mid n\right)-\beta\left(v\mid p_n,n\right) &= -\frac{\mathbb{X}_{\dagger}\left(\beta\left(v\mid p_n,n\right)\mid n\right)}{\xi'\left(\beta\left(v\mid p_n,n\right)\mid p_n,n\right)}-2\frac{\mathbb{X}_{\dagger}\left(b_\lambda\mid n\right)\mathbb{X}_{\dagger}'\left(b_\lambda\mid n\right)}{\left(\xi'\left(b_\lambda\mid p_n,n\right)+\lambda\mathbb{X}_{\dagger}'\left(b_\lambda\mid n\right)\right)^2}\\ &\quad +\frac{\mathbb{X}_{\dagger}^2\left(b_\lambda\mid n\right)\left(\xi''\left(b_\lambda\mid p_n,n\right)+\lambda\mathbb{X}_{\dagger}''\left(b_\lambda\mid n\right)\right)}{\left(\xi'\left(b_\lambda\mid p_n,n\right)+\lambda\mathbb{X}_{\dagger}'\left(b_\lambda\mid n\right)\right)^3}, \qquad \text{(S79)}\end{aligned}$$

where $b_\lambda\coloneqq\left(\xi\left(\cdot\mid p_n,n\right)+\lambda\mathbb{X}_{\dagger}\left(\cdot\mid n\right)\right)^{-1}\left(v\right)$. By (S71), (S72), (S73), (S79) and the mean value theorem, we have

$$\begin{aligned}\mathbb{I}_{\dagger}\left(G\left(\widetilde{\beta}_{\dagger}\left(v\mid n\right)\mid n\right)-G\left(\beta\left(v\mid p_n,n\right)\mid n\right)\right) &= \mathbb{I}_{\dagger}\cdot g\left(\beta\left(v\mid p_n,n\right)\mid n\right)\left(\widetilde{\beta}_{\dagger}\left(v\mid n\right)-\beta\left(v\mid p_n,n\right)\right)\\ &\quad +O\left(\frac{\log\left(L\right)}{Lh}\right)\end{aligned}$$

and

$$\mathbb{I}_{\dagger}\left(\widetilde{\beta}_{\dagger}\left(v\mid n\right)-\beta\left(v\mid p_n,n\right)\right)=-\mathbb{I}_{\dagger}\cdot\frac{\mathbb{X}_{\dagger}\left(\beta\left(v\mid p_n,n\right)\mid n\right)}{\xi'\left(\beta\left(v\mid p_n,n\right)\mid p_n,n\right)}+O\left(\frac{\log\left(L\right)}{Lh}\right).$$

It is easy to check that

$$\mathbb{I}_\dagger \cdot \mathbb{X}_\dagger\left(\beta\left(v \mid p_n, n\right) \mid n\right) = \mathbb{I}_\dagger \cdot \frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right) \mathbb{H}_\dagger\left(\beta\left(v \mid p_n, n\right) \mid n\right)}{(n-1)\, g^2\left(\beta\left(v \mid p_n, n\right) \mid n\right)} + O\left(\frac{\log(L)}{Lh}\right).$$

By these results and (S78), we have

$$\begin{aligned}
\widetilde{F}_\dagger^*\left(v \mid p_n\right) &= -\mathbb{I}_\dagger \cdot \frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right) \mathbb{H}_\dagger\left(\beta\left(v \mid p_n, n\right) \mid n\right)}{(n-1)\, g\left(\beta\left(v \mid p_n, n\right) \mid n\right) \xi'\left(\beta\left(v \mid p_n, n\right) \mid p_n, n\right)} \\
&\quad + \mathbb{I}_\dagger \cdot G\left(\beta\left(v \mid p_n, n\right) \mid n\right) + \mathbb{I}_\dagger \cdot \frac{\mathbb{G}_\dagger\left(\widetilde{\beta}_\dagger\left(v \mid n\right), n\right)}{r_n} + \left(1 - \mathbb{I}_\dagger\right) \widetilde{F}_\dagger^*\left(v \mid p_n\right) + O\left(\frac{\log(L)}{Lh}\right).
\end{aligned}$$

Then, by this result and

$$\widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right) \mid n\right) = \frac{\widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right), n\right)}{r_n} + \frac{\widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right), n\right)}{r_n}\left(\frac{r_n}{\widehat{r}_n^\dagger} - 1\right),$$

we can write

$$\begin{aligned}
\widetilde{F}_\dagger^*\left(v \mid p_n\right) &= -\frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right) \widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right), n\right)}{(n-1)\, \xi'\left(\beta\left(v \mid p_n, n\right) \mid p_n, n\right) g\left(\beta\left(v \mid p_n, n\right), n\right)} \\
&\quad + V_1^\dagger + V_2^\dagger + V_3^\dagger + V_4^\dagger + V_5^\dagger + O\left(\frac{\log(L)}{Lh}\right),
\end{aligned} \tag{S80}$$

