ROBUSTNESS, INFINITESIMAL NEIGHBORHOODS, AND MOMENT RESTRICTIONS

YUICHI KITAMURA, TAISUKE OTSU, AND KIRILL EVDOKIMOV

Very Preliminary

Abstract. This paper explores robust estimation of parameters identified by a set of moment restrictions. Suppose the econometrician observes data generated from a perturbed version of the probability distribution that corresponds to the true parameter value. Such perturbation can be regarded as a consequence of data errors, misspecification and other factors, following the literature of robust statistical estimation. There are two aspects in assessing robustness properties of an estimator. One is about bias, that is, the effect of the perturbation of the data generating mechanism on the behavior of the limit of the estimator. The other is about dispersion, often measured by the asymptotic variance. As far as one considers global perturbation, the former factor typically dominates, thereby making the latter a second order issue. An alternative approach is to consider the effect of local perturbation within shrinking topological neighborhoods of the original probability distribution, so that both bias and variance matter asymptotically. Such analysis, put loosely, enables the researcher to assess robustness in terms of asymptotic mean squared error (MSE). This paper derives asymptotic optimality results in moment restriction models along this line of analysis.

1. Introduction

To be added.

2. Preliminaries

Consider a probability measure $P_0 \in \mathcal{M}$, where $\mathcal{M}$ is the set of all probability measures on the Borel σ-field $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ of $\mathcal{X} \subseteq \mathbb{R}^d$. Let $g : \mathcal{X} \times \Theta \to \mathbb{R}^m$ be a vector of moment functions parametrized

Date: This Version: June 12, 2008.

Keywords: Empirical Likelihood, Hellinger Distance, Robust Estimation.

JEL Classification Number: C14.

We thank participants at the 2007 Winter Meetings of the Econometric Society and seminars at Harvard-MIT and Vanderbilt for valuable comments. Kitamura acknowledges financial support from the National Science Foundation via grants SES-0241770 and SES-0551271. Otsu acknowledges financial support from the National Science Foundation (SES-0720961).

by a p-dimensional vector θ which resides in $\Theta \subset \mathbb{R}^p$. The function g satisfies:

$$E_{P_0}\left[g\left(x,\theta_0\right)\right] = \int g\left(x,\theta_0\right) dP_0 = 0, \quad \theta_0 \in \Theta. \tag{2.1}$$

The econometrician wishes to estimate the unknown θ_0.

Suppose a random sample $\{x_i\}_{i=1}^n$ generated from $P_0^{\otimes n}$ is observed. Our focus is on robust estimation of θ_0 when observations are drawn from a (locally) perturbed version P_0, not P_0 itself. For the time being, however, consider the above model where no such perturbation is added to the measure P_0. Under this "classical" setting, various estimator for θ_0 are available, including GMM (Hansen (1982)), the empirical likelihood (EL) estimator and its variants. This paper is concerned with an estimator, which can be viewed as the Minimum Hellinger Distance Estimator (MHDE) by Beran (1977) applied to the moment restriction model (2.1). In particular, it will be shown to have asymptotic optimal robust properties in subsequent sections. The Hellinger distance between two probability measures is defined as follows:

Definition 2.1. Let P and Q be probability measures with densities p and q with respect to a dominating measure ν. The Hellinger distance between P and Q is then given by:

$$H\left(P,Q\right) = \left\{\int \left(p^{1/2} - q^{1/2}\right)^2 d\nu\right\}^{1/2} = \left\{2 - 2\int p^{1/2}q^{1/2}d\nu\right\}^{1/2}.$$

It is often convenient to use the standard notation in the literature that does not explicitly refer to the dominating measure. Then the above definition becomes:

$$H\left(P,Q\right) = \left\{\int \left(dP^{1/2} - dQ^{1/2}\right)^2\right\}^{1/2} = \left\{2 - 2\int dP^{1/2}dQ^{1/2}\right\}^{1/2}.$$

In what follows we show some results concerning the Hellinger distance that are useful in understanding the robustness theorems in the following sections.

Definition 2.2. Let P and Q be probability measures with densities p and q with respect to a dominating measure ν. The α-divergence from Q to P is given by

$$I_\alpha\left(P,Q\right) = \frac{1}{\alpha(1-\alpha)}\int \left(1 - \left(\frac{p}{q}\right)^\alpha\right) q d\nu, \quad \alpha \in \mathbb{R}.$$

If P is not absolutely continuous respect to Q, then $\int \mathbb{I}\left\{p > 0, q = 0\right\} d\nu > 0$, and as a consequence $I_\alpha\left(P,Q\right) = \infty$ for $\alpha \geq 1$. A similar argument shows that $I_\alpha\left(P,Q\right) = \infty$ if $Q \not\ll P$ and $\alpha \leq 0$. Note that I_α is well-defined for $\alpha = 0$ by taking the limit $\alpha \to 0$ in the definition. Indeed, L'Hospital's Rule implies that

$$\lim_{\alpha \to 0} I_\alpha\left(P,Q\right) = \int \log\left(\frac{p}{q}\right) q d\nu$$

(with the above convention for the case where $P \not\ll Q$), giving rise to the well-known Kullback-Leibler (KL) divergence measure. The case with $\alpha = 1$ corresponds to the KL divergence with the roles of P and Q reversed. Note that the above definitions imply that the α-divergence includes the Hellinger distance as a special case, in the sense that

$$H^2(P,Q) = \frac{1}{2} I_{\frac{1}{2}}(P,Q).$$

Lemma 2.1. *For probability measures P and Q,*

$$\max(\alpha, 1-\alpha)\, I_\alpha(P,Q) \geq \frac{1}{2} I_{\frac{1}{2}}(P,Q)$$

for every $\alpha \in \mathbb{R}$.

Proof. [**Proof of Lemma 2.1**] We first show the claim for $\alpha < \frac{1}{2}$, that is,

$$(1-\alpha)\, I_\alpha(P,Q) - \frac{1}{2} I_{\frac{1}{2}}(P,Q) \geq 0. \tag{2.2}$$

Let $H_\alpha(x) = \frac{1}{\alpha}(1 - x^\alpha) - 2\left(1 - x^{\frac{1}{2}}\right), \quad 0 \leq x \leq \infty$, then the above inequality becomes

$$\int H_\alpha\left(\frac{p}{q}\right) q d\nu \geq 0. \tag{2.3}$$

Note

$$\frac{d}{dx} H_\alpha(x) = -x^{\alpha-1} + x^{-\frac{1}{2}} \begin{cases} > 0 & \text{if } x > 1 \\ = 0 & \text{if } x = 1 \\ < 0 & \text{if } x < 1. \end{cases}$$

The above holds for the case with $\alpha = 0$ as well, since $H_0(x) = -\log x - 2\left(1 - x^{\frac{1}{2}}\right)$. Moreover, $H_\alpha(1) = 0$. Therefore $H_\alpha(x) \geq 0$ for all $x \geq 0$, and the desired inequality (2.3) follows immediately. Next, we prove the case with $\alpha > \frac{1}{2}$, that is,

$$\alpha I_\alpha(P,Q) \geq \frac{1}{2} I_{\frac{1}{2}}(P,Q).$$

Let $\beta = 1 - \alpha < \frac{1}{2}$, then the above inequality becomes

$$(1-\beta)\, I_{1-\beta}(P,Q) \geq \frac{1}{2} I_{\frac{1}{2}}(P,Q). \tag{2.4}$$

By (2.2) and the symmetry of the Hellinger distance,

$$(1-\beta) I_\beta(Q,P) \geq \frac{1}{2} I_{\frac{1}{2}}(Q,P) = \frac{1}{2} I_{\frac{1}{2}}(P,Q).$$

But the equality $I_{1-\beta}(P,Q) = I_\beta(Q,P)$ holds for every $\beta \in \mathbb{R}$, and (2.4) follows. □

Remark 2.1. Lemma 2.1 has some implications on a neighborhood system generated by the Hellinger distance. Consider the following neighborhood of a probability measure P whose radius in terms of I_α is $\delta > 0$:

$$B_{I_\alpha}(P,\delta) = \left\{ Q \in \mathcal{M} : \sqrt{I_\alpha(Q,P)} \le \delta \right\}.$$

Lemma 2.1 implies that

$$I_\alpha(P,Q) \ge \frac{1}{2\left(\left(\frac{1}{2}+L\right) \vee \left(\frac{1}{2}+U\right)\right)} I_{\alpha_0}(P,Q)$$

holds for every $\alpha \in \left[\frac{1}{2}-L, \frac{1}{2}+U\right]$ where $L, U > 0$ determine the lower and upper bounds for the range of α, if $\alpha_0 = \frac{1}{2}$. It is easy to verify that this statement holds only if $\alpha_0 = \frac{1}{2}$. Now, define

$$K(L,U) = \left(\frac{1}{2}+L\right) \vee \left(\frac{1}{2}+U\right),$$

then by the above inequality

$$\cup_{\alpha \in \left[\frac{1}{2}-L,\frac{1}{2}+U\right]} B_{I_\alpha}(P_0,\delta) \subset B_{I_{1/2}}\left(P_0, \sqrt{2K(L,U)}\delta\right) = H\left(P_0, 2\sqrt{K(L,U)}\delta\right). \tag{2.5}$$

That is, the union of the I_α-based neighborhoods over $\alpha \in \left[\frac{1}{2}-L, \frac{1}{2}+U\right]$ is covered by the Hellinger neighborhood $B_{I_{1/2}}$ with a "margin" given by the multiplicative constant $2\sqrt{K(L,U)}$. (2.5) is important, since in what follows we consider robustness of estimators against perturbation of P_0 within its neighborhood, and it is desirable to use a neighborhood that is sufficiently large to accommodate a large class of perturbations. The inclusion relationship shows that the Hellinger-based neighborhood covers other neighborhood systems based on $I_\alpha, \alpha \in \left[\frac{1}{2}-L, \frac{1}{2}+U\right]$ if the radii are chosen appropriately. It is easy to verify that (2.5) does not hold if the Hellinger distance $I_{\frac{1}{2}}$ is replaced by $I_\alpha, \alpha \ne \frac{1}{2}$, showing the special status of the Hellinger distance among the α-divergence family.

Beran (1977), considering a parametric model, proposed an estimator that minimizes the Hellinger distance between a model-based probability measure (from the parametric family) and a nonparametric probability measure estimate. This is known as the Minimum Hellinger Distance Estimator (MHDE). This method has been regarded as a robust procedure in the literature of parametric model estimation. An application of the MHDE procedure to the moment condition model (2.1) yields a computationally simple procedure as follows. Let P_n denote the empirical measure of observations $\{x_i\}_{i=1}^n$. P_n is an appropriate model-free estimator in our construction of the MHDE. Define

$$\mathcal{P}_\theta = \left\{ P \in \mathcal{M} : \int g(x,\theta)\, dP = 0 \right\}$$

then the MHDE, denoted by $\hat{\theta}$, is defined to be a parameter value that solves the optimization problem

$$\inf_{\theta\in\Theta}\inf_{P\in\mathcal{P}_\theta} H\left(P,P_n\right).$$

By convex duality theory (Kitamura (2006)), the objective function has the following representation:

$$\inf_{P\in\mathcal{P}_\theta} H\left(P,P_n\right)=\max_{\gamma\in\mathbb{R}^m}-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\gamma' g\left(x_i,\theta\right)}$$

Therefore the MHDE $\hat{\theta}=\arg\min_{\theta\in\Theta}\max_{\gamma\in\mathbb{R}^m}-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\gamma' g(x_i,\theta)}$ is easy to compute.

It is of interest to replace $H\left(\cdot,\cdot\right)$ in the above definition with the α-divergence. It is easy to see that the resulting family of estimators $\hat{\theta}_\alpha=\arg\min_{\theta\in\Theta}\max_{\gamma\in\mathbb{R}^m}-\frac{1}{n}\sum_{i=1}^{n}\rho_\alpha\left(\gamma' g\left(x_i,\theta\right)\right)$ with $\alpha\in\mathbb{R}$ corresponds to the so-called Generalized Empirical Likelihood (GEL) estimator discussed by Newey and Smith (2004). It is obvious that the MHDE $\hat{\theta}$ corresponds to $\hat{\theta}_\alpha$ with $\alpha=\left\{\frac{1}{2}\right\}$. Asymptotic properties of the (G)EL estimators under the current setting are well-understood (see, for example, Newey and Smith (2004)). Let $G=E_{P_0}\left[\partial g\left(x,\theta_0\right)/\partial\theta'\right]$, $\Omega=E_{P_0}\left[g\left(x,\theta_0\right)g\left(x,\theta_0\right)'\right]$, and $\Sigma=G'\Omega^{-1}G$. Then

$$\sqrt{n}\left(\hat{\theta}_\alpha-\theta_0\right)\overset{d}{\to}N\left(0,\Sigma^{-1}\right). \tag{2.6}$$

Therefore $\hat{\theta}_\alpha$, which includes the MHDE $\hat{\theta}$ as a special case, is a semiparametrically efficient estimator. Some alternative asymptotic efficiency criteria suggest that, among many those estimators, EL ($\hat{\theta}_\alpha$ with $\alpha=0$) has some optimality properties (Kitamura (2006)) assuming that observations are drawn according to P_0 in (2.1). In what follows, however, we argue that MHDE has asymptotic optimal robust properties if observations are drawn from a perturbed version of P_0.

3. Robustness

We now analyze robustness of the MHDE $\hat{\theta}$. Observe that the estimator $\hat{\theta}$ can be interpreted as a mapping of the empirical measure P_n. In other words, for each realization of P_n, we can compute the estimate by $\hat{\theta}$. To make the dependence explicit, let us denote $\hat{\theta}=T\left(P_n\right)$. Although we are interested in the properties of the mapping $T:\mathcal{M}\to\Theta$, the value $T\left(P\right)$ may not exist for some $P\in\mathcal{M}$. Therefore, we consider the mapping defined by a trimmed moment function:

$$\bar{T}\left(Q\right)=\arg\min_{\theta\in\Theta}\left\{\inf_{P\in\bar{\mathcal{P}}_\theta} H\left(P,Q\right)\right\},$$

where $\{m_n\}_{n\in\mathbb{N}}$ is a sequence of positive numbers satisfying $m_n \to \infty$ as $n \to \infty$, and

$$\bar{\mathcal{P}}_\theta = \left\{ P \in \mathcal{M} : \int g(x,\theta)\, \mathbb{I}\{x \in \mathcal{X}_n\}\, dP = 0 \right\},$$
$$\mathcal{X}_n = \left\{ x \in \mathcal{X} : \sup_{\theta\in\Theta} |g(x,\theta)| \le m_n \right\},$$

with the indicator function $\mathbb{I}\{\cdot\}$ and the Euclidean norm $|\cdot|$, i.e., $\mathcal{X}_n$ is a trimming set to bound the moment function and $\bar{\mathcal{P}}_\theta$ is a set of probability measures satisfying the bounded moment condition $E_P[g(x,\theta)\,\mathbb{I}\{x \in \mathcal{X}_n\}] = 0$. Lemma 6.1 (i) guarantees that for each $n \in \mathbb{N}$ and $Q \in \mathcal{M}$ the value $\bar{T}(Q)$ exists.

Let $\tau : \Theta \to \mathbb{R}$ be a transformation of the parameter. We first focus on the estimation problem of the transformed parameter $\tau(\theta_0)$ and investigate the behavior of the bias term $\tau \circ \bar{T}(Q) - \tau(\theta_0)$ in a ($\sqrt{n}$-shrinking) Hellinger ball with radius $r > 0$ around the true probability measure

$$B_H\left(P_0, r/\sqrt{n}\right) = \left\{ Q \in \mathcal{M} : H(Q, P_0) \le r/\sqrt{n} \right\}.$$

In particular, we compare the maximum bias of $\tau \circ \bar{T}$ over the Hellinger ball $B_H(P_0, r/\sqrt{n})$ with that of the alternative mapping $\tau \circ T_a$. We impose the following assumptions. Let $\mathcal{N}$ be an open neighborhood around θ_0.

Assumption 3.1. *Assume that*

- **(i):** $\{x_i\}_{i=1}^n$ *is iid;*
- **(ii):** Θ *is compact;*
- **(iii):** $\theta_0 \in int(\Theta)$ *is a unique solution to* $E_{P_0}[g(x,\theta)] = 0$*;*
- **(iv):** *for each* $\theta \in \Theta$, $g(x,\theta)$ *is continuous for all* $x \in \mathcal{X}$*;*
- **(v):** $E_{P_0}[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta] < \infty$ *for some* $\eta > 2$, $E_{P_0}\left[\sup_{\theta\in\mathcal{N}} |g(x,\theta)|^4\right] < \infty$, $g(x,\theta)$ *is continuously differentiable a.s. in* $\mathcal{N}$, $\sup_{x\in\mathcal{X}_n,\theta\in\mathcal{N}} |\partial g(x,\theta)/\partial\theta'| = o\left(n^{1/2}\right)$, *and* $E_{P_0}\left[\sup_{\theta\in\mathcal{N}} |\partial g(x,\theta)/\partial\theta'|^2\right] < \infty$*;*
- **(vi):** G *has the full column rank and* Ω *is positive definite;*
- **(vii):** $\{m_n\}_{n\in\mathbb{N}}$ *satisfies* $m_n \to \infty$, $n m_n^{-\eta} \to 0$, *and* $n^{-1/2} m_n^{1+\epsilon} = O(1)$ *for some* $0 < \epsilon < 2$ *as* $n \to \infty$*;*[1]
- **(viii):** τ *is continuously differentiable at* θ_0.

