ROBUST ESTIMATION OF MOMENT CONDITION MODELS WITH WEAKLY DEPENDENT DATA

KIRILL EVDOKIMOV, YUICHI KITAMURA, AND TAISUKE OTSU

Abstract. This paper considers robust estimation of moment condition models with time series data. Researchers frequently use moment condition models in dynamic econometric analysis. They are particularly useful when one wishes to avoid fully parameterizing the dynamics in the data. It is nevertheless desirable to use an estimation method that is robust against deviations from the model assumptions. For example, measurement errors can contaminate observations and thereby lead to such deviations. This is an important issue for time series data: in addition to conventional sources of mismeasurement, it is known that an inappropriate treatment of seasonality can cause serially correlated measurement errors. Efficiency is also a critical issue since time series sample sizes are often limited. This paper addresses these problems. Our estimator has three features: (i) it achieves an asymptotic optimal robust property, (ii) it treats time series dependence nonparametrically by a data blocking technique, and (iii) it is asymptotically as efficient as the optimally weighted GMM if indeed the model assumptions hold. A small scale simulation experiment suggests that our estimator performs favorably compared to other estimators including GMM, thereby supporting our theoretical findings.

1. Introduction

It is a common practice in empirical economics to estimate a dynamic economic model based on moment restrictions it implies. Moment condition-based estimation is often computationally convenient; the GMM estimator Hansen (1982) is a prime example. It is argued that a moment condition model imposes only mild assumptions and thereby enabling the researcher to conduct robust analysis, especially when economic theory provides little guidance for dynamic specifications. Also, GMM is generally regarded as a robust procedure. The last notion, however, deserves further investigation.

Date: This Version: November 23, 2009.

Keywords: Blocking, Generalized Empirical Likelihood, Hellinger Distance, Strong Mixing, Robustness.

JEL Classification Number: C14.

The research of the first author is supported by the Carl Arvid Anderson Fellowship. The research of the second author is supported by the National Science Foundation grants SES-0551271 and SES-0851759. The research of the third author is supported by the National Science Foundation grant SES-0720961.

Indeed, this paper demonstrates that an alternative estimator, which is termed the blockwise minimum Hellinger distance estimator (the blockwise MHDE), possesses a desirable robust property. The GMM estimator does not share this property, and our experimental result indicates that the latter can be sensitive to deviations from the model assumptions.

We now introduce some notation to formalize our problem concerning robustness. Consider a measurable space $(\Omega, \mathcal{F})$. Throughout this paper we consider time series of $\mathcal{X}$-valued random variables, where $\mathcal{X} \in \mathbb{R}^d$, and define $\mathcal{X}^\infty = \mathcal{X} \times \mathcal{X} \times \ldots$. Let $\mathcal{A}^\infty$ signify the Borel σ-algebra on $\mathcal{X}^\infty$. A measurable function $X^\infty : \Omega \to \mathcal{X}^\infty$ determines an infinite sequence $X^\infty(\omega) = (\ldots X_{-1}(\omega), X_0(\omega), X_1(\omega)\ldots)$ for a given $\omega \in \Omega$. For a probability measure P defined on $(\Omega, \mathcal{F})$, let the notation $P^{(k,t)}$ signify the $d \times k$-dimensional marginal distribution of $(X_t(\omega), \ldots, X_{t+k-1}(\omega))$ under P. If the process X^∞ that obeys P_0 is strictly stationary, $P^{(k,t)}$ does not depend on t, and the notation $P^{(k)}$ is used to denote it.

Let $g : \mathcal{X} \times \Theta \to \mathbb{R}^m$ be a vector-valued function parameterized by a p-dimensional vector θ which resides in $\Theta \subset \mathbb{R}^p$. Let P_0 be a probability measure on the measurable space $(\Omega, \mathcal{F})$, and suppose a random sequence X^∞ is strictly stationary under P_0. Moreover, suppose a model restriction of the following form holds for P_0:

$$
\begin{aligned}
E_{P_0}\left[g\left(X_t, \theta_0\right)\right] &= \int g\left(X_t(\omega), \theta_0\right) P_0(d\omega) \\
&= 0, \quad \theta_0 \in \Theta.
\end{aligned} \tag{1.1}
$$

The goal of the econometrician is to estimate the unknown θ_0. Note that the parameter θ_0 is identified by the marginal distribution of X_t only.

The model (1.1) imposes only mild restrictions on P_0, both in terms of distributional assumptions and dynamic specifications. It is, nevertheless, realistic to assume that the data observed by the researcher is drawn from a probability measure that is *not* P_0 in the model (1.1), but its perturbed version, due to, say, measurement errors. Let Q denote such a "perturbed" probability measure. The econometrician observes data $(x_1, \ldots, x_n)$, n consecutive values in a realization of the random element X^∞ that obeys Q, and calculates an estimator $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$. The goal is to obtain an estimator whose deviation from θ_0 (which corresponds to P_0) remains stable as far as Q is reasonably close to P_0.

This paper develops a formal theory of robust estimation for moment condition model with dependent data. There is a vast literature on robust methods in econometrics and statistics. A line of research that is highly relevant to the current paper is initiated by a seminal paper by Beran (1977). It considers robust estimation of parametric models with IID data, and shows that the minimum

Hellinger distance estimator (MHDE) has desirable properties. The parametric MHDE is robust in the sense that it is relatively insensitive to perturbations in the density that generates observations. Moreover, in the absence of such perturbations it is asymptotically equivalent to the ML estimator and therefore asymptotically efficient, at least to the first order. Thus the MHDE is robust and asymptotically efficient at the same time. Further theoretical developments on this finding can be found, for example, in Donoho and Liu (1988) and Rieder (1994). Kitamura, Otsu, and Evdokimov (2009) consider the moment condition model as presented above, under the assumption that the data is IID. They develop a robustness theory that deals with the semiparametric nature of the moment condition model, and show that the MHDE applied to the moment restriction model (the moment restriction MHDE) possesses an asymptotic minimax optimal robustness property. Also, analogous to Beran's result for the parametric MHDE, the moment restriction MHDE remains to be semiparametrically efficient in the absence of perturbations. Thus the moment restriction MHDE is robust and efficient in a semiparametric sense.

The present paper extends the above research on robustness to time series data. This is a practically important problem. For example, in addition to conventional mis-measurements, it has been pointed out that an inadequate seasonal adjustment yields serially correlated measurement errors that are very hard to deal with (see, Ashley and Vaughgan, 1986, for example). In spite of this, robust estimation has been mainly studied in the IID context. For dependent data, the literature has focused on parametric time series models Martin and Yohai (1986) or location parameter estimation in Gaussian time series with infinite dimensional correlation matrix (Andrews, 1982, 1988). The model considered here is semiparametric as it does not make distributional assumptions, and it also involves nonparametric treatments of dependence. This problem poses novel and important theoretical challenges. For example, robustness analysis as developed, for example, by Bickel (1981), Beran (1977, 1984) and Rieder (1994) requires a definition of infinitesimal neighborhoods (of probability measures) against which one wishes to remain robust. This has been considered extensively in the literature for IID data, though an appropriate its extension to weakly dependent data is by no means obvious. Our analysis of optimal robustness also entails intricate technical problems: for example, an appropriate least favorable distributions is an important building block of our minimax optimality theory, and obtaining it under dependence and blocking calls for new techniques. Needless to say, derivations of asymptotic distributions require appropriate treatments of dependence as well. The current paper addresses these problems.

2. The Estimator

As in Andrews (1982) and Kitamura, Otsu, and Evdokimov (2009), the notion of MHDE plays a central role in this paper. The Hellinger distance between two probability measures is defined as follows:

Definition 2.1. Let P and Q be probability measures on $\mathcal{X}^s := \otimes_{i=1}^s \mathcal{X}$, $s \geq 1$, with densities p and q with respect to a dominating measure ν. The Hellinger distance between P and Q is then given by:

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

One may rewrite the above as:

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

which is convenient as it avoids an explicit use of the dominating measure. Note that the above definition can be used to define the distance between two s-dimensional joint distributions for an arbitrary s, and the dimensionality s is treated implicitly in the notation.

The Hellinger distance H yields a natural method for estimating θ_0 in (1.1). This is straightforward to see, at least when the data is IID. Suppose $\{x_t\}_{t=1}^n$ is an IID sequence with each x_t distributed according to a measure μ_0 defined on $\mathcal{X}$, so that $P_0 = \mu_0^{\otimes n}$. Under this extra assumption (1.1) becomes

$$\int_{\mathcal{X}} g(x, \theta_0) d\mu_0 = 0. \tag{2.1}$$

Consider the following population problem:

$$v(\theta) := \min_{\tilde{\mu} \ll \mu_0} H(\tilde{\mu}, \mu_0) \quad s.t. \int g(x,\theta) d\tilde{\mu} = 0, \int d\tilde{\mu} = 1.$$

An application of convex duality yields

$$v(\theta) = \max_{\gamma} - \int \frac{1}{1 + \gamma' g(x,\theta)} d\mu_0$$

(see, for example, Kitamura (2006) for details). But if θ_0 is identified in (2.1), minimizing $v(\theta)$) over $\theta \in \Theta$ leads to $\theta_0 = \operatorname{argmin}_{\theta \in \Theta} v(\theta)$. In sum,

$$\theta_0 = \operatorname*{argmin}_{\theta} \max_{\gamma} - \int \frac{1}{1 + \gamma' g(z,\theta)} d\mu_0 = \operatorname*{argmin}_{\theta} \max_{\gamma} - E_{\mu_0} \left[\frac{1}{1 + \gamma' g(x,\theta)} \right].$$

Form a natural sample analogue to define:

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \max_{\gamma} -\frac{1}{n}\sum_{t=1}^{n}\frac{1}{1+\gamma' g(x_t,\theta)}.$$

This is the moment restriction MHDE considered by Kitamura, Otsu, and Evdokimov (2009) in an IID setting. If the data is dependent, however, it is less efficient than the optimally weighted GMM when the model assumption holds for the data. A way to deal with this issue fully nonparametrically is data blocking (see Kitamura, 1997 and Kitamura and Stutzer, 1997 for applications of data blocking in empirical likelihood type estimators). Consider data blocks $\{b_j\}_{j=1}^{n_B}$ of length M, where $b_j = \left(x_{(j-1)L+1},\ldots,x_{(j-1)L+M}\right) \in \mathcal{X}^M$, $\mathcal{X}^M = \mathcal{X}^{\otimes M}$, $n_B = \lfloor (n-M)/L \rfloor + 1$, and $\lfloor\cdot\rfloor$ denotes the integer part of $\cdot$. The integer L $(1 \le L \le M)$ is the distance between starting points of blocks. Define the "smoothed moment function" $\phi(b_j,\theta) = \frac{1}{\sqrt{M}}\sum_{l=1}^{M} g\left(x_{(j-1)L+l},\theta\right)$, $j = 1,\ldots,n_B$. In addition, define the empirical measure on the blocks $P_n^{(M)}$ as

$$P_n^{(M)} = \frac{1}{n_B}\sum_{j=1}^{n_B}\delta_{\left(X_{(j-1)L+1},\ldots,X_{(j-1)L+M}\right)}.$$

Applying the moment restricted MHDE to smoothed moment functions, one obtains

$$\hat{\theta}_H = \underset{\theta\in\Theta}{\text{argmin}} \max_{\gamma\in\mathbb{R}^m} -E_{P_n^{(M)}}\left[\frac{1}{1+\gamma'\phi(b,\theta)}\right] = \underset{\theta\in\Theta}{\text{argmin}} \max_{\gamma\in\mathbb{R}^m} -\frac{1}{n_B}\sum_{t=1}^{n_B}\frac{1}{1+\gamma'\phi(b_j,\theta)}. \tag{2.2}$$

This will be called the blockwise MHDE in this paper. Note that it can be seen as a mapping of (empirical) probability measure on blocks of length M to the parameter space, i.e. $\hat{\theta}_H = T\left(P_n^{(M)}\right)$, where $T(\cdot)$ is defined by (2.2). This estimator enjoys a nice asymptotic efficiency property if the model assumption holds for the observations, in the sense that the data obeys the law P_0 that satisfies (1.1). In this ideal scenario it is easy to show that $\sqrt{n}(\hat{\theta}_H - \theta_0) \xrightarrow{d} N(0,\Sigma)$, $\Sigma = (G'\Omega G)^{-1}$, $G = E_{P_0}[\frac{\partial}{\partial\theta'}g(z_t,\theta_0)]$, $\Omega = \sum_{t=-\infty}^{\infty} E_{P_0} g(z_t,\theta_0) g(z_{t-j},\theta_0)'$ under mild regularity conditions. The blockwise MHDE is therefore as efficient as the optimally weighted GMM in the absence of data perturbation. The subsequent sections show that it has desirable robustness properties as well. The blockwise MHDE is, therefore, robust and efficient under weak dependence.

3. Main Results

The focus of this paper is estimation of the parameter θ when the data are generated by a locally perturbed version of the probability measure P_0 that satisfies the model (1.1). In particular, we seek for an estimator that has small asymptotic MSE as far as the probability law of the data stays within a shrinking neighborhood of P_0. Since we study dependent data, an appropriate definition of a local neighborhoods that takes dependence into accounts needs to be developed. The following definition of neighborhoods is suitable for the development of our robustness theory in time series.

Definition 3.1. Let P be a probability measure on the measurable space $(\Omega, \mathcal{F})$. For a positive δ, we let $\mathcal{B}(P,\delta)$ denote the set of all probability measures Q that satisfies the following three conditions:

(i) $H\left(Q^{(t,M)}, P^{(t,M)}\right) \le \delta$ for each t,

(ii) A process X^∞ that obeys Q is strong mixing with α-mixing coefficients $\alpha(k)$ satisfying $\sum_{k=1}^{\infty} \alpha(k)^{1-2/\eta} < \infty$ for $\eta > 2$ defined in Assumption 3.1 (v) below.

(iii) for each t, $E_Q\left[\sup_{\theta\in\Theta} |g(X_t,\theta)|^\eta\right] < \infty$ for $\eta > 2$ defined in Assumption 3.1 (v) below.

Let M such that $M/n \to 0$ as $n \to \infty$. Sequences of local neighborhoods of the form $\mathcal{B}(P_0, r\sqrt{M/n})$ for $r > 0$ are used throughout our theoretical analysis in this section. We consider the effect of perturbations of P_0 within $\mathcal{B}(P_0, r\sqrt{M/n})$, that is, we analyze the maximum MSE of estimators when the probability law Q of the data varies within $\mathcal{B}(P_0, r\sqrt{M/n})$. Some comments are in order:

(1) The neighborhood $\mathcal{B}(P_0, r\sqrt{M/n})$ shrinks as n increases, because we assume that $M/n \to 0$ as $n \to \infty$. Its local nature has the effect of balancing the stochastic orders of the bias and standard error of an estimator, thereby allowing comparison of estimators according to their MSE.

(2) In the above setup, the distance between probability laws is defined by Hellinger distance between the M-dimensional marginal distributions of the probability laws, where M grows with the sample size n. An increase in block length M is "balanced" by the factor M in the radius of the neighborhood $\mathcal{B}(P_0, r\sqrt{M/n})$. Since the blocks length M is growing with n, the distance measure $H\left(Q^{(t,M)}, P^{(t,M)}\right)$ in (i) incorporates further information about dependence in the process as n increases.

(3) We do not assume the perturbed measure Q to be stationary, therefore finite dimensional distributions on different blocks may differ (though we will impose that the process X^∞ under P_0 is strictly stationary: see Assumption 3.1).

(4) We note that the local neighborhood system $\{\mathcal{B}(P_0, r\sqrt{M/n}), n \in \mathbb{N}\}$ introduced above has some connections with other definitions of neighborhood systems used in the robust estimation literature. Beran (1977,1978,1980) investigates robust estimation of *parametric* models in cross-sectional settings using the "'standard" definition of Hellinger neighborhood. Suppose the statistical model is given by $\{P_\theta\}_{\theta\in\Theta}$ where Θ is a finite dimensional parameter space.

