ROBUSTNESS, INFINITESIMAL NEIGHBORHOODS, AND MOMENT RESTRICTIONS

YUICHI KITAMURA, TAISUKE OTSU, AND KIRILL EVDOKIMOV

Abstract. This paper is concerned with robust estimation under moment restrictions. A moment restriction model is semiparametric and distribution-free, therefore it imposes mild assumptions. Yet it is reasonable to expect that the probability law of observations may have some deviations from the ideal distribution as modeled by the moment restriction model, due to various factors such as measurement errors. It is then sensible to seek an estimation procedure that are robust against slight perturbation in the probability measure that generates observations. This paper uses local perturbations within shrinking topological neighborhoods in its development of large sample theory, so that both bias and variance matter asymptotically. The main result shows that there exists a computationally convenient estimator that achieves optimal minimax robust properties. In addition, it is semiparametrically efficient if the original probability law remain unperturbed.

1. Introduction

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

$$E_{P_0}[g(x, \theta_0)] = \int g(x, \theta_0)\, dP_0 = 0, \quad \theta_0 \in \Theta. \tag{1.1}$$

The moment condition model (1.1) is semiparametric and distribution-free, therefore imposes mild assumptions. Nevertheless, it is reasonable to expect that the probability law of observations may have some deviations from the restriction under the moment condition model. It is then sensible to seek for

Date: This Version: November 18, 2011.
Keywords: Empirical Likelihood, Hellinger Distance, Robust Estimation.

JEL Classification Number: C14.

We thank participants at the CIREQ Conference on GMM, the 2007 Winter Meetings of the Econometric Society, the 2007 Netherlands Econometrics Study Group Annual Conference, and seminars at Boston University, Chicago Booth, Harvard-MIT, LSE, Ohio State, Seoul National University, Texas A&M, the University of Tokyo, Vanderbilt, and Wisconsin for valuable comments. We acknowledge financial support from the National Science Foundation via grants SES-0241770, SES-0551271 and SES-0851759 (Kitamura) and SES-0720961 (Otsu).

estimation and testing procedures that are robust against slight perturbations in the observed data, or more technically, perturbations in the probability measure that generates observations. This notion of robustness can be illustrated as follows. Suppose $\theta(P)$ is the solution of the moment condition model with respect to θ, when the marginal probability distribution is given by P. In this notation, the "true" value θ_0, in which the econometrician is interested, is given by $\theta_0 = \theta(P_0)$. Suppose, however, observations $x_1, ..., x_n$ are not drawn according to P_0, but its "perturbed" version P instead. This can be attributed to various factors, including measurement errors or data contamination, but many other violations of the (semiparametric) model assumptions are realistic concerns in the estimation of the model. The goal of robust estimation is to obtain an estimator $\hat{\theta} = \hat{\theta}(x_1, ..., x_n)$ that is not sensitive to such perturbations, so that the deviation of the estimated value $\hat{\theta}$ from and the parameter value of interest $\theta_0 = \theta(P_0)$ remains stable. The deviation can be decomposed as:

$$\hat{\theta} - \theta_0 = [\hat{\theta} - \theta(P)] + [\theta(P) - \theta(P_0)]. \tag{1.2}$$

In the asymptotic MSE calculation presented below, the expectation of the square of the term in the first square bracket contributes to the variance of the estimator, whereas the second corresponds to the bias. An estimator that achieves small MSE uniformly in P over a neighborhood of P_0 is desirable.

Asymptotic theory of robust estimation when the model is parametric has been considered extensively in the literature: see Rieder (1994) for a comprehensive survey. In a pioneering paper, Beran (1977) discusses "robust and efficient" estimation of parametric models. Suppose $P_\theta, \theta \in \Theta \subset \mathbb{R}^k$ is a parametric family of probability measures. Observations are drawn from a probability law P, which may not be a member of the parametric family. Let p_θ and p denote the densities associated with the probability measures P_θ and P. It is well-known that the parametric MLE procedure corresponds to minimizing the objective function $\rho = \int \log(p/p_\theta) p dx$. Beran points out that a small change in the density p can lead to a large change in the objective function ρ (note the log in ρ), implying the non-robustness of the MLE $\hat{\theta} = \operatorname{argmin} \rho$. He shows that the parametric Minimum Hellinger distance estimator (MHDE) is "robust and efficient," in the sense that (i) it has an asymptotic minimax robust property and (ii) it is asymptotically efficient when the model assumption is satisfied, i.e. when the sample is generated from $P_0 = P_{\theta_0}$, where θ_0 is the true value of the parameter of interest. Let $H(P_\theta, P) = \sqrt{\int (p_\theta^{1/2}(x) - p^{1/2}(x))^2 dx}$ denote the Hellinger distance between P_θ and P (a slightly more general definition of the Hellinger distance is given in the next section). The MHDE for a parametric model is implemented as follows:

$$
\begin{aligned}
\hat{\theta} &= \underset{\theta}{\operatorname{argmin}}\, h(P_\theta, \hat{P}) \\
&= \underset{\theta}{\operatorname{argmin}} \int (p_\theta^{1/2}(x) - \hat{p}^{1/2}(x))^2 dx
\end{aligned}
$$

where $\hat{p}$ is a nonparametric density estimator, such as a kernel density estimator, for P and $\hat{P}$ is the corresponding estimator for the probability measure of x. The MHDE is asymptotically equivalent to MLE and thus efficient if the model assumption is satisfied. One can replace the Hellinger distance with other divergence measures such as the Kolmogorov-Smirnov distance, which would make the corresponding minimum divergence estimator even more robust, but it would incur efficiency loss. The parametric MHDE has been studied extensively and applied to various models.

The parametric MHDE has theoretical advantages and excellent finite sample performance documented by numerous simulation studies, but it has some limitation as well. It requires the nonparametric density estimator when at least some components of z are continuously distributed. This makes its practical application inconvenient, and is problematic when x is high dimensional, due to the curse of dimensionality. This also necessitates the evaluation of the integral $\int (p_\theta^{1/2}(x) - \hat{p}^{1/2}(x))^2 dx$. This would either involve numerical integration or an approximation by an empirical average with inverse density weighting using a nonparametric density estimator. The former can be hard to implement, and the latter may have undesirable effects in finite samples. This paper aims at developing robust methods for moment restriction models, by applying the MHDE framework. The resulting estimator is asymptotically as efficient as optimally weighted GMM when the model assumptions hold, and at the same time it enjoys an optimal minimax robust property. The implementation of the estimator is easy, and unlike its parametric predecessor, it requires neither nonparametric density estimation nor evaluation of integration.

2. Preliminaries

The econometrician wishes to estimate the unknown θ_0 in (1.1). Suppose a random sample $\{x_i\}_{i=1}^n$ generated from P is observed. As discussed in Section 1, our focus is on robust estimation of θ_0 when the probability measure P, from which the observations are drawn, is a (locally) perturbed version of P_0, not P_0 itself. There exists an extensive literature concerning the estimation of (1.1) under the "classical" setting where data are indeed drawn from P_0. Many estimators 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 MHDE applied to the moment restriction model (1.1). 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:*

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

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:

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

Here we show some results concerning the Hellinger distance that are useful in understanding the robustness theorems in the next section.

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*

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

If P is not absolutely continuous respect to Q, then $\int \mathbb{I}\{p > 0, q = 0\} d\nu > 0$, and as a consequence $I_\alpha(P,Q) = \infty$ for $\alpha \geq 1$. A similar argument shows that $I_\alpha(P,Q) = \infty$ 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

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

(with the above convention for the case where $P \not\ll Q$), giving rise to the well-known Kullback-Leibler (KL) divergence measure from Q to P. 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 $\alpha \in \mathbb{R}$.

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 $\delta > 0$:

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

Lemma 2.1 implies that

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

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

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

then by the above inequality

$$\cup_{\alpha \in \left[\frac{1}{2}-L,\frac{1}{2}+U\right]} B_{I_\alpha}(P_0,\delta) \subset B_{I_{1/2}}\left(P_0, \sqrt{2C(L,U)}\delta\right). \tag{2.1}$$

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 $\sqrt{2C(L,U)}$. (2.1) 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 (2.1) does not hold if the Hellinger distance $I_{\frac{1}{2}}$ is replaced by $I_\alpha, \alpha \neq \frac{1}{2}$, showing the special status of the Hellinger distance among the α-divergence family.

Remark 2.2. Lemma 2.1 is a statement for every pair of measures (P,Q), thus it holds even if $P \not\ll Q$ or $P \not\ll Q$. On the other hand, it is useful to consider the behavior of I_α when one of the two measures is not absolutely continuous with respect to the other. Consider a sequence of probability measures $\left\{P^{(n)}\right\}_{n\in\mathbb{N}}$. Suppose $I_\alpha\left(P^{(n)},P_0\right) \to 0$ for an $\alpha \in \mathbb{R}$, then $I_{\alpha'}\left(P^{(n)},P_0\right) \to 0$ for every $\alpha' \in (0,1)$. But the reverse (i.e. reversing the roles of α and α') is not true. If $P^{(n)}, n \in \mathbb{N}$ are not absolutely continuous respect to P_0, $I_{\alpha'}\left(P^{(n)},P_0\right) = \infty$ for every $\alpha' \geq 1$ even if $\rho_\alpha\left(P^{(n)},P_0\right) \to 0$ for $\alpha \in (0,1)$ (and a similar argument holds for $\alpha' \leq 0$). This shows that I_α-based neighborhoods with $\alpha \notin (0,1)$ are too small: there are measures that are outside of $B_{I_\alpha}(P_0,\delta), \alpha \notin (0,1)$ no matter how large δ is, or how close they are to P_0 in terms of, say, the Hellinger distance H.

Remark 2.3. The inequality in Lemma 2.1 might be of interest on its own as it generalizes many inequalities in the literature. For $\alpha = 1$ or 0, it corresponds to the well-known inequality between the KL divergence and the Hellinger distance

$$H(P,Q)^2 \leq K(P,Q), \tag{2.2}$$

see, for example, Pollard (2002), p.62. Another commonly used definition of divergence between probability measures is the χ^2 distance. It is given, if $P \ll Q \ll \nu$, by $\chi^2(P,Q) = \int \frac{(p-q)^2}{q} d\nu$, and it is shown that

$$H(P,Q)^2 \leq \chi^2(P,Q) \tag{2.3}$$

(Reiss (1989)). This is implied by Lemma 2.1 by letting $\alpha = 2$. Proposition 3.1 in Zhang (2006) is closer to our result in terms of its generality; it shows that $\max(\alpha, 1-\alpha) I_\alpha(P,Q) \geq \frac{1}{2} I_{\frac{1}{2}}(P,Q)$ holds for $\alpha \in [0,1]$, which covers (2.2) but not (2.3)[1]. Lemma 2.1 shows that this type of inequality holds for all $\alpha \in \mathbb{R}$.

Beran (1977), considering a parametric model, proposed MHDE that minimizes the Hellinger distance between a model-based probability measure (from the parametric family) and a nonparametric probability measure estimate. An application of the MHDE procedure to the moment condition model (1.1) 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. Let

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

and

$$\mathcal{P} = \cup_{\theta \in \Theta} \mathcal{P}_\theta, \tag{2.4}$$

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

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

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

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

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

[1]Zhang (2006) also derives a lower bound for the Hellinger distance in terms of I_α.