1

ALETNATIVELY:: $m_n \sim n^a$ (i.e., $\frac{m_n}{n^a} \to 1$ as $n \to \infty$) with $1/\eta < a < 1/2$;

Assumption 3.1 (i)-(vi) are standard in the literature of the GMM. Assumption 3.1 (iii) is a global identification condition of the true parameter θ_0. Assumption 3.1 (iv) is required to guarantee the continuity of the mapping $\bar{T}(Q)$ in $Q \in \mathcal{M}$ for each $n \in \mathbb{N}$. Assumption 3.1 (v) contains the smoothness and boundedness conditions for the moment function and its derivatives. This assumption is stronger than the one to derive the asymptotic distribution in (2.6). Assumption 3.1 (vi) is a local identification condition for θ_0. This assumption guarantees that the asymptotic variance matrix Σ^{-1} exists. Assumption 3.1 (vii) is on the trimming parameter m_n. If $m_n = n^a$, this assumption is satisfied for $1/\eta < a < 1/2$. Define the projection of the measure Q to the space $\bar{\mathcal{P}}_\theta$ as

$$\bar{P}_{\theta,Q} = \arg\min_{P \in \bar{\mathcal{P}}_\theta} H(P,Q).$$

From the proof of Lemma 6.1 (i), the projection $\bar{P}_{\theta,Q}$ exists for each $n \in \mathbb{N}$, $\theta \in \Theta$, and $Q \in \mathcal{M}$. Assumption 3.1 (viii) is a standard requirement for the parameter transformation τ. The following theorem shows the optimality of the mapping $\tau \circ \bar{T}$ in terms of bias.

Theorem 3.1. *Suppose that Assumption 3.1 holds. If an alternative mapping $T_a : \mathcal{M} \to \Theta$ satisfies*

$$\sqrt{n}\left(\tau \circ T_a\left(\bar{P}_{\theta_0+t/\sqrt{n},P_0}\right) - \tau(\theta_0)\right) = \left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' t + o(1), \tag{3.1}$$

for each $t \in \mathbb{R}^p$, then for each $r > 0$,

$$\liminf_{n\to\infty} \sup_{Q \in B_H(P_0, r/\sqrt{n})} n\left(\tau \circ T(Q) - \tau(\theta_0)\right)^2 \geq 4r^2 B^* = \lim_{n\to\infty} \sup_{Q \in B_H(P_0, r/\sqrt{n})} n\left(\tau \circ \bar{T}(Q) - \tau(\theta_0)\right)^2, \tag{3.2}$$

where $B^* = \left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)' \Sigma^{-1} \left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)$.

Note that the mapping $\bar{T}$ satisfies the restriction (3.1) (because $\bar{T}\left(\bar{P}_{\theta_0+t/\sqrt{n},P_0}\right) = \theta_0 + t/\sqrt{n}$). Therefore, this theorem says that in a class of mappings that satisfies (3.1), the mapping $\bar{T}$ has the smallest maximum bias over the set $B_H(P_0, r/\sqrt{n})$. The (trimmed version of) the Hellinger-based mapping $\bar{T}$ is therefore optimally robust in a minimax sense. Also, if the moment function $\sup_{\theta\in\Theta} |g(x,\theta)|$ is bounded a.s., then we do not need the trimming term $\mathbb{I}\{x \in \mathcal{X}_n\}$ and the mapping for the minimum Hellinger distance estimator T has the above optimal robust property.

We now turn to the analysis of (the supremum of) the mean squared error (MSE) of the minimum Hellinger distance estimator and other competing estimators. Let $\{P_{s,\zeta} : s \in [0,\epsilon]\}$ with some $\epsilon > 0$ be a parametric submodel in the moment restriction model $\mathcal{P} = \cup_{\theta\in\Theta}\mathcal{P}_\theta$, which satisfies $P_{0,\zeta} = P_0$ and is differentiable in quadratic mean at $s = 0$ with score function ζ. A collection of all score functions is called the tangent set $\dot{\mathcal{P}}_{P_0}$. We say that an estimator $\tau \circ T_a(P_n)$ is regular for estimating $\tau(\theta_0) = \tau \circ T_a(P_0)$ if there exists a probability measure M such that

$$\sqrt{n}\left(\tau \circ T_a(P_n) - \tau \circ T_a\left(P_{1/\sqrt{n},\zeta}\right)\right) \xrightarrow{d} M, \quad \text{under } P_{1/\sqrt{n},\zeta}, \tag{3.3}$$

for every $\zeta \in \dot{\mathcal{P}}_{P_0}$. We obtain the following optimal MSE property of the minimum Hellinger distance estimator $\hat{\theta} = T(P_n)$.

Theorem 3.2. *Suppose that Assumption 3.1 holds. Then the following holds for each $r > 0$:*

(i): *If an alternative mapping T_a satisfies (3.1) and $\tau \circ T_a(P_n)$ is regular for estimating $\tau(\theta_0) = \tau \circ T_a(P_0)$, then*

$$\lim_{b\to\infty} \liminf_{n\to\infty} \sup_{Q\in B_H\left(P_0, r/\sqrt{n}\right)} \int b \wedge n\left(\tau \circ T_a(P_n) - \tau(\theta_0)\right)^2 dQ^{\otimes n} \geq \left(1+4r^2\right) B^*,$$

(ii): *the MHDE satisfies*

$$\lim_{b\to\infty} \lim_{n\to\infty} \sup_{Q\in B_H\left(P_0, r/\sqrt{n}\right)} \int b \wedge n\left(\tau \circ \bar{T}(P_n) - \tau(\theta_0)\right)^2 dQ^{\otimes n} = \left(1+4r^2\right) B^*,$$

or

$$\lim_{b\to\infty} \lim_{n\to\infty} \sup_{Q\in B_H\left(P_0, r/\sqrt{n}\right)\cap\left\{Q: E_Q\left[\sup_{\theta\in\Theta}|g(x,\theta)|^\eta\right]<\infty\right\}} \int b \wedge n\left(\tau \circ T(P_n) - \tau(\theta_0)\right)^2 dQ^{\otimes n} \leq \left(1+4r^2\right) B^*,$$

We next derive the Hajek-Le Cam-type convolution theorem. We define a class of regular estimators.

Definition 3.1. The mapping $S_n : (\mathcal{X}^n, \mathcal{B}(\mathcal{X}^n)) \to (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$ is called a *regular* estimator for $\bar{T}$ if for each $r > 0$, and sequence $Q_n \in B_H(P_0, r/\sqrt{n})$, $\sqrt{n}\left(S_n - \bar{T}(Q_n)\right) \xrightarrow{d} M$ under Q_n, where M is some probability distribution on $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$.

The convolution theorem is stated as follows.

Theorem 3.3. *Suppose that Assumption 3.1 holds with $nm_n^{1-\eta} \to 0$. Then*

(i): *for any regular estimator S_n, the limiting distribution M is represented as*

$$M = M_0 * N\left(0, \Sigma^{-1}\right),$$

where M_0 satisfies

$$\sqrt{n}\left(S_n - \theta_0 + \sqrt{n}\Sigma^{-1}\int \Lambda dP_n\right) \to_d M_0 \quad \text{under } P_0,$$

(ii): *$\bar{T}(P_n)$ has the limiting distribution:*

$$M^* = N\left(0, \Sigma^{-1}\right),$$

and

$$\sqrt{n}\left(\bar{T}(P_n) - \theta_0 + \sqrt{n}\Sigma^{-1}\int \Lambda dP_n\right) = o_p(1) \quad \text{under } P_0.$$

Let $\bar{\mathbb{R}}$ be the extended real line $[-\infty, \infty]$. We consider the following loss functions.

Assumption 3.2. *The loss function $\ell : \bar{\mathbb{R}}^p \to [0, \infty]$ is (i) symmetric subconvex (i.e., for all $z \in \mathbb{R}^p$ and $c \in \mathbb{R}$, $\ell(z) = \ell(-z)$ and $\{z \in \mathbb{R}^p : \ell(z) \leq c\}$ is convex); (ii) upper semicontinuous at infinity; and (iii) continuous on $\bar{\mathbb{R}}^p$.*

Let $\mathcal{S}$ be a set of all estimators, i.e., a set of all $\bar{\mathbb{R}}^p$-valued measurable functions. We now present an optimality result for the modified minimum Hellinger estimator $\bar{T}(P_n)$.

Theorem 3.4. *Suppose that Assumptions 3.1 and 3.2 hold. Then*

(i):

$$\lim_{r\to\infty} \liminf_{n\to\infty} \inf_{S_n \in \mathcal{S}} \sup_{Q \in B_H\left(P_0, r/\sqrt{n}\right)} \int \ell\left(\sqrt{n}\left(S_n - \bar{T}(Q)\right)\right) dQ^{\otimes n} \geq \int \ell dN\left(0, \Sigma^{-1}\right),$$

(ii): *for every $b > 0$ and $r > 0$, $\bar{T}(P_n)$ satisfies:*

$$\lim_{n\to\infty} \sup_{Q \in B_H\left(P_0, r/\sqrt{n}\right)} \int b \wedge \ell\left(\sqrt{n}\left(\bar{T}(P_n) - \bar{T}(Q)\right)\right) dQ^{\otimes n} = \int b \wedge \ell dN\left(0, \Sigma^{-1}\right).$$

More results to be added.

4. Conclusion

To be added.

5. Appendix

This Appendix presents the proofs of some of the results presented in the previous sections.

Notation. Let $C > 0$ be a generic positive constant, $\|\cdot\|$ is the L_2-metric,

$$
\begin{aligned}
\theta_n &= \theta_0 + t/\sqrt{n}, \quad \bar{T}_{Q_n} = \bar{T}(Q_n), \quad \bar{T}_{P_n} = \bar{T}(P_n), \\
g_n(x,\theta) &= g(x,\theta)\mathbb{I}\{x \in \mathcal{X}_n\}, \quad R_n(Q,\theta,\gamma) = -\int \frac{1}{(1+\gamma' g_n(x,\theta))} dQ, \\
G &= E_{P_0}\left[\partial g(x,\theta_0)/\partial\theta'\right], \quad \Omega = E_{P_0}\left[g(x,\theta_0) g(x,\theta_0)'\right], \\
\Lambda_n &= G'\Omega^{-1} g_n(x,\theta_0), \quad \Lambda = G'\Omega^{-1} g(x,\theta_0), \\
\psi_{n,Q_n} &= -2\left(\int \Lambda_n \Lambda_n' dQ_n\right)^{-1} \int \Lambda_n \left\{dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2}\right\} dQ_n^{1/2},
\end{aligned}
$$

5.1. **Proof of Theorem 3.1.** The proof is based on Rieder (1994, proof of Theorem 5.3.5). We first derive the lower bound B^*. Pick any $r > 0$ and $t \in \mathbb{R}^p$. From the definition of the Hellinger distance and the triangle inequality,

$$
H\left(\bar{P}_{\theta_n,P_0}, P_0\right) \leq \left\| d\bar{P}_{\theta_n,P_0}^{1/2} - dP_0^{1/2} + \frac{1}{2\sqrt{n}} t'\Lambda_n dP_0^{1/2} \right\| + \frac{1}{2\sqrt{n}} \left\| t'\Lambda_n dP_0^{1/2} \right\|. \tag{5.1}
$$

From the convex duality of partially finite programming (Borwein and Lewis (1993)), the Radon-Nikodym derivative $d\bar{P}_{\theta,Q}/dQ$ is written as

$$
\frac{d\bar{P}_{\theta,Q}}{dQ} = \frac{1}{\left(1 + \gamma_n(\theta,Q)' g_n(x,\theta)\right)^2}, \tag{5.2}
$$

for each $n \in \mathbb{N}$, $\theta \in \Theta$, and $Q \in \mathcal{M}$, where $\gamma_n(\theta,Q)$ solves

$$
0 = \int \frac{g_n(x,\theta)}{\left(1+\gamma_n(\theta,Q)' g_n(x,\theta)\right)^2} dQ = \int g_n(x,\theta)\left\{1 - 2\gamma_n(\theta,Q)' g_n(x,\theta) + \varrho_n(x,\theta,Q)\right\} dQ,
$$

where

$$
\varrho_n(x,\theta,Q) = \frac{3\left(\gamma_n(\theta,Q)' g_n(x,\theta)\right)^2 + 2\left(\gamma_n(\theta,Q)' g_n(x,\theta)\right)^3}{\left(1+\gamma_n(\theta,Q)' g_n(x,\theta)\right)^2}.
$$

Thus, if $\int g_n(x,\theta) g_n(x,\theta)' dQ$ is invertible, $\gamma_n(\theta,Q)$ is written as

$$
\gamma_n(\theta,Q) = \frac{1}{2}\left(\int g_n(x,\theta) g_n(x,\theta)' dQ\right)^{-1} \int g_n(x,\theta) dQ + \left(\int g_n(x,\theta) g_n(x,\theta)' dQ\right)^{-1} \int \varrho_n(x,\theta,Q) g_n(x,\theta) dQ. \tag{5.3}
$$

We now show that the first term of (5.1) is $o\left(n^{-1/2}\right)$. Note that $E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right]$ is invertible for all $n \in \mathbb{N}$ large enough by Lemma 6.4 (i) and Assumption 3.1 (vi). So, using (5.2), (5.3), and the triangle inequality, the first term of (5.1) is written as

$$\begin{aligned}
& \left\| d\bar{P}_{\theta_n,P_0}^{1/2} - dP_0^{1/2} + \frac{1}{2\sqrt{n}} t' \Lambda_n dP_0^{1/2} \right\| \\
= & \left\| -\frac{\gamma_n(\theta_n,P_0)' g_n(x,\theta_n)}{\left(1+\gamma_n(\theta_n,P_0)' g_n(x,\theta_n)\right)} dP_0^{1/2} + \frac{1}{2\sqrt{n}} t' G' \Omega^{-1} g_n(x,\theta_0) dP_0^{1/2} \right\| \\
\leq & \left\| -\frac{1}{2} E_{P_0}[g_n(x,\theta_n)]' E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right]^{-1} g_n(x,\theta_n) dP_0^{1/2} + \frac{1}{2\sqrt{n}} t' G' \Omega^{-1} g_n(x,\theta_0) dP_0^{1/2} \right\| \\
& + \left\| \frac{1}{2} \frac{\gamma_n(\theta_n,P_0)' g_n(x,\theta_n)}{1+\gamma_n(\theta_n,P_0)' g_n(x,\theta_n)} E_{P_0}[g_n(x,\theta_n)]' E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right]^{-1} g_n(x,\theta_n) dP_0^{1/2} \right\| \\
& + \left\| \frac{\left(\int \varrho_n(x,\theta_n,P_0) g_n(x,\theta_n) dP_0\right)' E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right]^{-1} g_n(x,\theta_n)}{\left(1+\gamma_n(\theta_n,P_0)' g_n(x,\theta_n)\right)} dP_0^{1/2} \right\| \\
= & \; T_1 + T_2 + T_3,
\end{aligned}$$

for all $n \in \mathbb{N}$ large enough. For T_1, the triangle inequality and a Taylor expansion around $t = 0$ yield

$$\begin{aligned}
T_1 \leq & \left\| -\frac{1}{2} E_{P_0}[g_n(x,\theta_n)]' \left(E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right]^{-1} - \Omega^{-1} \right) g_n(x,\theta_n) dP_0^{1/2} \right\| \\
& + \left\| -\frac{1}{2} E_{P_0}[g_n(x,\theta_n)]' \Omega^{-1} g_n(x,\theta_n) dP_0^{1/2} \right\| \\
& + \left\| -\frac{1}{2\sqrt{n}} t' \left(E_{P_0}\left[\frac{\partial g_n\left(x,\theta_0+\dot{t}/\sqrt{n}\right)}{\partial t'} \right] - G \right)' \Omega^{-1} g_n(x,\theta_n) dP_0^{1/2} \right\| \\
& + \left\| -\frac{1}{2\sqrt{n}} t' G' \Omega^{-1} \left\{ g_n(x,\theta_n) - g_n(x,\theta_0) \right\} dP_0^{1/2} \right\| \\
= & \; o\left(n^{-1/2}\right),
\end{aligned}$$

where $\dot{t}$ is a point on the line joining 0 and t, and the equality follows from Lemma 6.4 (i). For T_2, Lemma 6.4 implies

$$T_2 \leq o(1) \left\| E_{P_0}[g_n(x,\theta_n)]' E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right]^{-1} g_n(x,\theta_n) dP_0^{1/2} \right\| = o\left(n^{-1/2}\right).$$