Beran considers estimation of $\theta_0 \in \Theta$ from a random sample drawn from a probability measure Q that satisfies $H\left(Q, P_{\theta_0}\right) \leq r/\sqrt{n}$ for all n. Beran (1982) considers a similar problem with i.n.i.d. data, by introducing a definition of contamination neighborhood appropriate for nonidentical distributions. Kitamura, Otsu, and Evdokimov (2009) consider robust estimation when data are i.i.d. draws from a perturbed probability law of a *semiparametric* model, using a Hellinger-based neighborhood system as used in Beran (1977,1978,1980). Andrews (1988), in a weak dependence setting, considers estimation of location parameter with data being perturbations of a Gaussian stochastic process. Due to his interest in location parameter Andrews only assumes that *marginal* distribution of the stochastic process lie in a neighborhood shrinking at the $\sqrt{n}$ rate and imposes weak restrictions on the perturbations of the dependence structure of the process. The current paper differs from Andrews (1988) as it considers general moment condition models, and, moreover, It seeks robustness within neighborhoods defined for joint distributions of stochastic processes over time by considering M-dimensional distribution with $M \to \infty$, in contrast to neighborhoods defined by (one-period) marginals in Andrews (1988).

(5) Kitamura, Otsu, and Evdokimov (2009) discuss the possibility of replacing using other divergence measure to define an alternartive neighborhood system. In particular, the so-called α-divergence is a reasonable candidate since well-known measures of divergence as its special cases, such as relative entropy (or the Kullback-Leibler divergence), the Hellinger distance and the χ^2 distance. Kitamura, Otsu, and Evdokimov (2009) show the Hellinger distance-based neighborhood system has a unique property that it is in some sense "largest" among neighborhood systems based on other α-divergence. Since the goal of robust estimation is to guard against a large set of perturbations, this seems to justify the use of the Hellinger distance for constructing neighborhoods. See Kitamura, Otsu, and Evdokimov (2009) for a further discussion on this matter.

(6) Condition (ii) imposes a mixing condition Q. This does not seem to follow directly from (i) and Assumption 3.1 (i), which is a mixing condition on P_0, even for small r.

Let $\tau : \Theta \to \mathbb{R}$ be a transformation of the parameter. We study the estimation problem of the transformed parameter $\tau\left(\theta_0\right)$, as in Rieder (1994). Transforming the vector valued θ to a scalar $\tau(\theta)$ is convenient in calculating MSE's in our main theorem, which compares the asymptotic MSE of the blockwise MHDE $T\left(P_n^{(M)}\right) = \hat{\theta}_H$ with that of an alternative estimator $T_a = T_a\left(\{x_1, \ldots, x_n\}\right)$, in particular their maximum values of the MSEs as the probability of law of data varies over $\mathcal{B}(P_0, r\sqrt{M/n})$.

We impose the following assumptions. Let $\mathcal{U}$ be an open neighborhood around θ_0.

Assumption 3.1. *The following conditions hold:*

- **(i):** *The process $X^{\infty}(\omega)$ under the probability measure P_0 strictly stationary and α-mixing, with its α-mixing coefficients $\alpha(k)$ satisfying $\sum_{k=1}^{\infty} \alpha(k)^{1-2/\eta} < \infty$, where η is defined in (v) below;*
- **(ii):** *$\Theta \subset \mathbb{R}^p$ is compact;*
- **(iii):** *$\theta_0 \in int(\Theta)$ is a unique solution to $E_{P_0}[g(X_t,\theta)] = 0$;*
- **(iv):** *for each $\theta \in \Theta$, $g(x,\theta)$ is continuous for all $x \in \mathcal{X}$;*
- **(v):** *$E_{P_0}[\sup_{\theta\in\Theta}|g(X_t,\theta)|^{\eta}] < \infty$ for some $\eta \in (2,4]$, $E_{P_0}\left[\sup_{\theta\in\mathcal{U}}|g(X_t,\theta)|^4\right] < \infty$, $g(x,\theta)$ is continuously differentiable a.s. in $\mathcal{U}$, $E_{P_0}\left[\sup_{\theta\in\mathcal{U}}|\partial g(X_t,\theta)/\partial\theta'|^2\right] < \infty$, and $\sup_{x\in\mathcal{X}_n^1,\theta\in\mathcal{U}}|\partial g(x,\theta)/\partial\theta'| \le o\left(n^{1/2}\right)$, where $\mathcal{X}_n^1$ is defined in the Appendix;*
- **(vi):** *G has the full column rank and Ω is positive definite;*
- **(vii):** *$M = O(n^{\alpha})$, $0 < \alpha < \frac{2\eta-\eta^2}{2(\eta^2-1)}$. Everywhere M is implicitly assumed to depend on n.*
- **(viii):** *τ is continuously differentiable at θ_0.*

Assumption 3.1 (i)-(vi) are standard in the literature of the GMM. Assumption (i) is a regularity condition needed to guarantee that a Central Limit Theorem holds. Assumption 3.1 (iii) is a global identification condition of the true parameter θ_0. Assumption 3.1 (v) contains the smoothness and boundedness conditions for the moment function and its derivatives. This is stronger than the assumptions needed to derive the standard asymptotic normality result without data perturbation. Assumption 3.1 (vi) is a local identification condition for θ_0. This assumption guarantees that the asymptotic variance matrix Σ^{-1} is well defined. Assumption 3.1 (iv) is imposed to guarantee the continuity of the truncated MHDE mapping of block-measures $Q^{(M)}$ to Θ that are used in the proof of main results; see Appendix for the details. Assumption 3.1 (vii) restricts the rate of growth of block length with the sample size. This restriction allows introduction of a trimming sequence m_n, which plays an important role in the theoretical arguments[1]. Assumption (vii) is only a sufficient condition; we give a more general, but more complicated condition in the Appendix. Assumption 3.1 (viii) is a standard requirement for the parameter transformation τ.

In addition we need some regularity conditions on the alternative estimators $T_a : \{X_1,\dots,X_n\} \to \Theta$. We assume that an estimator T_a satisfies the following property:

[1]However, no trimming is needed for the estimation procedure.

Assumption 3.2. *There exists a sequence of functions $\psi_n(x)$ such that for every $r > 0$, every $\xi \in \mathbb{R}^p$, and every sequence $\{Q_n\}_{n\in\mathbb{N}}$,*

$$Q_n \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right) \cap \left\{P : E_P\left[g\left(X_t, \theta_0 + \xi/\sqrt{n}\right)\right] = 0 \text{ for all } t\right\}$$

the following holds

$$\sqrt{n}\left(T_a\left(\{X_1, \ldots, X_n\}\right) - \theta_0\right) - \xi - \frac{1}{\sqrt{n}}\sum \varphi_n\left(X_t\right) \to_d 0, \quad \text{under } Q_n, \tag{3.1}$$

where $E_{Q_n}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^n \varphi_n(X_t)\right] \to 0$ for all t, and $E_{Q_n}\left[\frac{1}{n}\sum_{t=1}^n\sum_{\tau=1}^n \varphi_n(X_t)\varphi_n(X_\tau)'\right] \to A_{\varphi\varphi'}$, where $A_{\varphi\varphi'} - \Sigma^{-1}$is a positive-semidefinite matrix.

The above assumption is an asymptotic linearity condition, and satisfied by standard estimators. The condition $A_{\varphi\varphi'} \geq \Sigma^{-1}$ is reasonable and holds for the optimal GMM/CUE and appropriate blockwise versions of GEL estimators.

The following assumption is only used to derive the minmax bound. It is not needed to show the properties of $T\left(P_n^{(M)}\right)$.

Assumption 3.3. *The following conditions hold:*

(i): *All the components of X_t are continuously distributed;*

(ii): $\eta > 4$.

Assumption 3.3 (i) is permits using integral transform as a part of the proof. This may be relaxed at the expense of extra complexity in the proofs. Assumption 3.3 (ii) is strong. Section 6.1 of Appendix introduces a trimming sequence $m_n \to \infty$ and trimmed moment condition function $\phi_n(b, \theta)$ such that $|\phi_n(b, \theta)| \leq m_n$ for all b. On the one hand, the trimming sequence should diverge fast enough, so that $|E_{P_0}[\phi_n(B, \theta_0)]| = o\left(\sqrt{M/n}\right)$, i.e. the moment condition based on the $\phi_n(B, \theta)$ is close enough to the original moment condition (1.1). On the other hand, the behaviour of $\phi_n(B, \theta)$ should not be driven by the tail events, so m_n should not grow too fast. To guarantee the compatibility of these two requirements we impose the condition that $\sup_{\theta\in\Theta}|g(X_t, \theta)|$ has more than four moments bounded (under P_0). Note that no trimming is necessary if moment condition function is bounded.

Our main result is the following optimal MSE property of the blockwise MHDE estimator.

Theorem 3.2. *Suppose that Assumption 3.1 holds. Define $B^* = \left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)'\Sigma^{-1}\left(\frac{\partial\tau(\theta_0)}{\partial\theta}\right)$. Then the following holds for each $r > 0$:*

(i): If an alternative estimator T_a satisfies regularity Assumptions 3.2 and Assumption 3.3 holds then

$$\lim_{\kappa\to\infty}\liminf_{n\to\infty}\sup_{Q\in\mathcal{B}\left(P_0,r\sqrt{M/n}\right)}\int\kappa\wedge n\left(\tau\circ T_a\left(\{x_1,\ldots,x_n\}\right)-\tau\left(\theta_0\right)\right)^2dQ\geq\left(1+4r^2\right)B^*.$$

(ii): Moreover, the blockwise MHDE estimator satisfies

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

Discussion:

(1) Part (i) of the theorem derives the minmax bound for the (truncated) Mean Squared Error (MSE) of any estimator satisfying (3.1). Part (ii) of the theorem shows that the bound of part (i) is actually tight and that blockwise MHDE estimator attains it. Since Σ is positive definite from Assumption 3.1 (vi), the lower bound $\left(1+4r^2\right)B^*$ is positive and finite.

(2) Parameter κ guarantees that the loss function is bounded, i.e. the theorem takes truncated MSE as a loss function. Without an upper bound the MSE may be infinite, prohibiting any meaningful comparison. This use of asymptotic truncation scheme is standard in the literature of robust estimation. The fact that $\kappa\to\infty$ in the limit theory allows the truncation parameter to be arbitrarily large.

(3) The theorem does not require stationarity of the perturbed measure Q_n. Only the true measure P_0 is assumed to be stationary. Measure Q_n may, for example, be nonstationary if the data contains seasonal measurement error. Alternatively, for data covering large time periods it is possible that the measurement of the first observations is different from the measurement error in the last observations, for instance, one may think that the variance of measurement error decreases with time due to improvement in accounting techniques.

(4) The theorem concerns estimation of the true value θ_0, not of a pseudo-true value. It therefore differs from the results in White (1982), Kitamura (1998), Kitamura (2002), and Schennach (2007).

The proof of Theorem 3.2 consists of the following steps. We first obtain the maximum bias of $\tau\circ T_a$ over the neighborhoods $\mathcal{B}\left(P_0^{(\infty)},r\sqrt{M/n}\right)$. Second, we use this maximum bias to calculate the lower bound for maximum MSE over $\mathcal{B}\left(P_0^{(\infty)},r\sqrt{M/n}\right)$. Then, we introduce trimmed blockwise MHDE $\bar{T}\left(\cdot\right)$ and show that it achieves the lower bounds of bias and MSE on $\mathcal{B}\left(P_0^{(\infty)},r\sqrt{M/n}\right)$ derived earlier. Finally, we show that the difference between MSE of trimmed estimator $\bar{T}\left(P_n^{(M)}\right)$ and MSE of blockwise MHDE $T\left(P_n^{(M)}\right)$ is negligible and hence blockwise MHDE $T\left(P_n^{(M)}\right)$ achieves the lower bound.

4. Monte-Carlo Experiment

4.1. **Experiment 1.** Our experimental designs are based on the nonlinear moment condition model of Hall and Horowitz (1996). The data are a bivariate trajectory of the stochastic process $(X_t, Z_t)_{t=1}^n$, where

$$X_t = \frac{1}{1-\alpha^2}\sum_{j=0}^{\infty}\alpha^j u_{t-j}^x, \quad Z_t = \frac{1}{1-\alpha^2}\sum_{j=0}^{\infty}\alpha^j u_{t-j}^z, \tag{4.1}$$

$$(u_t^x, u_t^z)' \sim i.i.d.\ N\left(0, \frac{0.4^2}{1-\alpha^2} I_2\right), \tag{4.2}$$

where I_2 denotes the 2×2 identity matrix. Thus, X_t and Z_t are independent $AR(1)$ processes with autocorrelation $\alpha = 0.75$ (the initial values are taken to be $(u_0^x, u_0^z)' \sim N\left(0, 0.4^2 I_2\right)$). Define:

$$g(x, z; \theta) = (\exp\{-0.72 - \theta(x+z) + 3z\} - 1)\begin{pmatrix} 1 \\ z \end{pmatrix},$$

then the moment restriction $E\left[g\left(X_t, Z_t; \theta_0\right)\right] = 0$ identifies $\theta_0 = 3$.

Our first Monte-Carlo experiment considers how various estimators perform in the presence of infrequent but relatively large measurement error. Here I_2 denotes the 2×2 identity matrix. We assume that the true data generating process is (4.1), however econometrician only observes $(\widetilde{X}_t, \widetilde{Z}_t)$, where

$$(\widetilde{X}_t, \widetilde{Z}_t) = \begin{cases} (X_t, Z_t) & \text{with probability } 0.95, \\ (X_t, Z_t) + c \cdot \xi_t & \text{with probability } 0.05. \end{cases}$$

Where ξ_t is a 1×2 random vector of independent zero mean components, which may be interpreted as measurement error. We are going to compare the results of estimation using the two-step generalized method of moments (GMM) of Hansen (1982), the continuous updating GMM (CUE) of Hansen, Heaton, and Yaron (1996) with optimal weighting, the blockwise empirical likelihood estimator (EL) of Kitamura (1997) (which is a blockwise version of the EL estimator as in Qin and Lawless (1994), Imbens, Spady, and Johnson (1998), and Owen (2001)), the time-smoothed exponential tilting estimator (ET) of Kitamura and Stutzer (1997), and the blockwise minimum Hellinger distance estimator (MHDE) (2.2). Results are based on 5000 replications for each specification. Each estimator is is obtained by minimizing its criterion function on a fine grid over $\Theta = [0, 10]$. As discussed earlier, EL, MHDE, and ET use block moment conditions, with fully overlapping blocks of length M. Correspondingly, for GMM and CUE estimators the weighting matrix is taken to be the inverse of HAC covariance matrix of Newey and West (1987) with Bartlett kernel and $M-1$ lags. In the experiments with $n = 100$ observations $M = 5$ and $M = 10$ are considered. When $n = 400$, block lengths $M = 10$ and $M = 20$ are considered. The results of estimation are presented in Tables 1 and 2.

The data generating process corresponding to the first row of Table 1 has $c = 0$ and hence represents the true model (4.1). For each scenario we report the Root Mean Squared Error (RMSE) and empirical probabilities $\Pr\left\{|\hat{\theta} - \theta_0| > 1.0\right\}$ ($\Pr\left\{|\hat{\theta} - \theta_0| > 0.5\right\}$ in Table 2 due to a larger number of observations) for each estimator. Confirming the theoretical findings of Newey and Smith (2004) and Kitamura and Otsu (2005) EL is superior on the basis of both criteria. At the same time the Minimum Hellinger Distance estimator is only marginally inferior to Empirical Likelihood estimator. We find that in a wide range of circumstances the finite sample criterion function of MHDE is very close to the criterion function of the EL. ET is inferior to both EL and MHDE, although only marginally. Two-step GMM is less efficient than the EL and MHDE, especially with the larger sample. The results of the Continuous Updating GMM estimator are inferior to all the other methods. Even with a restricted parameter space Θ we find that finite sample criterion function of CUE frequently has global minimum on the boundaries of Θ. Such a poor performance of CUE in estimation of nonlinear models was noticed earlier by Hansen, Heaton, and Yaron (1996, see p.272, Figure 5).

In the presence of measurement error EL, MHDE, and ET still outperform GMM and CUE, often by a large margin. When measurement error is small EL may still outperform MHDE and ET due to higher order properties of EL. When measurement becomes large none of the three methods seems to dominate the other. Note that Theorem 3.2 does not imply that MHDE should be optimal in all situations, but it rather shows its minimax property in terms of asymptotic MSE.