It is easy to verify that we can obtain the MHDE as a Generalized Empirical Likelihood (GEL) estimator by letting $\gamma = -1/2$ in equation (2.6) of Newey and Smith (2004). Asymptotic properties of the (G)EL estimators for θ_0 in (1.1), when data drawn from P_0 are observed are well-understood (see, for example, Kitamura and Stutzer (1997), Smith (1997), Imbens, Spady, and Johnson (1998) and Newey and Smith (2004)). Let $G = E_{P_0}\left[\partial g(x,\theta_0)/\partial\theta'\right]$, $\Omega = E_{P_0}\left[g(x,\theta_0)\, g(x,\theta_0)'\right]$, and $\Sigma = G'\Omega^{-1}G$. Then

$$\sqrt{n}\left(\hat{\theta}_{\mathrm{GEL}} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma^{-1}\right). \tag{2.5}$$

It follows that the MHDE and other GEL estimators are semiparametrically efficient in the absence of data perturbation. At the same time, the MHDE possesses a distinct property of being asymptotic optimal robust if observations are drawn from a perturbed version of P_0, as we shall see in the next section.

3. Robust Estimation Theory

We now analyze robustness of the MHDE $\hat{\theta}$. Define a functional

$$T(P) = \arg\min_{\theta\in\Theta}\max_{\gamma\in\mathbb{R}^m} -\int \frac{1}{1+\gamma' g(x,\theta)}dP$$

then the MHDE can be interpreted as the value of functional T evaluated the empirical measure P_n. In other words, each realization of P_n completely determines the value of the MHDE $\hat{\theta}$. To make the dependence explicit, we write $\hat{\theta} = T(P_n)$, and study properties of the mapping $T : \mathcal{M} \to \Theta$. This definition of $T(\cdot)$, however, causes a technical difficulty when the distribution of $g(x,\theta)$ is unbounded for some $\theta \in \Theta$ and $P \in \mathcal{M}$. To overcome this technical difficulty, we introduce the following mapping defined by a trimmed moment function:

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

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

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

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 $E_P[g(x,\theta)\,\mathbb{I}\{x \in \mathcal{X}_n\}] = 0$.

Let $\tau : \Theta \to \mathbb{R}$ be a possibly nonlinear transformation of the parameter. We first focus on the estimation problem of the transformed scalar 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 P_0

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

The transformation τ to a scalar, as used by Rieder (1994), is convenient in calculating squared biases and MSE's. One may, for example, let $\tau(\theta) = c'\theta$ using a constant p-vector c. Lemma A.1 (ii) guarantees that for each $r > 0$, the value $\bar{T}(Q)$ exists for all $Q \in B_H(P_0, r/\sqrt{n})$ and all n large enough.

Assumption 3.1. *Suppose the following conditions hold:*

(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): $g(x,\theta)$ *is continuous over* Θ *at each* $x \in \mathcal{X}$*;*

(v): $E_{P_0}[\sup_{\theta\in\Theta}|g(x,\theta)|^\eta] < \infty$ *for some* $\eta > 2$*, and there exists a neighborhood* $\mathcal{N}$ *around* θ_0 *such that* $E_{P_0}\left[\sup_{\theta\in\mathcal{N}}|g(x,\theta)|^4\right] < \infty$*,* $g(x,\theta)$ *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* $m_n \to \infty$*,* $nm_n^{-\eta} \to 0$*, and* $n^{-1/2}m_n^{1+\epsilon} = O(1)$ *for some* $0 < \epsilon < 2$ *as* $n \to \infty$*;*

(viii): τ *is continuously differentiable at* θ_0*.*

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 under P_0. Assumption 3.1 (iv) ensures 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 (2.5). 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 $m_n \sim n^a$, this assumption is satisfied for $1/\eta < a < 1/2$. Assumption 3.1 (viii) is a standard requirement for the parameter transformation τ. To characterize a class of estimators to be compared with the MHDE, we introduce the following definition.

Definition 3.1. *Let $T_a(P_n)$ be an estimator of θ_0 based on a mapping $T_a : \mathcal{M} \to \Theta$. Also, let $P_{\theta,\zeta}$ be a regular parametric submodel (see, Bickel, Klassen, Ritov, and Wellner (1993, p. 12) or Newey (1990)) of $\mathcal{P}$ in* (2.4) *such that $P_{\theta_0,0} = P_0$ and $P_{\theta_0+t/\sqrt{n},\zeta_n} \in B_H(P_0, r/\sqrt{n})$ holds for $\zeta_n = O\left(n^{-1/2}\right)$ eventually.*

(i): *T_a is called **Fisher consistent** if for every $\{P_{\theta_n,\zeta_n}\}_{n\in\mathbb{N}}$ and $t \in \mathbb{R}^p$,*

$$\sqrt{n}\left(T_a\left(P_{\theta_0+t/\sqrt{n},\zeta_n}\right) - \theta_0\right) \to t. \tag{3.1}$$

(ii): *T_a is called **regular** for θ_0 if for every $\{P_{\theta_n,\zeta_n}\}_{n\in\mathbb{N}}$ with $(\theta_n', \zeta_n')' = (\theta_0', 0)' + O(n^{-1/2})$, there exists a probability measure M such that*

$$\sqrt{n}\left(T_a(P_n) - T_a(P_{\theta_n,\zeta_n})\right) \stackrel{d}{\to} M, \quad \text{under } P_{\theta_n,\zeta_n}, \tag{3.2}$$

where the measure M does not depend on the sequence $(\theta_n', \zeta_n')'$.

Both conditions are weak and satisfied by GMM, (G)EL and other standard estimators. For example, the mapping T_a for the continuous updating GMM estimator (CUE) is given by

$$T_{CUE}(P) = \operatorname*{argmin}_{\theta\in\Theta} \left[\int g(x,\theta)dP\right]' \left[\int g(x,\theta)g(x,\theta)dP\right]^{-1} \left[\int g(x,\theta)dP\right],$$

and under Assumption 3.1 $T_{CUE}(P_{\theta_0+t/\sqrt{n}}, \zeta_n) = \theta_0 + t/\sqrt{n}$ for large n. CUE therefore trivially satisfies (3.1). The regularity condition (3.2) is standard in the literature of semiparametric efficiency; see, for example, Bickel, Klassen, Ritov, and Wellner (1993).

The following theorem shows the optimal robustness of the (trimmed) MHDE in terms of its maximum bias.

Theorem 3.1. Suppose that Assumption 3.1 holds.

(i): For every T_a which is Fisher consistent,

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

for each $r > 0$, where $B^* = \left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)' \Sigma^{-1}\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)$.

(ii): The mapping $\bar{T}$ is Fisher consistent and satisfies

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

for each $r > 0$.

Remark 3.1. The above result is concerned with deterministic properties of T_a and T. $T_a(Q)$ and $T(Q)$ can be regarded as the (probability) limit of the estimators $T_a(P_n)$ and $T(P_n)$ under Q, and therefore the terms evaluated here correspond to the bias of each estimators due to the deviation of Q from P_0. The theorem says that in the class of all mappings that are Fisher consistent, the mapping $\bar{T}$ has the smallest maximum bias over the set $B_H\left(P_0, r/\sqrt{n}\right)$. The (trimmed version of) the Hellinger-based mapping $\bar{T}$ is therefore optimally robust asymptotically in a minimax sense. The term $4r^2B^*$ provides a sharp lower bound for maximum squared bias, and it is attained by $\bar{T}$.

Remark 3.2. The theorem is concerned with the trimmed version of the MHDE. It avoids the complications associated with the existence of $T(Q)$ for certain Q's. If the support of $\sup_{\theta\in\Theta}|g(x,\theta)|$ is bounded under every $Q \in B_H(P_0, r/\sqrt{n})$ for large enough n (e.g. if the moment function g is bounded), then we do not need the trimming term $\mathbb{I}\{x \in \mathcal{X}_n\}$. In this case the mapping T without trimming has the above optimal robust property.

Remark 3.3. The index n in the statement of Theorem 3.1 simply parameterizes how close $Q \in B_H(P_0, r/\sqrt{n})$ and P_0 are, and does not have to be interpreted as the sample size. The next theorem, however, is concerned with MSE's and the index n represents the sample size there.

The next theorem is our main result, which is concerned with (the supremum of) the MSE of the minimum Hellinger distance estimator $\hat{\theta} = T(P_n)$ and other competing estimators. Let

$$\bar{B}_H\left(P_0, r/\sqrt{n}\right) = B_H\left(P_0, r/\sqrt{n}\right) \cap \left\{Q \in \mathcal{M} : E_Q\left[\sup_{\theta\in\Theta}|g(x,\theta)|^\eta\right] < \infty\right\}. \tag{3.3}$$

We use the notation $P^{\otimes n}$ to denote the n-fold product measure of a probability measure P.

Theorem 3.2. Suppose that Assumption 3.1 holds.

(i): For every Fisher consistent and regular mapping T_a,

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

for each $r > 0$.

(ii): The mapping T is Fisher consistent and regular, and the MHDE $\hat{\theta} = T(P_n)$ satisfies

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

for each $r > 0$.

Remark 3.4. This theorem establishes an asymptotic minimax optimality property of the MHDE in terms of MSE among all the estimators that satisfies the two conditions in Definition 3.1. Note that the expression $\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int b\wedge n\left(\tau\circ T_a\left(P_n\right)-\tau\left(\theta_0\right)\right)^2 dQ^{\otimes n}$ is the maximum finite sample MSE of $T_a(P_n)$. Thus our criterion for evaluating T_a (and T) is the limit of its maximum finite sample MSE. Taking the supremum over B_H before letting n go to infinity is important for capturing finite sample robustness properties. The method of calculating the truncated MSE first, then letting $b\to\infty$, is standard in the literature of robust estimation but also used in general contexts; see, for example, Bickel (1981) and LeCam and Yang (1990). Once again, we are able to derive a sharp lower bound for the maximum MSE and show that it is achieved by the MHDE $\hat{\theta}=T\left(P_n\right)$.

Remark 3.5. Unlike in Theorem 3.1, optimality is achieved by the untrimmed version of the MHDE. Note that $T(P_n)$ exists for large n under Assumption 3.1, in contrast to our discussion in Remark 3.2 on Theorem 3.1. Theorem 3.2, however, restricts the robustness neighborhood by an extra requirement as in (3.3). This is useful in showing that the untrimmed MHDE achieves the lower bound.