For T_3, we have

$$\begin{aligned}
T_3 \leq & \; O(1) \left| \int \left(\gamma_n(\theta_n,P_0)' g_n(x,\theta_n)\right)^2 g_n(x,\theta_n) dP_0 \right| \\
\leq & \; O(m_n) \left| \gamma_n(\theta_n,P_0)' \left(\int g_n(x,\theta_n) g_n(x,\theta_n)' dP_0 \right) \gamma_n(\theta_n,P_0) \right| \\
\leq & \; O\left(n^{-1} m_n\right) = o\left(n^{-1/2}\right),
\end{aligned}$$

where the first inequality follows from Lemma 6.4 (ii), the third inequality follows from Lemma 6.4, and the equality follows from Assumption 3.1 (vii). Combining these results, the first term of (5.1) is

$o\left(n^{-1/2}\right)$ and

$$\lim_{n\to\infty} nH\left(\bar{P}_{\theta_n,P_0},P_0\right)^2=\lim_{n\to\infty}\frac{1}{4}t'\left(\int\Lambda_n\Lambda_n' dP_0\right)t=\frac{1}{4}t'\Sigma t,$$

for all $t\in\mathbb{R}^p$, where the second equality follows from Lemma 6.4 (i). This implies that

$$\bar{P}_{\theta_n,P_0}\in B_H\left(P_0,r/\sqrt{n}\right) \tag{5.4}$$

for any t satisfying $\frac{1}{4}t'\Sigma t\le r^2-\epsilon$ for any $\epsilon\in\left(0,r^2\right)$ and n large enough. From (3.1), T_a satisfies

$$\lim_{n\to\infty} n\left(\tau\circ T_a\left(\bar{P}_{\theta_n,P_0}\right)-\tau\left(\theta_0\right)\right)^2=\left(\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'t\right)^2, \tag{5.5}$$

for each $t\in\mathbb{R}^p$. Therefore, similar to Rieder (1994, eq. (56) on p. 180), the lower bound of the maximum oscillation of T_a is obtained as

$$\begin{aligned}
&\liminf_{n\to\infty}\sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)} n\left(\tau\circ T_a\left(Q\right)-\tau\left(\theta_0\right)\right)^2\\
\ge\;& \liminf_{n\to\infty}\sup_{\left\{t\in\mathbb{R}^p:\bar{P}_{\theta_n,P_0}\in B_H\left(P_0,r/\sqrt{n}\right)\right\}} n\left(\tau\circ T_a\left(\bar{P}_{\theta_n,P_0}\right)-\tau\left(\theta_0\right)\right)^2\\
\ge\;& \max_{\left\{t\in\mathbb{R}^p:\frac{1}{4}t'\Sigma t\le r^2-\epsilon\right\}}\left(\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'t\right)^2\\
=\;& 4\left(r^2-\epsilon\right)\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'\Sigma^{-1}\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)=4\left(r^2-\epsilon\right)B^*,
\end{aligned} \tag{5.6}$$

for any $\epsilon\in\left(0,r^2\right)$, where the first inequality follows from the set inclusion relationship, the second inequality follows from (5.4) and (5.5), and the first equality follows from the Kuhn-Tucker theorem. Since ϵ can be arbitrarily small the minimum bound is $4r^2B^*$. Since Σ is positive definite from Assumption 3.1 (vi), the lower bound B^* is positive and finite.

[This part is also based on the proof of Rieder (1994, Theorem 5.3.5)] We now show that the limit of the maximum oscillation of the mapping $\tau\circ\bar{T}$ attains the bound B^*. A Taylor expansion of $\tau\circ\bar{T}_{Q_n}$ around $\bar{T}_{Q_n}=\theta_0$, Lemmas 6.1 (ii) and 6.2, and Assumption 3.1 (viii) imply that for each sequence $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$ and $r>0$,

$$\begin{aligned}
&\sqrt{n}\left(\tau\circ\bar{T}_{Q_n}-\tau\left(\theta_0\right)\right)=-\sqrt{n}\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'\Sigma^{-1}\int\Lambda_n dQ_n+o\left(1\right)\\
=\;& -\sqrt{n}\nu_0\int\Lambda_n\left\{dQ_n^{1/2}-dP_0^{1/2}\right\}dQ_n^{1/2}-\sqrt{n}\nu_0\int\Lambda_n dP_0^{1/2}\left\{dQ_n^{1/2}-dP_0^{1/2}\right\}+o\left(1\right),
\end{aligned}$$

where we denote $\nu_0 = \left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' \Sigma^{-1}$. From the triangle inequality,

$$\begin{aligned} & n\left(\tau \circ \bar{T}_{Q_n} - \tau(\theta_0)\right)^2 \\ \leq\ & n \left\{ \begin{array}{l} \left|\nu_0 \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dQ_n^{1/2}\right|^2 + \left|\nu_0 \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dP_0^{1/2}\right|^2 \\ +2\left|\nu_0 \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dQ_n^{1/2}\right| \left|\nu_0 \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dP_0^{1/2}\right| \end{array} \right\} + o(1) \\ =\ & n\left\{A_1 + A_2 + 2A_3\right\} \end{aligned}$$

For A_1, observe that

$$A_1 \leq \left|\int \nu_0 \Lambda_n \Lambda_n' \nu_0' dQ_n\right| \left|\int \left\{dQ_n^{1/2} - dP_0^{1/2}\right\}^2\right| \leq B^* \frac{r^2}{n} + o\left(n^{-1}\right),$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from Lemma 6.5 (i) and $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$. Similarly, we have $A_2 \leq B^* \frac{r^2}{n}$. From these results, A_3 satisfies

$$A_3 \leq \sqrt{B^* \frac{r^2}{n} + o\left(n^{-1}\right)} \sqrt{B^* \frac{r^2}{n}} = B^* \frac{r^2}{n} + o\left(n^{-1}\right).$$

Combining these terms,

$$\limsup_{n \to \infty} n\left(\tau \circ \bar{T}_{Q_n} - \tau(\theta_0)\right)^2 \leq 4r^2 B^*,$$

for any sequence $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$ and $r > 0$. Since $B_H\left(P_0, r/\sqrt{n}\right)$ is compact for each $n \in \mathbb{N}$ and $r > 0$, we have

$$\limsup_{n \to \infty} \sup_{Q_n \in B_H\left(P_0, r/\sqrt{n}\right)} n\left|\tau \circ \bar{T}_{Q_n} - \tau(\theta_0)\right|^2 \leq 4r^2 B^*.$$

5.2. **Proof of Theorem 3.2. Proof of (i).** Pick any $\epsilon \in \left(0, r^2\right)$. From $\lim_{n \to \infty} nH\left(\bar{P}_{\theta_n, P_0}, P_0\right)^2 = \frac{1}{4} t' \Sigma t$, we have $\left\{\bar{P}_{\theta_n, P_0} \in \mathcal{M} : \frac{1}{4} t' \Sigma t \leq r^2 - \epsilon\right\} \subset B_H\left(P_0, r/\sqrt{n}\right)$ for all n large enough. Thus,

$$\begin{aligned} & \lim_{b \to \infty} \liminf_{n \to \infty} \sup_{Q \in B_H\left(P_0, r/\sqrt{n}\right)} \int b \wedge n\left(\tau \circ T_a(P_n) - \tau(\theta_0)\right)^2 dQ^{\otimes n} \\ \geq\ & \lim_{b \to \infty} \liminf_{n \to \infty} \sup_{Q \in \left\{\bar{P}_{\theta_n, P_0} \in \mathcal{M} : \frac{1}{4} t' \Sigma t \leq r^2 - \epsilon\right\}} \int b \wedge n\left(\tau \circ T_a(P_n) - \tau(\theta_0)\right)^2 dQ^{\otimes n} \\ \geq\ & \lim_{b \to \infty} \liminf_{n \to \infty} \int b \wedge n\left(\tau \circ T_a(P_n) - \tau(\theta_0)\right)^2 d\bar{P}_{\theta_n, P_0}^{\otimes n}, \end{aligned} \tag{5.7}$$

for each $\left\{t \in \mathbb{R}^p : \frac{1}{4} t' \Sigma t \leq r^2 - \epsilon\right\}$. Note that from the proof of Theorem 3.1, we can show that the submodel $P_{s,-\Lambda}$ satisfying $P_{1/\sqrt{n},-\Lambda} = \bar{P}_{\theta_n, P_0}$ for each $n \in \mathbb{N}$ is differentiable in quadratic mean at $s = 0$ with score function $-\Lambda$. Thus, the convolution theorem (van der Vaart (1998, Theorem 25.20)) implies that for each $\left\{t \in \mathbb{R}^p : \frac{1}{4} t' \Sigma t \leq r^2 - \epsilon\right\}$,

$$\sqrt{n}\left(\tau \circ T_a(P_n) - \tau \circ T_a\left(\bar{P}_{\theta_n, P_0}\right)\right) \xrightarrow{d} dM_0 * N\left(0, B^*\right) \quad \text{under } \bar{P}_{\theta_n, P_0}, \tag{5.8}$$

for some probability measure M_0, which does not depend on t. Let t^* be a solution of $\max_{\left\{t\in\mathbb{R}^p:\frac{1}{4}t'\Sigma t\leq r^2-\epsilon\right\}}\left(\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)'t\right)^2$ satisfying $\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)'t^*\int\xi dM_0*N\left(0,B^*\right)\geq 0$. Such a choice is possible because the integral $\int\xi dM_0*N\left(0,B^*\right)$ does not depend on t. Therefore,

$$
\begin{aligned}
&\lim_{b\to\infty}\liminf_{n\to\infty}\int b\wedge n\left(\tau\circ T_a\left(P_n\right)-\tau\left(\theta_0\right)\right)^2 d\bar{P}^{\otimes n}_{\theta_0+t^*/\sqrt{n},P_0}\\
=\;&\lim_{b\to\infty}\int b\wedge\left(\xi+\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'t^*\right)^2 dM_0*N\left(0,B^*\right)\\
=\;&\int\xi^2 dM_0*N\left(0,B^*\right)+\left(\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'t^*\right)^2+2\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'t^*\int\xi dM_0*N\left(0,B^*\right)\\
\geq\;&\int\xi^2 dN\left(0,B^*\right)+4\left(r^2-\epsilon\right)B^*\\
=\;&\left\{1+4\left(r^2-\epsilon\right)\right\}B^*, \qquad (5.9)
\end{aligned}
$$

where the first equality follows from (3.1), (5.8), and the continuous mapping theorem, the second equality follows from the monotone convergence theorem, the inequality follows from Anderson's lemma (see, e.g., van der Vaart (1998, Lemma 8.5)) and the definition of t^*. From (5.7) and (5.9), the conclusion is obtained.

Proof of (ii). Pick any $r>0$. Observe that

$$
\begin{aligned}
&\limsup_{n\to\infty}\sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)}\int b\wedge n\left(\tau\circ T\left(P_n\right)-\tau\left(\theta_0\right)\right)^2 dQ^{\otimes n}\\
\leq\;&\limsup_{n\to\infty}\sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)}\int b\wedge n\left(\tau\circ T\left(P_n\right)-\tau\circ\bar{T}\left(P_n\right)\right)^2 dQ^{\otimes n}\\
&+2\limsup_{n\to\infty}\sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)}\int b\wedge\left\{n\left|\tau\circ T\left(P_n\right)-\tau\circ\bar{T}\left(P_n\right)\right|\left|\tau\circ\bar{T}\left(P_n\right)-\tau\left(\theta_0\right)\right|\right\}dQ^{\otimes n}\\
&+\limsup_{n\to\infty}\sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)}\int b\wedge n\left(\tau\circ\bar{T}\left(P_n\right)-\tau\left(\theta_0\right)\right)^2 dQ^{\otimes n}\\
=\;&A_1+2A_2+A_3,
\end{aligned}
$$

for each $b > 0$, where the inequality follows from the triangle inequality and $b \wedge (c_1 + c_2) \leq b \wedge c_1 + b \wedge c_2$ for any $c_1, c_2 \geq 0$. For A_1,

$$\begin{aligned} A_1 &= \limsup_{n\to\infty} \sup_{Q\in B_H\left(P_0, r/\sqrt{n}\right)} \left\{ \int_{(x_1,\ldots,x_n)\in\mathcal{X}_n^n} b \wedge n \left(\tau \circ T(P_n) - \tau \circ \bar{T}(P_n)\right)^2 dQ^{\otimes n} \right. \\ &\quad \left. + \int_{(x_1,\ldots,x_n)\notin\mathcal{X}_n^n} b \wedge n \left(\tau \circ T(P_n) - \tau \circ \bar{T}(P_n)\right)^2 dQ^{\otimes n} \right\} \\ &\leq b \times \limsup_{n\to\infty} \sup_{Q\in B_H\left(P_0, r/\sqrt{n}\right)} \int_{(x_1,\ldots,x_n)\notin\mathcal{X}_n^n} dQ^{\otimes n} \\ &\leq b \times \limsup_{n\to\infty} \sup_{Q\in B_H\left(P_0, r/\sqrt{n}\right)} nE_Q \left[\mathbb{I}\left\{ \sup_{\theta\in\Theta} |g(x,\theta)| \geq m_n \right\} \right]. \end{aligned}$$

If $E_Q\left[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta\right] < \infty$ for all $Q \in B_H\left(P_0, r/\sqrt{n}\right)$, then the Markov inequality and Assumption 3.1 (vii) imply that

$$A_1 \leq b \times \limsup_{n\to\infty} \sup_{Q\in B_H\left(P_0, r/\sqrt{n}\right)} n m_n^{-\eta} E_Q \left[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta \right] \to 0.$$

However, in general, we only have

$$\begin{aligned} & E_Q \left[\mathbb{I}\left\{ \sup_{\theta\in\Theta} |g(x,\theta)| \geq m_n \right\} \right] \\ = & \int \mathbb{I}\left\{ \sup_{\theta\in\Theta} |g(x,\theta)| \geq m_n \right\} \left\{ dQ^{1/2} - dP_0^{1/2} \right\}^2 + 2 \int \mathbb{I}\left\{ \sup_{\theta\in\Theta} |g(x,\theta)| \geq m_n \right\} dP_0^{1/2} \left\{ dQ^{1/2} - dP_0^{1/2} \right\} \\ & + E_{P_0} \left[\mathbb{I}\left\{ \sup_{\theta\in\Theta} |g(x,\theta)| \geq m_n \right\} \right] \\ \leq & \frac{r^2}{n} + 2\sqrt{E_{P_0} \left[\mathbb{I}\left\{ \sup_{\theta\in\Theta} |g(x,\theta)| \geq m_n \right\} \right]} \frac{r}{\sqrt{n}} + E_{P_0} \left[\mathbb{I}\left\{ \sup_{\theta\in\Theta} |g(x,\theta)| \geq m_n \right\} \right] \\ \leq & \frac{r^2}{n} + 2\sqrt{E_{P_0} \left[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta \right]} m_n^{-\eta/2} \frac{r}{\sqrt{n}} + E_{P_0} \left[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta \right] m_n^{-\eta}, \end{aligned}$$

for all $Q \in B_H\left(P_0, r/\sqrt{n}\right)$. Thus, from Assumption 3.1 (v) and (vii),

$$A_1 \leq br^2.$$

For A_2, a similar argument to A_1 yields

$$\begin{aligned}
A_2 &= \limsup_{n\to\infty} \sup_{Q\in B_H(P_0,r/\sqrt{n})} \left\{ \int_{(x_1,\dots,x_n)\in\mathcal{X}_n^n} b\wedge\left\{n\left|\tau\circ T(P_n)-\tau\circ\bar{T}(P_n)\right|\left|\tau\circ\bar{T}(P_n)-\tau(\theta_0)\right|\right\} dQ^{\otimes n} \right.\\
&\quad \left. +\int_{(x_1,\dots,x_n)\notin\mathcal{X}_n^n} b\wedge\left\{n\left|\tau\circ T(P_n)-\tau\circ\bar{T}(P_n)\right|\left|\tau\circ\bar{T}(P_n)-\tau(\theta_0)\right|\right\} dQ^{\otimes n} \right\}\\
&\leq b\times\limsup_{n\to\infty} \sup_{Q\in B_H(P_0,r/\sqrt{n})} \int_{(x_1,\dots,x_n)\notin\mathcal{X}_n^n} dQ^{\otimes n}\\
&\leq b\times\limsup_{n\to\infty} \sup_{Q\in B_H(P_0,r/\sqrt{n})} nE_Q\left[\mathbb{I}\left\{\sup_{\theta\in\Theta}|g(x,\theta)|\geq m_n\right\}\right]\\
&\leq br^2.
\end{aligned}$$

We now show that

$$A_3 \leq \left(1+4r^2\right)B^*,$$

as $b\to\infty$. Pick any $b>0$. Note that the mapping $f_{b,n}(Q)=\int b\wedge n\left(\tau\circ\bar{T}(P_n)-\tau(\theta_0)\right)^2 dQ^{\otimes n}$ is continuous in $Q\in B_H(P_0,r/\sqrt{n})$ under the Hellinger distance for each n, and the set $B_H(P_0,r/\sqrt{n})$ is compact under the Hellinger distance for each n. Thus, there exists $\tilde{Q}_{b,n}\in B_H(P_0,r/\sqrt{n})$ such that $\sup_{Q\in B_H(P_0,r/\sqrt{n})} f_n(Q)=f_n\left(\tilde{Q}_{b,n}\right)$ for each n. Then we have

$$\begin{aligned}
A_3 &= \limsup_{n\to\infty} \sup_{Q\in B_H(P_0,r/\sqrt{n})} \int b\wedge n\left(\tau\circ\bar{T}(P_n)-\tau(\theta_0)\right)^2 dQ^{\otimes n}\\
&= \limsup_{n\to\infty} \int b\wedge n\left(\tau\circ\bar{T}(P_n)-\tau(\theta_0)\right)^2 d\tilde{Q}_{b,n}^{\otimes n}\\
&= \int b\wedge\left(z+\tilde{t}_b\right)^2 dN(0,B^*)\\
&\leq B^*+\tilde{t}_b^2\\
&\leq \left(1+4r^2\right)B^*,
\end{aligned}$$

where $\tilde{t}_b=\limsup_{n\to\infty}\sqrt{n}\left(\tau\circ\bar{T}\left(\tilde{Q}_{b,n}\right)-\tau(\theta_0)\right)$, the third equality follows from Lemma 6.8 and the continuous mapping theorem, the first inequality follows from $b\wedge c\leq c$ and a direct calculation, and the second inequality follows from Theorem 3.1. Therefore, the conclusion is obtained.