The column labeled "%f" is concerned with a computation issue of EL, MHDE, and ET. In finite samples, it is possible that there exists no value of θ such that the zero vector is not contained in the convex hull of vectors $\{\phi(b_j, \theta)\}_{j=1}^{n_B}$. This is a situation where observations are providing strong evidence against the validity of the moment condition model (1.1). EL, MHDE, and ET are not well-defined in this case as they seek for probability weights that make the weighted average of ϕ's zero. The current simulation experiment discards such a replication in calculations of the summary statistics. Column "%f" of Tables 1 and 2 reports the percentage of the occurence of such replications. As can be seen from Table 1, such cases are practically negligible in most cases, though it increases for some case with a large c, especially in the case of the $-\chi_1^2$ measurement error specification.

		RMSE					$Pr\left\{\left\|\hat{\theta}-\theta_0\right\|>0.5\right\}$					
c	ξ_{tj}	EL	MHDE	ET	GMM	CUE	EL	MHDE	ET	GMM	CUE	%f
						$n=100,\ M=5$						
0		0.764	0.886	1.046	0.933	3.406	0.109	0.121	0.137	0.210	0.370	0.02
0.5	N	0.694	0.800	0.908	0.873	3.162	0.095	0.104	0.117	0.176	0.321	0.00
1	N	0.673	0.758	0.845	0.959	2.891	0.110	0.116	0.127	0.254	0.314	0.00
2	N	0.929	0.906	0.924	1.440	2.601	0.308	0.261	0.245	0.593	0.413	0.18
0.5	χ_1^2	0.732	0.873	1.062	0.915	3.300	0.106	0.117	0.133	0.200	0.354	0.02
1	χ_1^2	0.649	0.729	0.879	0.835	3.116	0.087	0.095	0.106	0.168	0.309	0.00
2	χ_1^2	0.618	0.672	0.736	0.878	2.901	0.074	0.080	0.086	0.219	0.282	0.00
0.5	$-\chi_1^2$	0.714	0.812	0.949	0.947	3.138	0.110	0.120	0.134	0.219	0.337	0.00
1	$-\chi_1^2$	0.773	0.808	0.939	1.130	2.878	0.163	0.155	0.165	0.344	0.341	0.22
2	$-\chi_1^2$	1.021	0.979	1.015	1.523	2.616	0.308	0.250	0.241	0.536	0.398	1.92
0.5	t_3	0.741	0.834	0.984	0.924	3.195	0.111	0.119	0.134	0.205	0.344	0.00
1	t_3	0.721	0.834	0.940	0.993	2.987	0.115	0.124	0.133	0.250	0.321	0.12
2	t_3	0.841	0.864	0.962	1.296	2.636	0.219	0.198	0.198	0.458	0.361	0.78
						$n=100,\ M=10$						
0		0.806	0.946	1.089	0.899	3.197	0.121	0.133	0.145	0.200	0.354	0.36
0.5	N	0.740	0.846	0.980	0.858	2.977	0.104	0.114	0.127	0.174	0.307	0.12
1	N	0.719	0.812	0.919	0.952	2.704	0.114	0.127	0.141	0.256	0.297	0.10
2	N	0.946	0.935	0.956	1.451	2.471	0.317	0.276	0.259	0.601	0.411	0.30
0.5	χ_1^2	0.811	0.924	1.067	0.892	3.112	0.119	0.126	0.138	0.192	0.339	0.16
1	χ_1^2	0.711	0.814	0.947	0.824	2.905	0.097	0.106	0.116	0.161	0.288	0.10
2	χ_1^2	0.632	0.702	0.761	0.888	2.775	0.077	0.085	0.096	0.232	0.267	0.04
0.5	$-\chi_1^2$	0.751	0.855	0.975	0.925	2.922	0.122	0.137	0.147	0.213	0.317	0.18
1	$-\chi_1^2$	0.811	0.875	0.979	1.127	2.760	0.171	0.169	0.176	0.342	0.333	0.28
2	$-\chi_1^2$	1.045	1.017	1.054	1.517	2.492	0.315	0.258	0.247	0.538	0.390	2.48
0.5	t_3	0.840	0.951	1.043	0.912	3.033	0.121	0.130	0.143	0.199	0.329	0.18
1	t_3	0.785	0.888	1.018	0.975	2.795	0.125	0.133	0.141	0.246	0.302	0.30
2	t_3	0.868	0.892	0.960	1.289	2.511	0.226	0.202	0.201	0.459	0.352	0.98

TABLE 1. In the second column (ξ_{tj}) N, χ_1^2, $-\chi_1^2$, and t_3 denote, correspondingly, $N(0,1)$,$(\chi_1^2-1)/\sqrt{2}$,$-(\chi_1^2-1)/\sqrt{2}$, and Student-$t_3/\sqrt{3}$ distributions of ξ_{tj}.

		RMSE					$Pr\left\{\left\|\hat{\theta}-\theta_0\right\|>1.0\right\}$					
c	ξ_{tj}	EL	MHDE	ET	GMM	CUE	EL	MHDE	ET	GMM	CUE	%f
						$n=400,\ M=10$						
0		0.286	0.289	0.295	0.402	2.848	0.081	0.082	0.085	0.112	0.245	0.00
0.5	N	0.285	0.289	0.295	0.402	2.745	0.078	0.079	0.082	0.112	0.235	0.00
1	N	0.280	0.284	0.287	0.368	2.560	0.071	0.071	0.072	0.101	0.211	0.00
2	N	0.470	0.452	0.443	0.678	2.747	0.302	0.274	0.260	0.482	0.416	0.00
0.5	χ_1^2	0.282	0.286	0.292	0.397	2.748	0.075	0.077	0.081	0.109	0.232	0.00
1	χ_1^2	0.269	0.271	0.274	0.366	2.702	0.060	0.061	0.063	0.091	0.216	0.00
2	χ_1^2	0.272	0.272	0.274	0.344	2.579	0.060	0.061	0.063	0.094	0.203	0.00
0.5	$-\chi_1^2$	0.286	0.291	0.298	0.419	2.666	0.077	0.079	0.081	0.114	0.228	0.00
1	$-\chi_1^2$	0.366	0.359	0.356	0.585	2.461	0.151	0.142	0.140	0.258	0.276	0.00
2	$-\chi_1^2$	0.700	0.635	0.607	1.153	2.270	0.425	0.357	0.332	0.652	0.484	0.54
0.5	t_3	0.289	0.292	0.298	0.422	2.772	0.075	0.076	0.081	0.112	0.236	0.04
1	t_3	0.312	0.309	0.313	0.475	2.490	0.086	0.086	0.089	0.144	0.220	0.08
2	t_3	0.497	0.468	0.453	0.799	2.377	0.251	0.221	0.207	0.409	0.350	0.36
						$n=400,\ M=20$						
0		0.295	0.299	0.306	0.395	2.708	0.090	0.091	0.094	0.113	0.236	0.00
0.5	N	0.293	0.297	0.306	0.390	2.597	0.085	0.086	0.088	0.113	0.226	0.00
1	N	0.285	0.289	0.292	0.361	2.433	0.079	0.079	0.078	0.101	0.200	0.00
2	N	0.470	0.453	0.443	0.697	2.639	0.302	0.275	0.263	0.495	0.408	0.00
0.5	χ_1^2	0.290	0.295	0.300	0.376	2.614	0.082	0.083	0.086	0.104	0.220	0.00
1	χ_1^2	0.276	0.278	0.282	0.354	2.579	0.065	0.066	0.068	0.090	0.206	0.00
2	χ_1^2	0.276	0.277	0.278	0.341	2.438	0.066	0.066	0.067	0.096	0.192	0.00
0.5	$-\chi_1^2$	0.293	0.299	0.309	0.407	2.568	0.082	0.083	0.085	0.110	0.222	0.00
1	$-\chi_1^2$	0.372	0.364	0.364	0.586	2.329	0.156	0.148	0.146	0.266	0.270	0.00
2	$-\chi_1^2$	0.709	0.641	0.612	1.169	2.245	0.432	0.362	0.333	0.657	0.484	0.54
0.5	t_3	0.296	0.299	0.310	0.415	2.661	0.081	0.082	0.088	0.112	0.228	0.04
1	t_3	0.316	0.315	0.325	0.470	2.383	0.090	0.093	0.093	0.146	0.214	0.06
2	t_3	0.499	0.469	0.455	0.803	2.317	0.251	0.222	0.211	0.414	0.341	0.46

TABLE 2. In the second column (ξ_{tj}) N, χ_1^2, $-\chi_1^2$, and t_3 denote, correspondingly, $N(0,1)$,$(\chi_1^2-1)/\sqrt{2}$,$-(\chi_1^2-1)/\sqrt{2}$, and Student-$t_3/\sqrt{3}$ distributions of ξ_{tj}.

4.2. **Experiment 2.** The independent measurement error model of the previous subsection may be somewhat restrictive, since in practice measurement errors may be correlated with the original data. To explore this and other forms of deviations from the model assumption, the following experiment studies the effects of various perturbations of the data generating process (4.1)-(4.2).

Note that the joint distribution $Q^{(M)}$ of the data block $B = (X_1, \ldots, X_M, Z_1, \ldots, Z_M)'$ of length M is fully determined by the bivariate distribution of the disturbances $(u_t^x, u_t^z)'$. The model (4.1)-(4.2) assumes that vector $(u_t^x, u_t^z)'$ has normal distribution with zero means and covariance matrix $\Sigma_0 = \frac{0.4^2}{1-\alpha^2} I_2$, i.e. $(u_t^x, u_t^z)'$ has independent components with equal variance. Following the notation introduced in Section 3, let $P_0^{(M)}$ denote the distribution of B under (4.1)-(4.2). To investigate the performance of the estimators we would like to consider various small perturbations of this probabilistic model. One way to build a family of such perturbations is to allow the components of the random vector $(u_t^x, u_t^z)'$ to have unequal variances and to be correlated, i.e. to have bivariate normal distribution with the covariance matrix

$$\Sigma_{(\delta,\rho)} = \frac{0.4^2}{1-\alpha^2} \begin{pmatrix} (1+\delta)^2 & \rho(1+\delta) \\ \rho(1+\delta) & 1 \end{pmatrix},$$

which is a perturbation of the matrix Σ_0 when δ and ρ are small. The form of covariance matrix is chosen so that $V[X_t]/V[Z_t] = V[u_t^x]/V[u_t^z] = (1+\delta)^2$ and the correlation $Corr(X_t, Z_t) = Corr(u_t^x, u_t^z) = \rho$. Note that $\Sigma_{(0,0)} = \Sigma_0$.

Each pair of parameters (δ, ρ) corresponds to a probability distribution on the block B; we denote this distribution by $P_{(\delta,\rho)}^{(M)}$. Note that $P_{(0,0)}^{(M)} = P_0^{(M)}$ and hence measure $P_{(\delta,\rho)}^{(M)}$ can be seen as a perturbed version of the measure $P_0^{(M)}$.

The idea here is to investigate finite sample properties of the estimators as the data distribution $P_{(\delta,\rho)}^{(M)}$ varies around the measure $P_0^{(M)}$ keeping an approximately constant distance from it. Lemma 6.11 in the Appendix shows that Hellinger distance between the true and perturbed probability measures on the block is $H\left(P_0^{(M)}, P_{(\delta,\rho)}^{(M)}\right) \approx \sqrt{M/4}\left(2\delta^2 + \rho^2\right)^{1/2}$ for small ρ and δ. Therefore values of (δ, ρ) that satisfy $c^2 = 2\delta^2 + \rho^2$ for some constant c are considered. All such values of (δ, ρ) can be conveniently represented as $\rho = c\cos(2\pi\omega)$ and $\delta = \left(c/\sqrt{2}\right)\sin(2\pi\omega)$ for $\omega \in [0, 1)$. In the Monte Carlo experiment we let $c = 0.1\sqrt{2}$, $\alpha = 0.75$, and $n = 100$. Estimation is performed with fully overlapping blocks of length $M = 5$.

We consider 64 different designs indexed by $j \in \{0, \ldots, 63\}$. In the j-th design we set $\omega_j = j/64$, and hence $\rho_j = 0.1\sqrt{2}\cos(2\pi\omega_j)$, $\delta = 0.1\sin(2\pi\omega_j)$. For each design 1,000 replications are obtained,

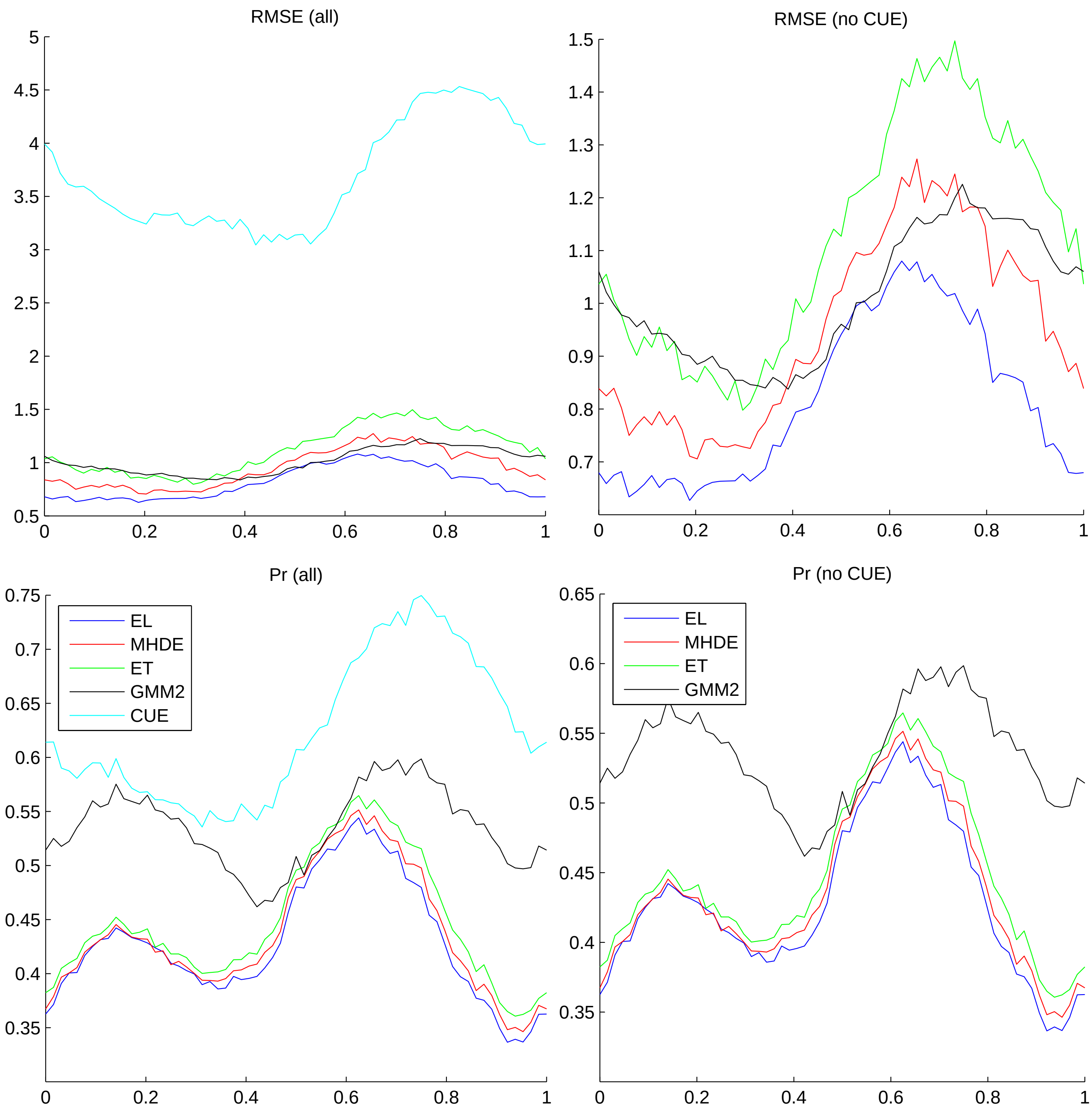

FIGURE 4.1. Local Neighborhood of the True Model. "Pr" denotes $Pr\left\{\left|\hat{\theta}-\theta_0\right|>0.5\right\}$.

and the RMSE and $Pr\left\{\left|\hat{\theta}-\theta_0\right|>0.5\right\}$ for each estimator are calculated. The results are presented in Figure 1. On the left panel, each curve represents the RMSE of a particular estimator as a function of ω_j. On the right panel, the empirical probabilities of each estimator deviating from $\theta_0=3$ by more than 0.5 are presented.[2]

As can be seen from the graphs, the RMSE of CUE is several times larger than those of the other estimators. To provide better insights of the relative performance of the other estimators, the right column of the graphs presents the same curves as the top row, except the results for CUE estimator are excluded from the picture. Interestingly, RMSE of ET is in most cases significantly larger than that of EL or MHDE and is comparable to that of GMM. However, the deviation probability $Pr\left\{\left|\hat{\theta}-\theta_0\right|>0.5\right\}$ is almost identical for EL, MHDE, and ET. This result seems to be attributable to the fact that the distribution of ET estimator has heavier tails than EL, the same problem that causes CUE's extremely poor performance in our nonlinear estimation problem.