Remark 3.6. Theorem 3.2 proves that the MHDE is asymptotically optimally robust over a sequence of infinitesimal neighborhoods. Note that the Hellinger neighborhood over which the maximum of MSE is taken is nonparametric, in the sense that potential deviations from P_0 cannot be indexed by a finite dimensional parameter. That is, our robustness concept demands uniform robustness over a nonparametric, infinitesimal neighborhood. The use of infinitesimal neighborhoods where the radius of the Hellinger ball shrinks at the rate $n^{1/2}$ is useful in balancing the magnitude of bias and variance in our asymptotics. If one uses a fixed, global neighborhood, then the bias term would dominate the behavior of estimators. This may fail to provide a good approximation of finite sample behavior in actual applications, since in reality it would be reasonable to be concerned with both the stochastic fluctuation of estimators and their deterministic bias due to, say, data contamination. We note that there is a related but distinct literature on the asymptotics theory when the model is globally misspecified, as in White (1982), who considered parametric MLE. Kitamura (1998) and Kitamura (2002) offer such analysis for conditional and unconditional moment condition models. Moreover, Schennach (2007) provides novel and potentially very useful results of EL estimators and its variants in misspecified moment condition models. We regard our paper as a complement to, rather than a substitute for the results obtained in these papers. There are fundamental differences between the characteristics of the problems the current paper considers and those of the papers on misspecification. First, our object of interest is θ_0, not a pseudo-true value, as we consider data perturbation rather than model misspecification. Second, the nature of our analysis is local and therefore the parameter value θ_0 in (1.1) is still identified asymptotically. Third, as noted above, we consider uniform robustness over a nonparametric neighborhood. The papers cited above consider pointwise problems. Therefore our approach deals with phenomena that are very different from the ones analyzed in the literature of misspecified models.

Remark 3.7. We have seen in Remark 2.1 that the Hellinger neighborhood B_H has nice and distinct properties, in particular the inclusion relationship (2.1). The Hellinger neighborhood B_H is commonly used in the literature of robust estimation (of parametric models); see, for example, Beran (1977), Bickel (1981), and Rieder (1994). We should note, however, that other neighborhood systems have been used in the literature as well. For example, one may replace the Hellinger distance H with the Kolmogorov-Smirnov (KS) distance in the definition of B_H. As Beran (1984) notes, however, that in order to guarantee robustness in the Kolmogorov-Smirnov neighborhood system one needs

> "to use minimum distance estimates based on the Kolmogorov-Smirnov metric or a distance weaker than the Kolmogorov-Smirnov metric ... The general principle here is that the estimation distance be no stronger than the distance describing the contamination neighborhood...."

Donoho and Liu (1988) develop a general theory of the above point. What this means is that an estimator that is robust against perturbations within Kolmogorov-Smirnov neighborhoods has to be minimizing the KS (or weaker) distance. The "minimum KS estimator" for the moment restriction model would be indeed robust, but it cannot be semiparametrically efficient when the model assumption holds. Therefore, unlike the moment restriction MHDE, the estimator is not "robust and efficient." Another drawback is its computation, since, unlike the moment restriction MHDE, no convenient algorithm to minimize the Kolmogorov-Smirnov distance under the moment restriction is known in the literature. It should be noted that the moment restriction MHDE is efficient in the sense that it achieves the semiparametric efficiency bound. It does not have the desirable higher order properties of EL (Newey and Smith (2004)) or the ETEL estimator proposed by Schennach (2007).

The above MSE theorem conveniently summarizes the desirable robustness properties of the MHDE in terms of both (deterministic) bias and variance. It has, however, some limitations. First, its minimaxity result is obtained within Fisher consistent and regular estimators. While these requirements are weak, it might be of interest to expand the class of estimators. More importantly, implicit in the MSE-based analysis is that we are interested in L^2-loss. One may wish to use other types of loss functions, however, and it is of interest to see whether the above minimax results can be extended to a larger class of loss. The next theorem addresses these two issues. Of course, the MSE has an advantage of subsuming the bias and the variance in one measure. To deal with general loss functions, the next theorem focuses on the risk of estimators around a Fisher-consistent mapping evaluated at the perturbed measure Q. This can be regarded as calculating the risk of the first bracket of the decomposition (1.2), that is, the stochastic part of the deviation of the estimator from the parameter of interest θ_0.

Let $\mathcal{S}$ be a set of all estimators, that is, the set of all $\bar{\mathbb{R}}^p$-valued measurable functions. We now investigate robust risk properties of this large class of estimators. The loss function we consider satisfies the following weak requirements.

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 $\bar{\mathbb{R}}^p$.*

We now present an optimal risk property for the MHDE.

Theorem 3.3. Suppose that Assumptions 3.1 and 3.2 hold.

(i): For every Fisher consistent mapping T_a,

$$\lim_{b\to\infty}\lim_{r\to\infty}\liminf_{n\to\infty}\inf_{S_n\in\mathcal{S}}\sup_{Q\in\bar{B}_H\left(P_0,r/\sqrt{n}\right)}\int b\wedge\ell\left(\sqrt{n}\left(S_n-\tau\circ T_a\left(Q\right)\right)\right)dQ^{\otimes n}\geq\int\ell dN\left(0,B^*\right).$$

(ii): The mapping T is Fisher consistent and the MHDE $\hat{\theta} = T(P_n)$ satisfies

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

for all $r > 0$.

Note that Theorem 3.3(ii) remains valid if $T(P_n)$ is replaced by $\bar{T}(P_n)$. This theorem shows that he MHDE is once again optimally robust even for the general risk criterion, and this holds in the class of essentially all possible estimators. As noted above, the result is concerned with the stochastic component of the decomposition (1.2). Theorem 3.1 has already established that the MHDE is optimal in terms of its bias, that is, the deterministic part of the decomposition (1.2) in the second bracket. The latter result does not depend on a specific loss function. Thus the MHDE enjoys general optimal robust properties under a quite general setting, both in terms of the stochastic component and the deterministic component. Note that analyzing these two parts separately is common in the literature of robust statistics: see, for example, Rieder (1994).

4. Simulation

The purpose of this section is to examine the robustness properties of the MHDE and other well-known estimators such as GMM using Monte Carlo simulations. MATLAB is used for computation throughout the experiments. The sample size n is 100 for all designs, and we ran 5000 replications for each design.

The baseline simulation design in this experiment follows that of Hall and Horowitz (1996). We then "contaminate" the simulated data to explore robustness of estimators. More specifically, let $x = (x_1, x_2)' \sim N\left(0, 0.4^2 I_2\right)$. This normal law corresponds to P_0 in the preceding sections. The specification of the moment function g is

$$g\left(x,\theta\right) = \left(\exp\left\{-0.72 - \theta\left(x_1 + x_2\right) + 3x_2\right\} - 1\right)\begin{pmatrix} 1 \\ x_2 \end{pmatrix}.$$

The moment condition $\int g(x,\theta)dP_0 = 0$ is uniquely solved at $\theta_0 = 3$. The goal is to estimate this value using the above specification of g when the original DGP is perturbed into different directions. More specifically, we use $x \sim N\left(0, \Sigma_{(\delta,\rho)}\right)$, where

$$\Sigma_{(\delta,\rho)} = 0.4^2 \begin{pmatrix} (1+\delta)^2 & \rho\left(1+\delta\right) \\ \rho\left(1+\delta\right) & 1 \end{pmatrix}.$$

The unperturbed case thus corresponds to $\delta = \rho = 0$. In the simulation we set $\rho = 0.1\sqrt{2}\cos\left(2\pi\omega\right)$ and $\delta = 0.1\sin\left(2\pi\omega\right)$ and let ω vary over $\omega_j = j/64, j = 0, ..., 63$. This yields 64 different designs, for each of them 5000 replications is performed and RMSE and $Pr\left\{\left|\hat{\theta} - \theta_0\right| > 0.5\right\}$ are calculated. We consider the following estimators: empirical likelihood (EL), MHDE, exponential tilting (ET), and GMM (GMM2). GMM2 is calculated following the standard two step procedure where the initial estimate is obtained from identity weighting. We also simulated the continuously updated GMM estimator (CUE). CUE's performance is extremely sensitive to data perturbations considered here; its RMSE is much higher than that of the other estimators. For convenience we only plot the results EL, MHDE, ET, and GMM2 here, while the graphs that include the results for CUE are available upon request from the authors. The results are presented in Figure 1. In the upper panel, each curve represents the RMSE of a particular estimator as a function of ω_j. The lower panel (labeled "Pr") displays the simulated probability of an estimator deviating from the target $\theta_0 = 3$ by more 0.5.

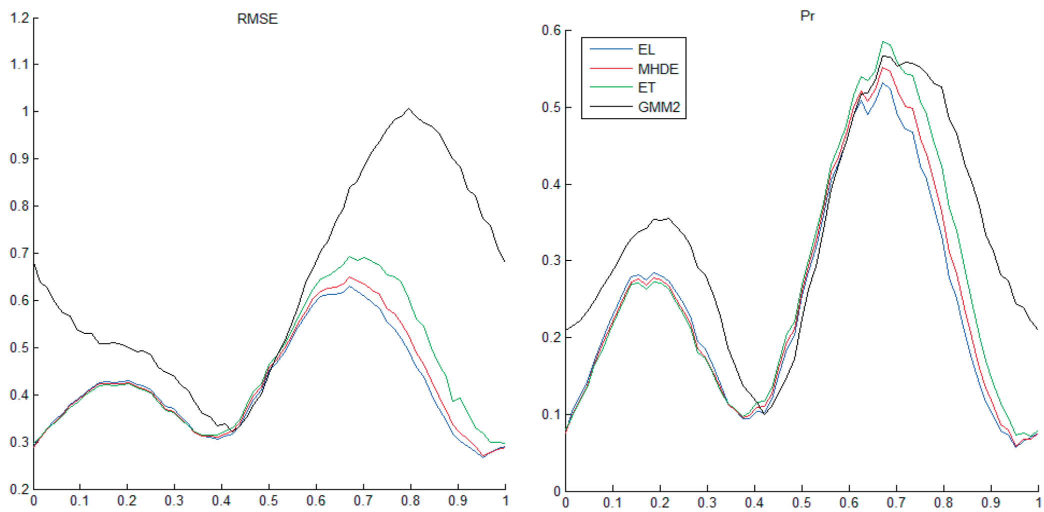

FIGURE 1. Local Neighborhood of the True Model. "Pr" denotes $Pr\left\{\left|\hat{\theta}-\theta_0\right|>0.5\right\}$.

While RMSE is a potentially informative measure, it can be a highly misleading as some of the estimators may not have finite moments. We thus focus on the results for deviation probabilities. We plotted the results in graphs to visualize the relative rankings of the estimators more clearly. We see that GMM2 is affected by perturbations much more than EL, MHDE and ET except for the values of ω's between 0.4 and 0.6, where the performance of the four estimators are rather close. ET seems to perform a little worse than MHDE and EL.

One needs be cautious in drawing conclusions based on limited simulation experiments as presented here. Nevertheless, it appears that two general features emerge form our results. First, the GMM type estimators (two step GMM and CUE) tend to be highly sensitive to data perturbations. Applying Beran's (1977) logic that connects the robustness of estimators to the forms of their objective functions, this may be attributed to the fact the GMM objective function is quadratic and therefore tends to react sensitively to the added noises. Second, EL, MHDE and ET are relatively well-behaved, and their rankings, not surprisingly, vary depending on the simulation design. The performance of MHDE, however, seems more stable compared with that of EL or ET: EL and ET exhibits more instability in throughout the different perturbation designs. Note that EL, MHDE, ET and CUE correspond to the GEL estimator with $\gamma=-1,-\frac{1}{2},0,1$ in equation (2.6) of Newey and Smith (2004). Given the good theoretical robustness property of the MHDE and the proximity of EL and ET in terms of their γ values, it is interesting to observe the reasonably robust behavior of EL and ET. Note that CUE, whose behavior is quite different from that of the MHDE and thus highly non-robust, has $\gamma = 1$, a value that is much higher than the optimally robust $\gamma = -1/2$ of the MHDE.

