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. These models 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 conditions. Moment condition-based estimation is often computationally convenient; the GMM estimator (Hansen, 1982) is a prime example. It is argued that moment condition models impose only mild assumptions and thereby enable 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. Indeed, this paper

*Department of Economics, Princeton University, Princeton, NJ 08544.

†Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520.

‡Department of Economics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, U.K.

E-mail addresses: kevdokim@princeton.edu, yuichi.kitamura@yale.edu, t.otsu@lse.ac.uk.

Date: August, 2017.

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

JEL Classification Number: C14.

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.

Let us introduce some notation to formalize our problem concerning robustness. Consider a measurable space $(\Omega, \mathcal{F})$. Throughout this paper we consider a 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 $X^\infty(\omega) = \{X_t(\omega)\}_{t=-\infty}^{\infty}$ for each $\omega \in \Omega$. 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 complete space of full trajectories $(\Omega, \mathcal{F})$, and suppose a random sequence X^∞ is strictly stationary under P_0. Moreover, suppose a moment restriction of the following form holds for P_0:

$$E_{P_0}[g(X_t, \theta_0)] = \int g(X_t(\omega), \theta_0) P_0(d\omega) = 0, \tag{1.1}$$

at some $\theta_0 \in \Theta$. 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 MHDE has desirable properties. This 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 maximum likelihood 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 (2013) 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. Dependent data introduces new challenges into the analysis. For instance, the study of Kitamura, Otsu and Evdokimov (2013) employs Le Cam-type results but no such results are known for the case of dependent data, hence a different approach is needed.

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 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 not 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 paper addresses these problems.

2. The Estimator

As in Andrews (1982) and Kitamura, Otsu and Evdokimov (2013), 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 $\mathcal{X}^s := \otimes_{i=1}^s \mathcal{X}$ with densities p and q with respect to a dominating measure ν. The Hellinger distance between P and Q is then given by

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

One may rewrite the above as

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

which is convenient as it avoids an explicit use of the dominating measure.[1] 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 $\mathcal{X}$. Under this extra assumption (1.1) becomes

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

Let $\mu \ll \mu_0$ mean a measure μ is absolutely continuous with respect to μ_0. Consider the following population problem:

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

An application of convex duality yields

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

(see, for example, Kitamura (2007) for details). But if θ_0 is identified in (2.1), minimizing $v(\theta)$ over $\theta \in \Theta$ leads to $\theta_0 = \arg\min_{\theta \in \Theta} v(\theta)$. In sum,

$$\theta_0 = \arg\min_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{1 + \gamma' g(x,\theta)} d\mu_0.$$

[1] Note that the Hellinger distance does not depend on the choice of the dominating measure (see e.g., Pollard, 2002, p. 61).

Form a natural sample analogue, we define

$$\hat{\theta} = \arg\min_{\theta\in\Theta}\max_{\gamma\in\mathbb{R}^m} -\frac{1}{n}\sum_{t=1}^{n}\frac{1}{1+\gamma' g(X_t,\theta)}.$$

This is the moment restriction MHDE for the 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 $B_j = (X_{(j-1)L+1}, \ldots, X_{(j-1)L+M})$, $n_B = \lfloor (n-M)/L \rfloor + 1$, and $\lfloor\cdot\rfloor$ denotes the integer part of $\cdot$. The positive integer $L \leq M$ is the distance between starting points of blocks. Define the "smoothed moment function" $\phi(B_j,\theta) = M^{-1/2}\sum_{l=1}^{M} g(X_{(j-1)L+l},\theta)$, $j = 1,\ldots,n_B$. In addition, by using the Dirac measure δ, define the empirical measure $P_n^{(M)}$ on the blocks as

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

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

$$\hat{\theta}_H = \arg\min_{\theta\in\Theta}\max_{\gamma\in\mathbb{R}^m} -\int \frac{1}{1+\gamma'\phi(b,\theta)}dP_n^{(M)} = \arg\min_{\theta\in\Theta}\max_{\gamma\in\mathbb{R}^m} -\frac{1}{n_B}\sum_{j=1}^{n_B}\frac{1}{1+\gamma'\phi(B_j,\theta)}. \tag{2.2}$$

This will be called the blockwise MHDE in this paper. Note that it can be seen as a mapping of the empirical probability measure on blocks of length M to the parameter space, i.e. $\hat{\theta}_H = T(P_n^{(M)})$, 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 satisfying (1.1). In this ideal scenario it is easy to show that $\sqrt{n}(\hat{\theta}_H - \theta_0) \xrightarrow{d} N(0,\Sigma^{-1})$, where $\Sigma = G'V^{-1}G$, $G = E_{P_0}[\partial g(X_t,\theta_0)/\partial\theta']$, and $V = \sum_{j=-\infty}^{\infty} E_{P_0}[g(X_t,\theta_0)g(X_{t-j},\theta_0)']$ 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.

In contrast to the MHDE in Kitamura, Otsu and Evdokimov (2013) for the IID data, the blockwise MHDE requires to choose a smoothing constant, the block length M. We note that this smoothing is introduced to recover the asymptotic efficiency under the ideal measure P_0, and is analogous to smoothing for the optimal GMM weight under dependent data. Therefore, this smoothing is different from the one for density estimation in Beran's (1977) parametric MHDE. For the choice of M, we can apply the conventional selection methods for the heteroskedasticity autocorrelation consistent (HAC) covariance matrix estimator (e.g., Andrews, 1991, and Newey and West, 1994). Although it is beyond the scope of this paper, another interesting direction is to extend our MSE result in Theorem (3.2) below to accommodate the higher order term so that the optimal M can be chosen to minimize the worst MSE over local neighborhoods around P_0.

To implement the blockwise MHDE (2.2), we can apply the nested optimization method as in computation of the empirical likelihood estimator (Chapter 12 of Owen, 2001, and Kitamura, 2007). In particular, we prepare a subroutine for the inner loop optimization to evaluate the profile objective function $\ell_H(\theta) = \max_{\gamma \in \mathbb{R}^m} -\frac{1}{n_B}\sum_{j=1}^{n_B} \frac{1}{1+\gamma'\phi(B_j,\theta)}$ for each θ. Note that this inner loop optimization is with respect to γ, where the Jacobian and Hessian are

$$\frac{1}{n_B}\sum_{j=1}^{n_B} \frac{\phi(B_j,\theta)}{\{1+\gamma'\phi(B_j,\theta)\}^2} \quad \text{and} \quad -\frac{2}{n_B}\sum_{j=1}^{n_B} \frac{\phi(B_j,\theta)\phi(B_j,\theta)'}{\{1+\gamma'\phi(B_j,\theta)\}^3},$$

respectively. Thus, Newton type methods typically work. Then the blockwise MHDE can be computed by the outer loop optimization, i.e., minimization of the profile objective function $\ell_H(\theta)$ for θ.

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 local neighborhoods needs to take dependence into account.

To motivate our choice of neighborhood, consider the so-called α-divergence family for probability measures P and Q on the s-fold product space $\mathcal{X}^s$ with densities p and q with respect to a dominating measure ν:

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

for $\alpha \in \mathbb{R}$. The cases of $\alpha = 0, 1$ are defined by taking the limits using L'Hospital's rule: I_1 and I_0 correspond to the well-known Kullback-Leibler (KL) divergence measure from P to Q and Q to P, respectively.[2] The α-divergence includes the Hellinger distance as a special case, in the sense that $H^2(P,Q) = I_{1/2}(P,Q)/2$.