5. Appendix

Notation. Let $C>0$ be a generic positive constant, $\|\cdot\|$ is the L_2-metric on the appropriate space. When measure P is stationary the time subscript t is unnecessary and is omitted. Also, for a finite dimensional measure $P^{(k,t)}$ of P, we sometimes omit the superscript (k,t) when it is clear from the context which finite dimensional measure is used. In the proofs, we also use the following notation:

$$
\begin{aligned}
\theta_n &= \theta_0+\xi/\sqrt{n}, \quad b=(x_1,\ldots,x_M)\in\mathcal{X}^M, \quad \phi_n(b,\theta)=\phi(b,\theta)\mathbb{I}\left\{b\in\mathcal{X}_n^M\right\}, \\
\Lambda &= \sqrt{M}G'\Omega^{-1}\phi(b,\theta_0), \quad \Lambda_n=\sqrt{M}G'\Omega^{-1}\phi_n(b,\theta_0), \\
R_n\left(Q^{(M)},\theta,\gamma\right) &= -\int\frac{1}{1+\gamma'\phi_n(b,\theta)}dQ^{(M)}, \quad \bar{P}^{(M)}_{\theta,Q^{(M)}}=\arg\min_{P^{(M)}\in\bar{P}^{(M)}_\theta,P^{(M)}\ll Q^{(M)}}H\left(P^{(M)},Q^{(M)}\right) \\
\Lambda &= \sqrt{M}G'\Omega^{-1}\phi(b,\theta_0), \quad \Lambda_n=\sqrt{M}G'\Omega^{-1}\phi_n(b,\theta_0), \\
\psi_{n,Q_n^{(M)}} &= -2\left(\int\Lambda_n\Lambda_n'dQ_n^{(M)}\right)^{-1}\int\Lambda_n\left\{dQ_n^{1/2}-d\bar{P}^{1/2}_{\theta_0,Q_n}\right\}dQ_n^{1/2}, \\
B_H\left(P_0^{(M)},\delta\right) &= \left\{Q^{(M)}\in\mathcal{M}^{(M)}:H\left(Q^{(M)},P_0^{(M)}\right)\le\delta\right\},
\end{aligned}
$$

where $\mathcal{M}^{(M)}$ is the set of all probability measures on the Borel σ-field $\left(\mathcal{X}^M,\mathcal{B}\left(\mathcal{X}^M\right)\right)$, while $\mathcal{X}_n^M$ and $\bar{T}(\cdot)$ are defined in subsection 5.1 below. Let $\mathcal{M}_S$ denote the set of all probability measures under

[2]Note that design with $\omega=1$ coincides with the $\omega=0$ design and hence the graphs are closed loops.

which the process $X^{\infty}(\omega)$ is strictly stationary. As usual, everywhere x_t and b_j denote realizations of X_t and B_j. The abbreviation w.p.a.1 should be read as "with probability approaching 1 as $n \to \infty$", UWLLN stands for Uniform Weak Law of Large Numbers, see for example Andrews (1987) or Potscher and Prucha (1989), and CLT denotes the Central Limit Theorem of Herrndorf (1984).

The first subsection of this Appendix introduces trimmed estimators that are used in the proofs. The second subsection gives several important lemmas that are used in the third section to prove Theorem 3.2. Auxiliary lemmas are given in Section 6.

5.1. **Trimming.** An essential tool of the proofs is the following mapping from $\mathcal{M}^{(M)}$ to Θ defined by a trimmed moment function:

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

where $\bar{\mathcal{P}}_\theta^{(M)}$ is the set of the M-dimensional measures $P^{(M)}$ of all probability measures P in $\bar{\mathcal{P}}_\theta$, where the latter set is defined as

$$\begin{aligned}
\bar{\mathcal{P}}_\theta &= \left\{P \in \mathcal{M}_S : \int \phi(b,\theta)\,\mathbb{I}\left\{b \in \mathcal{X}_n^M\right\} dP^{(M)} = 0\right\}, \\
\mathcal{X}_n^k &= \left\{(x_1,\ldots,x_k) \in \mathcal{X}^k : \sup_{\theta\in\Theta}\left|g(x_i,\theta)\right| \le m_n,\ i = 1,\ldots,k\right\},
\end{aligned}$$

where $\{m_n\}_{n\in\mathbb{N}}$ is a sequence of positive numbers satisfying $m_n \to \infty$ as $n \to \infty$, $\mathbb{I}\{\cdot\}$ is the indicator function, and $|\cdot|$ is the Euclidean norm, i.e., $\mathcal{X}_n^M$ is a trimming set to bound the moment function. Thus the set $\bar{\mathcal{P}}_\theta$ is the collection of probability measures satisfying the bounded moment condition $E_{P^{(M)}}\left[\phi(B,\theta)\,\mathbb{I}\left\{B \in \mathcal{X}_n^M\right\}\right] = 0$. Trimming is needed to guarantee the existence of the mapping $\bar{T}\left(Q^{(M)}\right)$. Lemma 6.1 (i) shows that for each $n \in \mathbb{N}$ and $Q \in \mathcal{M}_S$ the value $\bar{T}\left(Q^{(M)}\right)$ exists. To simplify the notation below we sometimes denote $\bar{T}_Q = \bar{T}(Q)$ for a measure Q when $\bar{T}(Q)$ is well-defined.

We may take the trimming sequence $\{m_n\}_{n\in\mathbb{N}}$ to satisfy $0 < \liminf_{n\to\infty} m_n/n^\beta \le \limsup_{n\to\infty} m_n/n^\beta < \infty$,

$$\frac{1}{2(\eta-1)} + \frac{\alpha}{\eta} < \beta < \min\left\{\frac{1}{2} - \alpha, \frac{1}{\eta}\right\}, \tag{5.1}$$

where α is from Assumption 3.1 (vii). Note that the restrictions imposed on α by Assumption 3.1 (vii) guarantee existence of β that satisfies (5.1). Assumption 3.1 (vii) together with (5.1) are sufficient to guarantee that

$$\max\left\{M^{1-1/\eta}m_n^{1-\eta}n^{1/2}, M^{34}m_n^{-3}n^{1/2}, Mm_nn^{-1/2}, nm_n^{-\eta}, M^3n^{-1}\right\}\to 0, \tag{5.2}$$

which is used in the proofs below.

5.2. **Key Lemmas.** We introduce measure $\widetilde{P}_{\theta_n}$ that will be used in the subsequent proofs.

Definition 5.1. For any $\xi\in\mathbb{R}$ and n large enough (so that $\theta_n\in\Theta$) define a nonstationary process Z_t in the following way. For each integer k, $\left(Z'_{kM+1},\dots,Z'_{(k+1)M}\right)'=\Upsilon^n\left(\left(X'_{kM+1},\dots,X'_{(k+1)M}\right)'\right)$, where for any Md dimensional random vector $\gamma=(\gamma_1,...,\gamma_{Md})'$, the r-th component of the vector transformation Υ^n is defined as

$$\Upsilon^n_r(\gamma)=\begin{cases}G^{-1}_{r|}\left(F_{r|}\left(\gamma_r|\gamma_{r-1},\dots,\gamma_1\right)\big|\left(\Upsilon^n_{r-1}(\gamma),\dots,\Upsilon^n_1(\gamma)\right)\right), & \text{if } r>1,\\ G_1^{-1}\left(F_1\left(\gamma_1\right)\right), & \text{if } r=1,\end{cases}$$

where $F_{r|}\left(\gamma_r|\gamma_{r-1},\dots,\gamma_1\right)$ is the cumulative distribution function of the r-th component of $\left(X'_{kM+1},\dots,X'_{(k+1)M}\right)'$, conditional on the first $(r-1)$ components of this vector. Thus $F_{|r}$ is fully defined by the cumulative distribution function $F(\cdot)$ of $\left(X'_{kM+1},\dots,X'_{(k+1)M}\right)'$, which corresponds to the the probability measure $P_0^{(M)}$. Similarly, $G^{-1}_{r|}$ is the inverse (in the first argument) of the conditional cumulative distribution function $G_{r|}\left(\gamma_r|\gamma_{r-1},\dots,\gamma_1\right)$, which is defined by the probability measure $\widetilde{P}^{(M,1)}_{\theta_n}$,

$$\frac{d\widetilde{P}^{(M,1)}_{\theta_n}}{dP_0^{(M)}}(b)=\frac{1+\zeta_n'\phi_n(b,\theta_n)}{\int\left(1+\zeta_n'\phi_n(b,\theta_n)\right)dP_0^M(b)}, \tag{5.3}$$

where $b=(x_1,\dots,x_M)$, and $\zeta_n=-E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\phi_n(B,\theta_n)'\right]^{-1}E_{P_0^{(M)}}\left[\phi(B,\theta_n)\right]$. Denote the probability measure of the nonstationary process Z_t by $\widetilde{P}_{\theta_n}$.

Lemma 5.2. *Suppose that Assumptions 3.1 and 3.3 hold. Then, for each $r>0$, any $\epsilon\in\left(0,r^2\right)$, and all n large enough, probability measure $\widetilde{P}_{\theta_n}$ satisfies*

$$\widetilde{P}_{\theta_n}\in\mathcal{B}\left(P_0,r\sqrt{M/n}\right),$$

when $\frac{1}{4}\xi'\Sigma\xi\le r^2-\epsilon$.

Proof. 1. We first derive $H\left(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\right)$. Denote $f_n(b,\theta_n,\zeta_n) = d\widetilde{P}_{\theta_n}^{(M,1)}(b)/dP_0^{(M)}$. By Taylor expansion

$$H\left(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\right) = \left\| \zeta_n' \frac{\partial f_n(b,\theta_n,0)}{\partial \zeta_n}^{1/2} dP_0^{1/2} + \frac{1}{2}\zeta_n' \frac{\partial f_n\left(b,\theta_n,\dot{\zeta}_n\right)}{\partial\zeta_n\partial\zeta_n'}^{1/2} \zeta_n dP_0^{1/2} \right\|, \tag{5.4}$$

where each element of $\dot{\zeta}_n$ is between the corresponding element of ζ_n and 0. Then

$$\frac{\partial f_n(b,\theta_n,0)}{\partial\zeta_n}^{1/2} = \frac{1}{2}\left(\phi_n(b,\theta_n) - E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]\right),$$

and

$$\begin{aligned}
&\frac{\partial f_n\left(b,\theta_n,\dot{\zeta}_n\right)}{\partial\zeta_n\partial\zeta_n'}^{1/2} \\
= & -\frac{1}{4}\left(1+\zeta_n'\phi_n(b,\theta_n)\right)^{-3/2}\left(1+\zeta_n' E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]\right)^{-1/2}\phi_n(b,\theta_n)\,\phi_n(b,\theta_n)' \\
& +\frac{3}{4}\left(1+\zeta_n'\phi_n(b,\theta_n)\right)^{1/2}\left(1+\zeta_n' E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]\right)^{-5/2} E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right] E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]' \\
& -\frac{1}{4}\left(1+\zeta_n'\phi_n(b,\theta_n)\right)^{-1/2}\left(1+\zeta_n' E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]\right)^{-3/2} \\
& \quad \times\left(\phi_n(b,\theta_n)\, E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]' + E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]\phi_n(b,\theta_n)'\right).
\end{aligned}$$

Lemma 6.4 and condition (5.2) imply that $\zeta_n = O\left(\sqrt{M/n}\right)$ and $\sup_{b\in\mathcal{X}^M}\zeta_n'\phi_n(b,\theta_n) = o(1)$, and hence it can be shown that

$$\begin{aligned}
\frac{n}{M}H\left(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\right)^2 &= \frac{n}{M}\left\|\frac{1}{2}\zeta_n'\left(\phi_n(b,\theta_n) - E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\right]\right) dP_0^{1/2}\right\|^2 + o(1) \\
&= \frac{1}{4}\frac{n}{M}\zeta_n' E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\,\phi_n(B,\theta_n)'\right]\zeta_n + o(1).
\end{aligned}$$

Therefore, from Taylor expansion

$$\begin{aligned}
\frac{n}{M}H\Big(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\Big)^2 = & \frac{1}{4}\frac{n}{M}E_{P_0^{(M)}}\left[\phi(B,\theta_n)'\right]\left(E_{P_0^{(M)}}\left[\phi_n(B,\theta_n)\,\phi_n(B,\theta_n)'\right]^{-1} - \Omega^{-1}\right)E_{P_0^{(M)}}\left[\phi(B,\theta_n)\right] \\
& +\frac{1}{4}\sqrt{\frac{n}{M}}E_{P_0^{(M)}}\left[\phi(B,\theta_n)'\right]\Omega^{-1}\frac{1}{\sqrt{M}}\left(\sqrt{n}E\left[\partial\phi_n\left(b,\dot{\theta}_n\right)/\partial\theta'\right]\theta_n - \sqrt{M}G\xi\right) \\
& +\frac{1}{4}\frac{1}{\sqrt{M}}\left(\sqrt{n}E\left[\partial\phi_n\left(b,\dot{\theta}_n\right)/\partial\theta'\right]\theta_n - \sqrt{M}G\xi\right)'\Omega^{-1}G\xi + \frac{1}{4}\xi'\Sigma\xi + o(1) \\
= & \frac{1}{4}\xi'\Sigma\xi + o(1),
\end{aligned}$$

where each element of vector $\dot{\theta}_n$ is between the corresponding element of θ_n and 0 and the second equality follows from Lemma 6.4. Therefore, for any $\epsilon > 0$, the condition $H\left(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\right) \leq r\sqrt{M/n}$ holds if $\frac{1}{4}\xi'\Sigma\xi + \epsilon < r^2$.

2. According to the Definition 3.1 we need to show that M-dimensional probability measures $\widetilde{P}_{\theta_n}^{(M,t)}$ on $(X_t, \ldots, X_{t+M-1})$ satisfy $H\left(\widetilde{P}_{\theta_n}^{(M,t)}, P_0^{(M)}\right) \leq r\sqrt{M/n}$ for all t. Note that measures $\widetilde{P}_{\theta_n}^{(M,t+Mk)}$ are the same for all integer k and hence we only need to consider measures $\widetilde{P}_{\theta_n}^{(M,t)}$ for $t = 1, \ldots, M$. We have shown above that $H\left(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\right) = O\left(\sqrt{M/n}\right)$. We will now show that $H\left(\widetilde{P}_{\theta_n}^{(M,t)}, P_0^{(M)}\right) = H\left(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\right) + o\left(\sqrt{M/n}\right)$ for $t = 2, \ldots, M$.