5.3. **Proof of Theorem 3.3. Proof of (i).** The proof is based on Rieder (1994, proof of Theorem 4.3.2). The proof is split into several steps.

Step 1: Analyze the simple perturbations $Q_n(\zeta,t)$ for bounded scores. [See Rieder (1994, proof of Theorem 4.3.2)] Let

$$\begin{aligned} L_\infty^p &= \left\{\text{all } \mathbb{R}^p\text{-valued bounded measurable functions on } \mathbb{R}^d\right\}, \\ Z_\infty^p(P_0) &= \{\zeta \in L_\infty^p : E_{P_0}\zeta = 0\} \text{ (space of bounded tangents at } P_0). \end{aligned}$$

Consider the simple perturbations of P_0 along ζ:

$$\frac{dQ_n(\zeta,t)}{dP_0} = 1 + \frac{1}{\sqrt{n}}t'\zeta.$$

Pick any $\zeta \in Z_\infty^p(P_0)$ satisfying

$$\det E_{P_0}\left[\Lambda\zeta'\right] \neq 0.$$

From Assumption 3.1 (vi), $E_{P_0}[\zeta\zeta']$ is positive definite. Pick any $t \in \mathbb{R}^p$. From the definition of $dQ_n(\zeta,t)/dP_0$, we have the Hellinger differentiability of $Q_n(\zeta,t)$, i.e.,

$$\begin{aligned} &\sqrt{n}\left\|dQ_n(\zeta,t)^{1/2} - dP_0^{1/2} - \frac{1}{2\sqrt{n}}t'\zeta dP_0^{1/2}\right\| \\ = &\sqrt{n}\left\|\left\{\left(1+\frac{1}{\sqrt{n}}\dot{t}'\zeta\right)^{-1/2} - 1\right\}\frac{1}{2\sqrt{n}}t'\zeta dP_0^{1/2}\right\| \to 0, \end{aligned} \tag{5.10}$$

where the equality follows from a Taylor expansion around $t = 0$ and $\dot{t}$ is a point on the line joining 0 and t, and the convergence follows from $\zeta \in Z_\infty^p(P_0)$. This implies

$$\begin{aligned} &\sqrt{n}\left\|dQ_n(\zeta,t)^{1/2} - dP_0^{1/2}\right\| \\ \leq\ &\sqrt{n}\left\|dQ_n(\zeta,t)^{1/2} - dP_0^{1/2} - \frac{1}{2\sqrt{n}}t'\zeta dP_0^{1/2}\right\| + \sqrt{n}\left\|\frac{1}{2\sqrt{n}}t'\zeta dP_0^{1/2}\right\| \\ =\ &o(1) + \frac{1}{2}t'E_{P_0}\left[\zeta\zeta'\right]t. \end{aligned}$$

Since ζ is bounded, $Q_n(\zeta,t) \in B_H(P_0, r/\sqrt{n})$ for some $r > 0$. Since S_n is regular,

$$\sqrt{n}\left(S_n - \bar{T}(Q_n(\zeta,t))\right) \xrightarrow{d} M \quad \text{under } Q_n(\zeta,t). \tag{5.11}$$

Also we have

$$\begin{aligned} &\sqrt{n}\left(\bar{T}(Q_n(\zeta,t)) - \theta_0\right) \\ = &-\sqrt{n}\Sigma^{-1}\int \Lambda_n dQ_n(\zeta,t) + o(1) \\ = &-\sqrt{n}\Sigma^{-1}E_{P_0}[\Lambda_n] - \Sigma^{-1}E_{P_0}\left[\Lambda_n\zeta'\right]t + o(1) \\ = &-\Sigma^{-1}E_{P_0}\left[\Lambda\zeta'\right]t + o(1), \end{aligned} \tag{5.12}$$

for each $\zeta \in Z_\infty^p(P_0)$ and $t \in \mathbb{R}^p$, where the first equality follows from $Q_n(\zeta,t) \in B_H(P_0, r/\sqrt{n})$ and Lemma 6.2, the second equality follows from the definition of $Q_n(\zeta,t)$, and the third equality follows from Lemma 6.4 (i) and $E_{P_0}[\Lambda_n\zeta'] \to E_{P_0}[\Lambda\zeta']$ as $n \to \infty$ (by applying the same argument as the proof of Lemma 6.4 (i)).

Step 2: Derive asymptotic normality of $Q_n(\zeta,t)$. [See Rieder (1994, Lemma 4.2.4)] From the definition of $Q_n(\zeta,t)$ and the L^2-differentiability of $Q_n(\zeta,t)$ in (5.10), we can apply a standard likelihood expansion (e.g., Rieder (1994, Lemma 4.2.4)), that is

$$\log \frac{dQ_n^{\otimes n}(\zeta,t)}{dP_0^{\otimes n}} = t'\frac{1}{\sqrt{n}}\sum_{i=1}^n \zeta(x_i) - \frac{1}{2}t'E_{P_0}\left[\zeta\zeta'\right]t + o_p(1) \quad \text{under } P_0. \tag{5.13}$$

This implies that $\{Q_n(\zeta,t)\}_{n\in\mathbb{N}}$ is asymptotically normal with the asymptotic variance $E_{P_0}[\zeta\zeta']$ and the asymptotic score $n^{-1}\sum_{i=1}^n \zeta(x_i)$ under P_0.

Step 3: Apply a convolution theorem. [See Rieder (1994, Lemma 4.2.4)] From (5.12),

$$\begin{aligned}\sqrt{n}\left(S_n - \bar{T}(Q_n(\zeta,t))\right) &= \sqrt{n}(S_n-\theta_0) - \sqrt{n}\left(\bar{T}(Q_n(\zeta,t)) - \theta_0\right)\\ &= \sqrt{n}(S_n-\theta_0) + \Sigma^{-1}E_{P_0}\left[\Lambda\zeta'\right]t + o(1). \end{aligned}\tag{5.14}$$

Combining the above result with (5.11) implies that the statistic $\sqrt{n}(S_n-\theta_0)$ is regular for estimating the parameter $-\Sigma^{-1}E_{P_0}[\Lambda\zeta']t$ with the limiting distribution M under $Q_n(\zeta,t)$. Note that $\{Q_n(\zeta,t)\}_{n\in\mathbb{N}}$ is asymptotically normal from Step 2. Therefore, we can apply a Hájek's convolution theorem for normal experiments (e.g., Rieder (1994, Theorem 3.2.3)), i.e., the limiting distribution M is represented as

$$M = M_0(\zeta) * N(0, \Xi(\zeta)), \tag{5.15}$$

where $\Xi(\zeta) = \Sigma^{-1}E_{P_0}[\Lambda\zeta']E_{P_0}[\zeta\zeta']^{-1}E_{P_0}[\Lambda\zeta']'\Sigma^{-1}$ and the distribution $M_0(\zeta)$ satisfies

$$\sqrt{n}(S_n-\theta_0) + \Sigma^{-1}E_{P_0}\left[\Lambda\zeta'\right]E_{P_0}\left[\zeta\zeta'\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \zeta(x_i) \xrightarrow{d} M_0(\zeta) \quad \text{under } P_0. \tag{5.16}$$

This holds for any $\zeta \in Z_\infty^p(P_0)$.

Step 4: Generalize to L_2^p score functions. [See Rieder (1994, second half of proof of Theorem 4.3.2)] Let $M^F(u)$ and $M_0^F(\zeta,u)$ be the Fourier transformations of M and $M_0(\zeta)$, respectively. From (5.15),

$$M^F(u)\exp\left(\frac{1}{2}u'\Xi(\zeta)u\right) = M_0^F(\zeta,u). \tag{5.17}$$

Consider $\tilde{\zeta} \in L_\infty^p$ such that $\tilde{\zeta} \to \Sigma^{-1}\Lambda \in L_2^p$ under the L_2-norm. Then we have $\Xi\left(\tilde{\zeta}\right) \to \Sigma^{-1}$, and the continuity theorem yields that $M_0\left(\tilde{\zeta}\right)$ converges weakly to M_0 with the Fourier transform

$$M_0^F(u) = M^F(u)\exp\left(\frac{1}{2}u'\Sigma^{-1}u\right).$$

Therefore, for any regular estimator, the limiting distribution M is written as

$$M = M_0 * N\left(0, \Sigma^{-1}\right). \tag{5.18}$$

Therefore, the first statement of Part (i) is proved. The second statement of Part (i) follows from

$$\begin{aligned}
&\Pr\left\{\left|\Sigma^{-1}E_{P_0}\left[\Lambda\tilde{\zeta}'\right]E_{P_0}\left[\tilde{\zeta}\tilde{\zeta}'\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{\zeta}(x_i) - \Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\Lambda(x_i)\right| > \epsilon : P_0\right\} \\
\leq\ &\frac{1}{\epsilon^2}E_{P_0}\left[\left|\Sigma^{-1}E_{P_0}\left[\Lambda\tilde{\zeta}'\right]E_{P_0}\left[\tilde{\zeta}\tilde{\zeta}'\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{\zeta}(x_i) - \Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\Lambda(x_i)\right|^2\right] \to 0,
\end{aligned}$$

where the inequality follows from the Markov inequality and the convergence follows from $\tilde{\zeta} \to \Sigma^{-1}\Lambda$ under the L_2-norm. Therefore, the conclusion is obtained.

Proof of (ii). This is implied from Lemma 6.8.

5.4. **Proof of Theorem 3.4. Proof of (i).** The proof is based on Rieder (1994, proof of Theorem 4.3.4). Pick any $\zeta \in Z_\infty^p(P_0)$ satisfying $\det E_{P_0}[\Lambda\zeta'] \neq 0$. From Assumption 3.2, the loss function ℓ is uniformly continuous on $\bar{\mathbb{R}}^p$ (note that $\bar{\mathbb{R}}^p$ is compact). Thus, from (5.14), the following uniform convergence holds for each $c > 0$:

$$\sup_{|t|\leq c}\left|\ell\left(\sqrt{n}\left\{S_n - \bar{T}\left(Q_n(\zeta,t)\right)\right\}\right) - \ell\left(\sqrt{n}\left\{S_n - \theta_0\right\} + \Sigma^{-1}E_{P_0}\left[\Lambda\zeta'\right]t\right)\right| \to 0.$$

This implies

$$\inf_{S_n\in\mathcal{S}}\sup_{|t|\leq c}\int\ell\left(\sqrt{n}\left\{S_n - \bar{T}\left(Q_n(\zeta,t)\right)\right\}\right)dQ_n^{\otimes n}(\zeta,t) - \inf_{S_n\in\mathcal{S}}\sup_{|t|\leq c}\int\ell\left(\sqrt{n}\left\{S_n - \theta_0\right\} + \Sigma^{-1}E_{P_0}\left[\Lambda\zeta'\right]t\right)dQ_n^{\otimes n}(\zeta,t) \to 0,$$

for each $c > 0$. Since $Q_n(\zeta,t)$ is asymptotically normal from (5.13), we can apply Rieder (1994, Theorem 3.3.8), i.e.,

$$\lim_{c\to\infty}\liminf_{n\to\infty}\inf_{S_n\in\mathcal{S}}\sup_{|t|\leq c}\int\ell\left(\sqrt{n}\left\{S_n - \bar{T}\left(Q_n(\zeta,t)\right)\right\}\right)dQ_n^{\otimes n}(\zeta,t) \geq \int\ell dN\left(0,\Xi(\zeta)\right), \tag{5.19}$$

where $\Xi(\zeta) = \Sigma^{-1}E_{P_0}[\Lambda\zeta']E_{P_0}[\zeta\zeta']^{-1}E_{P_0}[\Lambda\zeta']'\Sigma^{-1}$. Note that (5.19) holds for any $\zeta \in Z_\infty^p(P_0)$. Consider $\tilde{\zeta} \in L_\infty^p$ such that $\tilde{\zeta} \to \Sigma^{-1}\Lambda \in L_\infty^2$ under the L_2-norm. Then we have $\Xi\left(\tilde{\zeta}\right) \to \Sigma^{-1}$, and the pdf of $N\left(0,\Xi\left(\tilde{\zeta}\right)\right)$ converges to the pdf of $N\left(0,\Sigma^{-1}\right)$ at each $x\in\mathcal{X}$. Therefore, Fatou's lemma implies that

$$\liminf_{\tilde{\zeta}\to_{L_2}\Sigma^{-1}\Lambda}\int \ell dN\left(0,\Xi\left(\tilde{\zeta}\right)\right)\geq\int \ell dN\left(0,\Sigma^{-1}\right). \tag{5.20}$$

On the other hand, comparing the variances of $\Sigma^{-1}\Lambda$ and $\xi=\Sigma^{-1}E_{P_0}\left[\Lambda\zeta'\right]E_{P_0}\left[\zeta\zeta'\right]^{-1}\zeta$, we have

$$\begin{aligned} 0 &\leq Var\left(\Sigma^{-1}\Lambda-\xi\right)=\Sigma^{-1}-\Sigma^{-1}E_{P_0}\left[\Lambda\zeta'\right]E_{P_0}\left[\zeta\zeta'\right]^{-1}E_{P_0}\left[\zeta\Lambda'\right]\Sigma^{-1}\\ &= Var\left(\Sigma^{-1}\Lambda\right)-Var\left(\xi\right). \end{aligned}$$

This implies

$$\int \ell dN\left(0,\Sigma^{-1}\right)\geq\int \ell dN\left(0,\Xi\left(\tilde{\zeta}\right)\right), \tag{5.21}$$

for $\tilde{\zeta}\in Z^p_\infty\left(P_0\right)$. From (5.19), (5.20), and (5.21),

$$\lim_{c\to\infty}\liminf_{n\to\infty}\inf_{S_n\in\mathcal{S}}\sup_{|t|\leq c}\int \ell\left(\sqrt{n}\left\{S_n-\bar{T}\left(Q_n\left(\tilde{\zeta},t\right)\right)\right\}\right)dQ_n^{\otimes n}\left(\tilde{\zeta},t\right)\geq\int \ell dN\left(0,\Sigma^{-1}\right). \tag{5.22}$$

Now, from $Q_n\left(\zeta,t\right)\in B_H\left(P_0,r/\sqrt{n}\right)$ (see Step 1 in the proof of Theorem 3.3),

$$\begin{aligned} &\lim_{r\to\infty}\liminf_{n\to\infty}\inf_{S_n\in\mathcal{S}}\sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)}\int \ell\left(\sqrt{n}\left\{S_n-\bar{T}\left(Q\right)\right\}\right)dQ^{\otimes n}\\ \geq\ &\lim_{c\to\infty}\liminf_{n\to\infty}\inf_{S_n\in\mathcal{S}}\sup_{|t|\leq c}\int \ell\left(\sqrt{n}\left\{S_n-\bar{T}\left(Q_n\left(\tilde{\zeta},t\right)\right)\right\}\right)dQ_n^{\otimes n}\left(\tilde{\zeta},t\right)\geq\int \ell dN\left(0,\Sigma^{-1}\right), \end{aligned}$$

where the first inequality follows from a set inclusion relationship, and the second inequality follows from (5.22). The conclusion is obtained.