Define the corresponding I_α-divergence balls around a probability measure P with radius $\delta > 0$:

$$B_\alpha(P,\delta) = \{Q : \sqrt{I_\alpha(Q,P)} \le \delta\}.$$

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

Kitamura, Otsu and Evdokimov (2013, Lemma 2.1) implies that

$$\cup_{\alpha\in[1/2-L,1/2+U]} B_{I_\alpha}(P,\delta) \subset B_{I_{1/2}}(P, \sqrt{2C(L,U)}\delta), \tag{3.1}$$

for any constants $L > 0$, $U > 0$, and $C(L,U) := (1/2 + \max\{L,U\})^{-1}$. Therefore, any I_α-based neighborhood for $\alpha \in [1/2 - L, 1/2 + U]$ is covered by the Hellinger neighborhood $B_{I_{1/2}}$ with a larger radius $2\sqrt{C(L,U)}$. From the inclusion relationship (3.1), the Hellinger neighborhood $B_{I_{1/2}}$is large enough to cover other neighborhood systems based on I_α with $\alpha \in [1/2 - L, 1/2 + U]$ for an appropriately large radius. On the other hand, the relation (3.1) does not hold if the Hellinger distance $I_{1/2}$ on the right hand side is replaced by other divergences I_α with $\alpha \neq 1/2$. This shows a distinguishing feature of the Hellinger distance in the α-divergence family.

Since the goal of robust estimation is to guard against a large set of perturbations, the above motivates using Hellinger distance for constructing neighborhoods. However, the above result only applies to distributions of random vectors. It is not clear how to extend the notion of α-divergence or Hellinger distance to stochastic processes. Yet, the neighborhoods we consider need to capture not only the potential perturbations of the marginal distribution of X_t, but also the perturbations of the dependence structure of the time series. To take into account the dependence aspects of the stochastic process $\{X_t\}$ we consider the Hellinger distance on expanding blocks.

Let us introduce some additional notation. 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 is strictly stationary, $P^{(k,t)}$ does not depend on t, and the notation $P^{(k)}$ is used to denote it. The following definition of neighborhoods is suitable for the development of our robustness theory for weakly dependent data (recall that P_0 is assumed to be strictly stationary).

Definition 3.1. For any $r > 0$ and $n \in \mathbb{N}$, let $\mathcal{B}_n(r)$ be the set of all probability measures Q that satisfy the following three conditions:

- **(i):** $H(Q^{(1,t)}, P_0^{(1)}) \leq r/\sqrt{n}$ for each t,
- **(ii):** for any pair of integers (t, t') with $|t - t'| \leq M$, the bivariate marginals of $(X_t(\omega), X_{t'}(\omega))$ implied from Q and P_0 have the Hellinger distance less than a_n with $a_n \to 0$,
- **(iii):** 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;
- **(iv):** for each t, $E_Q[\sup_{\theta\in\Theta} |g(X_t, \theta)|^\eta] < \infty$ for $\eta > 2$ defined in Assumption 3.1 (v) below.

Let $M \to \infty$ be such that $M/n \to 0$ as $n \to \infty$. Sequences of local neighborhoods of the form $\mathcal{B}_n(r)$ are used throughout our theoretical analysis in this section. We consider the effect of perturbations of P_0 within $\mathcal{B}_n(r)$, that is, we analyze the maximum MSE of estimators when the probability law Q of the data varies within $\mathcal{B}_n(r)$. Note that the true parameter θ_0 and the true probability measure P_0 do not depend on the sample size.

The neighborhood $\mathcal{B}_n(r)$ 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.

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

Note that we do not assume the perturbed measure Q to be stationary. Therefore, the finite dimensional distributions on different blocks may differ (although we will impose that the process X^∞ under P_0 is strictly stationary; see Assumption 3.1). Condition (ii) imposes a mixing condition on Q. This does not seem to follow directly from (i) and Assumption 3.1 (i), which is a mixing condition on P_0.

The local neighborhood system $\{\mathcal{B}_n(r)\}_{n\in\mathbb{N}}$ 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 $H(Q, P_{\theta_0}) \leq r/\sqrt{n}$ 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 (2013) consider robust estimation when data are IID 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. Moreover, this paper 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).

One may also be interested in considering distances that result in even larger neighborhoods than the Hellinger distance allows, such as Kolmogorov-Smirnov distance. However, an estimator that is robust to such a wide variety of perturbations will be less efficient than the GMM estimator when the data does not contain perturbations. In contrast, blockwise MHDE estimator of this paper is asymptotically as efficient as the optimally weighted GMM when the model assumptions hold. Thus, blockwise MHDE possesses an optimal robustness property without sacrificing efficiency.

Let $\tau : \Theta \to \mathbb{R}$ be a possibly nonlinear transformation of the parameter. One may, for example, choose $\tau(\theta) = c'\theta$ for a constant vector c. We study the estimation problem of the transformed parameter $\tau(\theta_0)$, 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 with that of alternative estimators.

We impose the following assumptions.

Assumption 3.1. *The following conditions hold:*

- **(i):** *The process X^∞ under the probability measure P_0 is strictly stationary and α-mixing with the 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 $\eta > 2$, $g(x, \theta)$ is continuously differentiable a.s. in an open neighborhood $\mathcal{U}$ around θ_0, $E_{P_0}[\sup_{\theta\in\mathcal{U}} |g(X_t, \theta)|^4] < \infty$, $E_{P_0}[\sup_{\theta\in\mathcal{U}} |\partial g(X_t, \theta)/\partial\theta'|^\eta] < \infty$, and $\sup_{x\in\mathcal{X}_n,\theta\in\mathcal{U}} |\partial g(x, \theta)/\partial\theta'| \le o(n^{1/2})$, where $\mathcal{X}_n$ is defined in the Appendix;*
- **(vi):** *G has the full column rank and V is positive definite;*
- **(vii):** *M is implicitly assumed to depend on n, and satisfies $M = O(n^\alpha)$ for $0 < \alpha < \frac{\eta^2-2\eta}{2(\eta^2-1)}$;*
- **(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.[3] 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 $T_a(X_1, \ldots, X_n)$. We assume that an estimator T_a satisfies the following property:

Assumption 3.2. *There exists a sequence of functions $\varphi_n : \mathcal{X} \to \mathbb{R}^p$ such that for every $r > 0$, $\xi \in \mathbb{R}^p$, and sequence $\{Q_n\}_{n \in \mathbb{N}}$ satisfying*

$$Q_n \in \mathcal{B}_n(r) \cap \{P : E_P[g(X_t, \theta_0 + n^{-1/2}\xi)] = 0 \text{ for all } t\},$$

the following holds

$$\sqrt{n}\{T_a(X_1, \ldots, X_n) - \theta_0\} - \xi - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varphi_n(X_t) \xrightarrow{p} 0, \quad \text{under } Q_n, \tag{3.2}$$

where $E_{Q_n}[n^{-1/2} \sum_{t=1}^{n} \varphi_n(X_t)] \to 0$ for all t, and $E_{Q_n}[n^{-1} \sum_{t=1}^{n} \sum_{\tau=1}^{n} \varphi_n(X_t)\varphi_n(X_\tau)']$ converges to a positive definite matrix $A_{\varphi\varphi'}$ such that $A_{\varphi\varphi'} - \Sigma^{-1}$ is positive semidefinite.