To shorten the expressions denote $g^n(\cdot) = g_n(\cdot, \theta_n)/\sqrt{M}$. Let $P_0^{(x_1,\ldots,x_{2M})\backslash b}$ be the M-dimensional distribution of $(x_1, \ldots x_{t-1}, x_{M+t}, \ldots, x_{2M})$. For any $t = 1, \ldots, M$ and any $b = (x_t, \ldots, x_{t+M-1}) \in \mathcal{X}_n^M$ we have

$$\frac{d\widetilde{P}_{\theta_n}^{(M,t)}}{dP_0^{(M)}}(b) = \int A(x_1, \ldots, x_{2M})\, dP_0^{(x_1,\ldots,x_{2M})\backslash b},$$

where

$$\begin{aligned}
A(x_1, \ldots, x_{2M}) &= \frac{\left(1 + \zeta_n' \sum_{j=1}^{M} g^n(x_j)\right)\left(1 + \zeta_n' \sum_{j=M+1}^{2M} g^n(x_j)\right)}{\left(1 + M\zeta_n' E_{P_0}[g^n(X_1)]\right)^2} \\
&= \frac{1 + \zeta_n' \sum_{j=1}^{2M} g^n(x_j) + \sum_{j=1}^{M}\sum_{l=M+1}^{2M} \zeta_n' g^n(x_j)\, g^n(x_l)\, \zeta_n}{\left(1 + M\zeta_n' E_{P_0}[g^n(X_1)]\right)^2} \\
&= \frac{1 + \zeta_n' \sum_{j=t}^{t+M-1} g^n(x_j)}{1 + M\zeta_n' E_{P_0}[g^n(X_1)]}\left(1 + M\zeta_n' E_{P_0}[g^n(X_1)]\right)^{-1} \\
&\quad + \frac{\zeta_n'\left(\sum_{j=1}^{t-1} g^n(x_j) + \sum_{j=t+M}^{2M} g^n(x_j)\right) + \sum_{j=1}^{M}\sum_{l=M+1}^{2M} \zeta_n' g^n(x_j)\, g^n(x_l)\, \zeta_n}{\left(1 + M\zeta_n' E_{P_0}[g^n(X_1)]\right)^2}
\end{aligned}$$

Then Lemma 6.4 and condition (5.2) imply that $\zeta_n = O\left(\sqrt{M/n}\right)$, $E_{P_0}[g^n(X_1)] = O\left(\sqrt{1/(nM)}\right)$, and that

$$\frac{d\widetilde{P}_{\theta_n}^{(M,t)}}{dP_0^{(M)}}(b) = \frac{1 + \zeta_n' \sum_{j=t}^{t+M-1} g^n(x_j)}{1 + M\zeta_n' E_{P_0}[g^n(X_1)]} + o\left(\sqrt{M/n}\right).$$

Therefore, $H\left(\widetilde{P}_{\theta_n}^{(M,t)}, P_0^{(M)}\right)$

$$= \left\| \left(\frac{1+\zeta_n' \sum_{j=t}^{t+M-1} g^n(x_j)}{1+M\zeta_n' E_{P_0}[g^n(X_1)]} + o\left(\sqrt{M/n}\right) \right)^{1/2} dP_0^{1/2} - dP_0^{1/2} \right\| = H\left(\widetilde{P}_{\theta_n}^{(M,1)}, P_0^{(M)}\right) + o\left(\sqrt{M/n}\right),$$

where the last equality follows from the definition of $\widetilde{P}_{\theta_n}$ and the triangle inequality.

3. Condition (ii) of Definition 3.1 is satisfied since for large k, $\widetilde{\alpha}_n(k) \leq \widetilde{\alpha}_n(k-2M)$.

4. We now check condition (iii) of Definition 3.1,

$$\begin{aligned}
& E_{\widetilde{P}_{\theta_n}}\left[\sup_{\theta\in\Theta} |g(X_1,\theta)|^\eta\right] \\
= & \int \sup_{\theta\in\Theta} |g(x_1,\theta)|^\eta \frac{1+\zeta_n'\phi_n(b,\theta_n)}{\left(1+\zeta_n' E_{P_0}[\phi_n(B,\theta_n)]\right)^2} dP_0 \\
\leq & \sup_{b\in\mathcal{X}_n^M} \left| \frac{1+\zeta_n'\phi_n(b,\theta_n)}{\left(1+\zeta_n' E_{P_0}[\phi_n(B,\theta_n)]\right)^2} \right| \int \sup_{\theta\in\Theta} |g(X_1,\theta)|^\eta dP_0
\end{aligned}$$

where the equality follows from the definition of $\widetilde{P}_{\theta_n}$ and the second inequality follows from Assumption 3.1 (v) and the fact that $\sup_{b\in\mathcal{X}^M} \zeta_n'\phi_n(b,\theta_n) = o(1)$.

Therefore, conditions (i-iii) of Definition 3.1 are satisfied, which concludes the proof.

Lemma 5.3. *Suppose that Assumption 3.1 holds. Then, for each $r>0$,*

$$\limsup_{n\to\infty} \sup_{Q^{(M)}\in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)} n\left(\tau\circ\bar{T}\left(Q^{(M)}\right) - \tau(\theta_0)\right)^2 \leq 4r^2 B^*. \tag{5.5}$$

Proof. A Taylor expansion of $\tau\circ\bar{T}_{Q_n^{(M)}}$ around $\bar{T}_{Q_n^{(M)}} = \theta_0$, Lemmas 6.1 (ii) and 6.2, and Assumption 3.1 (viii) imply that for each sequence $Q_n^{(M)} \in B_H\left(P_0^{(M)}, r\sqrt{n/M}\right)$ and $r>0$,

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

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

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

For A_1, observe that

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

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

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

Combining these terms,

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

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

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

5.3. **Proof of Theorem 3.2. Proof of (i).** Pick any $\epsilon \in \left(0, r^2\right)$ and take

$$\bar{\xi} = 2\sqrt{r^2 - \epsilon} \left[\left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' \Sigma^{-1} \left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right) \right]^{-1/2} \Sigma^{-1} \frac{\partial \tau(\theta_0)}{\partial \theta}$$

Then $\frac{1}{4}\bar{\xi}'\Sigma\bar{\xi} = r^2 - \epsilon$, and hence $\widetilde{P}^{\infty}_{\theta_0 + \bar{\xi}/\sqrt{n}} \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)$ for all n large enough by Lemma 5.2. Also, $E_{\widetilde{P}^{(1,t)}_{\theta_0+\bar{\xi}/\sqrt{n}}}\left[g\left(X_t, \theta_0 + \bar{\xi}/\sqrt{n}\right)\right] = 0$ from (5.3). Hence, $\widetilde{P}_{\theta_n}$ satisfies the conditions imposed on measure Q_n in Assumption 3.2. Then we have

$$\begin{aligned}
&\lim_{\kappa\to\infty} \liminf_{n\to\infty} \sup_{Q \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)} \int \kappa \wedge n \left(\tau \circ T_a\left(\{x_1, \ldots, x_n\}\right) - \tau(\theta_0)\right)^2 dQ \\
\geq\ & \lim_{\kappa\to\infty} \liminf_{n\to\infty} \int \kappa \wedge n \left(\tau \circ T_a\left(\{x_1, \ldots, x_n\}\right) - \tau(\theta_0)\right)^2 d\widetilde{P}^{(n,1)}_{\theta_0+\bar{\xi}/\sqrt{n}} \\
=\ & \lim_{\kappa\to\infty} \liminf_{n\to\infty} \int \kappa \wedge n \left(\left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' \left(\bar{\xi} + \frac{1}{\sqrt{n}} \sum \varphi_n(X_t)\right) \right)^2 d\widetilde{P}^{(n,1)}_{\theta_0+\bar{\xi}/\sqrt{n}} \\
=\ & \left(\left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' \bar{\xi} \right)^2 + \left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' A_{\varphi\varphi'} \left(\frac{\partial \tau(\theta_0)}{\partial \theta}\right)' \\
\geq\ & \left\{1 + 4\left(r^2 - \epsilon\right)\right\} B^*
\end{aligned}$$

where the first inequality follows from $\widetilde{P}_{\theta_n} \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)$, the first equality follows from the assumption on T_a, Taylor expansion of $\tau \circ T_a$ around $T_a = \theta_0$, and the continuous mapping theorem, the second equality follows from Assumption 3.2, and the second inequality follows from the fact that

$A_{\varphi\varphi'} - \Sigma^{-1}$ is positive-semidefinite and a direct calculation. Since ϵ can be arbitrarily small, we obtain the conclusion.

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

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

for each $\kappa > 0$, where the inequality follows from the triangle inequality and $\kappa \wedge (c_1 + c_2) \leq \kappa \wedge c_1 + \kappa \wedge c_2$ for any $c_1, c_2 \geq 0$. Denote $\mathcal{X}_n^n = \{(x_1, \ldots, x_n) \in \mathcal{X}^n : \sup_{\theta\in\Theta} |g(x_t, \theta)| \leq m_n,\ t = 1, \ldots, n\}$. For A_1, Markov's inequality yields

$$\begin{aligned}
A_1 &= \limsup_{n\to\infty} \sup_{Q\in\mathcal{B}\left(P_0, r\sqrt{M/n}\right)} \left\{\int_{(x_1,\ldots,x_n)\in\mathcal{X}_n^n} \kappa \wedge n\left(\tau \circ T\left(P_n^{(M)}\right) - \tau \circ \bar{T}\left(P_n^{(M)}\right)\right)^2 dQ \right.\\
&\quad \left. + \int_{(x_1,\ldots,x_n)\notin\mathcal{X}_n^n} \kappa \wedge n\left(\tau \circ T\left(P_n^{(M)}\right) - \tau \circ \bar{T}\left(P_n^{(M)}\right)\right)^2 dQ \right\} \\
&\leq \kappa \times \limsup_{n\to\infty} \sup_{Q\in\mathcal{B}\left(P_0, r\sqrt{M/n}\right)} \int_{(x_1,\ldots,x_n)\notin\mathcal{X}_n^n} dQ \\
&\leq \kappa \times \limsup_{n\to\infty} \sup_{Q\in\mathcal{B}\left(P_0, r\sqrt{M/n}\right)} m_n^{-\eta} \sum_{t=1}^n E_Q\left[\sup_{\theta\in\Theta} |g(X_t, \theta)|^\eta\right] \\
&\leq \kappa \times \lim_{n\to\infty} Cnm_n^{-\eta} = 0,
\end{aligned}$$

where the second inequality follows from Markov inequality, and the third inequality follows from Definition 3.1(iii). A similar argument proves that $A_2 = 0$.

Thus, it is sufficient to show that$A_3 \leq \left(1 + 4r^2\right) B^*$, as $\kappa \to \infty$. Pick any $\kappa > 0$. Consider the mapping $f_{\kappa,n}(Q) = \int \kappa \wedge n\left(\tau \circ \bar{T}\left(P_n^{(M)}\right) - \tau\left(\theta_0\right)\right)^2 dQ$. For any $\epsilon > 0$ and for all $n \in \mathbb{N}$ by definition of supremum there exists $\tilde{Q}_n \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)$ such that

$$\sup_{Q_n \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)} f_{\kappa,n}\left(Q_n\right) \leq f_{\kappa,n}\left(\tilde{Q}_n\right) + \ \epsilon/n$$

for each n. Then we have

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

where the third equality follows from Lemma 6.10 and the continuous mapping theorem, the second inequality follows from $\kappa \wedge c \leq c$ and a direct calculation. Here,

$$\tilde{t} = \sqrt{n}\left(\tau \circ \left(\frac{1}{n_B}\sum_{j=1}^{n_B} \bar{T}_{\tilde{Q}_n^{(M,(j-1)L+1)}}\right) - \tau\left(\theta_0\right)\right),$$

which satisfies,

$$\begin{aligned}
\tilde{t}^2 &\leq \limsup_{n\to\infty} \sup_{Q \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)} n\left(\tau \circ \left(\frac{1}{n_B}\sum_{j=1}^{n_B} \bar{T}_{Q^{(M,(j-1)L+1)}}\right) - \tau\left(\theta_0\right)\right)^2 \\
&= \limsup_{n\to\infty} \sup_{Q \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)} \left(\frac{1}{n_B}\sum_{j=1}^{n_B} \sqrt{n}\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'\left(\bar{T}_{Q^{(M,(j-1)L+1)}} - \theta_0\right) + o\left(1\right)\right)^2 \\
&\leq \limsup_{n\to\infty} \sup_{Q^{(M)} \in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)} \left(\sqrt{n}\left(\frac{\partial\tau\left(\theta_0\right)}{\partial\theta}\right)'\left(\bar{T}_{Q^{(M)}} - \theta_0\right) + o\left(1\right)\right)^2 \\
&\leq \limsup_{n\to\infty} \sup_{Q^{(M)} \in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)} n\left(\tau \circ \left(\bar{T}_{Q^{(M)}}\right) - \tau\left(\theta_0\right)\right)^2 \leq 4r^2 B^*,
\end{aligned}$$

where the equality follows from Lemma 6.1(ii) and Assumption 3.1 (viii), the second inequality follows from the inclusion relationship $Q_n^{(M,(j-1)L+1)} \in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)$ for all $j \in 1, \ldots, n_B$, which follows from Definition 3.1(i), and the third inequality follows from Lemma 6.2, Assumption 3.1 (viii), and the fact that $B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)$ is a compact for all n, and the last inequality follows from Lemma 5.3. Hence $A_3 \leq \left(1 + 4r^2\right) B^*$, which concludes the proof.

6. Auxiliary Lemmas

Lemma 6.1. *Suppose that Assumption 3.1 holds. For each n (and correspondingly M) denote the set of all probability measures on the Borel σ-field $\mathcal{B}\left(\mathcal{X}^M\right)$ on $\mathcal{X}^M$ by $\mathcal{M}^{(M)}$. Then*

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

Proof of (i). Pick any $n\in\mathbb{N}$, $j\in\mathbb{N}$, and $Q^{(M)}\in\mathcal{M}^{(M)}$. Denote $R_n\left(Q^{(M)},\theta\right)=\inf_{P\in\bar{\mathcal{P}}_\theta^{(M)}} H\left(P^{(M)},Q^{(M)}\right)$. Since $\phi_n(b,\theta)$ is bounded for each $n\in\mathbb{N}$ and $\theta\in\Theta$, the duality of partially finite programming (Borwein and Lewis (1993)) yields that for each $(Q,\theta)\in\mathcal{M}\times\Theta$,

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

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

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

$$\begin{aligned}\sup_{\theta\in\Theta}\left|E_{Q_n^{(M)}}\left[\phi_n(B,\theta)\right]-E_{P_0^{(M)}}\left[\phi(B,\theta)\right]\right| &\le \sup_{\theta\in\Theta}\left|E_{Q_n^{(M)}}\left[\phi_n(B,\theta)\right]-E_{P_0^{(M)}}\left[\phi_n(B,\theta)\right]\right| \\ &\quad+\sup_{\theta\in\Theta}\left|E_{P_0^{(M)}}\left[\phi(B,\theta)\,\mathbb{I}\left\{B\notin\mathcal{X}_n^M\right\}\right]\right|.\end{aligned} \tag{6.2}$$

The first term of (6.2) satisfies

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

where the first inequality follows from the triangle inequality, the second inequality follows from $Q_n^{(M)} \in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)$ and Cauchy-Schwarz inequality, and the equality follows from Assumption 3.1 (v) and condition (5.2). The second term of (6.2) satisfies

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|E_{P_0^{(M)}}\left[\phi\left(B,\theta\right)\mathbb{I}\left\{B\notin\mathcal{X}_n^M\right\}\right]\right| \\
\leq\;& \sqrt{M}\left(\int\sup_{\theta\in\Theta}\left|g\left(x,\theta\right)\right|^{\eta}dP_0^{(1)}\right)^{1/\eta}\left(M\int\mathbb{I}\left\{\left|g\left(x,\theta\right)\right|>m_n\right\}dP_0^{(1)}\right)^{(\eta-1)/\eta} \\
\leq\;& \sqrt{M}\left(E_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}\left|g\left(X_1,\theta\right)\right|^{\eta}\right]\right)^{1/\eta}\left(Mm_n^{-\eta}E_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}\left|g\left(X_1,\theta\right)\right|^{\eta}\right]\right)^{(\eta-1)/\eta}=o\left(\sqrt{M/n}\right),
\end{aligned}$$

where the first inequality follows from Holder's inequality and the definitions of $\phi\left(\cdot\right)$ and $\mathcal{X}_n^M$, the second inequality follows from Markov inequality, and the equality follows from Assumption 3.1 (v) and condition (5.2). Combining these results, we obtain the uniform convergence $\sup_{\theta\in\Theta}\left|E_{Q_n^{(M)}}\left[\phi_n\left(B,\theta\right)\right]-E_{P_0^{(M)}}\left[\phi\left(B,\theta\right)\right]\right|\to 0$. From the triangle inequality,

$$\left|E_{P_0^{(M)}}\left[\phi\left(B,\bar{T}_{Q_n^{(M)}}\right)\right]\right|\leq\left|E_{P_0^{(M)}}\left[\phi\left(b,\bar{T}_{Q_n^{(M)}}\right)\right]-E_{Q_n^{(M)}}\left[\phi_n\left(b,\bar{T}_{Q_n^{(M)}}\right)\right]\right|+\left|E_{Q_n^{(M)}}\left[\phi_n\left(b,\bar{T}_{Q_n^{(M)}}\right)\right]\right|\to 0.$$

The conclusion is obtained from Lemma 6.6 (i) and Assumption 3.1 (iii).