Proof of (ii). The asymptotic local uniform normality of $\bar{T}$ (Lemma 6.8) and the uniform continuity of ℓ over $\bar{\mathbb{R}}^k$ imply that

$$\lim_{n\to\infty}\sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)}\int b\wedge\ell\left(\sqrt{n}\left\{\bar{T}\left(P_n\right)-\bar{T}\left(Q\right)\right\}\right)dQ^{\otimes n}=\int b\wedge\ell dN\left(0,\Sigma^{-1}\right),$$

for each $b>0$ and $r>0$. The conclusion is obtained.

6. Auxiliary Lemmas

Lemma 6.1. *[Existence and consistency of $\bar{T}$] Suppose that Assumption 3.1 holds. Then*

- **(i):** $\bar{T}\left(Q\right)$ *exists for each* $n\in\mathbb{N}$ *and each* $Q\in\mathcal{M}$,
- **(ii):** $\bar{T}_{Q_n}\to\theta_0$ *as* $n\to\infty$ *for each sequence* $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$ *and each* $r>0$.

Proof of (i). The proof is based on Kitamura (2001, Lemma 1). See also Beran (1984, Handbook chapter, p. 744). Pick any $n \in \mathbb{N}$ and $Q \in \mathcal{M}$. Denote $R_n(Q,\theta) = \inf_{P \in \bar{\mathcal{P}}_\theta} H(P,Q)$. Since $g_n(x,\theta)$ is bounded a.s. for each $n \in \mathbb{N}$ and $\theta \in \Theta$, the duality of partially finite programming (Borwein and Lewis (1993)) yields that for each $(Q,\theta) \in \mathcal{M} \times \Theta$,

$$R_n(Q,\theta) = \max_{\gamma \in \mathbb{R}^m} R_n(Q,\theta,\gamma). \tag{6.1}$$

From Rockafeller (1970, Theorem 10.8) and Assumption 3.1 (iv), $R_n(Q,\theta)$ is continuous in $(Q,\theta) \in \mathcal{M} \times \Theta$ under the Lévy metric. This continuity also implies that for each $Q \in \mathcal{M}$, $R_n(Q,\theta)$ is continuous in $\theta \in \Theta$. Since Θ is compact (Assumption 3.1 (ii)), $\bar{T}(Q) = \arg\min_{\theta \in \Theta} R_n(Q,\theta)$ exists.

Proof of (ii). Pick any sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$. The proof is based on Newey and Smith (2004, proof of Theorem 3.1). From Lemma 6.6 (i), $\left| E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right]\right| \to 0$. From the triangle inequality,

$$\sup_{\theta \in \Theta} \left| E_{Q_n}[g_n(x,\theta)] - E_{P_0}[g(x,\theta)] \right| \leq \sup_{\theta \in \Theta} \left| E_{Q_n}[g_n(x,\theta)] - E_{P_0}[g_n(x,\theta)] \right| + \sup_{\theta \in \Theta} \left| E_{P_0}[g(x,\theta)\mathbb{I}\{x \notin \mathcal{X}_n\}] \right|. \tag{6.2}$$

The first term of (6.2) satisfies

$$\begin{aligned}
& \sup_{\theta \in \Theta} \left| E_{Q_n}[g_n(x,\theta)] - E_{P_0}[g_n(x,\theta)] \right| \\
\leq\; & \sup_{\theta \in \Theta} \left| \int g_n(x,\theta) \left\{ dQ_n^{1/2} - dP_0^{1/2} \right\}^2 \right| + 2 \sup_{\theta \in \Theta} \left| \int g_n(x,\theta)\, dP_0^{1/2} \left\{ dQ_n^{1/2} - dP_0^{1/2} \right\} \right| \\
\leq\; & m_n \frac{r^2}{n} + 2 \sqrt{E_{P_0}\left[\sup_{\theta \in \Theta} |g(x,\theta)|^2\right]} \frac{r}{\sqrt{n}} = O\left(n^{-1/2}\right),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from $Q_n \in B_H(P_0, r/\sqrt{n})$ and the Cauchy-Schwarz inequality, and the equality follows from Assumption 3.1 (v) and (vii). The second term of (6.2) satisfies

$$\begin{aligned}
& \sup_{\theta \in \Theta} \left| E_{P_0}[g(x,\theta)\mathbb{I}\{x \notin \mathcal{X}_n\}] \right| \\
\leq\; & \left(\int \sup_{\theta \in \Theta} |g(x,\theta)|^\eta \, dP_0 \right)^{1/\eta} \left(\int \mathbb{I}\{x \notin \mathcal{X}_n\}\, dP_0 \right)^{(\eta-1)/\eta} \\
\leq\; & \left(E_{P_0}\left[\sup_{\theta \in \Theta} |g(x,\theta)|^\eta\right] \right)^{1/\eta} \left(m_n^{-\eta} E_{P_0}\left[\sup_{\theta \in \Theta} |g(x,\theta)|^\eta\right] \right)^{(\eta-1)/\eta} = o\left(n^{-1/2}\right),
\end{aligned}$$

where the first inequality follows from the Hölder inequality, and the second inequality follows from the Markov inequality, and the equality follows from Assumption 3.1 (v) and (vii). Combining these results, we obtain the uniform convergence $\sup_{\theta\in\Theta}\left|E_{Q_n}\left[g_n\left(x,\theta\right)\right]-E_{P_0}\left[g\left(x,\theta\right)\right]\right|\to 0$. From the triangle inequality,

$$\left|E_{P_0}\left[g\left(x,\bar{T}_{Q_n}\right)\right]\right|\leq\left|E_{P_0}\left[g\left(x,\bar{T}_{Q_n}\right)\right]-E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)\right]\right|+\left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)\right]\right|\to 0.$$

The conclusion is obtained from Assumption 3.1 (iii).

Lemma 6.2. *[Differentiability of $\bar{T}$] Suppose that Assumption 3.1 holds. Then for each sequence $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$ and $r>0$,*[2]

$$\sqrt{n}\left(\bar{T}_{Q_n}-\theta_0\right)=-\sqrt{n}\Sigma^{-1}\int\Lambda_n dQ_n+o\left(1\right). \tag{6.3}$$

Proof. The proof is based on Rieder (1994, proof of Theorems 6.3.4 (and maybe Theorem 6.4.5)). Pick any $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$ and $r>0$. Observe that

$$\begin{aligned}
&\left\|dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\left(\bar{T}_{Q_n}-\theta_0\right)'\Lambda_n dQ_n^{1/2}\right\|^2\\
=&\left\|dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\Lambda_n dQ_n^{1/2}\right\|^2+\left\|\frac{1}{2}\left(\bar{T}_{Q_n}-\theta_0-\psi_{n,Q_n}\right)'\Lambda_n dQ_n^{1/2}\right\|^2\\
&+\left\{\int\left(dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\Lambda_n dQ_n^{1/2}\right)\Lambda_n' dQ_n^{1/2}\right\}\left(\bar{T}_{Q_n}-\theta_0-\psi_{n,Q_n}\right)\\
=&\left\|dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\Lambda_n dQ_n^{1/2}\right\|^2+\left\|\frac{1}{2}\left(\bar{T}_{Q_n}-\theta_0-\psi_{n,Q_n}\right)'\Lambda_n dQ_n^{1/2}\right\|^2, \qquad (6.4)
\end{aligned}$$

where the second equality follows from

$$\begin{aligned}
&\int\left\{dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\Lambda_n dQ_n^{1/2}\right\}\Lambda_n' dQ_n^{1/2}\\
=&\int\Lambda_n'\left\{dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}\right\}dQ_n^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\int\Lambda_n\Lambda_n' dQ_n=0.
\end{aligned}$$

[2]REPONSE TO COMMENT FOR 316. I changed the statement of this lemma from

$$\sqrt{n}\left(\bar{T}_{Q_n}-\theta_0\right)=-\sqrt{n}\Sigma^{-1}\int\Lambda dQ_n+o\left(1\right)$$

to

$$\sqrt{n}\left(\bar{T}_{Q_n}-\theta_0\right)=-\sqrt{n}\Sigma^{-1}\int\Lambda_n dQ_n+o\left(1\right).$$

Actually, this is sufficient for our purpose in the proofs of Theorems 3.1 and 3.4 (i). So we don't need to impose $\int\left|g\left(x,\theta_0\right)\right|^4 dQ_n$ here.

From the triangle inequality, the left hand side of (6.4) satisfies

$$
\begin{aligned}
& \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0\right)' \Lambda_n dQ_n^{1/2} \right\| \\
\leq\; & \left\| dQ_n^{1/2} - d\bar{P}_{\bar{T}_{Q_n},Q_n}^{1/2} \right\| + \left\| d\bar{P}_{\bar{T}_{Q_n},Q_n}^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0\right)' \Lambda_n dQ_n^{1/2} \right\| \\
\leq\; & \left\| dQ_n^{1/2} - d\bar{P}_{\bar{T}_{Q_n},Q_n}^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(n^{-1/2}\right) \\
=\; & \min_{\theta\in\Theta} \left\| dQ_n^{1/2} - d\bar{P}_{\theta,Q_n}^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(n^{-1/2}\right) \\
\leq\; & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0+\psi_{n,Q_n},Q_n}^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(n^{-1/2}\right) \\
\leq\; & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\| + \left\| -d\bar{P}_{\theta_0+\psi_{n,Q_n},Q_n}^{1/2} + d\bar{P}_{\theta_0,Q_n}^{1/2} - \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\| \\
& + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(n^{-1/2}\right) \\
=\; & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(\left|\psi_{n,Q_n}\right|\right) + o\left(n^{-1/2}\right),
\end{aligned}
$$

where the second inequality follows from Lemma 6.3 (i), the first equality and the third inequality follows from the definition of $\bar{T}_{Q_n}$, the fourth inequality follows from the triangle inequality, and the second equality follows from Lemma 6.3 (ii). Thus, from (6.4),

$$
\begin{aligned}
& \left| \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\|^2 + \left\| \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right)' \Lambda_n dQ_n^{1/2} \right\|^2 \right|^{1/2} \\
\leq\; & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(\left|\psi_{n,Q_n}\right|\right) + o\left(n^{-1/2}\right).
\end{aligned}
$$

This implies

$$
\begin{aligned}
& o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(\left|\psi_{n,Q_n}\right|\right) + o\left(n^{-1/2}\right) \\
\geq\; & \sqrt{\frac{1}{4}\left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right)' \int \Lambda_n \Lambda_n' dQ_n \left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right)} \\
\geq\; & C\left|\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right|, \qquad (6.5)
\end{aligned}
$$

for all n large enough, where the second inequality follows from Lemma 6.5 (i) and Assumption 3.1 (vi).

We now analyze ψ_{n,Q_n}. From the definition of ψ_{n,Q_n},

$$
\begin{aligned}
\psi_{n,Q_n} =\; & -2\left\{\left(\int \Lambda_n \Lambda_n' dQ_n\right)^{-1} - \Sigma^{-1}\right\} \int \Lambda_n \left\{dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2}\right\} dQ_n^{1/2} \\
& -2\Sigma^{-1} \int \Lambda_n \left\{dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2}\right\} dQ_n^{1/2}. \qquad (6.6)
\end{aligned}
$$

The second term of (6.6) satisfies

$$\begin{aligned}
& -2\Sigma^{-1}\int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} \right\} dQ_n^{1/2} \\
= & -2\Sigma^{-1}G'\Omega^{-1}\left(\int g_n(x,\theta_0) g_n(x,\theta_0)' dQ_n \right) \gamma_n(\theta_0,Q_n) \\
& +2\Sigma^{-1}G'\Omega^{-1}\left(\int \frac{\gamma_n(\theta_0,Q_n)' g_n(x,\theta_0)}{1+\gamma_n(\theta_0,Q_n)' g_n(x,\theta_0)} g_n(x,\theta_0) g_n(x,\theta_0)' \right) \gamma_n(\theta_0,Q_n) \\
= & -\Sigma^{-1}G'\Omega^{-1}\left\{ \int g_n(x,\theta_0) dQ_n + \frac{1}{2}\int \varrho_n(x,\theta_0,Q_n) g_n(x,\theta_0) dQ_n \right\} + o\left(n^{-1/2}\right) \\
= & -\Sigma^{-1}\int \Lambda_n dQ_n + o\left(n^{-1/2}\right),
\end{aligned}$$

where the first equality follows from (5.2), the second equality follows from (5.3) and Lemma 6.5, and the third equality follows from Lemma 6.5. From this and Lemma 6.5 (i), the first term of (6.6) is $o\left(n^{-1/2}\right)$. Therefore,

$$\sqrt{n}\psi_{n,Q_n} = -\sqrt{n}\Sigma^{-1}\int \Lambda_n dQ_n + o(1),$$

which also implies $|\psi_{n,Q_n}| = O\left(n^{-1/2}\right)$ (by Lemma 6.5 (i)). From (6.5),

$$\sqrt{n}\left(\bar{T}_{Q_n} - \theta_0\right) = \sqrt{n}\psi_{n,Q_n} + o\left(\sqrt{n}\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o(1).$$

By solving for $\sqrt{n}\left(\bar{T}_{Q_n} - \theta_0\right)$, the conclusion is obtained.

Lemma 6.3. *Suppose that Assumption 3.1 holds. Then for each sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$,*

(i): $\left\| d\bar{P}_{\bar{T}_{Q_n},Q_n}^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0\right)' \Lambda_n dQ_n^{1/2} \right\| = o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(n^{-1/2}\right)$,

(ii): $\left\| d\bar{P}_{\theta_0+\psi_{n,Q_n},Q_n}^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\| = o\left(|\psi_{n,Q_n}|\right) + o\left(n^{-1/2}\right)$,

Proof of (i). Denote $t_n = \bar{T}_{Q_n} - \theta_0$. Pick any $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$. From (5.2),

$$\begin{aligned}
& \left\| d\bar{P}_{\bar{T}_{Q_n},Q_n}^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}t_n'\Lambda_n dQ_n^{1/2} \right\| \\
\leq & \left\| \left\{ \gamma_n(\theta_0,Q_n)' g_n(x,\theta_0) - \gamma_n\left(\bar{T}_{Q_n},Q_n\right)' g_n\left(x,\bar{T}_{Q_n}\right) \right\} dQ_n^{1/2} + \frac{1}{2}t_n'\Lambda_n dQ_n^{1/2} \right\| \\
& + \left\| \begin{array}{c} \left\{ \gamma_n(\theta_0,Q_n)' g_n(x,\theta_0) - \gamma_n\left(\bar{T}_{Q_n},Q_n\right)' g_n\left(x,\bar{T}_{Q_n}\right) \right\} \\ \times \left\{ \frac{1}{\left(1+\gamma_n\left(\bar{T}_{Q_n},Q_n\right)' g_n\left(x,\bar{T}_{Q_n}\right)\right)\left(1+\gamma_n(\theta_0,Q_n)' g_n(x,\theta_0)\right)} - 1 \right\} dQ_n^{1/2} \end{array} \right\| = T_1 + T_2.
\end{aligned}$$

For T_2, Lemmas 6.5 and 6.6 imply

$$T_2 \leq o(1)\left\| \gamma_n\left(\bar{T}_{Q_n},Q_n\right)' g_n\left(x,\bar{T}_{Q_n}\right) dQ_n^{1/2} + \gamma_n(\theta_0,Q_n)' g_n(x,\theta_0) dQ_n^{1/2} \right\| = o\left(n^{-1/2}\right).$$

Thus, we focus on T_1. From (5.3),

$$\begin{aligned} T_1 &\leq \left\| \left\{ \begin{array}{l} -\frac{1}{2}\left(\int g_n\left(x,\bar{T}_{Q_n}\right)dQ_n\right)'\left(\int g_n\left(x,\bar{T}_{Q_n}\right)g_n\left(x,\bar{T}_{Q_n}\right)'dQ_n\right)^{-1}g_n\left(x,\bar{T}_{Q_n}\right) \\ +\frac{1}{2}\left(\int g_n\left(x,\theta_0\right)dQ_n\right)'\left(\int g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'dQ_n\right)^{-1}g_n\left(x,\theta_0\right)+\frac{1}{2}t_n'\Lambda_n \end{array} \right\} dQ_n^{1/2} \right\| \\ &\quad + \left\| \left\{ \left(\int \varrho_n\left(x,\theta_0,Q_n\right)g_n\left(x,\theta_0\right)dQ_n\right)'\left(\int g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'dQ_n\right)^{-1}g_n\left(x,\theta_0\right)\right\} dQ_n^{1/2} \right\| \\ &\quad + \left\| \left\{ \left(\int \varrho_n\left(x,\bar{T}_{Q_n},Q_n\right)g_n\left(x,\bar{T}_{Q_n}\right)dQ_n\right)'\left(\int g_n\left(x,\bar{T}_{Q_n}\right)g_n\left(x,\bar{T}_{Q_n}\right)'dQ_n\right)^{-1}g_n\left(x,\theta_0\right)\right\} dQ_n^{1/2} \right\| \\ &= T_{11}+T_{12}+T_{13}. \end{aligned}$$

Lemmas 6.5 and 6.6 imply that $T_{12} = o\left(n^{-1/2}\right)$ and $T_{13} = o\left(n^{-1/2}\right)$. Thus, we focus on T_{11}. Taylor expansions of $g_n\left(x,\bar{T}_{Q_n}\right)$ around $\bar{T}_{Q_n} = \theta_0$ yield

$$\begin{aligned} T_{11} &\leq \left\| \left\{ -\frac{1}{2}\left(\int g_n\left(x,\bar{T}_{Q_n}\right)dQ_n\right)' \left\{ \begin{array}{c} \left(\int g_n\left(x,\bar{T}_{Q_n}\right)g_n\left(x,\bar{T}_{Q_n}\right)'dQ_n\right)^{-1} \\ -\left(\int g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'dQ_n\right)^{-1} \end{array} \right\} g_n\left(x,\bar{T}_{Q_n}\right) \right\} dQ_n^{1/2} \right\| \\ &\quad + \left\| -\frac{1}{2}\left(\int g_n(x,\bar{T}_{Q_n})dQ_n\right)'\left(\int g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'dQ_n\right)^{-1}\left\{g_n\left(x,\bar{T}_{Q_n}\right)-g_n\left(x,\theta_0\right)\right\}dQ_n^{1/2} \right\| \\ &\quad + \left\| -\frac{1}{2}t_n'\left(\int \frac{\partial g_n\left(x,\dot{\theta}\right)}{\partial\theta'}dQ_n - G\right)'\left(\int g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'dQ_n\right)^{-1}g_n\left(x,\theta_0\right)dQ_n^{1/2} \right\| \\ &\quad + \left\| \frac{1}{2}t_n'G'\left\{\Omega^{-1}-\left(\int g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'dQ_n\right)^{-1}\right\}g_n\left(x,\theta_0\right)dQ_n^{1/2} \right\| \\ &= o\left(n^{-1/2}\right)+o\left(t_n\right), \end{aligned}$$

where $\dot{\theta}$ is a point on the line joining θ_0 and $\bar{T}_{Q_n}$, and the inequality follows from the triangle inequality and Lemmas 6.5 (i) and 6.6 (i).