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 next assumption is only used to derive the minmax bound. It is not needed to show the properties of the blockwise MHDE estimator $\hat{\theta}_H$.

Assumption 3.3. *(i) All components of X_t are continuously distributed; (ii) $\eta > 4$.*

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

Assumption 3.3 is restrictive and is used to construct appropriate least favorable distributions, which are important building blocks of our minimax optimality theory. These constructions turn out to be more complicated in the case of weakly dependent data than in the case of IID data. Assumption 3.3 (i) allows to use an integral transform in a part of the proof. It may be possible to relax this condition 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 $|E_{P_0}[\phi_n(B,\theta_0)]| = o(\sqrt{M/n})$, 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 behavior 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 the regularity Assumption 3.2 and Assumption 3.3 holds, then

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

(ii): The blockwise MHDE estimator $\hat{\theta}_H = T(P_n^{(M)})$ satisfies

$$\lim_{\kappa\to\infty}\lim_{n\to\infty}\sup_{Q\in\mathcal{B}_n(r)}\int \kappa\wedge n\{\tau(\hat{\theta}_H)-\tau(\theta_0)\}^2 dQ = (1+4r^2)B^*.$$

This theorem compares the asymptotic MSE of the blockwise MHDE $T(P_n^{(M)}) = \hat{\theta}_H$ with that of an alternative estimator $T_a = T_a(x_1,\ldots,x_n)$. In particular, the theorem compares the maximum values of their MSEs as the probability of law of data varies over $\mathcal{B}_n(r)$.

Part (i) of the theorem derives the minmax bound for the (truncated) Mean Squared Error (MSE) of any estimator satisfying (3.2). 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 $(1+4r^2)B^*$ is positive and finite.

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. That $\kappa \to \infty$ in the limit theory allows the truncation parameter to be arbitrarily large.

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 improvements in accounting techniques.

It is important to note that the theorem concerns estimation of the true value θ_0, not of a pseudo-true value under global misspecification. It therefore differs from the results in White (1982), Kitamura (1998), Kitamura (2002), and Schennach (2007). Also, Dahlhaus and Wefelmeyer (1996) studied efficient estimation of the pseudo-true values of globally misspecified parametric time series models.

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 $\mathcal{B}_n(r)$. Second, we use this maximum bias to calculate the lower bound for maximum MSE over $\mathcal{B}_n(r)$. Then, we introduce trimmed blockwise MHDE $\bar{T}(\cdot)$ and show that it achieves the lower bounds of bias and MSE on $\mathcal{B}_n(r)$ derived earlier. Finally, we show that the difference between MSE of trimmed estimator $\bar{T}(P_n^{(M)})$ and MSE of blockwise MHDE $T(P_n^{(M)})$ is negligible and hence blockwise MHDE $T(P_n^{(M)})$ achieves the lower bound.

In this paper, we focus on robustness of point estimation methods under moment conditions. It is natural to expect that analogous optimal robustness analysis can be done for hypothesis testing problems. For parametric models, the issue of optimal robust testing under local deviations from the model assumption was investigated by Rieder (1978) and Beran (1981). For moment condition models with IID data, Kitamura and Otsu (2010) demonstrated that the Wald type test based on the MHDE (without blocking) possesses desirable optimal robustness properties for testing the parameter hypothesis $H_0 : \tau(\theta_0) = 0$. First, Kitamura and Otsu (2010) showed that in certain 'regular' class of tests (which include the conventional GMM-based Wald, likelihood ratio, and score tests), the MHDE Wald test using the asymptotic chi-squared critical value has the smallest size distortion under local perturbations. For the IID case, this result is derived by applying the minimax optimal bias property of the MHDE (Kitamura, Otsu and Evdokimov, 2013, Theorem 3.1) to the context of hypothesis testing. For the time series case, we establish an analogous bias optimality of the blockwise MHDE in Lemma A.3. Therefore, although formal investigation is beyond the scope of this paper, a similar argument to Kitamura and Otsu (2010) will guarantee that the Wald test by using blockwise MHDE Wald test minimizes the worst size distortion under local perturbations over $\mathcal{B}_n(r)$. Second, for the IID data, Kitamura and Otsu (2010) also provided a Neyman-Pearson type power optimality result, i.e., under certain restriction on the size property, the MHDE Wald test minimizes the worst type II error probability over a locally perturbed set of measures around the conventional local alternative hypotheses. For the IID case, this result is established by applying the minimax bias and risk robustness properties of the MHDE (Kitamura, Otsu and Evdokimov, 2013, Theorems 3.1, and 3.3). In addition to the bias robustness in Lemma A.3, the risk robustness of the blockwise MHDE can be obtained by applying Lemma B.10. Thus, we can expect that the Neyman-Pearson type power optimality may be also established for the Wald blockwise MHDE for the time series case.

4. Monte-Carlo Experiments

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

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

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

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 $(u_0^x, u_0^z)' \sim N(0, 0.4^2 I_2)$). Define

$$g(x, z, \theta) = [\exp\{-0.72 - \theta(x+z) + 3z\} - 1](1, z)',$$

then the moment restriction $E[g(X_t, Z_t, \theta_0)] = 0$ 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) of equation (2.2). The results are based on 10000 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 are presented in Tables 1 and 2.

The data generating process corresponding to the first row of Table 1 has $c = 0$ and represents the true model (4.1). For each scenario we report the Root Mean Squared Error (RMSE) and the probabilities $\Pr\{|\hat{\theta} - \theta_0| > 1.0\}$ ($\Pr\{|\hat{\theta} - \theta_0| > 0.5\}$ in Table 2) 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 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 behavior 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 wide margin. When measurement error is small EL may still outperform MHDE and ET thanks to its higher order properties. When measurement errors become 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 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. The simulation experiment discards such replications in calculations of the summary statistics. Column "%f" of Tables 1 and 2 reports the percentage of such replications. As can be seen from Table 1, such cases are very rare in most cases, though become more likely for larger c, especially in the case of the $-\chi_1^2$ measurement errors.