5. Conclusion

In this paper we have explored the issue of robust estimation in a moment restriction model. The model is semiparametric and distribution-free, therefore imposes mild assumptions. Yet it is reasonable to expect that the probability law of observations may have some deviations from the ideal distribution as modeled by the moment restriction model. It is then sensible to seek estimation procedures that are robust against slight perturbations in the probability measure that generates observations, which can be caused by, for example, data contamination. Our main theoretical result shows that the minimum Hellinger distance estimator (MHDE) possesses optimal minimax robust properties. Moreover, it remains semiparametrically efficient when the model assumptions hold. Convenient numerical algorithms for its implementation are provided. Our simulation results indicate that GMM can be highly sensitive to data perturbations. The performance of the MHDE remains stable over a wide range of simulation designs, which is in accordance with our theoretical findings.

The results obtained in this paper are concerned with estimation, though it might be potentially possible to extend our robustness theory to parameter testing problems. It is of practical importance to consider robust methods for testing and confidence interval calculations so that the results of statistical inference for moment restriction models are reliable and not too sensitive to departures from model assumptions. Interestingly, there exists a literature on parametric robust inference based on the MHDE method. We plan to investigate robust testing procedure in moment condition models in our future research.

References

Beran, R. (1977): "Minimum Hellinger distance estimates for parametric models," *Annals of Statistics*, 5, 445–463.

——— (1984): "Minimum distance procedures," in *Handbook of Statistics*, ed. by P. Krishnaiah, and P. Sen. Elsevier Science, pp. 741–754.

Bickel, P., C. Klassen, Y. Ritov, and J. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins Press.

Bickel, P. J. (1981): "Quelques aspects de la statistique robuste," in *Ecole d'Eté de Probabilités de Saint Flour IX 1979*, ed. by P. Hennequin. Springer, pp. 1–72.

Borwein, J. M., and A. S. Lewis (1993): "Partially-finite programming in L_1 and the existence of maximum entropy estimates," *SIAM Journal of Optimization*, 3, 248–267.

Donoho, D., and R. Liu (1988): "The "automatic" robustness of minimum distance functionals," *Annals of Statistics*, 16, 552–586.

Hall, P., and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators," *Econometrica*, 64, 891–916.

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

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

Kitamura, Y. (1998): "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood," Working Paper, Department of Economics, University of Wisconsin.

——— (2002): "A Likelihood-based Approach to the Analysis of a Class of Nested and Non-nested Models," Working Paper, Department of Economics, University of Pennsylvania.

——— (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: Cambridge University Press, forthcoming.

Kitamura, Y., and M. Stutzer (1997): "An Information Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65(4), 861–874.

LeCam, L., and G. Yang (1990): *Asymptotics in Statistics: Some Basic Concepts*. Springer.

Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

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

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge.

Reiss, R.-D. (1989): *Approximate Distributions of Order Statistics*. Springer-Verlag.

Rieder, H. (1994): *Robust Asymptotic Statistics*. Springer-Verlag.

Schennach, S. M. (2007): "Point estimation with exponentially tilted empirical likelihood," *Annals of Statistics*, 35, 634–672.

Smith, R. J. (1997): "Alternative semi-parametric likelihood approaches to generalized method of moments estimation," *Economic Journal*, 107, 503–519.

van der Vaart, A. (1998): *Asymptotic Statistics*. Cambridge University Press.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

Zhang, T. (2006): "From ε-entropy to KL-entropy: Analysis of Minimum Information Complexity Density Estimation," *The Annals of Statistics*, 34, 2180–2210.

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, Princeton University, Princeton, NJ-08544.

E-mail address: kevdokim@princeton.edu

SUPPLEMENTAL APPENDIX FOR "ROBUSTNESS, INFINITESIMAL NEIGHBORHOODS, AND MOMENT RESTRICTIONS"

YUICHI KITAMURA, TAISUKE OTSU AND KIRILL EVDOKIMOV

A.1 Proofs of Main Results

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

A.1.1. **Proof of Lemma 2.1.** We first show the claim for $\alpha < \frac{1}{2}$, that is,

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

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

$$\int H_\alpha \left(\frac{p}{q}\right) q d\nu \geq 0. \tag{A.2}$$

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 (A.2) 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

$$(1-\beta) I_{1-\beta} (P,Q) \geq \frac{1}{2} I_{\frac{1}{2}} (P,Q). \tag{A.3}$$

By (A.1) 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 (A.3) follows.

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

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

A.1.2. **Proof of Theorem 3.1.**

A.1.2.1. *Proof of (i).* Pick arbitrary $r>0$ and $t\in\mathbb{R}^p$. Consider the following parametric submodel having the likelihood ratio

$$\frac{dP_{\theta_n,\zeta_n}}{dP_0} = \frac{1+\zeta_n' g_n(x,\theta_n)}{\int (1+\zeta_n' g_n(x,\theta_n))\, dP_0} = f(x,\theta_n,\zeta_n), \tag{A.4}$$

where

$$\zeta_n = -E_{P_0}\left[g(x,\theta_n)\, g_n(x,\theta_n)'\right]^{-1} E_{P_0}[g(x,\theta_n)].$$

Note that $P_{\theta_0,0} = P_0$, $P_{\theta_n,\zeta_n}\in\mathcal{P}_{\theta_n}$ (by the definition of ζ_n), and $\zeta_n = O\left(n^{-1/2}\right)$ (by the proof of Lemma A.4 (i)). Also, since $\sup_{x\in\mathcal{X}} |\zeta_n' g_n(x,\theta_n)| = O\left(n^{-1/2} m_n\right) = o(1)$, the likelihood ratio $\frac{dP_{\theta_n,\zeta_n}}{dP_0}$ is well-defined for all n large enough. So, for this submodel the mapping T_a must satisfy (3.1).
We now evaluate the Hellinger distance between P_{θ_n,ζ_n} and P_0. An expansion around $\zeta_n = 0$ yields

$$H(P_{\theta_n,\zeta_n}, P_0) = \left\| \zeta_n' \left.\frac{\partial f(x,\theta_n,\zeta_n)^{1/2}}{\partial\zeta_n}\right|_{\zeta_n=0} dP_0^{1/2} + \frac{1}{2}\zeta_n' \left.\frac{\partial^2 f(x,\theta_n,\zeta_n)^{1/2}}{\partial\zeta_n\partial\zeta_n'}\right|_{\zeta_n=\dot{\zeta}_n} \zeta_n dP_0^{1/2} \right\|,$$

where $\dot{\zeta}_n$ is a point on the line joining ζ_n and 0, and

$$\left.\frac{\partial f(x,\theta_n,\zeta_n)^{1/2}}{\partial\zeta_n}\right|_{\zeta_n=0} = \frac{1}{2}\left\{g_n(x,\theta_n) - E_{P_0}[g_n(x,\theta_n)]\right\},$$

$$\begin{aligned}
\frac{\partial^2 f(x,\theta_n,\zeta_n)^{1/2}}{\partial\zeta_n\partial\zeta_n'} &= -\frac{1}{4}\left(1+\zeta_n' g_n(x,\theta_n)\right)^{-3/2}\left(1+\zeta_n' E_{P_0}[g_n(x,\theta_n)]\right)^{-1/2} g_n(x,\theta_n)\, g_n(x,\theta_n)' \\
&\quad -\frac{1}{2}\left(1+\zeta_n' g_n(x,\theta_n)\right)^{-1/2}\left(1+\zeta_n' E_{P_0}[g_n(x,\theta_n)]\right)^{-3/2} g_n(x,\theta_n)\, E_{P_0}[g_n(x,\theta_n)]' \\
&\quad +\frac{3}{4}\left(1+\zeta_n' g_n(x,\theta_n)\right)^{1/2}\left(1+\zeta_n' E_{P_0}[g_n(x,\theta_n)]\right)^{-5/2} E_{P_0}[g_n(x,\theta_n)]\, E_{P_0}[g_n(x,\theta_n)]'.
\end{aligned}$$

Thus, a lengthy but straightforward calculation combined with Lemma A.4, $\zeta_n = O\left(n^{-1/2}\right)$, and $\sup_{x\in\mathcal{X}} |\zeta_n' g_n(x,\theta_n)| = o(1)$ implies

$$nH\left(P_{\theta_n,\zeta_n},P_0\right)^2 = n\left\|\frac{1}{2}\zeta_n'\left(g_n(x,\theta_n) - E_{P_0}\left[g_n(x,\theta_n)\right]\right)dP_0^{1/2}\right\|^2 + o(1) \to \frac{1}{4}t'\Sigma^{-1}t. \tag{A.5}$$

Based on this limit, a lower bound of the maximum bias of T_a is obtained as (see, Rieder (1994, eq. (56) on p. 180))

$$\begin{aligned}
&\liminf_{n\to\infty} \sup_{Q\in B_H\left(P_0,r/\sqrt{n}\right)} n\left(\tau\circ T_a(Q) - \tau(\theta_0)\right)^2 \\
\geq\ & \liminf_{n\to\infty} \sup_{\left\{t\in\mathbb{R}^p: P_{\theta_n,\zeta_n}\in B_H\left(P_0,r/\sqrt{n}\right)\right\}} n\left(\tau\circ T_a\left(P_{\theta_n,\zeta_n}\right) - \tau(\theta_0)\right)^2 \\
\geq\ & \max_{\left\{t\in\mathbb{R}^p:\frac{1}{4}t'\Sigma t\leq r^2-\epsilon\right\}} \left(\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)'t\right)^2 = 4\left(r^2-\epsilon\right)\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)'\Sigma^{-1}\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right),
\end{aligned}$$

for each $\epsilon\in\left(0,r^2\right)$, where the first inequality follows from the set inclusion relationship, the second inequality follows from (3.1) and (A.5), and the equality follows from the Kuhn-Tucker theorem. Since ϵ can be arbitrarily small, we obtain the conclusion.

A.1.2.2. *Proof of (ii).* Pick arbitrary $r>0$ and sequence $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$. We first show the Fisher consistency of $\bar{T}$. From Lemma A.2 (note: $P_{\theta_n,\zeta_n}\in B_H\left(P_0,r/\sqrt{n}\right)$ for all n large enough),

$$\begin{aligned}
\sqrt{n}\left(\bar{T}\left(P_{\theta_n,\zeta_n}\right)-\theta_0\right) &= -\sqrt{n}\Sigma^{-1}\int\Lambda_n dP_{\theta_n,\zeta_n} + o(1) \\
&= \Sigma^{-1}G'\Omega^{-1}\int \partial g\left(x,\dot{\theta}\right)/\partial\theta dP_{\theta_n,\zeta_n}t + o(1) \\
&\to t
\end{aligned}$$

for all n large enough, where $\dot{\theta}$ is a point on the line joining θ_n and θ_0, the second equality follows from $\int g(x,\theta_0)\mathbb{I}\{x\notin\mathcal{X}_n\}dP_{\theta_n,\zeta_n} = o\left(n^{-1/2}\right)$ (by a similar argument to (A.16)), $\int g(x,\theta_n)dP_{\theta_n,\zeta_n} = 0$ (by $P_{\theta_n,\zeta_n}\in\mathcal{P}_{\theta_n}$), and an expansion around $\theta_n=\theta_0$, and the convergence follows from the last statement of Lemma A.4 (i). Therefore, $\bar{T}$ is Fisher consistent.