Lemma 6.2. *Suppose that Assumption 3.1 holds. Then, for each $r>0$ and each sequence $Q_n^{(M)}\in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)$,*

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

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

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

where the second equality follows from

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

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

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

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

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

This implies

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

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

We now analyze $\psi_{n,Q_n^{(M)}}$. From the definition of $\psi_{n,Q_n^{(M)}}$,

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

Note that from the convex duality of partially finite programming (Borwein and Lewis (1993)), the Radon-Nikodym derivative the $d\bar{P}_{\theta,Q^{(M)}}^{(M)}/dQ^{(M)}$ can be written as

$$\frac{d\bar{P}_{\theta,Q^{(M)}}^{(M)}}{dQ^{(M)}} = \frac{1}{\left(1+\gamma_n\left(\theta,Q^{(M)}\right)'\phi_n\left(b,\theta\right)\right)^2}, \tag{6.7}$$

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

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

where

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

Thus, if $\int\phi_n\left(b,\theta\right)\phi_n\left(b,\theta\right)'dQ^{(M)}$ is invertible, $\gamma_n\left(\theta,Q^{(M)}\right)$ is written as

$$\begin{aligned}\gamma_n\left(\theta,Q^{(M)}\right) &= \frac{1}{2}\left(\int\phi_n\left(b,\theta\right)\phi_n\left(b,\theta\right)'dQ^{(M)}\right)^{-1}\int\phi_n\left(b,\theta\right)dQ^{(M)}\\ &\quad+\left(\int\phi_n\left(b,\theta\right)\phi_n\left(b,\theta\right)'dQ^{(M)}\right)^{-1}\int\varrho_n\left(x,\theta,Q^{(M)}\right)\phi_n\left(b,\theta\right)dQ^{(M)}.\end{aligned} \tag{6.8}$$

The second term of (6.6) is

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

where the first equality follows from (6.7), the second equality follows from (6.8) and Lemma 6.5, and the third equality follows from Lemma 6.5. Similarly, the first term of (6.6) is $o\left(n^{-1/2}\right)$. Therefore,

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

and $\left|\psi_{n,Q_n^{(M)}}\right| = O\left(n^{-1/2}\right)$ from Lemma 6.5. Then from (6.5),

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

By solving for $\sqrt{n}\left(\bar{T}_{Q_n^{(M)}} - \theta_0\right)$, the conclusion is obtained. The above also shows that $\bar{T}_{Q_n^{(M)}} - \theta_0 = O(1/\sqrt{n})$.

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

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

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

Proof of (i). Denote $t_n = \bar{T}_{Q_n^{(M)}} - \theta_0$. Pick any $Q_n^{(M)} \in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)$ and $r > 0$. From (6.7),

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

For T_2, Lemmas 6.5 and 6.6 imply

$$T_2 \leq o\left(1\right) \left\| \gamma_n\left(\bar{T}_{Q_n^{(M)}}, Q_n\right)' \phi_n\left(b, \bar{T}_{Q_n^{(M)}}\right) dQ_n^{1/2} + \gamma_n\left(\theta_0, Q_n^{(M)}\right)' \phi_n\left(b,\theta_0\right) dQ_n^{1/2} \right\| = o\left(\sqrt{\frac{M}{n}}\right).$$

Thus, we focus on T_1. From (6.8),

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

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

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

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

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

Lemma 6.4. *Suppose that Assumption 3.1 holds. Then, for each* $\xi \in \mathbb{R}^p$, $|E_{P_0}[\phi_n(B,\theta_0)]| = o\left(\sqrt{M/n}\right)$, $|E_{P_0}[\phi_n(B,\theta_n)]| = O\left(\sqrt{M/n}\right)$, $\left|E_{P_0}\left[\phi_n(B,\theta_n)\,\phi_n(B,\theta_n)'\right] - \Omega\right| = o(1)$, *and* $\left|E_{P_0}\left[\partial\phi_n(B,\theta_n)/\partial\theta'\right] - \sqrt{M}G\right| = o\left(\sqrt{M}\right)$.

Proof of the first statement. Observe that,

$$\begin{aligned} |E_{P_0}[\phi_n(B,\theta_0)]| &= \left| \int \frac{1}{\sqrt{M}} \sum_{i=1}^{M} g(x_i,\theta_0)\, \mathbb{I}\left\{ b \notin \mathcal{X}_n^M \right\} dP_0^{(M)} \right| \\ \leq & \left| \int |\phi(b,\theta_0)|\, \mathbb{I}\left\{ b \notin \mathcal{X}_n^M \right\} dP_0^{(M)} \right| \\ \leq & \sqrt{M} \left(\int |g(x,\theta_0)|^4 dP_0^{(1)} \right)^{1/4} \left(\int \mathbb{I}\left\{ b \notin \mathcal{X}_n^M \right\} dP_0^{(M)} \right)^{3/4} \\ \leq & \sqrt{M} \left(\int |g(x,\theta_0)|^4 dP_0^{(1)} \right)^{1/4} \left(M \int \mathbb{I}\left\{ \sup_{\theta\in\mathcal{U}} |g(x,\theta)| > m_n \right\} dP_0^{(1)} \right)^{3/4} \\ \leq & \sqrt{M} C \left(M m_n^{-4} E_{P_0^{(1)}}\left[\sup_{\theta\in\mathcal{U}} |g(X_t,\theta_0)|^4 \right] \right)^{3/4} = o\left(\sqrt{M/n}\right), \end{aligned} \tag{6.10}$$

where the second inequality follows from Holder's inequality, the third inequality follows from the definition of $\mathcal{X}_n^M$, the fourth inequality follows from Markov inequality, and the second equality follows from Assumption 3.1 (v) and condition (5.2).

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

$$\left|E_{P_0}\left[\phi_n\left(B,\theta_n\right)\right]\right| \leq \left|\int \phi\left(b,\theta_n\right)\mathbb{I}\left\{b \notin \mathcal{X}_n^M\right\} dP_0^{(M)}\right| + \left|E_{P_0}\left[\phi\left(B,\theta_n\right)\right]\right|. \tag{6.11}$$

By the same argument as (6.10), the first term of (6.11) is $o\left(\sqrt{M/n}\right)$ (note that $E_{P_0}\left[\left|g\left(X,\theta_n\right)\right|^{\eta}\right] < \infty$ from Assumption 3.1 (v)). The second term of (6.11) satisfies

$$\left|E_{P_0}\left[\phi\left(B,\theta_n\right)\right]\right| \leq E_{P_0}\left[\sqrt{M}\sup_{\theta\in\mathcal{U}}\left|\frac{\partial g\left(X_t,\theta\right)}{\partial\theta}\right|\right]\left|\frac{\xi}{\sqrt{n}}\right| = O\left(\sqrt{M/n}\right),$$

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

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

$$\begin{aligned}
& \left|E_{P_0}\left[\phi_n\left(B,\theta_n\right)\phi_n\left(B,\theta_n\right)'\right] - \Omega\right| \\
\leq\ & \left|E_{P_0}\left[\phi_n\left(B,\theta_n\right)\phi_n\left(B,\theta_n\right)'\right] - E_{P_0}\left[\phi\left(B,\theta_n\right)\phi\left(B,\theta_n\right)'\right]\right| \\
& + \left|E_{P_0}\left[\phi\left(B,\theta_n\right)\phi\left(B,\theta_n\right)'\right] - E_{P_0}\left[\phi\left(B,\theta_0\right)\phi\left(B,\theta_0\right)'\right]\right| \\
& + \left|E_{P_0}\left[\phi\left(B,\theta_0\right)\phi\left(B,\theta_0\right)'\right] - \Omega\right| \\
\leq\ & \left|\int \phi\left(b,\theta_n\right)\phi'\left(b,\theta_n\right)'\mathbb{I}\left\{b \notin \mathcal{X}_n^M\right\} dP_0\right| + o(1) + o(1) \\
\leq\ & M\sqrt{\int \left|g\left(x,\theta_n\right)\right|^4 dP_0}\sqrt{\int I\left\{b \notin \mathcal{X}_n^M\right\} dP_0} + o(1) \\
\leq\ & M^{3/2}C\sqrt{m_n^{-4}E_{P_0}\left[\sup_{\theta\in\mathcal{U}}\left|g\left(X_t,\theta\right)\right|^4\right]} + o(1) \\
\leq\ & M^{3/2}Cm_n^{-2} + o(1) = o(1),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the continuity of $g(x,\theta)$ at θ_0 and the definition of Ω, the third inequality follows from Cauchy-Schwarz inequality for n large enough, the fifth inequality follows from Markov inequality, and the last line follows from Assumption 3.1 (v) and (ix), and condition (5.2).

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

$$\begin{aligned}
& \left|E_{P_0}\left[\partial \phi_n\left(B, \theta_n\right) / \partial \theta^{\prime}\right]-E_{P_0}\left[\partial \phi\left(B, \theta_0\right) / \partial \theta^{\prime}\right]\right| \\
\leq & \left|E_{P_0}\left[\partial \phi_n\left(B, \theta_n\right) / \partial \theta^{\prime}\right]-E_{P_0}\left[\partial \phi\left(B, \theta_n\right) / \partial \theta^{\prime}\right]\right| \\
& +\sqrt{M}\left|E_{P_0}\left[\partial g\left(X_t, \theta_n\right) / \partial \theta^{\prime}\right]-E_{P_0}\left[\partial g\left(X_t, \theta_0\right) / \partial \theta^{\prime}\right]\right| \\
\leq \sqrt{M} & \sqrt{E_{P_0}\left[\sup _{\theta \in \mathcal{U}}\left|\frac{\partial g\left(X_t, \theta\right)}{\partial \theta^{\prime}}\right|^2\right]} \sqrt{m_n^{-\eta} E_{P_0}\left[\sup _{\theta \in \Theta}\left|g\left(X_t, \theta\right)\right|^\eta\right]}+o(\sqrt{M})=o(\sqrt{M}),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the triangle, Cauchy-Schwarz, and Markov inequalities and the continuity of $\partial g(x, \theta) / \partial \theta^{\prime}$ at θ_0, and the equality follows from Assumption 3.1 (v) and condition (5.2).

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

(i): $\left|E_{Q_n^{(M)}}\left[\phi_n\left(B, \theta_0\right)\right]\right|=O(\sqrt{M / n})$, *and* $\left|E_{Q_n^{(M)}}\left[\phi_n\left(B, \theta_0\right) \phi_n\left(B, \theta_0\right)^{\prime}\right]-\Omega\right|=o(1)$,

(ii): $\gamma_n\left(\theta_0, Q_n^{(M)}\right)=\arg \max _{\gamma \in \mathbb{R}^m}-\int \frac{1}{\left(1+\gamma^{\prime} \phi_n\left(b, \theta_0\right)\right)} d Q_n^{(M)}$ *exists for all n large enough, and* $\left|\gamma_n\left(\theta_0, Q_n^{(M)}\right)\right|=O(\sqrt{M / n})$, *and* $\sup _{b \in \mathcal{X}^M}\left|\gamma_n\left(\theta_0, Q_n^{(M)}\right)^{\prime} \phi_n\left(b, \theta_0\right)\right| \rightarrow 0$.

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

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

where the first and second inequalities follow from the triangle inequality and Lemma 6.4, the third inequality follows from Cauchy-Schwartz inequality and $Q_n^{(M)} \in B_H\left(P_0^{(M)}, r \sqrt{M / n}\right)$, and the equality follows from Assumption 3.1 (v) and condition (5.2).

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

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

where the first inequality follows from the triangle inequality, the second inequality follows from the triangle inequality and Lemma 6.8 (i), the third inequality follows from Cauchy-Schwarz inequality and $Q_n^{(M)} \in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)$, and the equality follows from Assumption 3.1 (v) and condition (5.2).

Proof of (ii). The proof is based on Newey and Smith (2004, proofs of Lemmas A.1-3). Pick any $\xi \in \mathbb{R}^p$. Define

$$\Gamma_n = \{\gamma \in \mathbb{R}^m : |\gamma| \leq a_n\} \tag{6.12}$$

with $a_n\sqrt{M}m_n \to 0$ and $a_n\sqrt{n/M} \to \infty$. Observe that

$$\sup_{\gamma\in\Gamma_n, b\in\mathcal{X}^M, \theta\in\Theta}\left|\gamma'\phi_n(b,\theta)\right| \leq a_n\sqrt{M}m_n \to 0. \tag{6.13}$$

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

$$\begin{aligned}
-1 &= R_n\left(Q_n^{(M)},\theta_n,0\right) \leq R_n\left(Q_n^{(M)},\theta_n,\tilde{\gamma}\right) \\
&= -1+\tilde{\gamma}'E_{Q_n^{(M)}}\left[\phi_n(B,\theta_n)\right]-\tilde{\gamma}'E_{Q_n^{(M)}}\left[\frac{\phi_n(B,\theta_n)\phi_n(B,\theta_n)'}{\left(1+\dot{\gamma}'\phi_n(B,\theta_n)\right)^3}\right]\tilde{\gamma} \\
&\leq -1+\tilde{\gamma}'E_{Q_n^{(M)}}\left[\phi_n(B,\theta_n)\right]-C\tilde{\gamma}'E_{Q_n^{(M)}}\left[\phi_n(B,\theta_n)\phi_n(B,\theta_n)'\right]\tilde{\gamma} \\
&\leq -1+|\tilde{\gamma}|\left|E_{Q_n^{(M)}}\left[\phi_n(B,\theta_n)\right]\right|-C|\tilde{\gamma}|^2, \qquad (6.14)
\end{aligned}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, the second inequality follows from (6.13), and the last inequality follows from Lemma 6.4 and Assumption 3.1 (vi). Thus, Lemma 6.5 (i) implies

$$C|\tilde{\gamma}| \leq \left|E_{Q_n^{(M)}}\left[\phi_n(B,\theta_n)\right]\right| = O\left(\sqrt{M/n}\right). \tag{6.15}$$

From $a_n\sqrt{M}n^{1/2} \to \infty$, $\tilde{\gamma}$ is an interior point of Γ_n and satisfies the first-order condition $\partial R_n\left(Q_n^{(M)},\theta_0,\tilde{\gamma}\right)/\partial\gamma = 0$ for all n large enough. Since $R_n\left(Q_n^{(M)},\theta_0,\gamma\right)$ is concave in γ for all n large enough, $\tilde{\gamma} = \arg\max_{\gamma\in\mathbb{R}^m} R_n\left(Q_n^{(M)},\theta_n,\gamma\right)$ for all n large enough. Thus, the first statement is obtained. Also, from (6.15), the second statement is obtained. Using condition (5.2), the third statement follows from

$$\sup_{b\in\mathcal{X}^M}\left|\gamma_n\left(\theta_n,Q_n^{(M)}\right)'\phi_n\left(b,\theta_n\right)\right| \le O\left(Mn^{-1/2}m_n\right) = o\left(1\right). \tag{6.16}$$

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

(i): $\left|E_{Q_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)\right]\right| = O\left(\sqrt{M/n}\right)$, $\left|E_{Q_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)'\right] - \Omega\right| = o\left(1\right)$, *and*
$\left|E_{Q_n^{(M)}}\left[\partial\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\right] - \sqrt{M}G\right| = o\left(\sqrt{M}\right)$,

(ii): $\gamma_n\left(\bar{T}_{Q_n^{(M)}},Q_n\right) = \arg\max_{\gamma\in\mathbb{R}^m} -\int\frac{1}{\left(1+\gamma'\phi_n\left(b,\bar{T}_{Q_n^{(M)}}\right)\right)}dQ_n$ *exists for all n large enough,*
$\left|\gamma_n\left(\bar{T}_{Q_n^{(M)}},Q_n\right)\right| = O\left(\sqrt{M/n}\right)$, *and* $\sup_{b\in\mathcal{X}^{(\mathcal{M})}}\left|\gamma_n\left(\bar{T}_{Q_n^{(M)}},Q_n\right)'\phi_n\left(b,\bar{T}_{Q_n^{(M)}}\right)\right| \to 0.$

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

$$\sup_{b\in\mathcal{X}^M,\theta\in\Theta}\left|\tilde{\gamma}'\phi_n\left(b,\theta\right)\right| \le Mm_n/\sqrt{n} \to 0. \tag{6.17}$$

Observe that,

$$
\begin{aligned}
& \left| E_{Q_n^{(M)}} \left[\phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right)' \right] \right| \\
\leq \; & \int \sup_{\theta \in \Theta} \left| \phi_n (b, \theta) \right|^2 \left\{ dQ_n^{1/2} - dP_0^{1/2} \right\}^2 + 2 \int \sup_{\theta \in \Theta} \left| \phi_n (b, \theta) \right|^2 dP_0^{1/2} \left\{ dQ_n^{1/2} - dP_0^{1/2} \right\} \\
& + E_{P_0} \left[\sup_{\theta \in \Theta} \left| \phi_n (B, \theta) \right|^2 \right] \\
\leq \; & M m_n^2 r^2 \frac{M}{n} + 2 m_n \sqrt{M} E \left(\sum_{j=1}^{M} E_{P_0} \left[\sup_{\theta \in \Theta} \left| g(X_1, \theta) g(X_j, \theta)' \right| \right] \right)^{1/2} r \sqrt{\frac{M}{n}} \\
& + \sum_{j=1}^{M} E_{P_0^{(M)}} \left[\sup_{\theta \in \Theta} \left| g(X_1, \theta) g(X_j, \theta)' \right| \right] \\
\leq \; & 2 \left(r^2 \frac{M^2 m_n^2}{n} + \sum_{j=1}^{M} 12 \alpha (j-1)^{1-2/\eta} E_{P_0} \left[\sup_{\theta \in \Theta} \left| g(X_1, \theta) \right|^{\eta} \right]^{2/\eta} \right), \qquad (6.18)
\end{aligned}
$$

and hence is bounded due to condition (5.2) and Assumption 3.1 (i,v). Here the first inequality follows from the triangle inequality, the second inequality follows from Cauchy-Schwartz inequality and $Q_n^{(M)} \in B_H \left(P_0^{(M)}, r\sqrt{M/n} \right)$, and the third inequality follows from Davydov (1968).