where

$$\begin{aligned}
V_1^\dagger &:= \left(1 - \mathbb{I}_\dagger\right) \frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right) \widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right), n\right)}{(n-1)\, \xi'\left(\beta\left(v \mid p_n, n\right) \mid p_n, n\right) g\left(\beta\left(v \mid p_n, n\right), n\right)} \\
V_2^\dagger &:= -\mathbb{I}_\dagger \cdot \frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right) \widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right), n\right)}{(n-1)\, \xi'\left(\beta\left(v \mid p_n, n\right) \mid p_n, n\right) g\left(\beta\left(v \mid p_n, n\right), n\right)}\left(\frac{r_n}{\widehat{r}_n^\dagger} - 1\right) \\
V_3^\dagger &:= \mathbb{I}_\dagger\left\{G\left(\beta\left(v \mid p_n, n\right) \mid n\right) + \frac{\eta_n\left(p_n, G\left(\beta\left(v \mid p_n, n\right) \mid n\right)\right)}{(n-1)\, \xi'\left(\beta\left(v \mid p_n, n\right) \mid p_n, n\right)}\right\} \\
V_4^\dagger &:= \mathbb{I}_\dagger \cdot \frac{\mathbb{G}_\dagger\left(\widetilde{\beta}_\dagger\left(v \mid n\right), n\right)}{r_n} \\
V_5^\dagger &:= \left(1 - \mathbb{I}_\dagger\right) \widetilde{F}_\dagger^*\left(v \mid p_n\right).
\end{aligned}$$

It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
\mathrm{E}_\dagger\left[\left(1 - \mathbb{I}_\dagger\right)^2 \widehat{g}_\dagger^2\left(\beta\left(v \mid p_n, n\right), n\right)\right] &\lesssim \sqrt{\Pr_\dagger\left[\mathbb{I}_\dagger = 0\right] \cdot \mathrm{E}_\dagger\left[\left(\widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right), n\right) - \widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right)\right)^4\right]} \\
&\quad + \Pr_\dagger\left[\mathbb{I}_\dagger = 0\right] \widehat{g}^2\left(\beta\left(v \mid p_n, n\right), n\right).
\end{aligned} \tag{S81}$$

By the Rosenthal inequality,

$$\mathrm{E}_\dagger\left[\left(\widehat{g}_\dagger\left(\beta\left(v \mid p_n, n\right), n\right) - \widehat{g}\left(\beta\left(v \mid p_n, n\right), n\right)\right)^4\right] = O_p\left((Lh)^{-2}\right).$$

By this result, $\Pr_\dagger[\mathbb{I}_\dagger = 0] = O_p(L^{-1})$ and (S81), we have $\mathrm{Var}_\dagger\left[V_1^\dagger\right] = O_p(L^{-1})$. It follows from similar arguments and $\mathrm{E}_\dagger\left[\left(r_n/\widehat{r}_n^\dagger - 1\right)^2\right] = O_p(L^{-1})$ that $\mathrm{Var}_\dagger\left[V_2^\dagger\right] = O_p(L^{-1})$. If follows from

$$\mathrm{Var}_\dagger[\mathbb{I}_\dagger] = \Pr_\dagger[\mathbb{I}_\dagger = 0] \cdot \Pr_\dagger[\mathbb{I}_\dagger = 1] = O_p(L^{-1})$$

that $\mathrm{Var}_\dagger\left[V_3^\dagger\right] = O_p(L^{-1})$. By Chernozhukov et al. (2014, Corollary 5.1), we have $\mathrm{E}\left[\|\mathbb{G}(\cdot,n)\|_{\mathcal{B}_n}\right] = O(L^{-1/2})$ and $\mathrm{E}_\dagger\left[\left\|\widehat{G}_\dagger(\cdot,n) - \widehat{G}(\cdot,n)\right\|_{\mathcal{B}_n}\right] = O_p(L^{-1/2})$. By these result and Ledoux and Talagrand (1991, Theorem 6.20), we have

$$\mathrm{Var}_\dagger\left[\mathrm{V}_4^\dagger\right] \lesssim \mathrm{E}_\dagger\left[\left(\|\mathbb{G}_\dagger(\cdot,n)\|_{\mathcal{B}_n}\right)^2\right] = O_p(L^{-1}).$$

It is easy to see that since $0 \le \widetilde{F}_\dagger^*(v \mid p_n) \le n/r_n$,

$$\mathrm{Var}_\dagger\left[V_5^\dagger\right] \le \mathrm{E}\left[(1-\mathbb{I}_\dagger)^2\left(\widetilde{F}_\dagger^*(v \mid p_n)\right)^2\right] \lesssim \Pr_\dagger[\mathbb{I}_\dagger = 0] = O_p(L^{-1}).$$