We next show (3.1). An expansion of $\tau\circ\bar{T}_{Q_n}$ around $\bar{T}_{Q_n}=\theta_0$, Lemmas A.1 (ii) and A.2, and Assumption 3.1 (viii) imply

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

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

$$\begin{aligned} & n\left(\tau \circ \bar{T}_{Q_n} - \tau(\theta_0)\right)^2 \\ \leq\ & n\left\{ \begin{array}{l} \left|\nu_0' \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dQ_n^{1/2}\right|^2 + \left|\nu_0' \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dP_0^{1/2}\right|^2 \\ +2\left|\nu_0' \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dQ_n^{1/2}\right| \left|\nu_0' \int \Lambda_n \left\{dQ_n^{1/2} - dP_0^{1/2}\right\} dP_0^{1/2}\right| \end{array} \right\} + o(1) \\ =\ & n\left\{A_1 + A_2 + 2A_3\right\} + o(1). \end{aligned}$$

For A_1, observe that

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

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from Lemma A.5 (i) and $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$. Similarly, we have $A_2 \leq B^* \frac{r^2}{n} + o\left(n^{-1}\right)$ and $A_3 \leq B^* \frac{r^2}{n} + o\left(n^{-1}\right)$. Combining these terms,

$$\limsup_{n\to\infty} n\left(\tau \circ \bar{T}_{Q_n} - \tau(\theta_0)\right)^2 \leq 4r^2 B^*, \tag{A.6}$$

for any sequence $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$ and $r > 0$. Pick any $r > 0$. Since the supremum $\sup_{Q\in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)} n\left(\tau \circ \bar{T}(Q) - \tau(\theta_0)\right)^2$ is finite for all n large enough (from Lemma A.1 (i)), there exists a sequence $Q_n^* \in B_H\left(P_0, r/\sqrt{n}\right)$ such that

$$\limsup_{n\to\infty} n\left(\tau \circ \bar{T}_{Q_n^*} - \tau(\theta_0)\right)^2 = \limsup_{n\to\infty} \sup_{Q\in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)} n\left(\tau \circ \bar{T}(Q) - \tau(\theta_0)\right)^2.$$

Therefore, the conclusion follows by (A.6).

A.1.3. **Proof of Theorem 3.2.**

A.1.3.1. *Proof of (i).* Pick arbitrary $\epsilon \in \left(0, r^2\right)$ and $r > 0$. Consider the parametric submodel P_{θ_n,ζ_n} defined in (A.4). The convolution theorem (Theorem 25.20 of van der Vaart (1998)) implies that for each $t \in \mathbb{R}^p$, there exists a probability measure M_0 which does not depend on t and satisfies

$$\sqrt{n}\left(\tau \circ T_a(P_n) - \tau \circ T_a\left(P_{\theta_n,\zeta_n}\right)\right) \xrightarrow{d} M_0 * N(0, B^*) \quad \text{under } P_{\theta_n,\zeta_n}. \tag{A.7}$$

Let

$$t^* = \arg\max_{\left\{t\in\mathbb{R}^p : \frac{1}{4}t'\Sigma t \leq r^2 - \epsilon\right\}} \left(\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)' t\right)^2 \quad \text{s.t.} \quad \left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)' t^* \int \xi dM_0 * N(0, B^*) \geq 0.$$

Since the integral $\int \xi dM_0 * N(0, B^*)$ does not depend on t, such t^* always exists. From $\frac{1}{4}t^{*\prime}\Sigma t^* \leq r^2 - \epsilon$ and (A.5), it holds that $P_{\theta_0+t^*/\sqrt{n},\zeta_n} \in B_H\left(P_0, r/\sqrt{n}\right)$ for all n large enough. Also, note that

$E_{P_{\theta_n,\zeta_n}}\left[\sup_{\theta\in\Theta}|g(x,\theta)|^{\eta}\right]<\infty$ for all n large enough (by $\sup_{x\in\mathcal{X}}|\zeta_n' g_n(x,\theta_n)|=o(1)$ and Assumption 3.1 (v)). Thus, $P_{\theta_0+t^*/\sqrt{n},\zeta_n}\in\bar{B}_H(P_0,r/\sqrt{n})$ for all n large enough, and we have

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

where the first equality follows from the Fisher consistency of T_a, (A.9), and the continuous mapping theorem, the second equality follows from the monotone convergence theorem, and the second inequality follows from the definition of t^*. Since ϵ can be arbitrarily small, we obtain the conclusion.

A.1.3.2. *Proof of (ii).* Pick arbitrary $r>0$ and $b>0$. Applying the inequality $b\wedge(c_1+c_2)\leq b\wedge c_1+b\wedge c_2$ for any $c_1,c_2\geq 0$,

$$\begin{aligned}
& \limsup_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int b\wedge n\left(\tau\circ T(P_n)-\tau(\theta_0)\right)^2 dQ^{\otimes n} \\
\leq\ & \limsup_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int b\wedge n\left(\tau\circ T(P_n)-\tau\circ\bar{T}(P_n)\right)^2 dQ^{\otimes n} \\
& +2\limsup_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int b\wedge\left\{n\left|\tau\circ T(P_n)-\tau\circ\bar{T}(P_n)\right|\left|\tau\circ\bar{T}(P_n)-\tau(\theta_0)\right|\right\}dQ^{\otimes n} \\
& +\limsup_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int b\wedge n\left(\tau\circ\bar{T}(P_n)-\tau(\theta_0)\right)^2 dQ^{\otimes n} \\
=\ & A_1+2A_2+A_3, \qquad \text{(A.8)}
\end{aligned}$$

For A_1,

$$\begin{aligned}
A_1 &\leq b\times\limsup_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int_{(x_1,\dots,x_n)\notin\mathcal{X}_n^n} dQ^{\otimes n} \\
&\leq b\times\limsup_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\sum_{i=1}^{n}\int_{x_i\notin\mathcal{X}_n} dQ \\
&\leq b\times\limsup_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})} n m_n^{-\eta}E_Q\left[\sup_{\theta\in\Theta}|g(x,\theta)|^{\eta}\right]=0, \qquad \text{(A.9)}
\end{aligned}$$

where the first inequality follows from $T(P_n) = \bar{T}(P_n)$ for all $(x_1, \ldots, x_n) \in \mathcal{X}_n^n$, the second inequality follows from a set inclusion relation, the third inequality follows from the Markov inequality, and the equality follows from Assumption 3.1 (vii) and $E_Q[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta] < \infty$ for all $Q \in \bar{B}_H(P_0, r/\sqrt{n})$. Similarly, we have $A_2 = 0$.

We now consider A_3. 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 $B_H(P_0, r/\sqrt{n})$ (not $\bar{B}_H(P_0, r/\sqrt{n})$) 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

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

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 first inequality follows from $\bar{B}_H(P_0, r/\sqrt{n}) \subseteq B_H(P_0, r/\sqrt{n})$, the second equality follows from Lemma A.8 (with $Q_n = \tilde{Q}_{b,n}$) and the continuous mapping theorem, the second inequality follows from $b \wedge c \leq c$ and a direct calculation, and the last inequality follows from Theorem 3.1 (ii). Combining these results, the conclusion is obtained.

A.1.4. **Proof of Theorem 3.3.**

A.1.4.1. *Proof of (i).* Consider the parametric submodel P_{θ_n,ζ_n} defined in (A.4). Since ℓ is uniformly continuous on $\bar{\mathbb{R}}^p$ (by Assumption 3.2) and T_a is Fisher consistent,

$$
b \wedge \ell\left(\sqrt{n}\left\{S_n - \tau \circ T_a\left(P_{\theta_n,\zeta_n}\right)\right\}\right) - b \wedge \ell\left(\sqrt{n}\left\{S_n - \tau(\theta_0)\right\} - \left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' t\right) \to 0,
$$

uniformly in t, $|t| < c$ and $\{S_n\}_{n\in\mathbb{N}}$ for each $c > 0$ and $b > 0$. Thus,

(A.10)
$$
\inf_{S_n\in\mathcal{S}} \sup_{|t|\leq c} \int b \wedge \ell\left(\sqrt{n}\left\{S_n - \tau \circ T_a\left(P_{\theta_n,\zeta_n}\right)\right\}\right) dP_{\theta_n,\zeta_n}^{\otimes n} = \inf_{R_n\in\mathcal{R}} \sup_{|t|\leq c} \int b \wedge \ell\left(R_n - \left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' t\right) dP_{\theta_n,\zeta_n}^{\otimes n} + o(1),
$$

for each $c > 0$, where $R_n = \sqrt{n}\{S_n - \tau(\theta_0)\}$ is a standardized estimator and $\mathcal{R} = \{\sqrt{n}\{S_n - \tau(\theta_0)\} : S_n \in \mathcal{S}\}$. By expanding the log likelihood ratio $\log \frac{dP^{\otimes n}_{\theta_n,\zeta_n}}{dP_0^{\otimes n}}$ around $\zeta_n = 0$,

$$\begin{aligned}
\log \frac{dP^{\otimes n}_{\theta_n,\zeta_n}}{dP_0^{\otimes n}} &= \zeta_n' \sum_{i=1}^n \{g_n(x_i,\theta_n) - E_{P_0}[g_n(x,\theta_n)]\} \\
&\quad - \frac{\zeta_n' \sum_{i=1}^n g_n(x_i,\theta_n) g_n(x_i,\theta_n) \zeta_n}{2\left(1 + \dot{\zeta}_n' g_n(x_i,\theta_n)\right)^2} + \frac{n\zeta_n' E_{P_0}[g_n(x,\theta_n)] E_{P_0}[g_n(x,\theta_n)]' \zeta_n}{2\left(1 + \ddot{\zeta}_n' \int g_n(x,\theta_n)\right)^2} \\
&= L_1 - L_2 + L_3.
\end{aligned}$$

where $\dot{\zeta}_n$ and $\ddot{\zeta}_n$ are points on the line joining ζ_n and 0. For L_1, an expansion of $g_n(x,\theta_n)$ (in ζ_n) around $\theta_n = \theta_0$ combined with Lemma A.4 (i) implies that under P_0,

$$L_1 = -t' G' \Omega^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g_n(x_i,\theta_n) - E_{P_0}[g_n(x,\theta_n)]\} + o_p(1).$$

Also, Lemma A.4 (i) and $\sup_{x\in\mathcal{X}} |\zeta_n' g_n(x,\theta_n)| = o(1)$ imply that under P_0,

$$L_2 \xrightarrow{p} \frac{1}{2} t' \Sigma t, \quad L_3 \to 0.$$

Therefore, in the terminology of Rieder (1994, Definition 2.2.9), the parametric model P_{θ_n,ζ_n} is asymptotically normal with the asymptotic sufficient statistic $-G'\Omega^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \{g_n(x_i,\theta_n) - E_{P_0}[g_n(x,\theta_n)]\}$ and the asymptotic covariance matrix Σ. Note that this is essentially the LAN (local asymptotic normality) condition introduced by LeCam. If P_{θ_n,ζ_n} is asymptotically normal in this sense, we can directly apply the result of the minimax risk bound by Rieder (1994, Theorem 3.3.8 (a)), that is

$$\text{(A.11)} \qquad \lim_{b\to\infty} \lim_{c\to\infty} \liminf_{n\to\infty} \inf_{S_n \in \mathcal{S}} \sup_{|t|\le c} \int b \wedge \ell \left(R_n - \left(\frac{\partial \tau(\theta_0)}{\partial \theta} \right)' t \right) dP^{\otimes n}_{\theta_n,\zeta_n} \ge \int \ell dN(0, B^*)$$