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

Note that the joint distribution $Q^{(M)}$ of the data block $B=(X_1,\dots,X_M,Z_1,\dots,Z_M)'$ 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 $\Sigma_0=0.4^2(1-\alpha^2)^{-1}I_2$, 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 at an approximately constant distance from it. One

		RMSE					$Pr\{\lvert\hat{\theta}-\theta_0\rvert > 1.0\}$					
c	ξ_{tj}	EL	MHDE	ET	GMM	CUE	EL	MHDE	ET	GMM	CUE	%f
						$n=100,\ M=5$						
0		0.745	0.884	1.063	0.933	3.338	0.114	0.125	0.140	0.208	0.359	0.01
0.5	N	0.695	0.797	0.964	0.879	3.094	0.103	0.112	0.126	0.188	0.323	0.00
1	N	0.695	0.763	0.879	0.966	2.838	0.109	0.117	0.126	0.257	0.305	0.00
2	N	0.936	0.923	0.949	1.434	2.465	0.316	0.275	0.260	0.594	0.411	0.25
0.5	χ_1^2	0.742	0.889	1.056	0.911	3.266	0.109	0.121	0.135	0.204	0.345	0.00
1	χ_1^2	0.637	0.731	0.879	0.834	3.027	0.082	0.091	0.103	0.162	0.298	0.00
2	χ_1^2	0.614	0.650	0.719	0.871	2.829	0.076	0.080	0.087	0.214	0.275	0.00
0.5	$-\chi_1^2$	0.735	0.860	0.991	0.944	3.098	0.119	0.132	0.144	0.227	0.331	0.00
1	$-\chi_1^2$	0.788	0.847	0.960	1.161	2.826	0.173	0.170	0.174	0.363	0.347	0.30
2	$-\chi_1^2$	1.011	0.990	1.041	1.497	2.580	0.290	0.247	0.241	0.527	0.390	2.27
0.5	t_3	0.735	0.851	1.007	0.931	3.164	0.114	0.123	0.137	0.211	0.339	0.01
1	t_3	0.717	0.794	0.911	0.996	2.897	0.122	0.126	0.132	0.252	0.310	0.18
2	t_3	0.872	0.878	0.940	1.299	2.568	0.228	0.203	0.198	0.457	0.348	0.68
						$n=100,\ M=10$						
0		0.833	0.993	1.139	0.917	3.111	0.124	0.134	0.148	0.201	0.340	0.19
0.5	N	0.738	0.869	1.014	0.862	2.906	0.110	0.121	0.132	0.181	0.308	0.13
1	N	0.729	0.822	0.918	0.961	2.677	0.118	0.128	0.136	0.260	0.293	0.06
2	N	0.957	0.961	0.992	1.443	2.378	0.320	0.282	0.271	0.597	0.408	0.29
0.5	χ_1^2	0.815	0.969	1.110	0.885	3.077	0.122	0.132	0.143	0.195	0.328	0.14
1	χ_1^2	0.686	0.804	0.938	0.817	2.800	0.090	0.102	0.114	0.156	0.277	0.04
2	χ_1^2	0.635	0.676	0.742	0.878	2.667	0.080	0.087	0.094	0.223	0.259	0.02
0.5	$-\chi_1^2$	0.799	0.943	1.070	0.934	2.911	0.131	0.140	0.152	0.225	0.319	0.08
1	$-\chi_1^2$	0.823	0.900	0.996	1.153	2.675	0.182	0.183	0.188	0.363	0.337	0.39
2	$-\chi_1^2$	1.047	1.014	1.069	1.497	2.458	0.305	0.258	0.250	0.525	0.388	2.29
0.5	t_3	0.808	0.931	1.053	0.907	2.979	0.124	0.134	0.146	0.200	0.323	0.09
1	t_3	0.776	0.864	0.973	0.989	2.714	0.130	0.135	0.143	0.251	0.297	0.29
2	t_3	0.898	0.922	0.969	1.298	2.429	0.236	0.212	0.210	0.457	0.340	0.77

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

		RMSE					$Pr\{\lvert\hat{\theta}-\theta_0\rvert > 0.5\}$					
c	ξ_{tj}	EL	MHDE	ET	GMM	CUE	EL	MHDE	ET	GMM	CUE	%f
						$n = 400,\ M = 10$						
0		0.292	0.294	0.309	0.409	2.668	0.082	0.084	0.089	0.111	0.229	0.00
0.5	N	0.282	0.286	0.296	0.400	2.679	0.074	0.075	0.078	0.104	0.222	0.00
1	N	0.281	0.281	0.290	0.383	2.452	0.070	0.071	0.074	0.103	0.203	0.00
2	N	0.473	0.455	0.444	0.678	2.609	0.305	0.276	0.263	0.486	0.409	0.00
0.5	χ_1^2	0.284	0.286	0.293	0.397	2.662	0.076	0.078	0.082	0.105	0.223	0.00
1	χ_1^2	0.278	0.279	0.284	0.377	2.586	0.067	0.069	0.072	0.092	0.206	0.00
2	χ_1^2	0.268	0.269	0.274	0.339	2.422	0.058	0.058	0.059	0.087	0.182	0.00
0.5	$-\chi_1^2$	0.290	0.292	0.298	0.412	2.618	0.083	0.084	0.086	0.117	0.227	0.00
1	$-\chi_1^2$	0.367	0.358	0.358	0.589	2.371	0.142	0.133	0.133	0.255	0.264	0.01
2	$-\chi_1^2$	0.713	0.643	0.612	1.167	2.213	0.432	0.365	0.337	0.656	0.481	0.58
0.5	t_3	0.286	0.287	0.294	0.407	2.677	0.078	0.079	0.081	0.110	0.225	0.01
1	t_3	0.317	0.312	0.315	0.476	2.418	0.095	0.092	0.092	0.148	0.219	0.05
2	t_3	0.494	0.461	0.447	0.800	2.297	0.237	0.210	0.198	0.403	0.329	0.42
						$n = 400,\ M = 20$						
0		0.300	0.310	0.317	0.401	2.533	0.088	0.091	0.094	0.111	0.218	0.00
0.5	N	0.290	0.294	0.305	0.384	2.545	0.079	0.081	0.084	0.101	0.212	0.00
1	N	0.285	0.291	0.297	0.374	2.326	0.074	0.075	0.077	0.105	0.194	0.00
2	N	0.473	0.455	0.448	0.693	2.527	0.302	0.278	0.265	0.501	0.404	0.00
0.5	χ_1^2	0.292	0.296	0.303	0.385	2.544	0.083	0.085	0.087	0.103	0.216	0.00
1	χ_1^2	0.285	0.287	0.292	0.363	2.465	0.076	0.076	0.078	0.093	0.198	0.00
2	χ_1^2	0.273	0.275	0.279	0.333	2.283	0.062	0.063	0.064	0.087	0.170	0.00
0.5	$-\chi_1^2$	0.296	0.299	0.313	0.401	2.487	0.087	0.088	0.092	0.115	0.215	0.00
1	$-\chi_1^2$	0.372	0.365	0.365	0.590	2.259	0.150	0.142	0.139	0.260	0.257	0.01
2	$-\chi_1^2$	0.723	0.649	0.621	1.182	2.165	0.437	0.370	0.344	0.662	0.479	0.56
0.5	t_3	0.293	0.296	0.305	0.398	2.553	0.083	0.084	0.087	0.107	0.216	0.01
1	t_3	0.324	0.319	0.323	0.472	2.305	0.099	0.099	0.099	0.149	0.210	0.06
2	t_3	0.499	0.465	0.451	0.809	2.204	0.242	0.216	0.203	0.412	0.323	0.43

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

can calculate that the Hellinger distance between the true and perturbed probability measures on the block is $H(P_0^{(M)}, P_{(\delta,\rho)}^{(M)}) \approx \sqrt{M/4}\sqrt{2\delta^2+\rho^2}$ for small ρ and δ. Therefore values of (δ, ρ) that satisfy $c^2 = \delta^2 + \rho^2/2$ for some constant c are considered. We consider 64 different designs indexed by $j \in \{0, \dots, 63\}$. In the j-th design we set $\omega_j = j/64$, $\delta_j = c\sin(2\pi\omega_j)$, $\rho_j = \sqrt{2}c\cos(2\pi\omega_j)$.[4] In the Monte Carlo experiment we set $c = 0.1$, $\alpha = 0.75$, and $n = 400$. Estimation is performed with fully overlapping blocks of length 10. For each design 10000 replications are computed.