It follows from the above results, (S77), (S80) and the Cauchy-Schwarz inequality that

$$\mathrm{Var}_\dagger\left[\widehat{F}_\dagger^*(v \mid p_n)\right] = \mathrm{Var}_\dagger\left[\frac{\eta_n(p_n, G(\beta(v \mid p_n, n) \mid n))\,\widehat{g}_\dagger(\beta(v \mid p_n, n), n)}{(n-1)\,\xi'(\beta(v \mid p_n, n) \mid p_n, n)\,g(\beta(v \mid p_n, n), n)}\right] + o_p\left((Lh)^{-1}\right). \quad \text{(S82)}$$

By simple calculation, we have

$$\begin{aligned}
&\mathrm{Var}_\dagger[\widehat{g}_\dagger(\beta(v \mid p_n, n), n)] \\
&\qquad = \frac{1}{L}\left\{\frac{1}{L}\sum_{l:N_l=n}\left(\sum_{i=1}^{N_l^*}\frac{1}{h}\mathcal{K}_1\left(B_{il}, \beta(v \mid p_n, n) \mid h, \underline{\widehat{b}}_n, \overline{\widehat{b}}_n\right)\right)^2 - \widehat{g}^2(\beta(v \mid p_n, n), n)\right\}.
\end{aligned}$$

The conclusion follows from this result, (S82), $\widehat{g}(\beta(v \mid p_n, n), n) \to_p g(\beta(v \mid p_n, n), n)$ and the fact that

$$\frac{1}{Lh}\sum_{l:N_l=n}\left(\sum_{i=1}^{N_l^*}\mathcal{K}_1\left(B_{il}, \beta(v \mid p_n, n) \mid h, \underline{\widehat{b}}_n, \overline{\widehat{b}}_n\right)\right)^2 \to_p g(\beta(v \mid p_n, n), n)\int K^2(u)\,du.$$

■

References

Bickel, P. J. and K. A. Doksum (2015). *Mathematical statistics: basic ideas and selected topics, volumes I-II package*. Chapman and Hall/CRC.

Chen, S. X. and T. Huang (2007). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics 35*(2), 265–282.

Chen, X., M. L. Gentry, T. Li, and J. Lu (2025). Identification and Inference in First-Price Auctions with Risk Averse Bidders and Selective Entry. *Review of Economic Studies*.

Cheng, M.-Y., J. Fan, and J. S. Marron (1997). On automatic boundary corrections. *The Annals of Statistics 25*(4), 1691 1708.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*(4), 1564 1597.

Dette, H., N. Neumeyer, and K. F. Pilz (2006). a simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 1 23.

Gentry, M. and T. Li (2014). Identification in Auctions With Selective Entry. *Econometrica 82*(1), 315–344.

Giné, E. and A. Guillou (1999). Laws of the Iterated Logarithm for Censored Data. *The Annals of Probability 27*(4), 2042–2067.

Giné, E. and A. Guillou (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincare (B) Probability and Statistics 38*(6), 907 921.

Giné, E. and R. Nickl (2015). *Mathematical foundations of infinite-dimensional statistical models*, Volume 40. Cambridge university press.

Guerre, E., I. Perrigne, and Q. Vuong (2000). Optimal Nonparametric Estimation of First price Auctions. *Econometrica 68*(3), 525 574.

Hahn, J. and Z. Liao (2021). Bootstrap Standard Error Estimates and Inference. *Econometrica 89*(4), 1963–1977.

Hansen, B. (2022). *Econometrics*. Princeton University Press.

Hickman, B. R. and T. P. Hubbard (2015). Replacing Sample Trimming with Boundary Correction in Nonparametric Estimation of First-Price Auctions. *Journal of Applied Econometrics 30*(5), 739 762.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing 3*(3), 135–146.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*, Volume 61. Springer.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: isoperimetry and processes.* Springer Science & Business Media.

Lejeune, M. and P. Sarda (1992). Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis 14*(4), 457–471.

Lewbel, A. (2007). A local generalized method of moments estimator. *Economics Letters 94*(1), 124–128.

Lewbel, A. (2019). The Identification Zoo: Meanings of Identification in Econometrics. *Journal of Economic Literature 57*(4), 835–903.

Ma, J., V. Marmer, and A. Shneyerov (2019). Inference for first-price auctions with Guerre, Perrigne, and Vuong's Estimator. *Journal of Econometrics*.

Ma, J., V. Marmer, A. Shneyerov, and P. Xu (2021). Monotonicity-constrained nonparametric estimation and inference for first-price auctions. *Econometric Reviews 40*(10), 944–982.

Marmer, V. and A. Shneyerov (2012). Quantile-based nonparametric inference for first-price auctions. *Journal of Econometrics 167*(2), 345–357.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Nolan, D. and D. Pollard (1987). U-Processes: Rates of Convergence. *The Annals of Statistics 15*(2), 780 799.

Severini, T. A. (2005). *Elements of distribution theory*. Number 17. Cambridge University Press.

Zincenko, F. (2024). Estimation and inference of seller's expected revenue in first-price auctions. *Journal of Econometrics 241*(1), 105734.