(see also Theorem 1 in LeCam and Yang (1990)). From (A.10) and (A.11),

$$\lim_{b\to\infty} \lim_{c\to\infty} \liminf_{n\to\infty} \inf_{S_n \in \mathcal{S}} \sup_{|t|\le c} \int b \wedge \ell \left(\sqrt{n} \{S_n - \tau \circ T_a(P_{\theta_n,\zeta_n})\} \right) dP^{\otimes n}_{\theta_n,\zeta_n} \ge \int \ell dN(0, B^*).$$

Finally, since $E_{P_{\theta_n,\zeta_n}}[\sup_{\theta\in\Theta} |g(x,\theta)|^\eta] < \infty$ for all n large enough (by $\sup_{x\in\mathcal{X}} |\zeta_n' g_n(x,\theta_n)| = o(1)$ and Assumption 3.1 (v)), we have $P_{\theta_n,\zeta_n} \in \bar{B}_H(P_0, r/\sqrt{n})$ for all t satisfying $\frac{1}{4} t' \Sigma t \le r^2 - \epsilon$ with any $\epsilon \in (0, r^2)$ and all n large enough. Therefore, the set inclusion relation yields

$$\begin{aligned}
&\lim_{b\to\infty} \lim_{r\to\infty} \liminf_{n\to\infty} \inf_{S_n \in \mathcal{S}} \sup_{Q \in \bar{B}_H(P_0, r/\sqrt{n})} \int b \wedge \ell \left(\sqrt{n} \{S_n - \tau \circ T_a(Q)\} \right) dQ^{\otimes n} \\
\ge\ &\lim_{b\to\infty} \lim_{c\to\infty} \liminf_{n\to\infty} \inf_{S_n \in \mathcal{S}} \sup_{|t|\le c} \int b \wedge \ell \left(\sqrt{n} \{S_n - \tau \circ T_a(P_{\theta_n,\zeta_n})\} \right) dP^{\otimes n}_{\theta_n,\zeta_n},
\end{aligned}$$

which implies the conclusion.

A.1.4.2. *Proof of (ii).* Pick arbitrary $r>0$ and $b>0$. Since $T(P_n)=\bar{T}(P_n)$ for all $(x_1,\ldots,x_n)\in\mathcal{X}_n^n$,

$$\begin{aligned}
&\lim_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int b\wedge\ell\left(\sqrt{n}\left\{\tau\circ T(P_n)-\tau\circ\bar{T}(Q)\right\}\right)dQ^{\otimes n}\\
\leq\ &\lim_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int_{(x_1,\ldots,x_n)\notin\mathcal{X}_n^n} b\wedge\ell\left(\sqrt{n}\left\{\tau\circ T(P_n)-\tau\circ\bar{T}(Q)\right\}\right)dQ^{\otimes n}\\
&+\lim_{n\to\infty}\sup_{Q\in\bar{B}_H(P_0,r/\sqrt{n})}\int_{(x_1,\ldots,x_n)\in\mathcal{X}_n^n} b\wedge\ell\left(\sqrt{n}\left\{\tau\circ\bar{T}(P_n)-\tau\circ\bar{T}(Q)\right\}\right)dQ^{\otimes n}.
\end{aligned}\tag{A.12}$$

An argument similar to (A.9) implies that the first term of (A.12) is zero. From $\mathcal{X}_n^n\subseteq\mathcal{X}^n$ and $\bar{B}_H(P_0,r/\sqrt{n})\subseteq B_H(P_0,r/\sqrt{n})$, the second term of (A.12) is bounded from above by

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

where the equality follows from Lemma A.8, the uniform continuity of ℓ over $\bar{\mathbb{R}}^p$, and compactness of $B_H(P_0,r/\sqrt{n})$ under the Hellinger distance. Let $b\to\infty$ and the conclusion follows.

A.2 Auxiliary Lemmas

Lemma A.1. *Suppose that Assumption 3.1 holds. Then*

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

Proof of (i). The proof is split into several steps. Let $\mathcal{G}(\theta,Q)$ be the convex hull of the support of $g(x,\theta)$ under $x\sim Q$.

In the first step, we show $0\in\text{int}\mathcal{G}(\theta_0,P_0)$. If $0\notin\mathcal{G}(\theta_0,P_0)$, then we have $E_{P_0}[g(x,\theta_0)]\neq 0$, which is a contradiction. Thus, it is enough to show that 0 is not on the boundary of $\mathcal{G}(\theta_0,P_0)$. Suppose 0 is indeed on the boundary of $\mathcal{G}(\theta_0,P_0)$. In this case, we have two cases: (a) there exists a constant m-vector $a\neq 0$ such that $a'g\geq 0$ for all $g\in\mathcal{G}(\theta_0,P_0)$ and $P_0\{g\in\mathcal{G}(\theta_0,P_0):a'g>0\}>0$, or (b) there exists $a\neq 0$ such that $a'g=0$ for all $g\in\mathcal{G}(\theta_0,P_0)$. For the case (a), we have $a'E_{P_0}[g(x,\theta_0)]>0$, which contradicts with $E_{P_0}[g(x,\theta_0)]=0$. For the case (b), we have $a'E_{P_0}\left[g(x,\theta_0)g(x,\theta_0)'\right]a=0$, which contradicts with Assumption 3.1 (vi).

In the second step, we show that for each $r>0$, there exists $\delta>0$ such that $0\in\text{int}\mathcal{G}(\theta,Q)$ for all $|\theta-\theta_0|\leq\delta$ and all $Q\in B_H(P_0,\delta)$. Pick any $r>0$. From the first step, we can find $m+1$ points $\{\tilde{g}_1,\ldots,\tilde{g}_{m+1}\}=\{g(\tilde{x}_1,\theta_0),\ldots,g(\tilde{x}_{m+1},\theta_0)\}$ in the support of $g(x,\theta_0)$ under $x\sim P_0$ such that 0 is interior of the convex hull of $\{\tilde{g}_1,\ldots,\tilde{g}_{m+1}\}$. From the property of the convex hull (Rockafeller, 1970, Corollary 2.3.1), we can take $c_r>0$ such that for any points $\{g_1,\ldots,g_{m+1}\}$ satisfying $|g_j-\tilde{g}_j|\leq c_r$

for $j = 1, \ldots, m+1$, the interior of the convex hull of $\{g_1, \ldots, g_{m+1}\}$ contains 0. Let us take any $j = 1, \ldots, m+1$. For the second step, it is sufficient to show that there exists $\delta_j > 0$ such that $Q\{|g(x,\theta) - \tilde{g}_j| \leq c_r\} > 0$ for all $|\theta - \theta_0| \leq \delta_j$ and all $Q \in B_H(P_0, \delta_j)$. Suppose this is false, i.e., for any $\delta_j > 0$, we can take a pair (Q_j, θ_j) such that $H(Q_j, P_0) \leq \delta_j$, $|\theta_j - \theta_0| \leq \delta_j$, and $Q_j\{|g(x,\theta_j) - \tilde{g}_j| \leq c_r\} = 0$. Then we have

$$\delta_j \geq H(Q_j, P_0) \geq \sqrt{\int_{\{|g(x,\theta_j)-\tilde{g}_j| \leq c_r\}} \left(\sqrt{dQ_j} - \sqrt{dP_0}\right)^2} = \sqrt{P_0\{|g(x,\theta_j) - \tilde{g}_j| \leq c_r\}}.$$

On the other hand, by Assumption 3.1 (iv), the dominated convergence theorem guarantees

$$P_0\{|g(x,\theta_j) - \tilde{g}_j| \leq c_r\} \to P_0\{|g(x,\theta_0) - \tilde{g}_j| \leq c_r\} > 0 \qquad \text{as } \theta_j \to \theta_0.$$

Since δ_j can be arbitrarily small, we have a contradiction. This completes the second step.

In the third step, we show that for each $r > 0$, there exists $\delta > 0$ such that $R_n(\theta, Q) = \inf_{P \in \bar{\mathcal{P}}_\theta, P \ll Q} H(P,Q)$ has a minimum on $\{\theta \in \Theta : |\theta - \theta_0| \leq \delta\}$ for all $Q \in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)$ and all n large enough. Let us take $\delta > 0$ to satisfy the conclusion of the second step. By Assumption 3.1 (iv), we can take N_δ to satisfy $\max_{1 \leq j \leq m+1} \sup_{\theta \in \Theta, |\theta - \theta_0| \leq \delta} |g(\tilde{x}_j, \theta)| \leq m_{N_\delta}$. Thus, letting $\mathcal{G}_n(\theta, Q)$ be the convex hull of the support of $g_n(x,\theta)$ under $x \sim Q$, the second step also guarantees that for each $r > 0$, there exists $\delta > 0$ such that $0 \in \text{int}\mathcal{G}_n(\theta, Q)$ for all $|\theta - \theta_0| \leq \delta$, all $Q \in B_H(P_0, \delta)$, and all $n \geq N_\delta$. Based on this, the convex duality result in Borwein and Lewis (1993, Theorem 3.4) implies $R_n(\theta, Q) = \sup_{\gamma \in \mathbb{R}^m} -\int \frac{1}{(1+\gamma' g_n(x,\theta))} dQ$ for all $|\theta - \theta_0| \leq \delta$, all $Q \in B_H(P_0, \delta)$, and all $n \geq N_\delta$. Since $\sup_{\gamma \in \mathbb{R}^m} -\int \frac{1}{(1+\gamma' g_n(x,\theta))} dQ$ is continuous at all θ with $|\theta - \theta_0| \leq \delta$ (by the maximum theorem), the Weierstrass theorem completes the third step.