The results are presented in Figure 4.1. The top panels plot RMSE of the estimators as function of ω_j. The bottom panels show the estimated probabilities $\Pr\{|\hat{\theta} - \theta_0| > 1.0\}$. As in the first experiment, RMSE and $\Pr\{|\hat{\theta} - \theta_0| > 1.0\}$ of CUE are much larger than those of other estimators. To provide better insights on the relative performance of other estimators, the right panels of the figure present the same plots as the left ones but exclude CUE. MHDE, EL, and ET outperform GMM. Interestingly, EL and MHDE are very close for all scenarios. ET is close to EL and MHDE although appears to be slightly less robust against a range of misspecifications.

Before closing this section it might be beneficial to discuss a possible interpretation of the simulation results in light of the main theoretical results such as Theorem 3.2. Consider minimizing the α-divergence in Definition 2.1, with the measure Q replaced by the blockwise empirical measure $P_n^{(M)}$ as we did in (2.2), subject to the moment constraint $\int \phi(b, \theta)dQ = 0, \theta \in \Theta$. This gives rise to a family of estimator indexed by α, including the blockwise MHDE as a special case of α being $\frac{1}{2}$, which is optimally robust according to our Theorem 3.2. Note that the value of the estimator that minimizes the α-divergence varies continuously with the value of α. Thus one expects that estimators with their α close to the optimal $\frac{1}{2}$ remain comparatively robust, and as α moves away from $\frac{1}{2}$ the corresponding estimator would grow increasingly susceptible to the effects of data contamination, which is the paper's major concern. The experimental results are consistent with this prediction: the MHDE ($\alpha = \frac{1}{2}$) performs well over a wide range of data generating processes in both experiments, and the same applies to the estimators with their corresponding values α relatively close to the optimal value $\alpha = \frac{1}{2}$, namely EL ($\alpha = 1$), and ET ($\alpha = 0$). On the other hand, the finite sample performance of CUE, which has the α value of 2 and thus quantitatively very different from the asymptotically optimal MHDE, is poor. As GMM and CUE are both based on a closely related quadratic measure, this also explains the erratic performance of GMM. Overall, the simulations provide strong support for the theoretical results obtained in Section 3.

---

[4]Note that design with $\omega = 1$ coincides with the $\omega = 0$ design, i.e., the graphs are closed loops.

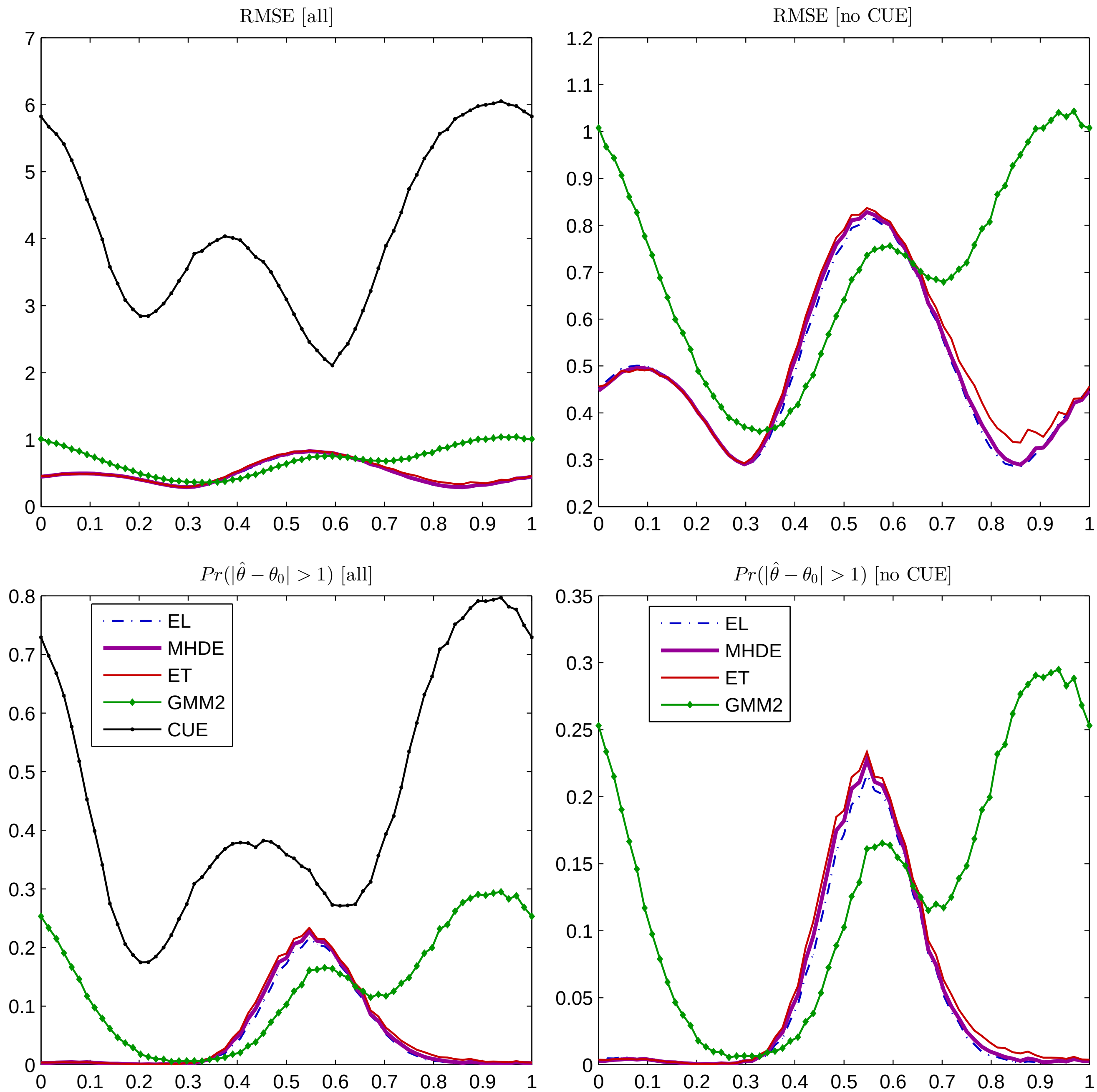

FIGURE 4.1. Local neighborhood of the true model.

5. CONCLUSION

This paper studied robust estimation of moment restriction models with time series data. Often the data used in empirical analysis is not ideal and is subject to errors, for instance due to data contamination or incorrect deseasonalization. In such cases, the distribution of data at hand is a perturbed version of the true data distribution. This paper studies robustness of a large class of estimation procedures to perturbations in the data generating probability measure. The main result of the paper is demonstrating that the blockwise MHDE possesses optimal minimax robust properties. The paper derives minimax lower bound of MSE risk and shows that the blockwise MHDE estimator achieves this bound. At the same time, blockwise MHDE is known to be semiparametrically efficient in the ideal scenario of error-free data. Thus, blockwise MHDE estimator is both robust and efficient. The Monte Carlo experiments suggest that GMM and Continuously Updated GMM are sensitive to data perturbations, while MHDE is not.