Proof of (ii). The proof is similar to that of Part (i).

Lemma 6.4. *Suppose that Assumption 3.1 hold. Then for each* $t \in \mathbb{R}^p$*,*

- **(i):** $\left|E_{P_0}\left[g_n\left(x,\theta_0\right)\right]\right| = o\left(n^{-1/2}\right)$, $\left|E_{P_0}\left[g_n\left(x,\theta_n\right)\right]\right| = O\left(n^{-1/2}\right)$, $\left|E_{P_0}\left[g_n\left(x,\theta_n\right)g_n\left(x,\theta_n\right)'\right]-\Omega\right| = o(1)$*, and* $\left|E_{P_0}\left[\partial g_n\left(x,\theta_n\right)/\partial\theta'\right]-G\right| = o(1)$*,*
- **(ii):** $\gamma_n\left(\theta_n,P_0\right) = \arg\max_{\gamma\in\mathbb{R}^m} -\int \frac{1}{(1+\gamma' g_n(x,\theta_n))}dP_0$ *exists for all* n *large enough,* $\left|\gamma_n\left(\theta_n,P_0\right)\right| = O\left(n^{-1/2}\right)$*, and* $\sup_{x\in\mathcal{X}}\left|\gamma_n\left(\theta_n,P_0\right)'g_n\left(x,\theta_n\right)\right| \to 0$.

Proof of (i). Proof of the first statement. Pick any $t \in \mathbb{R}^p$. Observe that

$$\begin{aligned}
|E_{P_0}[g_n(x,\theta_0)]| &= \left| \int g(x,\theta_0)\,\mathbb{I}\{x \notin \mathcal{X}_n\}\,dP_0 \right| \\
&\le \left(\int |g(x,\theta_0)|^\eta \, dP_0 \right)^{1/\eta} \left(\int \mathbb{I}\{x \notin \mathcal{X}_n\}\,dP_0 \right)^{(\eta-1)/\eta} \\
&\le \left(E_{P_0}\left[|g(x,\theta_0)|^\eta\right]\right)^{1/\eta} \left(m_n^{-\eta} E_{P_0}\left[\sup_{\theta\in\Theta} |g(x,\theta_0)|^\eta\right]\right)^{(\eta-1)/\eta} = o\left(n^{-1/2}\right), \qquad (6.7)
\end{aligned}$$

where the first inequality follows from the Hölder inequality, the second inequality follows from the Markov inequality, and the second equality follows from Assumption 3.1 (v) and (vii).

Proof of the second statement. Pick any $t \in \mathbb{R}^p$. From the triangle inequality,

$$|E_{P_0}[g_n(x,\theta_n)]| \le \left| \int g(x,\theta_n)\,\mathbb{I}\{x \notin \mathcal{X}_n\}\,dP_0 \right| + |E_{P_0}[g(x,\theta_n)]|. \qquad (6.8)$$

By the same argument as (6.7), the first term of (6.8) is $o\left(n^{-1/2}\right)$ (note that $E_{P_0}\left[|g(x,\theta_n)|^\eta\right] < \infty$ for all n large enough from Assumption 3.1 (v)). The second term of (6.8) satisfies

$$|E_{P_0}[g(x,\theta_n)]| \le E_{P_0}\left[\sup_{\theta\in\mathcal{N}} \left|\frac{\partial g(x,\theta)}{\partial\theta'}\right|\right] \left|\frac{t}{\sqrt{n}}\right| = O\left(n^{-1/2}\right),$$

for all n large enough, where the inequality follows from a Taylor expansion around $t = 0$ and Assumption 3.1 (iii), and the equality follows from Assumption 3.1 (v). Therefore, the conclusion is obtained.

Proof of the third statement. Pick any $t \in \mathbb{R}^p$. Observe that

$$\begin{aligned}
&\left|E_{P_0}\left[g_n(x,\theta_n)\,g_n(x,\theta_n)'\right] - E_{P_0}\left[g(x,\theta_0)\,g(x,\theta_0)'\right]\right| \\
\le\ &\left|E_{P_0}\left[g_n(x,\theta_n)\,g_n(x,\theta_n)'\right] - E_{P_0}\left[g(x,\theta_n)\,g(x,\theta_n)'\right]\right| + \left|E_{P_0}\left[g(x,\theta_n)\,g(x,\theta_n)'\right] - E_{P_0}\left[g(x,\theta_0)\,g(x,\theta_0)'\right]\right| \\
\le\ &\left(E_{P_0}\left[\sup_{\theta\in\mathcal{N}} |g(x,\theta)|^\eta\right]\right)^{1/(1+\delta)} \left(m_n^{-\eta} E_{P_0}\left[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta\right]\right)^{\delta/(1+\delta)} + o(1) = o(1),
\end{aligned}$$

for all n large enough and sufficiently small $\delta > 0$, where the first inequality follows from the triangle inequality, the second inequality follows from the Hölder and Markov inequalities and the continuity of $g(x,\theta)$ at θ_0, and the equality follows from Assumption 3.1 (v) and (vii).

Proof of the fourth statement. Pick any $t \in \mathbb{R}^p$. Observe that

$$\begin{aligned}
&\left|E_{P_0}\left[\partial g_n(x,\theta_n)/\partial\theta'\right] - E_{P_0}\left[\partial g(x,\theta_0)/\partial\theta'\right]\right| \\
\le\ &\left|E_{P_0}\left[\partial g_n(x,\theta_n)/\partial\theta'\right] - E_{P_0}\left[\partial g(x,\theta_n)/\partial\theta'\right]\right| + \left|E_{P_0}\left[\partial g(x,\theta_n)/\partial\theta'\right] - E_{P_0}\left[\partial g(x,\theta_0)/\partial\theta'\right]\right| \\
\le\ &\sqrt{E_{P_0}\left[\sup_{\theta\in\mathcal{N}} \left|\frac{\partial g(x,\theta)}{\partial\theta'}\right|^2\right]} \sqrt{m_n^{-\eta} E_{P_0}\left[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta\right]} + o(1) = o(1),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz and Markov inequalities and the the continuity of $\partial g(x,\theta)/\partial\theta'$ at θ_0, and the equality follows from Assumption 3.1 (v) and (vii).

Proof of (ii). The proof is based on Newey and Smith (2004, proofs of Lemmas A.1-3). Pick any $t \in \mathbb{R}^p$. Let $\Gamma_n = \{\gamma \in \mathbb{R}^m : |\gamma| \le a_n\}$ with $a_n m_n \to 0$ and $a_n n^{1/2} \to \infty$. Observe that

$$\sup_{\gamma\in\Gamma_n, x\in\mathcal{X}, \theta\in\Theta} \left|\gamma' g_n(x,\theta)\right| \le a_n m_n \to 0. \tag{6.9}$$

Since $R_n(P_0,\theta_n,\gamma)$ is twice continuously differentiable with respect to γ and Γ_n is compact, $\tilde{\gamma} = \arg\max_{\gamma\in\Gamma_n} R_n(P_0,\theta_n,\gamma)$ exists for each $n \in \mathbb{N}$. A Taylor expansion around $\tilde{\gamma} = 0$ yields

$$\begin{aligned} -1 &= R_n(P_0,\theta_n,0) \le R_n(P_0,\theta_n,\tilde{\gamma}) = -1 + \tilde{\gamma}' E_{P_0}[g_n(x,\theta_n)] - \tilde{\gamma}' E_{P_0}\left[\frac{g_n(x,\theta_n) g_n(x,\theta_n)'}{(1+\dot{\gamma}' g_n(x,\theta_n))^3}\right]\tilde{\gamma} \\ &\le -1 + \tilde{\gamma}' E_{P_0}[g_n(x,\theta_n)] - C\tilde{\gamma}' E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right]\tilde{\gamma} \\ &\le -1 + |\tilde{\gamma}|\,|E_{P_0}[g_n(x,\theta_n)]| - C|\tilde{\gamma}|^2, \end{aligned} \tag{6.10}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, the second inequality follows from (6.9), and the last inequality follows from Lemma 6.4 (i) and Assumption 3.1 (vi). Thus, Lemma 6.4 (i) implies

$$C|\tilde{\gamma}| \le |E_{P_0}[g_n(x,\theta_n)]| = O\left(n^{-1/2}\right). \tag{6.11}$$

From $a_n n^{1/2} \to \infty$, $\tilde{\gamma}$ is an interior point of Γ_n and satisfies the first-order condition $\partial R_n(Q_n,\theta_0,\tilde{\gamma})/\partial\gamma = 0$ for all n large enough. Since $R_n(Q_n,\theta_0,\gamma)$ is concave in γ for all n large enough, $\tilde{\gamma} = \arg\max_{\gamma\in\mathbb{R}^m} R_n(P_0,\theta_n,\gamma)$ for all n large enough. Thus, the first statement is obtained. Also, from (6.11), the second statement is obtained. Using Assumption 3.1 (vii), the third statement follows from

$$\sup_{x\in\mathcal{X}} \left|\gamma_n(\theta_n,P_0)' g_n(x,\theta_n)\right| \le O\left(n^{-1/2} m_n\right) = o(1). \tag{6.12}$$

Lemma 6.5. *Suppose that Assumption 3.1 holds. Then for each sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$,*

- **(i):** $|E_{Q_n}[g_n(x,\theta_0)]| = O\left(n^{-1/2}\right)$, *and* $\left|E_{Q_n}\left[g_n(x,\theta_0) g_n(x,\theta_0)'\right] - \Omega\right| = o(1)$,
- **(ii):** $\gamma_n(\theta_0,Q_n) = \arg\max_{\gamma\in\mathbb{R}^m} -\int \frac{1}{(1+\gamma' g_n(x,\theta_0))} dQ_n$ *exists for all n large enough, and* $|\gamma_n(\theta_0,Q_n)| = O\left(n^{-1/2}\right)$, *and* $\sup_{x\in\mathcal{X}} \left|\gamma_n(\theta_0,Q_n)' g_n(x,\theta_0)\right| \to 0$.

Proof of (i). Proof of the first statement. Pick any sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$. Observe that

$$\begin{aligned}
& |E_{Q_n}[g_n(x,\theta_0)]| \\
\leq\ & \left| \int g_n(x,\theta_0)\{dQ_n - dP_0\} \right| + |E_{P_0}[g_n(x,\theta_0)]| \\
\leq\ & \left| \int g_n(x,\theta_0)\left\{dQ_n^{1/2} - dP_0^{1/2}\right\}^2 \right| + 2\left| \int g_n(x,\theta_0)\, dP_0^{1/2}\left\{dQ_n^{1/2} - dP_0^{1/2}\right\} \right| + o\left(n^{-1/2}\right) \\
\leq\ & m_n \frac{r^2}{n} + 2E_{P_0}\left[|g(x,\theta_0)|^2\right]\frac{r}{\sqrt{n}} + o\left(n^{-1/2}\right) = O\left(n^{-1/2}\right),
\end{aligned}$$

where the first and second inequalities follow from the triangle inequality and Lemma 6.4 (i), the third inequality follows from the Cauchy-Schwarz inequality and $Q_n \in B_H(P_0, r/\sqrt{n})$, and the equality follows from Assumption 3.1 (v) and (vii).

Proof of the second statement. Pick any sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$. From the triangle inequality,

$$\begin{aligned}
& \left|E_{Q_n}\left[g_n(x,\theta_0)\, g_n(x,\theta_0)'\right] - E_{P_0}\left[g(x,\theta_0)\, g(x,\theta_0)'\right]\right| \\
\leq\ & \left|E_{Q_n}\left[g_n(x,\theta_0)\, g_n(x,\theta_0)'\right] - E_{P_0}\left[g_n(x,\theta_0)\, g_n(x,\theta_0)'\right]\right| + \left|E_{P_0}\left[g(x,\theta_0)\,(x,\theta_0)'\, \mathbb{I}\{x \notin \mathcal{X}_n\}\right]\right|. \qquad (6.13)
\end{aligned}$$

The first term of (6.13) satisfies

$$\begin{aligned}
& \left|E_{Q_n}\left[g_n(x,\theta_0)\, g_n(x,\theta_0)'\right] - E_{P_0}\left[g_n(x,\theta_0)\, g_n(x,\theta_0)'\right]\right| \\
\leq\ & \left| \int g_n(x,\theta_0)\, g_n(x,\theta_0)'\left\{dQ_n^{1/2} - dP_0^{1/2}\right\}^2 \right| + 2\left| \int g_n(x,\theta_0)\, g_n(x,\theta_0)'\, dP_0^{1/2}\left\{dQ_n^{1/2} - dP_0^{1/2}\right\} \right| \\
\leq\ & m_n^2 \frac{r^2}{n} + 2E_{P_0}\left[|g(x,\theta_0)|^4\right]\frac{r}{\sqrt{n}} = o(1),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality and $Q_n \in B_H(P_0, r/\sqrt{n})$, and the equality follows from Assumption 3.1 (v) and (vii). The second term of (6.13) satisfies

$$\begin{aligned}
& \left|E_{P_0}\left[g(x,\theta_0)\, g(x,\theta_0)'\, \mathbb{I}\{x \notin \mathcal{X}_n\}\right]\right| \\
\leq\ & \left(\int \left|g(x,\theta_0)\, g(x,\theta_0)'\right|^{1+\delta} dP_0 \right)^{1/(1+\delta)} \left(\int \mathbb{I}\{x \notin \mathcal{X}_n\}\, dP_0 \right)^{\delta/(1+\delta)} \\
\leq\ & \left(E_{P_0}\left[|g(x,\theta_0)|^{2+\delta}\right] \right)^{1/(1+\delta)} \left(m_n^{-\eta} E_{P_0}\left[|g(x,\theta_0)|^{\eta}\right] \right)^{\delta/(1+\delta)} = o(1),
\end{aligned}$$

for sufficiently small $\delta > 0$, where the first inequality follows from the Hölder inequality, the second inequality follows from the Markov inequality, and the equality follows from Assumption 3.1 (vii).