Finally, based on the third step, it is sufficient for the conclusion to show that for every $r > 0$, there exists $N \in \mathbb{N}$ such that $R_n(\theta_0, Q) < \inf_{\theta \in \Theta : |\theta - \theta_0| > \delta} R_n(\theta, Q)$ for all $n \geq N$ and all $Q \in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)$. Pick any $r > 0$. We first derive an upper bound of $R_n(\theta_0, Q) = \sup_{\gamma \in \mathbb{R}^m} -\int \frac{1}{(1+\gamma' g_n(x,\theta_0))} dQ$. From Lemma A.5 (ii), $\gamma_n(\theta_0, Q) = \arg\max_{\gamma \in \mathbb{R}^m} -\int \frac{1}{(1+\gamma' g_n(x,\theta_0))} dQ$ exists and $\sup_{x \in \mathcal{X}} \left|\gamma_n(\theta_0, Q)' g_n(x, \theta_0)\right| \leq \frac{1}{2}$ for all n large enough and all $Q \in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)$. Thus, by a second-order expansion around $\gamma_n(\theta_0, Q) = 0$, we have

$$R_n(\theta_0, Q) \quad \leq \quad -1 + \gamma_n(\theta_0, Q)' \int g_n(x, \theta_0)\, dQ.$$

Define $C^* = \inf_{\theta \in \Theta : |\theta - \theta_0| > \delta} |E_{P_0}[g(x,\theta)]|^2 / (1 + |E_{P_0}[g(x,\theta)]|) > 0$. From Lemma A.5 and $m_n n^{-1/2} \to 0$, it holds

$$m_n(R(\theta_0, Q) + 1) \leq m_n \left|\gamma_n(\theta_0, Q)' \int g_n(x, \theta_0)\, dQ\right| < \frac{C^*}{4}, \tag{A.13}$$

for all n large enough and all $Q \in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)$. We now derive a lower bound of $R_n(\theta, Q)$ with $|\theta - \theta_0| > \delta$. Pick any $\theta \in \Theta$ such that $|\theta - \theta_0| > \delta$, and take any n large enough and $Q \in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)$ to satisfy (A.13). If $0 \notin \mathcal{G}_n(\theta, Q)$, then $R_n(\theta, Q) = +\infty$. Thus, we concentrate on the case of $0 \in \mathcal{G}_n(\theta, Q)$, which guarantees $R_n(\theta, Q) = \sup_{\gamma \in \mathbb{R}^m} -\int \frac{1}{(1+\gamma' g_n(x,\theta))} dQ$ (Borwein and Lewis, 1993, Theorem 3.4). Let $\gamma_0(\theta) = E_{P_0}[g(x,\theta)] / (1 + |E_{P_0}[g(x,\theta)]|)$. Obseve that

$$\begin{aligned} R_n(\theta, Q) &\geq -\int \frac{1}{\left(1 + m_n^{-1}\gamma_0(\theta)' g_n(x,\theta)\right)} dQ \\ &= -1 + m_n^{-1}\gamma_0(\theta)' \int g_n(x,\theta)\, dQ - m_n^{-2} \int \frac{\left(\gamma_0(\theta)' g_n(x,\theta)\right)^2}{\left(1 + t(x)\, m_n^{-1}\gamma_0(\theta)' g_n(x,\theta)\right)^3} dQ, \end{aligned}$$

where the second equality follows from an expanison ($t(x) \in (0,1)$ for almost evevry x under Q). From a similar argument to Lemma A.5 with $\sup_{\theta\in\Theta} |\gamma_0(\theta)| \leq 1$ and $m_n \to \infty$,

$$\sup_{\theta\in\Theta} \left| \int g_n(x,\theta)\, dQ - \int g(x,\theta)\, dP_0 \right| \leq \frac{C^*}{4}, \qquad m_n^{-1} \sup_{\theta\in\Theta} \left| \int \frac{\left(\gamma_0(\theta)' g_n(x,\theta)\right)^2}{\left(1 + t_1(x)\, m_n^{-1}\gamma_0(\theta)' g_n(x,\theta)\right)^3} dQ \right| \leq \frac{C^*}{4},$$

for all n large enough and all $Q \in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)$. Combining these results and using the definition of C^*, we obtain

$$\inf_{\theta\in\Theta: |\theta-\theta_0|>\delta} m_n\left(R_n(\theta, Q) + 1\right) \geq \frac{C^*}{2}, \tag{A.14}$$

for all n large enough and all $Q \in B_H\left(P_0, \frac{r}{\sqrt{n}}\right)$. Therefore, (A.13) and (A.14) complete the proof of the final step.

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

(A.15)

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

The first term of (A.15) satisfies

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

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 (A.15) satisfies

$$
\begin{aligned}
& \sup_{\theta\in\Theta} \left| E_{P_0}\left[g(x,\theta)\, \mathbb{I}\{x \notin \mathcal{X}_n\}\right]\right| \\
\leq & \left(\int \sup_{\theta\in\Theta} |g(x,\theta)|^{\eta} dP_0 \right)^{1/\eta} \left(\int \mathbb{I}\{x\notin \mathcal{X}_n\} dP_0 \right)^{(\eta-1)/\eta} \\
\leq & \left(E_{P_0}\left[\sup_{\theta\in\Theta} |g(x,\theta)|^{\eta}\right]\right)^{1/\eta} \left(m_n^{-\eta} E_{P_0}\left[\sup_{\theta\in\Theta} |g(x,\theta)|^{\eta}\right]\right)^{(\eta-1)/\eta} = o\left(n^{-1/2}\right),
\end{aligned} \tag{A.16}
$$

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 $\sup_{\theta\in\Theta} |E_{Q_n}[g_n(x,\theta)] - E_{P_0}[g(x,\theta)]| \to 0$. Therefore, from the triangle inequality and $\left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)\right]\right| = O\left(n^{-1/2}\right)$ (Lemma A.6 (i)),

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

The conclusion follows from Assumption 3.1 (iii).

Lemma A.2. *Suppose that Assumption 3.1 holds. Then for each $r>0$ and sequence $Q_n \in B_H(P_0, r/\sqrt{n})$,*

$$
\sqrt{n}\left(\bar{T}_{Q_n} - \theta_0\right) = -\sqrt{n}\Sigma^{-1}\int \Lambda_n dQ_n + o(1). \tag{A.17}
$$

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

$$
\begin{aligned}
& \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0\right)' \Lambda_n dQ_n^{1/2} \right\|^2 \\
= & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\|^2 + \left\| \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right)' \Lambda_n dQ_n^{1/2} \right\|^2 \\
& + \left\{ \int \left(dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right) \Lambda_n' dQ_n^{1/2} \right\} \left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right) \\
= & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\|^2 + \left\| \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right)' \Lambda_n dQ_n^{1/2} \right\|^2,
\end{aligned} \tag{A.18}
$$

where the second equality follows from

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

The left hand side of (A.18) satisfies

$$
\begin{aligned}
& \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\left(\bar{T}_{Q_n} - \theta_0\right)' \Lambda_n dQ_n^{1/2} \right\| \\
\leq\ & \left\| dQ_n^{1/2} - d\bar{P}_{\bar{T}_{Q_n},Q_n}^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(n^{-1/2}\right) \\
\leq\ & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0+\psi_{n,Q_n},Q_n}^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(n^{-1/2}\right) \\
\leq\ & \left\| dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2} + \frac{1}{2}\psi_{n,Q_n}' \Lambda_n dQ_n^{1/2} \right\| + o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(\left|\psi_{n,Q_n}\right|\right) + o\left(n^{-1/2}\right),
\end{aligned} \tag{A.19}
$$

where the first inequality follows from the triangle inequality and Lemma A.3 (i), the second inequality follows from $\bar{T}_{Q_n} = \arg\min_{\theta\in\Theta} \left\| dQ_n^{1/2} - d\bar{P}_{\theta,Q_n}^{1/2} \right\|$, and the third inequality follows from the triangle inequality and Lemma A.3 (ii). From (A.18) and (A.19),

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

This implies

$$
\begin{aligned}
& o\left(\left|\bar{T}_{Q_n} - \theta_0\right|\right) + o\left(\left|\psi_{n,Q_n}\right|\right) + o\left(n^{-1/2}\right) \\
\geq\ & \sqrt{\frac{1}{4}\left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right)' \int \Lambda_n \Lambda_n' dQ_n \left(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right)} \geq C\left|\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}\right|,
\end{aligned} \tag{A.20}
$$

for all n large enough, where the second inequality follows from Lemma A.5 (i) and Assumption 3.1 (vi).

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

$$
\begin{aligned}
\psi_{n,Q_n} &= -2\left\{\left(\int \Lambda_n \Lambda_n' dQ_n\right)^{-1} - \Sigma^{-1}\right\} \int \Lambda_n \left\{dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2}\right\} dQ_n^{1/2} \\
& \quad -2\Sigma^{-1} \int \Lambda_n \left\{dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2}\right\} dQ_n^{1/2}.
\end{aligned} \tag{A.21}
$$

From this and Lemma A.5 (i), the first term of (A.21) is $o\left(n^{-1/2}\right)$. The second term of (A.21) satisfies

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

where the first equality follows from (A.22), the second equality follows from (A.23) and Lemma A.5, and the third equality follows from Lemma A.5. Therefore,

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

which also implies $|\psi_{n,Q_n}|=O\left(n^{-1/2}\right)$ (by Lemma A.5 (i)). Combining this with (A.20),

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

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

Lemma A.3. *Suppose that Assumption 3.1 holds. Then for each $r>0$ and sequence $Q_n\in B_H\left(P_0,r/\sqrt{n}\right)$,*

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

Proof of (i). 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

$$\frac{d\bar{P}_{\theta,Q}}{dQ}=\frac{1}{\left(1+\gamma_n\left(\theta,Q\right)'g_n\left(x,\theta\right)\right)^2}, \tag{A.22}$$

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

$$0=\int\frac{g_n\left(x,\theta\right)}{\left(1+\gamma_n\left(\theta,Q\right)'g_n\left(x,\theta\right)\right)^2}dQ=E_Q\left[g_n\left(x,\theta\right)\left\{1-2\gamma_n\left(\theta,Q\right)'g_n\left(x,\theta\right)+\varrho_n\left(x,\theta,Q\right)\right\}\right], \tag{A.23}$$

with

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

Denote $t_n = \bar{T}_{Q_n} - \theta_0$. Pick arbitrary $r > 0$ and sequence $Q_n \in B_H(P_0, r/\sqrt{n})$. From the triangle inequality and (A.22),

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

For T_2, Lemmas A.5 and A.6 imply $T_2 = o\left(n^{-1/2}\right)$. For T_1, the triangle inequality and (A.23) yield

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

Lemmas A.5 and A.6 imply that $T_{12} = o\left(n^{-1/2}\right)$ and $T_{13} = o\left(n^{-1/2}\right)$. For T_{11}, expansions of $g_n\left(x, \bar{T}_{Q_n}\right)$ around $\bar{T}_{Q_n} = \theta_0$ yield

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

where $\dot{\theta}$ is a point on the line joining θ_0 and $\bar{T}_{Q_n}$, and the equality follows from Lemmas A.5 (i) and A.6 (i).

Proof of (ii). Similar to the proof of Part (i) of this lemma.

Lemma A.4. *Suppose that Assumption 3.1 hold. Then for each* $t \in \mathbb{R}^p$*,*

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

Proof of (i). Proof of the first statement. The same argument as (A.16) with Assumption 3.1 (iii) yields the conclusion.

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

$$|E_{P_0}[g_n(x,\theta_n)]| \le |E_{P_0}[g(x,\theta_n)\mathbb{I}\{x \notin \mathcal{X}_n\}]| + |E_{P_0}[g(x,\theta_n)]|. \tag{A.24}$$

By the same argument as (A.16) and $E_{P_0}[|g(x,\theta_n)|^\eta] < \infty$ (from Assumption 3.1 (v)), the first term of (A.24) is $o(n^{-1/2})$. The second term of (A.24) 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). Combining these results, the conclusion is obtained.

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

$$\begin{aligned} & \left|E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right] - \Omega\right| \\ \le\ & \left|E_{P_0}\left[g_n(x,\theta_n) g_n(x,\theta_n)'\right] - E_{P_0}\left[g(x,\theta_n) g(x,\theta_n)'\right]\right| + \left|E_{P_0}\left[g(x,\theta_n) g(x,\theta_n)'\right] - \Omega\right|. \end{aligned}$$

The first term is $o(n^{-1/2})$ by the same argument as (A.16) and the second term converges to zero by the continuity of $g(x,\theta)$ at θ_0.