A Taylor expansion around $\tilde{\gamma} = 0$ yields $R_n \left(Q_n, \bar{T}_{Q_n^{(M)}}, \tilde{\gamma} \right)$

$$
\begin{aligned}
= \; & -1 + \tilde{\gamma}' E_{Q_n^{(M)}} \left[\phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \right] - \tilde{\gamma}' E_{Q_n^{(M)}} \left[\frac{\phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right)'}{\left(1 + \dot{\gamma}' \phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \right)^3} \right] \tilde{\gamma} \\
\geq \; & -1 + \sqrt{M/n} \left| E_{Q_n^{(M)}} \left[\phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \right] \right| - C \tilde{\gamma}' E_{Q_n^{(M)}} \left[\phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right)' \right] \tilde{\gamma} \\
\geq \; & -1 + \sqrt{M/n} \left| E_{Q_n^{(M)}} \left[\phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \right] \right| - CM/n, \qquad (6.19)
\end{aligned}
$$

for all n large enough, where $\dot{\gamma}$ is a point one the line joining 0 and $\tilde{\gamma}$, the first inequality follows from (6.17), and the second inequality follows from $\tilde{\gamma}'\tilde{\gamma} = M/n$ and (6.18). From the definitions of $\gamma_n \left(\bar{T}_{Q_n^{(M)}}, Q_n \right)$ and $\bar{T}_{Q_n^{(M)}}$,

$$
\begin{aligned}
& -1 + \sqrt{\frac{M}{n}} \left| E_{Q_n^{(M)}} \left[\phi_n \left(B, \bar{T}_{Q_n^{(M)}} \right) \right] \right| - C \frac{M}{n} \\
\leq \; & R_n \left(Q_n, \bar{T}_{Q_n^{(M)}}, \tilde{\gamma} \right) \leq R_n \left(Q_n, \bar{T}_{Q_n^{(M)}}, \gamma_n \left(\bar{T}_{Q_n^{(M)}}, Q_n \right) \right) \leq R_n \left(Q_n, \theta_0, \gamma_n \left(\theta_0, Q_n^{(M)} \right) \right), \qquad (6.20)
\end{aligned}
$$

where the first inequality follows from (6.19). From $\left|\gamma_n\left(\theta_0, Q_n^{(M)}\right)\right| = O\left(\sqrt{M/n}\right)$ and $\left|E_{Q_n^{(M)}}\left[\phi_n\left(B,\theta_0\right)\right]\right| = O\left(\sqrt{M/n}\right)$ (by Lemma 6.5 (ii) and (i)), similar to (6.14) we have $R_n\left(Q_n, \theta_0, \gamma_n\left(\theta_0, Q_n^{(M)}\right)\right)$

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

Combining (6.20) and (6.21), we have $\left|E_{Q_n^{(M)}}\left[\phi_n\left(B, \bar{T}_{Q_n^{(M)}}\right)\right]\right| = O\left(\sqrt{M/n}\right)$.

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

$$\begin{aligned}
&\left|E_{Q_n^{(M)}}\left[\phi_n\left(B, \bar{T}_{Q_n^{(M)}}\right)\phi_n\left(B, \bar{T}_{Q_n^{(M)}}\right)'\right] - E_{P_0^{(M)}}\left[\phi\left(B,\theta_0\right)\phi\left(B,\theta_0\right)'\right]\right| \\
\leq\ &\left|E_{Q_n^{(M)}}\left[\phi_n\left(B, \bar{T}_{Q_n^{(M)}}\right)\phi_n\left(B, \bar{T}_{Q_n^{(M)}}\right)'\right] - E_{P_0^{(M)}}\left[\phi_n\left(B, \bar{T}_{Q_n^{(M)}}\right)\phi_n\left(B, \bar{T}_{Q_n^{(M)}}\right)'\right]\right| \\
&+\left|E_{P_0^{(M)}}\left[\phi\left(B, \bar{T}_{Q_n^{(M)}}\right)\phi\left(B, \bar{T}_{Q_n^{(M)}}\right)'\mathbb{I}\left\{B \notin \mathcal{X}_n^M\right\}\right]\right| \\
&+\left|E_{P_0^{(M)}}\left[\phi\left(B, \bar{T}_{Q_n^{(M)}}\right)\phi\left(B, \bar{T}_{Q_n^{(M)}}\right)'\right] - E_{P_0^{(M)}}\left[\phi\left(B,\theta_0\right)\phi\left(B,\theta_0\right)'\right]\right|.
\end{aligned} \tag{6.22}$$

The first term of (6.22) satisfies

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

for all n large enough, where the first inequality follows from the triangle inequality, the second inequality follows from the consistency of $\bar{T}_{Q_n^{(M)}}$ (Lemma 6.1 (ii)) and Cauchy-Schwartz inequality, and $Q_n^{(M)} \in B_H\left(P_0^{(M)}, r\sqrt{M/n}\right)$, and the equality follows from Assumption 3.1 (v) and condition (5.2). The second term of (6.22) satisfies

$$\begin{aligned}
&\left|E_{P_0^{(M)}}\left[\phi\left(B, \bar{T}_{Q_n^{(M)}}\right)\phi\left(B, \bar{T}_{Q_n^{(M)}}\right)'\mathbb{I}\left\{B \notin \mathcal{X}_n^M\right\}\right]\right| \\
\leq\ & M^{3/2}\left(\int \sup_{\theta\in\mathcal{U}}\left|g\left(x,\theta\right)g\left(x,\theta\right)'\right|^2 dP_0\right)^{1/2}\left(\int \mathbb{I}\left\{\left|g\left(x,\theta\right)\right| > m_n\right\} dP_0\right)^{1/2} \\
\leq\ & M^{3/2}\left(E_{P_0}\left[\sup_{\theta\in\mathcal{U}}\left|g\left(X,\theta\right)\right|^4\right]\right)^{1/2}\left(m_n^{-4} E_{P_0}\left[\sup_{\theta\in\mathcal{U}}\left|g\left(X,\theta\right)\right|^4\right]\right)^{1/2} = o\left(1\right),
\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz inequality, the second inequality follows from Markov inequality, and the equality follows from Assumption 3.1 (v) and condition (5.2). the third term of (6.22) is $o(1)$ by the continuity of $g(x,\theta)$ at θ_0, consistency of $\bar{T}_{Q_n^{(M)}}$, Assumption 3.1 (v) and the dominated convergence theorem.

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

$$\begin{aligned}
& \left|E_{Q_n^{(M)}}\left[\partial\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\right] - E_{P_0^{(M)}}\left[\partial\phi\left(B,\theta_0\right)/\partial\theta'\right]\right| \\
\leq\; & \left|E_{Q_n^{(M)}}\left[\partial\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\right] - E_{P_0^{(M)}}\left[\partial\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\right]\right| \\
& + \left|E_{P_0^{(M)}}\left[\partial\phi\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\mathbb{I}\left\{B\notin\mathcal{X}_n^M\right\}\right]\right| \\
& + \left|E_{P_0^{(M)}}\left[\partial\phi\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\right] - E_{P_0^{(M)}}\left[\partial\phi\left(B,\theta_0\right)/\partial\theta'\right]\right|. \qquad (6.23)
\end{aligned}$$

The first term of (6.23) satisfies

$$\begin{aligned}
& \left|E_{Q_n^{(M)}}\left[\partial\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\right] - E_{P_0^{(M)}}\left[\partial\phi_n\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\right]\right| \\
\leq\; & \left|\int \partial\phi_n\left(b,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\left\{dQ_n^{1/2}-dP_0^{1/2}\right\}^2\right| + 2\left|\int \partial\phi_n\left(b,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta' dP_0^{1/2}\left\{dQ_n^{1/2}-dP_0^{1/2}\right\}\right| \\
\leq\; & \sup_{\theta\in\mathcal{U}}\left|\partial\phi_n\left(b,\theta\right)/\partial\theta'\right| r^2\frac{M}{n} + 2E_{P_0}\left[\sup_{\theta\in\mathcal{U}}\left|\partial\phi_n\left(b,\theta\right)/\partial\theta'\right|^2\right]^{1/2} r\sqrt{\frac{M}{n}} \\
=\; & O\left(\sqrt{M^3/n}\right) = o(1),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from Cauchy-Schwartz inequality, and the equality follows from Assumption 3.1 (v) and condition (5.2). The second term of (6.23) satisfies

$$\begin{aligned}
& \left|E_{P_0^{(M)}}\left[\partial\phi\left(B,\bar{T}_{Q_n^{(M)}}\right)/\partial\theta'\mathbb{I}\left\{B\notin\mathcal{X}_n^M\right\}\right]\right| \\
\leq M\; & \left(\int\sup_{\theta\in\mathcal{U}}\left|\partial g\left(x,\theta\right)/\partial\theta'\right|^2 dP_0\right)^{1/2}\left(\int\mathbb{I}\left\{\sup_{\theta\in\mathcal{U}}\left|g\left(x,\theta\right)\right|>m_n\right\}dP_0\right)^{1/2} \\
\leq\; & CM\left(m_n^{-4}E_{P_0}\left[\sup_{\theta\in\mathcal{U}}\left|g\left(X,\theta\right)\right|^4\right]\right)^{1/2} = o(1),
\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz inequality, the second inequality follows from Markov inequality, and the equality follows from Assumption 3.1 (v) and from condition (5.2). The third term of (6.23) is $o\left(\sqrt{M}\right)$ by the continuity of $\partial g(x,\theta)/\partial\theta'$ at θ_0, consistency of $\bar{T}_{Q_n^{(M)}}$, Assumption 3.1 (v) and the dominated convergence theorem. Therefore, the conclusion is obtained.

Proof of (ii). The proof is exactly as for Lemma 6.5 (ii) except using Lemma 6.6 (i) instead of Lemma 6.5 (i).

Lemma 6.7. *Suppose that Assumption 3.1 holds. Then, for each $r > 0$ and each sequence $Q_n \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)$, $\bar{T}_{P_n^{(M)}} \to_p \theta_0$ under Q_n.*

Proof. The proof is based on Newey and Smith (2004, proof of Theorem 3.1). From the triangle inequality,

$$\begin{aligned}\sup_{\theta\in\Theta}\left|E_{P_n^{(M)}}\left[\phi_n(B,\theta)\right]-E_{P_0^{(M)}}\left[\phi(B,\theta)\right]\right| &\leq \sup_{\theta\in\Theta}\left|E_{P_n^{(M)}}\left[\phi_n(B,\theta)\right]-\frac{1}{n_B}\sum_{j=1}^{n_B}E_{Q_n^{(M,(j-1)L+1)}}\left[\phi_n(B,\theta)\right]\right|\\ &\quad+\frac{1}{n_B}\sum_{j=1}^{n_B}\sup_{\theta\in\Theta}\left|E_{Q_n^{(M,(j-1)L+1)}}\left[\phi_n(B,\theta)\right]-E_{P_0^{(M)}}\left[\phi_n(B,\theta)\right]\right|\\ &\quad+\sup_{\theta\in\Theta}\left|E_{P_0^{(M)}}\left[\phi_n(B,\theta)\right]-E_{P_0^{(M)}}\left[\phi(B,\theta)\right]\right|\to_p 0\end{aligned}$$

where the convergence follows from a UWLLN for the first term, while the last two terms are shown to be $o(1)$ in the proof of Lemma 6.1 (ii) . From the first statement of Lemma 6.9 (i), $\left|E_{P_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)\right]\right| \xrightarrow{p} 0$. Thus, by the triangle inequality,

$$\left|E_{P_0^{(M)}}\left[\phi\left(B,\bar{T}_{P_n^{(M)}}\right)\right]\right| \leq \left|E_{P_0^{(M)}}\left[\phi\left(B,\bar{T}_{P_n^{(M)}}\right)\right]-E_{P_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)\right]\right|+\left|E_{P_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)\right]\right|\to_p 0.$$

The conclusion follows from Assumption 3.1 (iii).

Lemma 6.8. *Suppose that Assumption 3.1 holds. Then, for each $r > 0$ and each sequence $Q_n \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)$ the followings hold under Q_n:*

(i): $\left|E_{P_n^{(M)}}\left[\phi_n(B,\theta_0)\right]\right| = O_p\left(\sqrt{M/n}\right)$, $\left|E_{P_n^{(M)}}\left[\phi_n(B,\theta_0)\phi_n(B,\theta_0)'\right]-\Omega\right| = o(1)$,

(ii): $\gamma_n\left(\theta_0, P_n^{(M)}\right) = \arg\max_{\gamma\in\mathbb{R}^m} -\int \frac{1}{(1+\gamma'\phi_n(b,\theta_0))}dP_n^{(M)}$ *exists w.p.a.1,* $\left|\gamma_n\left(\theta_0, P_n^{(M)}\right)\right| = O_p\left(\sqrt{M/n}\right)$, *and* $\sup_{b\in\mathcal{X}^M}\left|\gamma_n\left(\theta_0, P_n^{(M)}\right)'\phi_n(b,\theta_0)\right| \to_p 0$.

Proof of (i). Observe that,

$$
\begin{aligned}
(6.24)\qquad & \left|E_{P_n^{(M)}}\left[\phi_n\left(B,\theta_0\right)\right]\right| \\
\le\ & \left|E_{P_n^{(M)}}\left[\phi_n\left(B,\theta_0\right)\right]-\frac{1}{n_B}\sum_{j=1}^{n_B}E_{Q_n^{(M,(j-1)L+1)}}\left[\phi_n\left(B,\theta_0\right)\right]\right| \\
& +\frac{1}{n_B}\sum_{j=1}^{n_B}\left|E_{Q_n^{(M,(j-1)L+1)}}\left[\phi_n\left(B,\theta_0\right)\right]\right| = O_p\left(\sqrt{M/n}\right)+O\left(\sqrt{M/n}\right),
\end{aligned}
$$

where the inequality follows from the triangle inequality, and the equality follows from the CLT and Lemma 6.5 (i).