Proof of (ii). The proof is based on Newey and Smith (2004, proofs of Lemmas A.1-3). Pick any sequence $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$ and $r>0$. Since $R_n\left(Q, \theta, \gamma\right)$ is twice continuously differentiable with respect to γ and Γ_n is compact, $\bar{\gamma} = \arg\max_{\gamma\in\Gamma_n} R_n\left(Q_n, \theta_0, \gamma\right)$ exists for each $n \in \mathbb{N}$. A Taylor expansion around $\bar{\gamma}=0$ yields

$$\begin{aligned} -1 &= R_n\left(Q_n, \theta_0, 0\right) \leq R_n\left(Q_n, \theta_0, \bar{\gamma}\right) = -1 + \bar{\gamma}' E_{Q_n}\left[g_n\left(x, \theta_0\right)\right] - \bar{\gamma}' E_{Q_n}\left[\frac{g_n\left(x, \theta_0\right) g_n\left(x, \theta_0\right)'}{\left(1+\dot{\gamma}' g_n\left(x, \theta_0\right)\right)^3}\right]\bar{\gamma} \\ &\leq -1 + \bar{\gamma}' E_{Q_n}\left[g_n\left(x, \theta_0\right)\right] - C\bar{\gamma}' E_{Q_n}\left[g_n\left(x, \theta_0\right) g_n\left(x, \theta_0\right)'\right]\bar{\gamma} \\ &\leq -1 + \left|\bar{\gamma}\right|\left|E_{Q_n}\left[g_n\left(x, \theta_0\right)\right]\right| - C\left|\bar{\gamma}\right|^2, \qquad (6.14) \end{aligned}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\bar{\gamma}$, the second inequality follows from (6.9), and the last inequality follows from Lemma 6.5 (i) and $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$. Thus, Lemma 6.5 (i) implies

$$C\left|\bar{\gamma}\right| \leq \left|E_{Q_n}\left[g_n\left(x, \theta_0\right)\right]\right| = O\left(n^{-1/2}\right). \qquad (6.15)$$

From $a_n n^{1/2} \to \infty$, $\bar{\gamma}$ is an interior point of Γ_n and satisfies the first-order condition $\partial R_n\left(Q_n, \theta_0, \bar{\gamma}\right)/\partial\gamma = 0$ for all n large enough. Since $R_n\left(Q_n, \theta_0, \gamma\right)$ is concave in γ for all n large enough, $\bar{\gamma} = \arg\max_{\gamma\in\mathbb{R}^m} R_n\left(Q_n, \theta_0, \gamma\right)$ for all n large enough. Thus, the first statement is obtained. Also, from (6.15), the second statement is obtained. The third statement follows from

$$\sup_{x\in\mathcal{X}}\left|\gamma_n\left(\theta_0, Q_n\right)' g_n\left(x, \theta_0\right)\right| \leq O\left(n^{-1/2} m_n\right) = o\left(1\right). \qquad (6.16)$$

Lemma 6.6. *Suppose that Assumption 3.1 holds. Then for each sequence* $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$ *and* $r>0$,

- **(i):** $\left|E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right]\right| = O\left(n^{-1/2}\right)$, $\left|E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right) g_n\left(x, \bar{T}_{Q_n}\right)'\right] - \Omega\right| = o\left(1\right)$, *and* $\left|E_{Q_n}\left[\partial g_n\left(x, \bar{T}_{Q_n}\right)/\partial\theta'\right] - G\right| = o\left(1\right)$,
- **(ii):** $\gamma_n\left(\bar{T}_{Q_n}, Q_n\right) = \arg\max_{\gamma\in\mathbb{R}^m} -\int \frac{1}{\left(1+\gamma' g_n\left(x, \bar{T}_{Q_n}\right)\right)} dQ_n$ *exists for all* n *large enough,* $\left|\gamma_n\left(\bar{T}_{Q_n}, Q_n\right)\right| = O\left(n^{-1/2}\right)$, *and* $\sup_{x\in\mathcal{X}}\left|\gamma_n\left(\bar{T}_{Q_n}, Q_n\right)' g_n\left(x, \bar{T}_{Q_n}\right)\right| \to 0$.

Proof of (i). Proof of the first statement. Pick any sequence $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$ and $r>0$. Define $\tilde{\gamma} = n^{-1/2} E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right] / \left|E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right]\right|$. Since $\left|\tilde{\gamma}\right| = n^{-1/2}$,

$$\sup_{x\in\mathcal{X}, \theta\in\Theta}\left|\tilde{\gamma}' g_n\left(x, \theta\right)\right| \leq n^{-1/2} m_n \to 0. \qquad (6.17)$$

Observe that

$$
\begin{aligned}
& \left| E_{Q_n} \left[g_n \left(x, \bar{T}_{Q_n} \right) g_n \left(x, \bar{T}_{Q_n} \right)' \right] \right| \\
\leq \; & \int \sup_{\theta \in \Theta} \left| g_n (x, \theta) \right|^2 \left\{ dQ_n^{1/2} - dP_0^{1/2} \right\}^2 + 2 \int \sup_{\theta \in \Theta} \left| g_n (x, \theta) \right|^2 dP_0^{1/2} \left\{ dQ_n^{1/2} - dP_0^{1/2} \right\} + E_{P_0} \left[\sup_{\theta \in \Theta} \left| g_n (x, \theta) \right|^2 \right] \\
\leq \; & m_n^2 \frac{r^2}{n} + 2 m_n \sqrt{ E_{P_0} \left[\sup_{\theta \in \Theta} \left| g (x, \theta) \right|^2 \right] } \frac{r}{\sqrt{n}} + E_{P_0} \left[\sup_{\theta \in \Theta} \left| g (x, \theta) \right|^2 \right] \leq C E_{P_0} \left[\sup_{\theta \in \Theta} \left| g (x, \theta) \right|^2 \right],
\end{aligned} \tag{6.18}
$$

for all n large enough, where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality and $Q_n \in B_H \left(P_0, r/\sqrt{n} \right)$, and the last inequality follows from Assumption 3.1 (v) and (vii). Thus, a Taylor expansion around $\tilde{\gamma} = 0$ yields

$$
\begin{aligned}
R_n \left(Q_n, \bar{T}_{Q_n}, \tilde{\gamma} \right) &= -1 + \tilde{\gamma}' E_{Q_n} \left[g_n \left(x, \bar{T}_{Q_n} \right) \right] - \tilde{\gamma}' E_{Q_n} \left[\frac{ g_n \left(x, \bar{T}_{Q_n} \right) g_n \left(x, \bar{T}_{Q_n} \right)' }{ \left(1 + \dot{\gamma}' g_n \left(x, \bar{T}_{Q_n} \right) \right)^3 } \right] \tilde{\gamma} \\
&\geq -1 + n^{-1/2} \left| E_{Q_n} \left[g_n \left(x, \bar{T}_{Q_n} \right) \right] \right| - C \tilde{\gamma}' E_{Q_n} \left[g_n \left(x, \bar{T}_{Q_n} \right) g_n \left(x, \bar{T}_{Q_n} \right)' \right] \tilde{\gamma} \\
&\geq -1 + n^{-1/2} \left| E_{Q_n} \left[g_n \left(x, \bar{T}_{Q_n} \right) \right] \right| - C n^{-1},
\end{aligned} \tag{6.19}
$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, the first inequality follows from (6.17), and the second inequality follows from $\tilde{\gamma}' \tilde{\gamma} = n^{-1}$ and (6.18). From the duality of partially finite programming (Borwein and Lewis (1993)), $\gamma_n \left(\bar{T}_{Q_n}, Q_n \right)$ and $\bar{T}_{Q_n}$ are written as

$$
\begin{aligned}
\gamma_n \left(\bar{T}_{Q_n}, Q_n \right) &= \arg \max_{\gamma \in \mathbb{R}^m} R_n \left(Q_n, \bar{T}_{Q_n}, \gamma \right), \\
\bar{T}_{Q_n} &= \arg \min_{\theta \in \Theta} R_n \left(Q_n, \theta, \gamma_n (\theta, Q_n) \right).
\end{aligned}
$$

Thus, from (6.19),

$$
\begin{aligned}
& -1 + n^{-1/2} \left| E_{Q_n} \left[g_n \left(x, \bar{T}_{Q_n} \right) \right] \right| - C n^{-1} \\
\leq \; & R_n \left(Q_n, \bar{T}_{Q_n}, \tilde{\gamma} \right) \leq R_n \left(Q_n, \bar{T}_{Q_n}, \gamma_n \left(\bar{T}_{Q_n}, Q_n \right) \right) \leq R_n \left(Q_n, \theta_0, \gamma_n (\theta_0, Q_n) \right).
\end{aligned} \tag{6.20}
$$

From $\left| \gamma_n (\theta_0, Q_n) \right| = O \left(n^{-1/2} \right)$ and $\left| E_{Q_n} \left[g_n (x, \theta_0) \right] \right| = O \left(n^{-1/2} \right)$ (by Lemmas 6.5 (ii) and 6.5 (i)), (6.14) yields

$$
R_n \left(Q_n, \theta_0, \gamma_n (\theta_0, Q_n) \right) \leq -1 + \left| \gamma_n (\theta_0, Q_n) \right| \left| E_{Q_n} \left[g_n (x, \theta_0) \right] \right| - C \left| \gamma_n (\theta_0, Q_n) \right|^2 = -1 + O \left(n^{-1} \right). \tag{6.21}
$$

Combining (6.20) and (6.21), we have $\left| E_{Q_n} \left[g_n \left(x, \bar{T}_{Q_n} \right) \right] \right| = O \left(n^{-1/2} \right)$.

Proof of the second statement. Pick any sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$. From the triangle inequality,

$$\begin{aligned}
& \left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right) g_n\left(x,\bar{T}_{Q_n}\right)'\right] - E_{P_0}\left[g\left(x,\theta_0\right) g\left(x,\theta_0\right)'\right]\right| \\
\leq & \left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right) g_n\left(x,\bar{T}_{Q_n}\right)'\right] - E_{P_0}\left[g_n\left(x,\bar{T}_{Q_n}\right) g_n\left(x,\bar{T}_{Q_n}\right)'\right]\right| \\
& + \left|E_{P_0}\left[g\left(x,\bar{T}_{Q_n}\right) g\left(x,\bar{T}_{Q_n}\right)' \mathbb{I}\left\{x \notin \mathcal{X}_n\right\}\right]\right| \\
& + \left|E_{P_0}\left[g\left(x,\bar{T}_{Q_n}\right) g\left(x,\bar{T}_{Q_n}\right)'\right] - E_{P_0}\left[g\left(x,\theta_0\right) g\left(x,\theta_0\right)'\right]\right|. \qquad (6.22)
\end{aligned}$$

The first term of (6.22) satisfies

$$\begin{aligned}
& \left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right) g_n\left(x,\bar{T}_{Q_n}\right)'\right] - E_{P_0}\left[g_n\left(x,\bar{T}_{Q_n}\right) g_n\left(x,\bar{T}_{Q_n}\right)'\right]\right| \\
\leq & \left|\int g_n\left(x,\bar{T}_{Q_n}\right) g_n\left(x,\bar{T}_{Q_n}\right)' \left\{dQ_n^{1/2} - dP_0^{1/2}\right\}^2\right| + 2\left|\int g_n\left(x,\bar{T}_{Q_n}\right) g_n\left(x,\bar{T}_{Q_n}\right) dP_0^{1/2} \left\{dQ_n^{1/2} - dP_0^{1/2}\right\}\right| \\
\leq & m_n^2 \frac{r^2}{n} + 2E_{P_0}\left[\sup_{\theta\in\mathcal{N}} \left|g\left(x,\theta\right)\right|^4\right] \frac{r}{\sqrt{n}} = o\left(1\right),
\end{aligned}$$

for all n large enough, where the first inequality follows from the triangle inequality, the second inequality follows from the consistency of $\bar{T}_{Q_n}$ (Lemma 6.1 (ii)) and the Cauchy-Schwarz inequality, and $Q_n \in B_H(P_0, r/\sqrt{n})$, and the equality follows from Assumption 3.1 (v) and (vii). The second term of (6.22) satisfies

$$\begin{aligned}
& \left|E_{P_0}\left[g\left(x,\bar{T}_{Q_n}\right) g\left(x,\bar{T}_{Q_n}\right)' \mathbb{I}\left\{x \notin \mathcal{X}_n\right\}\right]\right| \\
\leq & \left(\int \sup_{\theta\in\mathcal{N}} \left|g\left(x,\theta\right) g\left(x,\theta\right)'\right|^{1+\delta} dP_0\right)^{1/(1+\delta)} \left(\int \mathbb{I}\left\{x \notin \mathcal{X}_n\right\} dP_0\right)^{\delta/(1+\delta)} \\
\leq & \left(E_{P_0}\left[\sup_{\theta\in\mathcal{N}} \left|g\left(x,\theta\right)\right|^{2+\delta}\right]\right)^{1/(1+\delta)} \left(m_n^{-\eta} E_{P_0}\left[\sup_{\theta\in\Theta} \left|g\left(x,\theta\right)\right|^{\eta}\right]\right)^{\delta/(1+\delta)} = o\left(1\right),
\end{aligned}$$

for sufficiently small $\delta > 0$, where the first inequality follows from the Hölder inequality, the second inequality follows from the Markov inequality, and the equality follows from Assumption 3.1 (v) and (vii). By the continuity of $g(x,\theta)$ at θ_0 and the consistency of $\bar{T}_{Q_n}$, the third term of (6.22) is $o(1)$.

Proof of the third statement. Pick any sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$. From the triangle inequality,

$$\begin{aligned} & \left|E_{Q_n}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right] - E_{P_0}\left[\partial g\left(x,\theta_0\right)/\partial\theta'\right]\right| \\ \leq\ & \left|E_{Q_n}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right] - E_{P_0}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]\right| \\ & + \left|E_{P_0}\left[\mathbb{I}\left\{x \notin \mathcal{X}_n\right\}\partial g\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]\right| \\ & + \left|E_{P_0}\left[\partial g\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right] - E_{P_0}\left[\partial g\left(x,\theta_0\right)/\partial\theta'\right]\right|. \end{aligned} \tag{6.23}$$

The first term of (6.23) satisfies

$$\begin{aligned} & \left|E_{Q_n}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right] - E_{P_0}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]\right| \\ \leq\ & \left|\int \partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\left\{dQ_n^{1/2} - dP_0^{1/2}\right\}^2\right| + 2\left|\int \partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta' dP_0^{1/2}\left\{dQ_n^{1/2} - dP_0^{1/2}\right\}\right| \\ \leq\ & \sup_{\theta\in\mathcal{N}}\left|\partial g_n\left(x,\theta\right)/\partial\theta'\right|\frac{r^2}{n} + 2\sup_{\theta\in\mathcal{N}}\left|\partial g_n\left(x,\theta\right)/\partial\theta'\right|\frac{r}{\sqrt{n}} = o\left(1\right), \end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality, and the equality follows from Assumption 3.1 (v) and (vii). The second term of (6.23) satisfies

$$\begin{aligned} & \left|E_{P_0}\left[\mathbb{I}\left\{x \notin \mathcal{X}_n\right\}\partial g\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]\right| \\ \leq\ & \sqrt{E_{P_0}\left[\mathbb{I}\left\{x \notin \mathcal{X}_n\right\}\right]}\sqrt{E_{P_0}\left[\sup_{\theta\in\mathcal{N}}\left|\partial g\left(x,\theta\right)/\partial\theta'\right|^2\right]} \\ \leq\ & C\left(m_n^{-4}E_{P_0}\left[\sup_{\theta\in\mathcal{N}}\left|g\left(x,\theta\right)\right|^4\right]\right)^{1/2} = o\left(1\right), \end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, the second inequality follows from the Markov inequality, and the equality follows from Assumption 3.1 (v) and (vii). By the continuity of $\partial g(x,\theta)/\partial\theta'$ at θ_0 and the consistency of $\bar{T}_{Q_n}$, the third term of (6.23) is $o(1)$. Therefore, the conclusion is obtained.