Acknowledgements: The research of the first author was supported by the Cowles Foundation at Yale University and the G. C. Chow Econometric Research Program at Princeton University, and the National Science Foundation grant SES-1459993. The research of the second author was supported by the National Science Foundation grants SES-0551271, SES-0851759, and SES-1156266. The research of the third author was supported by the National Science Foundation grant SES-0720961 and the ERC Consolidator Grant (SNP 615882).

Appendix A. Main Proofs

Notation. Let $C > 0$ be a generic positive constant, $|\cdot|$ be the Euclidean norm, $\|\cdot\|$ be the L_2-metric on the appropriate space, and $\mathbb{I}\{\cdot\}$ be the indicator function. When a measure P is strictly stationary, the time subscript t for the marginal is unnecessary and 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. Let $\mathcal{M}$ denote the set of all probability measures that are defined on $\mathcal{A}^\infty$. Let $\mathcal{M}_S \subset \mathcal{M}$ denote the set of all probability measures under which the process $X^\infty(\omega)$ is strictly stationary. In the proofs, we also use the following notation:

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

where $\mathcal{M}^{(M)}$ is the set of all probability measures for the Borel σ-algebra of $\mathcal{X}^M$. 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 Pötscher and Prucha (1989), and CLT denotes the Central Limit Theorem of Herrndorf (1984).

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

A.1. **Trimming.** An essential tool of the proofs is the following mapping from $\mathcal{M}^{(M)}$ to Θ defined by a trimmed moment function:

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

where

$$
\begin{aligned}
\bar{\mathcal{P}}_{\theta}^{(M)} &= \left\{P^{(M)} \in \mathcal{M}_S^{(M)} : \int \phi_n(b,\theta) dP^{(M)} = 0\right\}, \\
\phi_n(b_j,\theta) &= \frac{1}{\sqrt{M}} \sum_{l=1}^{M} g_n(x_{(j-1)L+l},\theta) = \frac{1}{\sqrt{M}} \sum_{l=1}^{M} g(x_{(j-1)L+l},\theta)\mathbb{I}\{x_{(j-1)L+l} \in \mathcal{X}_n\}, \\
\mathcal{X}_n &= \left\{x \in \mathcal{X} : \sup_{\theta\in\Theta} |g(x,\theta)| \le m_n\right\},
\end{aligned}
$$

where $\{m_n\}_{n\in\mathbb{N}}$ is a sequence of positive numbers satisfying $m_n \to \infty$. Thus the set $\bar{\mathcal{P}}_{\theta}^{(M)}$ is the collection of probability measures satisfying the trimmed moment condition $E_{P^{(M)}}[\phi_n(B,\theta)] = 0$. Trimming is needed to guarantee existence of the mapping $\bar{T}(Q^{(M)})$. Lemma B.1 (i) shows that for each $n \in \mathbb{N}$ and $Q^{(M)} \in \mathcal{M}_S^{(M)}$ the value $\bar{T}(Q^{(M)})$ exists. To simplify the notation below we sometimes denote $\bar{T}_Q = \bar{T}(Q)$ for a measure Q.

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

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

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 (A.1). Assumption 3.1 (vii) together with (A.1) are sufficient to guarantee that

$$
\max\{M^{1-1/\eta} m_n^{1-\eta} n^{1/2}, M^{3/4} m_n^{-3} n^{1/2}, M m_n n^{-1/2}, n m_n^{-\eta}, M^3 n^{-1}\} \to 0, \tag{A.2}
$$

which is used in the proofs below.

A.2. **Key Lemmas.** Take any $r > 0$. The key to show Part (i) of Theorem 3.2 is to construct a strictly stationary probability measure $\widetilde{P}_{\theta_n}$ on $(\Omega,\mathcal{F})$ such that $\widetilde{P}_{\theta_n} \in \mathcal{B}_n(r)$ for all n large enough and its marginal $\widetilde{P}_{\theta_n}^{(1)}$ satisfies $E_{\widetilde{P}_{\theta_n}^{(1)}}[g(X,\theta_n)] = 0$ for all n large enough. Then the lower bound in Part (i) of Theorem 3.2 is given by evaluating the asymptotic MSE under this measure.

Definition A.1. For any $\xi \in \mathbb{R}^p$ and n large enough (so that $\theta_n \in \Theta$) define a stochastic process $\{Z_t\}$ in the following way. Let $\{X_t\}$ be a stochastic process generated from P_0 and define $\Upsilon(X_t)$, where for any $\gamma = (\gamma_1, \ldots, \gamma_d)'$, the r-th component of the vector transformation Υ is defined as

$$\Upsilon_r(\gamma) = \begin{cases} G_{r|}^{-1}(F_{r|}(\gamma_r|\gamma_{r-1}, \ldots, \gamma_1)|(\Upsilon_{r-1}(\gamma), \ldots, \Upsilon_1(\gamma))), & \text{if } r = 2, \ldots, d, \\ G_1^{-1}(F_1(\gamma_1)), & \text{if } r = 1, \end{cases}$$

where $F_{r|}(\gamma_r|\gamma_{r-1}, \ldots, \gamma_1)$ is the cumulative distribution function of the r-th component of X_t, conditional on the first $(r-1)$ components of this vector. Thus $F_{r|}$ is fully defined by the cumulative distribution function $F(\cdot)$ of X_t, which corresponds to the the probability measure $P_0^{(1)}$. Similarly, $G_{r|}^{-1}$ is the inverse (in the first argument) of the conditional cumulative distribution function $G_{r|}(\gamma_r|\gamma_{r-1}, \ldots, \gamma_1)$, which is defined by the probability measure $\widetilde{P}_{\theta_n}^{(1)}$ having the density

$$\frac{d\widetilde{P}_{\theta_n}^{(1)}}{dP_0^{(1)}}(x) = \frac{1 + \zeta_n' g_n(x, \theta_n)}{\int (1 + \zeta_n' g_n(x, \theta_n)) dP_0^{(1)}(x)}, \tag{A.3}$$

where $\zeta_n = -E_{P_0^{(1)}}[g(X, \theta_n) g_n(X, \theta_n)']^{-1} E_{P_0^{(1)}}[g(X, \theta_n)]$. Denote the probability measure of the process $\{\Upsilon(X_t)\}$ by $\widetilde{P}_{\theta_n}$.

Since P_0 generating $\{X_t\}$ is strictly stationary and satisfies the mixing condition in Definition 3.1, so is $\widetilde{P}_{\theta_n}$ generating $\{\Upsilon(X_t)\}$. Also, by construction, the marginal measure of $\Upsilon(X_t)$ is given by $\widetilde{P}_{\theta_n}^{(1)}$ and satisfies $E_{\widetilde{P}_{\theta_n}^{(1)}}[g(X, \theta_n)] = 0$.