Proof of the second statement. Pick any $r>0$ and any sequence $Q_n \in \mathcal{B}^{\infty}\left(P_0, r\sqrt{M/n}\right)$. From the triangle inequality,

$$
\begin{aligned}
& \left|\int \phi_n\left(b,\theta_0\right)\phi_n\left(b,\theta_0\right)' dP_n^{(M)}-\Omega\right| \\
\le\ & \left|\frac{1}{n_B}\sum_{j=1}^{n_B}E_{Q_n^{(M,(j-1)L+1)}}\left[\phi\left(B,\theta_0\right)\phi\left(B,\theta_0\right)'\right]-\Omega\right| \\
& +\left|\frac{1}{n_B}\sum_{j=1}^{n_B}\phi\left(B_j,\theta_0\right)\phi\left(B_j,\theta_0\right)'-\frac{1}{n_B}\sum_{j=1}^{n_B}E_{Q_n^{(M,(j-1)L+1)}}\left[\phi\left(B,\theta_0\right)\phi\left(B,\theta_0\right)'\right]\right| = o_p(1),
\end{aligned}
$$

where the first term is $o(1)$ by the triangle inequality and Lemma 6.5 (i) and the second term is $o_p(1)$ by UWLLN.

Proof of (ii). The proof is exactly as for Lemma 6.5 (ii) except using Lemma 6.8 (i) instead of Lemma 6.5 (i).

Lemma 6.9. *For each $r>0$ and each sequence $Q_n \in \mathcal{B}\left(P_0, r\sqrt{M/n}\right)$ the followings hold under Q_n:*

- **(i):** $\left|E_{P_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)\right]\right| = O_p\left(\sqrt{M/n}\right)$, $\left|E_{P_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)'\right]-\Omega\right| = o(1)$, *and* $\left|E_{P_n^{(M)}}\left[\partial\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)/\partial\theta'\right]-G\right| = o\left(\sqrt{M}\right)$,
- **(ii):** $\gamma_n\left(\bar{T}_{P_n^{(M)}}, P_n^{(M)}\right) = \arg\max_{\gamma\in\mathbb{R}^m} -\int \frac{1}{\left(1+\gamma'\phi_n\left(b,\bar{T}_{P_n^{(M)}}\right)\right)} dP_n^{(M)}$ *exists w.p.a.1,* $\left|\gamma_n\left(\bar{T}_{P_n^{(M)}}, P_n^{(M)}\right)\right| = O_p\left(\sqrt{M/n}\right)$, *and* $\sup_{b\in\mathcal{X}^M}\left|\gamma_n\left(\bar{T}_{P_n^{(M)}}, P_n^{(M)}\right)'\phi_n\left(b,\bar{T}_{P_n^{(M)}}\right)\right| \to_p 0.$

Proof of (i). By UWLLN,

$$\sup_{\theta\in\Theta}\left|E_{P_n^{(M)}}\left[\phi_n(B,\theta)\phi_n(B,\theta)'\right]-\frac{1}{n_B}\sum_{j=1}^{n_B}E_{Q_n^{(M,(j-1)L+1)}}\left[\phi_n(B,\theta)\phi_n(B,\theta)'\right]\right|\to_p 0, \tag{6.25}$$

Then, from the calculation of the last 3 lines of (6.18) and the triangle inequality, w.p.a.1 the following holds

$$\sup_{\theta\in\Theta}\left|E_{P_n^{(M)}}\left[\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)\phi_n\left(B,\bar{T}_{P_n^{(M)}}\right)'\right]\right|<CE_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}|g(X,\theta)|^\eta\right]^{2/\eta}.$$

From here the proof of the first statement is the same as for the first statement of Lemma 6.6 (i) except using Lemma 6.8 instead of Lemma 6.5.

The second statement follows from (6.25) and Lemma 6.6 (i). The third statement of the lemma follows from continuity $\partial\phi_n(x,\theta)/\partial\theta'$ at θ_0, Lemma 6.7 and Lemma 6.6 (i).

Proof of (ii). The proof is similar to the proof of Lemma 6.5 (ii) except using Lemma 6.9 (i) instead of Lemma 6.5 (i).

Lemma 6.10. *Suppose that Assumption 3.1 holds. Then, for each $r>0$ and each sequence $Q_n\in\mathcal{B}\left(P_0,r\sqrt{M/n}\right)$,*

$$\sqrt{n}\left(\bar{T}_{P_n^{(M)}}-\theta_0\right)=-\sqrt{n}(M\Sigma)^{-1}\int\Lambda_n dP_n^{(M)}+o_p(1)\quad\textit{under } Q_n, \tag{6.26}$$

$$\sqrt{n}\left(\bar{T}_{P_n^{(M)}}-\frac{1}{n_B}\sum_{j=1}^{n_B}\bar{T}_{Q_n^{(M,(j-1)L+1)}}\right)\to_d N\left(0,\Sigma^{-1}\right)\quad\textit{under } Q_n, \tag{6.27}$$

where $Q_n^{(M,(j-1)L+1)}$ is the M-dimensional measure on the j-th block, $j=1,\dots,n_B$.

Proof. The proof of (6.26) is similar to that of Lemma 6.2. Replace $Q_n^{(M)}$ with $P_n^{(M)}$ and use Lemmas 6.8 and 6.9 instead of Lemmas 6.5 and 6.6.

Now we prove (6.27). Lemma 6.2 shows that for any $Q_n\in\mathcal{B}\left(P_0,r\sqrt{M/n}\right)$ and for any block j,

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

Hence,

$$\sqrt{n}\left(\frac{1}{n_B}\sum_{j=1}^{n_B}\bar{T}_{Q_n^{(M,(j-1)L+1)}}-\theta_0\right)=-\sqrt{n}(M\Sigma)^{-1}\frac{1}{n_B}\sum_{j=1}^{n_B}\int\Lambda_n dQ_n^{(M,(j-1)L+1)}+o(1),$$

Subtracting the above from (6.26) one obtains

$$
\begin{aligned}
& \sqrt{n}\left(\bar{T}_{P_n^{(M)}} - \frac{1}{n_B}\sum_{j=1}^{n_B} \bar{T}_{Q_n^{(M,(j-1)L+1)}}\right) \\
= \quad & -\sqrt{n}\,(M\Sigma)^{-1}\left(\int \Lambda_n dP_n^{(M)} - \frac{1}{n_B}\sum_{j=1}^{n_B}\int \Lambda_n dQ_n^{(M,(j-1)L+1)}\right) \\
= \quad & -\sqrt{n}\,(M\Sigma)^{-1}\frac{1}{n_B}\sum_{j=1}^{n_B}\left(\Lambda_n(B_j) - \int \Lambda_n(b)\,dQ_n^{(M,(j-1)L+1)}\right) \\
= \quad & -\Sigma^{-1}G'\Omega^{-1}\frac{1+Mn_B/n}{M\sqrt{n}}\sum_{j=1}^{n_B}\left(\sqrt{M}\phi_n(B_j) - \int \sqrt{M}\phi_n(b)\,dQ_n^{(M,(j-1)L+1)}\right) \\
\to_d \quad & N\left(0,\Sigma^{-1}\right),
\end{aligned}
$$

where the second equality follows from the definition of $P_n^{(M)}$, third equality follows from the definition of the block empirical measure, and the convergence follows from the CLT and the fact that

$$
\begin{aligned}
& E_{Q_n}\left[\frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{n} g(X_t,\theta_0)\,g(X_k,\theta_0)'\right] - \Omega + O(M/n) \\
= \quad & E_{Q_n}\left[\frac{2}{n}\sum_{t=1}^{n-M}\sum_{m=1}^{M} g(X_t,\theta_0)\,g(X_{t+m},\theta_0)' + \frac{1}{n}\sum_{t=1}^{n} g(X_t,\theta_0)\,g(X_t,\theta_0)'\right] - \Omega + O(M/n) \\
& + \frac{1}{n}E_{Q_n}\left[\sum_{t=n-M+1}^{n}\sum_{k=1}^{n} g(X_t,\theta_0)\,g(X_{t+m},\theta_0)'\right] + E_{Q_n}\left[\frac{2}{n}\sum_{t=1}^{n-M}\sum_{m=M+1}^{n} g(X_t,\theta_0)\,g(X_{t+m},\theta_0)'\right] \\
\leq \quad & o(1) + \frac{2}{n}\sum_{t=1}^{n-M}\sum_{m=M+1}^{n} 12\alpha(m)^{1-2/\eta}\,E_{Q_n}\left[\sup_{\theta\in\Theta}|g(X_t,\theta)|^{\eta}\right]^{2/\eta} = o(1),
\end{aligned}
$$

where the $O(M/n)$ term accounts for the weighting of the first $M-1$ and last $M-1$ observations due to blocking, the first equality is a rearrangement of the sum, the seqond equality follows from the definition of Ω, an argument similar to the proof of the second statement of Lemma 6.5 (i), and the result of Davydov (1968), and the third equality follows from Definition 3.1 (ii) and (iii).

Lemma 6.11. *Suppose* $X = (X_1,\ldots,X_M)'$ *and* $Z = (Z_1,\ldots,Z_M)'$, *where* $X_t = \sum_{j=0}^{\infty}\alpha^j u_{t-j}^x$ *and* $Z_t = \sum_{j=0}^{\infty}\alpha^j u_{t-j}^z$ *with* $|\alpha| < 1$ *and*

$$
(u_t^x, u_t^z)' \sim_{i.i.d.} N\left((0,0)', \frac{1}{1-\alpha^2}\Sigma_{(\delta,\rho)}\right),\ \ \Sigma_{(\delta,\rho)} = s^2\begin{pmatrix} (1+\delta)^2 & \rho(1+\delta) \\ \rho(1+\delta) & 1 \end{pmatrix}
$$

For any given values of $\rho$and δ denote the probability measure of the $2M$-dimensional vector $(X', Z')'$ by $P_{(\delta,\rho)}$. If $\sqrt{\delta^2+\rho^2} \to 0$,

$$H\left(P_{(0,0)}, P_{(\delta,\rho)}\right)^2 = \frac{M}{4}\left(2\delta^2+\rho^2\right) + o\left(\delta^2+\rho^2\right).$$

Proof. Note that

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim N\left(0_{2M\times 1}, s^2 \cdot S \otimes A\right),$$

where $0_{2M\times 1}$ is $2M \times 1$ vector of zeros,

$$\begin{aligned} A &= \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{M-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{M-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{M-1} & \alpha^{M-2} & \alpha^{M-3} & \cdots & 1 \end{pmatrix}, \\ S &= \begin{pmatrix} (1+\delta)^2 & \rho(1+\delta) \\ \rho(1+\delta) & 1 \end{pmatrix}. \end{aligned}$$

The joint density of $(X', Z')'$ is hence

$$p_{(\delta,\rho)}(x,z) = \frac{1}{(2\pi)^M \sqrt{|S\otimes A|}} \exp\left\{-\frac{1}{2s^2}\left(x', z'\right)[S\otimes A]^{-1}\begin{pmatrix} x \\ z \end{pmatrix}\right\}.$$

Here and further$|B|$ denotes the determinant of any square matrix B. From the properties of Kronecker product $[S\otimes A]^{-1} = S^{-1}\otimes A^{-1}$, $|S\otimes A| = |S|^M |A|^2 = (1+\delta)^{2M}\left(1-\rho^2\right)^M |A|^2$, and $|I_2 \otimes A| = |A|^2$. Denote

$$\begin{aligned} l(\delta,\rho) &= \log p_{(\delta,\rho)}(x,z) \\ &= -\frac{M}{2}\log\left((1+\delta)^2\left(1-\rho^2\right)\right) - \frac{1}{2s^2}\left(x', z'\right)\left[S^{-1}\otimes A^{-1}\right]\begin{pmatrix} x \\ z \end{pmatrix} - M\log(2\pi) - \log|A|. \end{aligned}$$

Note that

$$
\begin{aligned}
\frac{\partial l(0,0)}{\partial \delta} &= -M + \frac{1}{2s^2}(x', z') \left[-\left(\frac{\partial}{\partial \delta} S^{-1} \right)\Big|_{\delta=\rho=0} \otimes A^{-1} \right] \begin{pmatrix} x \\ z \end{pmatrix} \\
&= -M + \frac{1}{2s^2}(x', z') \left[\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \otimes A^{-1} \right] \begin{pmatrix} x \\ z \end{pmatrix} \\
&= -M + x' \left(s^2 A \right)^{-1} x \\
\frac{\partial l(0,0)}{\partial \rho} &= \frac{1}{2s^2}(x', z') \left[-\left(\frac{\partial}{\partial \rho} S^{-1} \right)\Big|_{\delta=\rho=0} \otimes A^{-1} \right] \begin{pmatrix} x \\ z \end{pmatrix} \\
&= \frac{1}{2s^2}(x', z') \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes A^{-1} \right] \begin{pmatrix} x \\ z \end{pmatrix} \\
&= x' \left(s^2 A^{-1} \right) z.
\end{aligned}
$$

Define $\dot{l} = (\partial l(0,0)/\partial \delta, \partial l(0,0)/\partial \rho)'$ and Fisher information $I_0 = \int \dot{l}\dot{l}' dP_0$. A direct calculation yields

$$
H\left(P_{(0,0)}, P_{(\delta,\rho)} \right)^2 = \frac{1}{4}(\delta, \rho) I_0 \begin{pmatrix} \delta \\ \rho \end{pmatrix} + o\left(\delta^2 + \rho^2 \right).
$$

Since $(\delta, \rho)\,\dot{l} = -\delta M + \delta x' \left(s^2 A \right)^{-1} x + \rho x' \left(s^2 A^{-1} \right) z$,

$$
\begin{aligned}
(\delta, \rho) I_0 \begin{pmatrix} \delta \\ \rho \end{pmatrix} &= \int \left(-\delta M + \delta x' \left(s^2 A \right)^{-1} x + \rho x' \left(s^2 A^{-1} \right) z \right)^2 dP_0(x, z) \\
&= E\left[\left(-\delta M + \delta X'X + \rho X'Z \right)^2 \right],
\end{aligned}
$$

where $(X', Z')' \sim N(0, I_{2M})$. Thus,

$$
\begin{aligned}
(\delta, \rho) I_0 \begin{pmatrix} \delta \\ \rho \end{pmatrix} &= E\left[\delta^2 M^2 - 2\delta^2 M X'X + \delta^2 \left(X'X \right)^2 + \rho^2 \left(X'Z \right)^2 \right] \\
&= \delta^2 \left(M^2 - 2M^2 + 3M + M(M-1) \right) + \rho^2 M \\
&= 2M\delta^2 + M\rho^2.
\end{aligned}
$$

References

Andrews, D. W. (1982): "Robust and Asymptotically Efficient Estimation of Location in a Stationary Strong Mixing Gaussian Parametric Model," Cowles Foundation Discussion Papers 659, Cowles Foundation, Yale University.

——— (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

——— (1988): "Robust Estimation of Location in a Gaussian Parametric Model," *Advances in Econometrics*, 7, 3–44.

Ashley, R., and D. Vaughan (1986): "Measuring Measurement Error in Economic Time Series," *Journal of Business and Economic Statistics*, 4, 95–103.

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

——— (1978a): "Asymptotic Lower Bounds for Risk in Robust Estimation," *Annals of Statistics*, 8, 1252–1264.

——— (1978b): "An Efficient and Robust Adaptive Estimator of Location," *Annals of Statistics*, 6, 292–313.

——— (1982): "Robust Estimation in Models for Independent Non-Identically Distributed Data," *Annals of Statistics*, 10, 415–428.

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

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, pp. 1–72. Springer.

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.

Davydov, Y. (1968): "Convergence of Distributions Generated by Stationary Stochastic Processes," *Theory of Probability and its Applications*, 8, 675–683.

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

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

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

Herrndorf, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12, 141–153.

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. (1997): "Empirical likelihood methods with weakly dependent processes," *Annals of Statistics*, 25, 2084–2102.

——— (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 University Press, Cambridge, forthcoming.

Kitamura, Y., and T. Otsu (2005): "Minimax Estimation and Testing for Moment Condition Models via Large Deviations," Manuscript, Department of Economics, Yale University.

Kitamura, Y., T. Otsu, and K. Evdokimov (2009): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," Cowles Foundation Discussion Papers 1720, Cowles Foundation, Yale University, under revision for *Econometrica.*

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

Martin, R., and V. Yohai *(1986): "Influence Functionals for Time Series,"* Annals of Statistics*, 14, 781–818.*

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

Newey, W. K., and K. D. West *(1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,"* Econometrica*, 55(3), 703–708.*

Owen, A. *(2001):* Empirical Likelihood*. Chapman and Hall/CRC.*

Potscher, B. M., and I. R. Prucha *(1989): "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes,"* Econometrica*, 57(3), 675–683.*

Qin, J., and J. Lawless *(1994): "Empirical Likelihood and General Estimating Equations,"* Annals of Statistics*, 22, 300–325.*

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.*

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

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

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