Proof of (ii). THIS IS EXACTLY THE SAME PROOF AS FOR LEMMA 6.4 (ii) EXCEPT USING 6.6 (i) INSTEAD OF 6.4 (i). The proof is based on Newey and Smith (2004, proofs of Lemmas A.1-3). Pick any $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$. Since $R_n(Q,\theta,\gamma)$ is twice continuously differentiable with respect to γ and Γ_n is compact, $\bar{\gamma} = \arg\max_{\gamma\in\Gamma_n} R_n\left(Q_n, \bar{T}_{Q_n}, \gamma\right)$ exists for each $n \in \mathbb{N}$. A Taylor expansion around $\bar{\gamma} = 0$ yields

$$
\begin{aligned}
-1 &= R_n\left(Q_n, \bar{T}_{Q_n}, 0\right) \leq R_n\left(Q_n, \bar{T}_{Q_n}, \bar{\gamma}\right) = -1 + \bar{\gamma}' E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right] - \bar{\gamma}' E_{Q_n}\left[\frac{g_n\left(x, \bar{T}_{Q_n}\right) g_n\left(x, \bar{T}_{Q_n}\right)'}{\left(1 + \dot{\gamma}' g_n\left(x, \bar{T}_{Q_n}\right)\right)^3}\right]\bar{\gamma} \\
&\leq -1 + \bar{\gamma}' E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right] - C\bar{\gamma}' E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right) g_n\left(x, \bar{T}_{Q_n}\right)'\right]\bar{\gamma} \\
&\leq -1 + \left|\bar{\gamma}\right| \left|E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right]\right| - C\left|\bar{\gamma}\right|^2, \qquad (6.24)
\end{aligned}
$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\bar{\gamma}$, the second inequality follows from (6.9), and the last inequality follows from Lemma 6.6 (i) and Assumption 3.1 (vi). Thus, Lemma 6.6 (i) implies

$$
C\left|\bar{\gamma}\right| \leq \left|E_{Q_n}\left[g_n\left(x, \bar{T}_{Q_n}\right)\right]\right| = O\left(n^{-1/2}\right). \tag{6.25}
$$

From $a_n n^{1/2} \to \infty$, $\bar{\gamma}$ is an interior point of Γ_n and satisfies the first-order condition $\partial R_n\left(Q_n, \bar{T}_{Q_n}, \bar{\gamma}\right)/\partial\gamma = 0$ for all n large enough. Since $R_n\left(Q_n, \bar{T}_{Q_n}, \gamma\right)$ is concave in γ for all n large enough, we have $\bar{\gamma} = \arg\max_{\gamma \in \mathbb{R}^m} R_n\left(Q_n, \bar{T}_{Q_n}, \gamma\right)$ for all n large enough. Thus, the first statement is obtained. Also, from (6.25), the second statement is obtained. The third statement follows from

$$
\sup_{x \in \mathcal{X}} \left|\gamma_n\left(\bar{T}_{Q_n}, Q_n\right)' g_n\left(x, \bar{T}_{Q_n}\right)\right| \leq O\left(n^{-1/2} m_n\right) = o(1). \tag{6.26}
$$

Lemma 6.7. *[Consistency of $\bar{T}_{P_n}$] Suppose that Assumption 3.1 holds. Then for each sequence $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$ and $r > 0$, $\bar{T}_{P_n} \xrightarrow{p} \theta_0$ under Q_n.*

Proof. The proof is based on Newey and Smith (2004, proof of Theorem 3.1). From Assumption 3.1 (iii) and (v) (continuity of $g(x, \theta)$ in $\mathcal{N}$), it is sufficient to show that $\left|E_{P_0}\left[g\left(x, \bar{T}_{P_n}\right)\right]\right| \xrightarrow{p} 0$. Observe that

$$
\begin{aligned}
&\left|E_{P_0}\left[g\left(x, \bar{T}_{P_n}\right)\right]\right| \\
\leq\ &\left|E_{P_0}\left[\mathbb{I}\left\{x \notin \mathcal{X}_n\right\} g\left(x, \bar{T}_{P_n}\right)\right]\right| + \left|E_{P_0}\left[g_n\left(x, \bar{T}_{P_n}\right)\right] - E_{P_n}\left[g_n\left(x, \bar{T}_{P_n}\right)\right]\right| + \left|E_{P_n}\left[g_n\left(x, \bar{T}_{P_n}\right)\right]\right| \\
\leq\ &E_{P_0}\left[\mathbb{I}\left\{x \notin \mathcal{X}_n\right\} \sup_{\theta \in \Theta}\left|g(x, \theta)\right|\right] + \sup_{\theta \in \Theta}\left|E_{P_n}\left[g_n(x, \theta)\right] - E_{P_0}\left[g_n(x, \theta)\right]\right| + o_p(1), \qquad (6.27)
\end{aligned}
$$

where the first inequality follows from the triangle inequality, and the second inequality follows from Lemma 6.10 (i). The first term of (6.27) satisfies

$$\begin{aligned} & E_{P_0}\left[\mathbb{I}\left\{x \notin \mathcal{X}_n\right\} \sup_{\theta \in \Theta} \left|g\left(x,\theta\right)\right|\right] \\ \leq\ & \sqrt{E_{P_0}\left[\mathbb{I}\left\{x \notin \mathcal{X}_n\right\}\right]}\sqrt{E_{P_0}\left[\sup_{\theta\in\Theta}\left|g\left(x,\theta\right)\right|^2\right]} \\ \leq\ & \sqrt{m_n^{-1}E_{P_0}\left[\sup_{\theta\in\Theta}\left|g\left(x,\theta\right)\right|\right]}\sqrt{E_{P_0}\left[\sup_{\theta\in\Theta}\left|g\left(x,\theta\right)\right|^2\right]} \to 0, \end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, the second inequality follows from the Markov inequality, and the convergence follows from Assumption 3.1 (v) and (vii). The second term of (6.27) satisfies

$$\begin{aligned} & \sup_{\theta\in\Theta}\left|E_{P_n}\left[g_n\left(x,\theta\right)\right] - E_{P_0}\left[g_n\left(x,\theta\right)\right]\right| \\ \leq\ & \sup_{\theta\in\Theta}\left|E_{P_n}\left[g_n\left(x,\theta\right)\right] - E_{Q_n}\left[g_n\left(x,\theta\right)\right]\right| + \sup_{\theta\in\Theta}\left|E_{Q_n}\left[g_n\left(x,\theta\right)\right] - E_{P_0}\left[g_n\left(x,\theta\right)\right]\right| \\ =\ & \sup_{\theta\in\Theta}\left|\int g_n\left(x,\theta\right)\left\{dQ_n^{1/2} - dP_0^{1/2}\right\}^2 + 2\int g_n\left(x,\theta\right)dP_0^{1/2}\left\{dQ_n^{1/2} - dP_0^{1/2}\right\}\right| + o_p\left(1\right) \\ \leq\ & \sup_{\theta\in\Theta}\int\left|g_n\left(x,\theta\right)\right|\left\{dQ_n^{1/2} - dP_0^{1/2}\right\}^2 + 2\sup_{\theta\in\Theta}\int\left|g_n\left(x,\theta\right)\right|\left|dP_0^{1/2}\right|\left|dQ_n^{1/2} - dP_0^{1/2}\right| + o_p\left(1\right) \\ \leq\ & m_n\frac{r^2}{n} + 2m_n\frac{r}{\sqrt{n}} + o_p\left(1\right) \xrightarrow{p} 0, \end{aligned}$$

where the first inequality follows from the triangle inequality, the first equality follows from a uniform law of large numbers, the second inequality follows from the triangle inequality, the third inequality follows from the Cauchy-Schwarz inequality and $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$, and the convergence follows from Assumption 3.1 (vii). Combining these results, we have $\left|E_{P_0}\left[g\left(x,\bar{T}_{P_n}\right)\right]\right| \xrightarrow{p} 0$.

Lemma 6.8. *[Local uniform normality of $\bar{T}$] Suppose that Assumption 3.1 holds . Then for each sequence $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$ and $r > 0$,*

$$\sqrt{n}\left(\bar{T}_{P_n} - \theta_0\right) = -\sqrt{n}\Sigma^{-1}\int \Lambda_n dP_n + o_p\left(1\right) \quad \text{under } Q_n, \tag{6.28}$$

$$\sqrt{n}\left(\bar{T}_{P_n} - \bar{T}_{Q_n}\right) \xrightarrow{d} N\left(0, \Sigma^{-1}\right) \quad \text{under } Q_n. \tag{6.29}$$

Proof of (6.28). The proof is similar to that of Lemma 6.2. Replace Q_n with P_n and use Lemma 6.9 and Lemma 6.10 instead of Lemma 6.5 and 6.6.

Proof of (6.29). From (6.28) and Lemma 6.2,

$$\begin{aligned}&\sqrt{n}\left(\bar{T}_{P_n}-\bar{T}_{Q_n}\right)\\ =\quad&-\Sigma^{-1}G'\Omega^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{g_n\left(x_i,\theta_0\right)-E_{Q_n}\left[g_n\left(x,\theta_0\right)\right]\right\}+o_p\left(1\right).\end{aligned}$$

Thus, it is sufficient to check that we can apply a central limit theorem to the triangular array $\{g_n(x_i,\theta_0)\}_{1\le i\le n,n}$ (NEED A REFERENCE). Observe that

$$\begin{aligned}&E_{Q_n}\left[\left|g_n\left(x,\theta_0\right)\right|^{2+\epsilon}\right]\\ =\quad&\int\left|g_n\left(x,\theta_0\right)\right|^{2+\epsilon}\left\{dQ_n^{1/2}-dP_0^{1/2}\right\}^2+2\int\left|g_n\left(x,\theta_0\right)\right|^{2+\epsilon}dP_0^{1/2}\left\{dQ_n^{1/2}-dP_0^{1/2}\right\}+E_{P_0}\left[\left|g_n\left(x,\theta_0\right)\right|^{2+\epsilon}\right]\\ \le\quad&m_n^{2+\epsilon}\frac{r^2}{n}+2m_n^{1+\epsilon}E_{P_0}\left[\left|g\left(x,\theta_0\right)\right|^2\right]\frac{r}{\sqrt{n}}+E_{P_0}\left[\left|g\left(x,\theta_0\right)\right|^{2+\epsilon}\right]<\infty,\end{aligned}$$

for all n large enough, where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from Assumption 3.1 (v) and (vii). Therefore, the conclusion is obtained.

Lemma 6.9. *Suppose that Assumption 3.1 holds. Then for each sequence $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$ and $r>0$ the followings hold under Q_n:*

- **(i):** $\left|E_{P_n}\left[g_n\left(x,\theta_0\right)\right]\right|=O_p\left(n^{-1/2}\right)$, $\left|E_{P_n}\left[g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'\right]-\Omega\right|=o_p\left(1\right)$,
- **(ii):** $\gamma_n\left(\theta_0,P_n\right)=\arg\max_{\gamma\in\mathbb{R}^m}-\int\frac{1}{\left(1+\gamma'g_n\left(x,\theta_0\right)\right)}dP_n$ *exists a.s. for all n large enough,* $\left|\gamma_n\left(\theta_0,P_n\right)\right|=O_p\left(n^{-1/2}\right)$, *and* $\sup_{x\in\mathcal{X}}\left|\gamma_n\left(\theta_0,P_n\right)'g_n\left(x,\theta_0\right)\right|\xrightarrow{p}0$.

Proof of (i). Proof of the first statement. From the triangle inequality,

$$\left|E_{P_n}\left[g_n\left(x,\theta_0\right)\right]\right|\quad\le\quad\left|E_{P_n}\left[g_n\left(x,\theta_0\right)\right]-E_{Q_n}\left[g_n\left(x,\theta_0\right)\right]\right|+\left|E_{Q_n}\left[g_n\left(x,\theta_0\right)\right]\right|.$$

From the proof of (6.29) in Lemma 6.8, an application of central limit theorem implies that the first term is $O_p\left(n^{-1/2}\right)$. From Lemma 6.5 (i), the second term is $O\left(n^{-1/2}\right)$.

Proof of the second statement. From the triangle inequality,

$$\begin{aligned}&\left|E_{P_n}\left[g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'-\Omega\right]\right|\\ \le\quad&\left|E_{P_n}\left[g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'\right]-E_{Q_n}\left[g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'\right]\right|+\left|E_{Q_n}\left[g_n\left(x,\theta_0\right)g_n\left(x,\theta_0\right)'\right]-\Omega\right|.\end{aligned}$$

From a law of large numbers, the first term is $o_p(1)$. From Lemma 6.5 (i), the second term is $o(1)$.

Proof of (ii). THIS IS EXACTLY THE SAME PROOF AS FOR LEMMA 6.4 (ii) EXCEPT USING 6.9 (i) INSTEAD OF 6.4 (i). The proof is based on Newey and Smith (2004, proofs of Lemmas A.1-3). Pick any sequence $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$ and $r>0$. Since $R_n\left(Q,\theta,\gamma\right)$ is twice continuously differentiable with respect to γ and Γ_n is compact, $\bar{\gamma} = \arg\max_{\gamma\in\Gamma_n} R_n(P_n, \theta_0, \gamma)$ exists for each $n \in \mathbb{N}$. A Taylor expansion around $\bar{\gamma} = 0$ yields

$$\begin{aligned} -1 &= R_n(P_n, \theta_0, 0) \leq R_n(P_n, \theta_0, \bar{\gamma}) = -1 + \bar{\gamma}' E_{P_n}[g_n(x, \theta_0)] - \bar{\gamma}' E_{P_n}\left[\frac{g_n(x, \theta_0) g_n(x, \theta_0)'}{(1 + \dot{\gamma}' g_n(x, \theta_0))^3}\right]\bar{\gamma} \\ &\leq -1 + \bar{\gamma}' E_{P_n}[g_n(x, \theta_0)] - C\bar{\gamma}' E_{P_n}\left[g_n(x, \theta_0) g_n(x, \theta_0)'\right]\bar{\gamma} \\ &\leq -1 + |\bar{\gamma}| |E_{P_n}[g_n(x, \theta_0)]| - C|\bar{\gamma}|^2, \end{aligned}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\bar{\gamma}$, the second inequality follows from (6.9), and the last inequality follows from Lemma 6.9 (i). Thus, Lemma 6.9 (i) implies

$$C|\bar{\gamma}| \leq |E_{P_n}[g_n(x, \theta_0)]| = O_p\left(n^{-1/2}\right). \tag{6.30}$$

From $a_n n^{1/2} \to \infty$, $\bar{\gamma}$ is an interior point of Γ_n and satisfies the first-order condition $\partial R_n(P_n, \theta_0, \bar{\gamma})/\partial\gamma = 0$ for all n large enough. Since $R_n(P_n, \theta_0, \gamma)$ is concave in γ for all n large enough, $\bar{\gamma} = \arg\max_{\gamma\in\mathbb{R}^m} R_n(P_n, \theta_0, \gamma)$ for all n large enough. Thus, the first statement is obtained. From (6.30), the second statement is obtained. The third statement follows from

$$\sup_{x\in\mathcal{X}} \left|\gamma_n(\theta_0, P_n)' g_n(x, \theta_0)\right| \leq O_p\left(n^{-1/2} m_n\right) = o(1).$$

Lemma 6.10. *Suppose that Assumption 3.1 holds. Then for each sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r > 0$ the followings hold under Q_n:*

- **(i):** $\left|E_{P_n}\left[g_n\left(x, \bar{T}_{P_n}\right)\right]\right| = O_p\left(n^{-1/2}\right)$, $\left|E_{P_n}\left[g_n\left(x, \bar{T}_{P_n}\right) g_n\left(x, \bar{T}_{P_n}\right)'\right] - \Omega\right| = O_p\left(n^{-1/2}\right)$, *and* $\left|E_{P_n}\left[\partial g_n\left(x, \bar{T}_{P_n}\right)/\partial\theta'\right] - G\right| = o(1)$,,
- **(ii):** $\gamma_n\left(\bar{T}_{P_n}, P_n\right) = \arg\max_{\gamma\in\mathbb{R}^m} - \int \frac{1}{\left(1+\gamma' g_n\left(x, \bar{T}_{P_n}\right)\right)} dP_n$ *exists a.s. for all n large enough,* $\left|\gamma_n\left(\bar{T}_{P_n} P_n\right)\right| = O_p\left(n^{-1/2}\right)$, *and* $\sup_{x\in\mathcal{X}}\left|\gamma_n\left(\bar{T}_{P_n}, P_n\right)' g_n\left(x, \bar{T}_{P_n}\right)\right| \xrightarrow{p} 0$.

Proof of (i). Proof of the first statement. By a uniform law of large numbers

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|E_{P_n}\left[g_n(x, \theta) g_n(x, \theta)'\right] - E_{Q_n}\left[g_n(x, \theta) g_n(x, \theta)'\right]\right| = 0, \tag{6.31}$$

then from (6.18) and T,

$$\sup_{\theta\in\Theta}\left|E_{P_n}\left[\phi_n\left(b, \bar{T}_{Q_n}\right)\phi_n\left(b, \bar{T}_{Q_n}\right)'\right]\right| < CE_{P_0}\left[\sup_{\theta\in\Theta}|g(x, \theta)|^2\right].$$

From here the proof of the first statement is exactly as for the first statement of Lemma 6.6 (i) except using Lemma 6.9 (i) instead of Lemma 6.6 (i). Replace Q_n with P_n and use Lemma 6.9 and instead of

Lemma 6.5. Note that we cannot use Lemma 6.7 to show the above bound, since the proof of Lemma 6.7 relies on the first statement of this lemma.

Proof of the second statement follows from (6.31) and Lemma 6.6 (i).

Proof of the third statement. By a uniform law of large numbers,

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|E_{P_n}\left[\partial g_n\left(x,\theta\right)/\partial\theta\right]-E_{Q_n}\left[\partial g_n\left(x,\theta\right)/\partial\theta\right]\right|=0,$$

and hence the conclusion follows from Lemma 6.6 (i).

Proof of (ii). The proof is similar to that of the third statement of Lemma 6.6 (i).

References

Hansen, L. P. (1982): "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica*, 50, 1029–1054.

Kitamura, Y. (2006): "Empirical Likelihood Methods in Econometrics: Theory and Practice," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge University Press, Cambridge, forthcoming.

Newey, W. K., and R. J. Smith (2004): "Higher order properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

Cowles Foundation for Research in Economics, Yale University, New Haven, CT-06520.
E-mail address: yuichi.kitamura@yale.edu

Cowles Foundation for Research in Economics, Yale University, New Haven, CT-06520.
E-mail address: taisuke.otsu@yale.edu

Department of Economics, Yale University, New Haven, CT-06520.
E-mail address: kirill.evdokimov@yale.edu