Proof of the fourth statement. Similar to the proof of the third statement.

Proof of (ii). Pick an arbitrary $t \in \mathbb{R}^p$. Let $\Gamma_n = \{\gamma \in \mathbb{R}^m : |\gamma| \le a_n\}$ with a positive sequence $\{a_n\}_{n\in\mathbb{N}}$ satisfying $a_n m_n \to 0$ and $a_n n^{1/2} \to \infty$. Observe that

$$\sup_{\gamma\in\Gamma_n, x\in\mathcal{X}, \theta\in\Theta} \left|\gamma' g_n(x,\theta)\right| \le a_n m_n \to 0. \tag{A.25}$$

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

$$\begin{aligned} -1 &= R_n(P_0, \theta_n, 0) \leq R_n(P_0, \theta_n, \tilde{\gamma}) = -1 + \tilde{\gamma}' E_{P_0}[g_n(x, \theta_n)] - \tilde{\gamma}' E_{P_0}\left[\frac{g_n(x, \theta_n) g_n(x, \theta_n)'}{(1 + \dot{\gamma}' g_n(x, \theta_n))^3}\right]\tilde{\gamma} \\ &\leq -1 + \tilde{\gamma}' E_{P_0}[g_n(x, \theta_n)] - C\tilde{\gamma}' E_{P_0}\left[g_n(x, \theta_n) g_n(x, \theta_n)'\right]\tilde{\gamma} \\ &\leq -1 + |\tilde{\gamma}| \left|E_{P_0}[g_n(x, \theta_n)]\right| - C|\tilde{\gamma}|^2, \qquad \text{(A.26)} \end{aligned}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, the second inequality follows from (A.25), and the last inequality follows from Lemma A.4 (i) and Assumption 3.1 (vi). Thus, Lemma A.4 (i) implies

$$C|\tilde{\gamma}| \leq \left|E_{P_0}[g_n(x, \theta_n)]\right| = O\left(n^{-1/2}\right). \qquad \text{(A.27)}$$

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 and the first statement is obtained. Thus, the second statement is obtained from (A.27). The third statement follows from (A.27) and Assumption 3.1 (vii).

Lemma A.5. *Suppose that Assumption 3.1 holds. Then for each $r > 0$ and sequence $Q_n \in B_H(P_0, r/\sqrt{n})$,*

- **(i):** $\left|E_{Q_n}[g_n(x, \theta_0)]\right| = O\left(n^{-1/2}\right)$, *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* $\sup_{x \in \mathcal{X}} \left|\gamma_n(\theta_0, Q_n)' g_n(x, \theta_0)\right| = o(1)$.

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

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

where the first and second inequalities follow from the triangle inequality and Lemma A.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 arbitrary $r > 0$ and sequence $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$. From the triangle inequality,

$$\begin{aligned}
&\left|E_{Q_n}\left[g_n\left(x,\theta_0\right) g_n\left(x,\theta_0\right)'\right] - \Omega\right| \\
\leq\; & \left|E_{Q_n}\left[g_n\left(x,\theta_0\right) g_n\left(x,\theta_0\right)'\right] - E_{P_0}\left[g_n\left(x,\theta_0\right) g_n\left(x,\theta_0\right)'\right]\right| + \left|E_{P_0}\left[g\left(x,\theta_0\right) g\left(x,\theta_0\right)' \mathbb{I}\left\{x \notin \mathcal{X}_n\right\}\right]\right|.
\end{aligned} \tag{A.28}$$

The first term of the RHS of (A.28) satisfies

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

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality and $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$, and the equality follows from Assumption 3.1 (v) and (vii). The second term of (A.28) satisfies

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

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 (vii).

Proof of (ii). Similar to the proof of Lemma A.4 (ii). Repeat the same argument with $R_n\left(Q_n, \theta_0, \gamma\right)$ instead of $R_n\left(P_0, \theta_n, \gamma\right)$.

Lemma A.6. *Suppose that Assumption 3.1 holds. Then for each* $r > 0$ *and sequence* $Q_n \in B_H\left(P_0, r/\sqrt{n}\right)$,

- **(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* $\sup_{x\in\mathcal{X}} \left|\gamma_n\left(\bar{T}_{Q_n}, Q_n\right)' g_n\left(x, \bar{T}_{Q_n}\right)\right| = o\left(1\right)$.

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

$$\sup_{x\in\mathcal{X},\theta\in\Theta} \left|\tilde{\gamma}' g_n\left(x,\theta\right)\right| \leq n^{-1/2} m_n \to 0. \tag{A.29}$$

Observe that

(A.30)

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

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, an expansion around $\tilde{\gamma}=0$ yields

$$\begin{aligned}
R_n\left(Q_n,\bar{T}_{Q_n},\tilde{\gamma}\right) &= -1+\tilde{\gamma}'E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)\right]-\tilde{\gamma}'E_{Q_n}\left[\frac{g_n\left(x,\bar{T}_{Q_n}\right)g_n\left(x,\bar{T}_{Q_n}\right)'}{\left(1+\dot{\gamma}'g_n\left(x,\bar{T}_{Q_n}\right)\right)^3}\right]\tilde{\gamma} \\
&\geq -1+n^{-1/2}\left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)\right]\right|-C\tilde{\gamma}'E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)g_n\left(x,\bar{T}_{Q_n}\right)'\right]\tilde{\gamma} \\
&\geq -1+n^{-1/2}\left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)\right]\right|-Cn^{-1}, \qquad \text{(A.31)}
\end{aligned}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, the first inequality follows from (A.29), and the second inequality follows from $\tilde{\gamma}'\tilde{\gamma}=n^{-1}$ and (A.30). 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 $\gamma_n\left(\bar{T}_{Q_n},Q_n\right)=\arg\max_{\gamma\in\mathbb{R}^m}R_n\left(Q_n,\bar{T}_{Q_n},\gamma\right)$ and $\bar{T}_{Q_n}=\arg\min_{\theta\in\Theta}R_n\left(Q_n,\theta,\gamma_n\left(\theta,Q_n\right)\right)$. Therefore, from (A.31),

$$\begin{aligned}
&-1+n^{-1/2}\left|E_{Q_n}\left[g_n\left(x,\bar{T}_{Q_n}\right)\right]\right|-Cn^{-1} \\
\leq\ &R_n\left(Q_n,\bar{T}_{Q_n},\tilde{\gamma}\right)\leq R_n\left(Q_n,\bar{T}_{Q_n},\gamma_n\left(\bar{T}_{Q_n},Q_n\right)\right)\leq R_n\left(Q_n,\theta_0,\gamma_n\left(\theta_0,Q_n\right)\right). \qquad \text{(A.32)}
\end{aligned}$$

By a similar argument to (A.26) combined with $\left|\gamma_n\left(\theta_0,Q_n\right)\right|=O\left(n^{-1/2}\right)$ and $\left|E_{Q_n}\left[g_n\left(x,\theta_0\right)\right]\right|=O\left(n^{-1/2}\right)$ (by Lemma A.5), we have

(A.33)

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

From (A.32) and (A.33), the conclusion follows.

Proof of the second statement. Similar to the proof of the second statement of Lemma A.5 (i).

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

$$\begin{aligned} \left|E_{Q_n}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]-G\right| &\leq \left|E_{Q_n}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]-E_{P_0}\left[\partial g_n\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]\right| \\ &\quad +\left|E_{P_0}\left[\mathbb{I}\left\{x\notin\mathcal{X}_n\right\}\partial g\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]\right|+\left|E_{P_0}\left[\partial g\left(x,\bar{T}_{Q_n}\right)/\partial\theta'\right]-G\right|. \end{aligned} \tag{A.34}$$

The first term of (A.34) satisfies

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

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality, and the equality follows from Assumption 3.1 (v) and (vii). The second term of (A.34) is $o(1)$ by the same argument as (A.16). The third term of (A.34) is $o(1)$ by the continuity of $\partial g(x,\theta)/\partial\theta'$ at θ_0 and Lemma A.1 (ii). Therefore, the conclusion is obtained.

Proof of (ii). Similar to the proof of Lemma A.4 (ii). Repeat the same argument with $R_n\left(Q_n,\bar{T}_{Q_n},\gamma\right)$ instead of $R_n(P_0,\theta_n,\gamma)$.

Lemma A.7. *Suppose that Assumption 3.1 holds. Then for each sequence $Q_n \in B_H(P_0, r/\sqrt{n})$ and $r>0$, $\bar{T}_{P_n} \xrightarrow{p} \theta_0$ under Q_n.*

Proof. Similar to the proof of Lemma A.1 (i).

Lemma A.8. *Suppose that Assumption 3.1 holds . Then for each $r>0$ and sequence $Q_n \in B_H(P_0, r/\sqrt{n})$,*

$$\begin{aligned} \sqrt{n}\left(\bar{T}_{P_n}-\theta_0\right) &= -\sqrt{n}\Sigma^{-1}\int\Lambda_n dP_n+o_p(1) \quad \textit{under } Q_n, \\ \sqrt{n}\left(\bar{T}_{P_n}-\bar{T}_{Q_n}\right) &\xrightarrow{d} N\left(0,\Sigma^{-1}\right) \quad \textit{under } Q_n. \end{aligned}$$

Proof. The proof of the first statement is similar to that of Lemma A.2 (replace Q_n with P_n and use Lemmas A.9 and A.10 instead of Lemmas A.5 and A.6). For the second statement, Lemma A.2 and the first statement imply

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

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

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

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 A.9. *Suppose that Assumption 3.1 holds. Then for each $r>0$ and sequence $Q_n\in B_H(P_0,r/\sqrt{n})$, the followings hold under Q_n:*

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

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

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

The first term is $O_p\left(n^{-1/2}\right)$ by the central limit theorem for the triangular array $\{g_n(x_i,\theta_0)\}_{1\le i\le n,n}$. The second term is $O\left(n^{-1/2}\right)$ by Lemma A.5 (i).

Proof of the second statement. From the triangle inequality,

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

From a law of large numbers, the first term is $o_p(1)$. From Lemma A.5 (i), the second term is $o(1)$.

Proof of (ii). Similar to the proof of Lemma A.4 (ii) except using Lemma A.9 (i) instead of Lemma A.4 (i).

Lemma A.10. *Suppose that Assumption 3.1 holds. Then for each $r>0$ and sequence $Q_n\in B_H(P_0,r/\sqrt{n})$, 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* $\left|E_{P_n}\left[\partial g_n\left(x,\bar{T}_{P_n}\right)/\partial\theta'\right]-G\right|=o_p(1)$,

(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* $\sup_{x\in\mathcal{X}}\left|\gamma_n\left(\bar{T}_{P_n}, P_n\right)' g_n\left(x, \bar{T}_{P_n}\right)\right| = o_p(1)$.

Proof of (i). Similar to the proof of Lemma A.6 (i).

Proof of (ii). Similar to the proof of Lemma A.6 (ii).

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, Princeton University, Princeton, NJ-08544.

E-mail address: `kevdokim@princeton.edu`