Lemma A.2. *Suppose that Assumptions 3.1 and 3.3 hold. Then for all $r > 0$ and $\epsilon \in (0, r^2)$ satisfying $\frac{1}{4}\xi' G' V_1^{-1} G \xi \leq r^2 - \epsilon$, the probability measure $\widetilde{P}_{\theta_n}$ satisfies $\widetilde{P}_{\theta_n} \in \mathcal{B}_n(r)$ for all n large enough.*

Proof. Pick any $r > 0$ and $\epsilon \in (0, r^2)$ such that $\frac{1}{4}\xi' G' V_1^{-1} G \xi \leq r^2 - \epsilon$. First, we show that $\widetilde{P}_{\theta_n}$ satisfies Definition 3.1 (i) for all n large enough. Denote $f_n(x, \zeta_n) = \sqrt{\frac{d\widetilde{P}_{\theta_n}^{(1)}}{dP_0^{(1)}}(x)}$. By Taylor expansion,

$$H(\widetilde{P}_{\theta_n}^{(1)}, P_0^{(1)}) = \left\| \zeta_n' \frac{\partial f_n(x, 0)}{\partial \zeta_n} \sqrt{dP_0^{(1)}} + \frac{1}{2} \zeta_n' \frac{\partial f_n(x, \dot{\zeta}_n)}{\partial \zeta_n \partial \zeta_n'} \zeta_n \sqrt{dP_0^{(1)}} \right\|, \tag{A.4}$$

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

$$\frac{\partial f_n(x, 0)}{\partial \zeta_n} = \frac{1}{2}\{g_n(x, \theta_n) - E_{P_0^{(1)}}[g_n(X, \theta_n)]\},$$

and

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

By modifying the proof of Lemma B.4 (for the case of $M=1$), we obtain $\zeta_n = O(n^{-1/2})$ and thus $\sup_{x\in\mathcal{X}} \zeta_n' g_n(x,\theta_n) = o(1)$ (due to condition (A.2). So, Taylor expansions and Lemma B.4 (adapted for the case of $M=1$), we have

$$\begin{aligned}
nH(\widetilde{P}_{\theta_n}^{(1)}, P_0^{(1)})^2 &= n\left\| \frac{1}{2}\zeta_n'\{g_n(x,\theta_n) - E_{P_0^{(1)}}[g_n(X,\theta_n)]\}\sqrt{dP_0^{(1)}} \right\|^2 + o(1) \\
&= \frac{n}{4} E_{P_0^{(1)}}[g(X,\theta_n)]' V_1^{-1} E_{P_0^{(1)}}[g(X,\theta_n)] + o(1) \\
&= \frac{1}{4}\xi' E_{P_0^{(1)}}\left[\frac{\partial g(X,\dot{\theta}_n)}{\partial \theta'}\right]' V_1^{-1} E_{P_0^{(1)}}\left[\frac{\partial g(X,\dot{\theta}_n)}{\partial \theta'}\right]\xi + o(1) \\
&= \frac{1}{4}\xi' G' V_1^{-1} G\xi + o(1),
\end{aligned}$$

where each element of vector $\dot{\theta}_n$ is between the corresponding element of θ_n and 0. Therefore, $\widetilde{P}_{\theta_n}$ satisfies Definition 3.1 (i) for all n large enough.

Next, we show that $\widetilde{P}_{\theta_n}$ satisfies Definition 3.1 (ii) for all n large enough. Pick any t and s satisfying $|t-s| \leq M$. For the bivariate marginals $\widetilde{P}_{\theta_n}^{t,s}$ and $P_0^{t,s}$ on (X_t, X_s) under $\widetilde{P}_{\theta_n}$ and P_0, the Hellinger distance is characterized by

$$H(\widetilde{P}_{\theta_n}^{t,s}, P_0^{t,s})^2 = \left\| \sqrt{f(\Upsilon^{-1}(x_t), \Upsilon^{-1}(x_s))(\Upsilon^{-1})'(x_t)(\Upsilon^{-1})'(x_s)} - \sqrt{f(x_t,x_s)} \right\|^2,$$

where $f(x_t,x_s)$ is the bivariate density of (X_t,X_s) under P_0, and $\Upsilon^{-1}(x) = F^{-1}\left(\frac{F(x)+\zeta_n'\int^x g_n(a,\theta_n)f(a)da}{1+\zeta_n' E_{P_0}[g_n(X,\theta_n)]}\right)$. Thus, expansions around $\zeta_n = 0$ yield $H(\widetilde{P}_{\theta_n}^{t,s}, P_0^{t,s})^2 = O(\zeta_n^2)$. Since $\zeta_n = O(n^{-1/2})$ (by Lemma B.4 adapted for the case of $M=1$) and $a_n n^{1/2} \to \infty$, Definition 3.1 (ii) is satisfied for all n large enough.

Third, due to the construction of $\widetilde{P}_{\theta_n}$(i.e., $\{\Upsilon(X_t)\} \sim \widetilde{P}_{\theta_n}$ and $\{X_t\}$ is generated from strictly stationary P_0), $\widetilde{P}_{\theta_n}$ satisfies the mixing condition in Definition 3.1 (iii).

Finally, we check Definition 3.1 (iv):

$$E_{\widetilde{P}_{\theta_n}}\left[\sup_{\theta\in\Theta}|g(X_t,\theta)|^\eta\right] \leq \sup_{x\in\mathcal{X}_n}\left|\frac{1+\zeta_n' g_n(x,\theta_n)}{\{1+\zeta_n' E_{P_0}[g_n(X,\theta_n)]\}^2}\right| E_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}|g(X,\theta)|^\eta\right] < \infty,$$

where the second inequality follows from Assumption 3.1 (v) and the fact that $\sup_{x\in\mathcal{X}_n}\zeta_n' g_n(x,\theta_n) = o(1)$. Therefore, the conclusion is obtained.

Lemma A.3. *Suppose that Assumption 3.1 holds. Then for each* $r>0$,

$$\limsup_{n\to\infty}\sup_{Q\in\mathcal{B}_n(r)} n\{\tau\circ\bar{T}(Q^{(M)})-\tau(\theta_0)\}^2 \leq 4r^2\tau_\theta'\Sigma_1^{-1}\tau_\theta^2. \tag{A.5}$$

Proof. A Taylor expansion of $\tau\circ\bar{T}_{Q_n^{(M)}}$ around $\bar{T}_{Q_n^{(M)}}=\theta_0$, Lemmas B.1 (ii) and B.2, and Assumption 3.1 (viii) imply that for each sequence $Q_n\in\mathcal{B}_n(r)$ and $r>0$,

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

where we denote $\nu'=\tau_\theta'\Sigma^{-1}$. From the triangle inequality,

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

For A_1, observe that

$$A_1\leq\left|\int\nu'\Lambda_n\Lambda_n'\nu dQ_n^{(M)}\right|\left\|\sqrt{dQ_n^{(M)}}-\sqrt{dP_0^{(M)}}\right\|_2^2\leq B^*r^2\frac{M^2}{n}+o\left(\frac{M^2}{n}\right),$$

where the first inequality follows from Cauchy-Schwarz inequality, and the second inequality follows from Lemma B.5 (i) and $Q\in\mathcal{B}_n(r)$. Similarly, we have $A_2\leq B^*r^2\frac{M}{n}$. From these results, A_3 satisfies

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

Combining these terms, we obtain $n\{\tau\circ\bar{T}_{Q_n^{(M)}}-\tau(\theta_0)\}^2\leq 4r^2B^*+o(1)$ for each sequence $Q_n\in\mathcal{B}_n(r)$ and $r>0$. Since $Q^{(M)}$ implied from $Q\in\mathcal{B}_n(r)$‘ belongs to a compact of measures for each $n\in\mathbb{N}$ and $r>0$, the conclusion follows.

A.3. **Proof of Theorem 3.2. Proof of (i).** Let $\Sigma_1 = G'V_1G$. Pick any $\epsilon \in (0, r^2)$ and take

$$\bar{\xi} = 2\sqrt{r^2 - \epsilon}(\tau_\theta' \Sigma_1^{-1} \tau_\theta)^{-1/2} \Sigma_1^{-1} \tau_\theta.$$

Then $\frac{1}{4}\bar{\xi}'\Sigma_1\bar{\xi} = r^2 - \epsilon$, and hence $\widetilde{P}_{\theta_0+\bar{\xi}/\sqrt{n}} \in \mathcal{B}_n(r)$ for all n large enough by Lemma A.2. Also, $E_{\widetilde{P}^{(1,t)}_{\theta_0+\bar{\xi}/\sqrt{n}}}[g(X_t, \theta_0 + \bar{\xi}/\sqrt{n})] = 0$. Hence, $\widetilde{P}_{\theta_n}$ satisfies the conditions imposed on measure Q_n in Assumption 3.2. Then we have

$$\begin{aligned}
& \lim_{\kappa\to\infty} \liminf_{n\to\infty} \sup_{Q\in\mathcal{B}_n(r)} \int \kappa \wedge n\{\tau \circ T_a(x_1, \ldots, x_n) - \tau(\theta_0)\}^2 dQ \\
\geq\ & \lim_{\kappa\to\infty} \liminf_{n\to\infty} \int \kappa \wedge n\{\tau \circ T_a(x_1, \ldots, x_n) - \tau(\theta_0)\}^2 d\widetilde{P}^{(n,1)}_{\theta_0+\bar{\xi}/\sqrt{n}} \\
=\ & \lim_{\kappa\to\infty} \liminf_{n\to\infty} \int \kappa \wedge \left(\tau_\theta' \left\{\bar{\xi} + \frac{1}{\sqrt{n}} \sum_{t=1}^n \varphi_n(x_t)\right\}\right)^2 d\widetilde{P}^{(n,1)}_{\theta_0+\bar{\xi}/\sqrt{n}} \\
=\ & (\tau_\theta'\bar{\xi})^2 + \tau_\theta' A_{\varphi\varphi'} \tau_\theta \\
\geq\ & 4(r^2 - \epsilon)\tau_\theta' \Sigma_1^{-1} \tau_\theta^2 + B^*,
\end{aligned}$$

where the first inequality follows from $\widetilde{P}_{\theta_n} \in \mathcal{B}_n(r)$, 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,

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

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

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

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 $A_3 \leq 4(r^2-\epsilon)\tau_\theta'\Sigma_1^{-1}\tau_\theta^2 + B^*$ as $\kappa\to\infty$. Pick any $\kappa > 0$. Consider the mapping $f_{\kappa,n}(Q) = \int \kappa \wedge n\{\tau\circ\bar{T}(P_n^{(M)}) - \tau(\theta_0)\}^2 dQ$. For any $\epsilon > 0$ and for all $n\in\mathbb{N}$ by definition of supremum there exists $\tilde{Q}_n \in \mathcal{B}_n(r)$ such that

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

for each n. Then we have

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

where the second equality follows from Lemma B.10 and the continuous mapping theorem, the second inequality follows from $\kappa\wedge c\leq c$ and a direct calculation. Here

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

which satisfies

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

where the equality follows from Lemma B.1 (ii) and Assumption 3.1 (viii), the second inequality follows from Jensen's inequality, the third inequality follows from the inclusion relationship $Q^{(M,(j-1)L+1)}\in B_H(P_0^{(M)},r\sqrt{M/n})$ for all $j\in 1,\dots,n_B$ by Definition 3.1 (i), the fourth inequality follows from Lemma B.2, Assumption 3.1 (viii), and the fact that $B_H(P_0^{(M)},r\sqrt{M/n})$ is a compact for all n, and the last inequality follows from Lemma A.3. Hence $A_3\leq(1+4r^2)B^*$, which concludes the proof.

Appendix B. Auxiliary Lemmas

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

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

Proof of (i). Pick any $n\in\mathbb{N}$ and $Q^{(M)}\in\mathcal{M}^{(M)}$. Denote $R_n(Q^{(M)},\theta)=\inf_{P\in\bar{\mathcal{P}}_\theta^{(M)}} H(P^{(M)},Q^{(M)})$. 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 $R_n(Q^{(M)},\theta)=\max_{\gamma\in\mathbb{R}^m} R_n(Q^{(M)},\theta,\gamma)$ for each $(Q^{(M)},\theta)\in\mathcal{M}^{(M)}\times\Theta$. From Rockafeller (1970, Theorem 10.8) and Assumption 3.1 (iv), $R_n(Q^{(M)},\theta)$ is continuous in $(Q^{(M)},\theta)\in\mathcal{M}^{(M)}\times\Theta$ under the Levy metric. This continuity also implies that for each $Q^{(M)}\in\mathcal{M}^{(M)}$, $R_n(Q^{(M)},\theta)$ is continuous in $\theta\in\Theta$. Since Θ is compact (Assumption 3.1 (ii)), $\bar{T}(Q^{(M)})=\arg\min_{\theta\in\Theta} R_n(Q^{(M)},\theta)$ exists.

Proof of (ii). Pick any $r>0$ and any sequence $Q_n\in\mathcal{B}_n(r)$. The proof is based on Newey and Smith (2004, proof of Theorem 3.1). From Lemma B.6 (i), $|E_{Q_n^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})]|\to 0$. From the triangle inequality,

$$
\begin{aligned}
&\sup_{\theta\in\Theta}|E_{Q_n^{(M)}}[\phi_n(B,\theta)]-E_{P_0^{(M)}}[\phi(B,\theta)]|\\
\leq\;&\sup_{\theta\in\Theta}|E_{Q_n^{(M)}}[\phi_n(B,\theta)]-E_{P_0^{(M)}}[\phi_n(B,\theta)]|+\sup_{\theta\in\Theta}|E_{P_0^{(M)}}[\phi_n(B,\theta)-\phi(B,\theta)]|:=T_1+T_2.
\end{aligned}
$$

The first term T_1 satisfies

$$
\begin{aligned}
T_1 &\leq \frac{1}{\sqrt{M}}\sum_{j=1}^{M}\sup_{\theta\in\Theta}\left|\int g_n(X_j,\theta)\left(\sqrt{dQ_n^{(1)}}-\sqrt{dP_0^{(1)}}\right)^2\right|\\
&\quad+\frac{2}{\sqrt{M}}\sum_{j=1}^{M}\sup_{\theta\in\Theta}\left|\int g_n(X_j,\theta)\sqrt{dP_0^{(1)}}\left(\sqrt{dQ_n^{(1)}}-\sqrt{dP_0^{(1)}}\right)\right|\\
&\leq m_n\sqrt{M}\frac{r^2}{n}+2\sqrt{M}\sqrt{E_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}|g(X,\theta)|^2\right]}\frac{r}{\sqrt{n}}=o(1), \qquad \text{(B.1)}
\end{aligned}
$$

where the first inequality follows from the triangle inequality, the second inequality follows from $Q_n\in\mathcal{B}_n(r)$ and Cauchy-Schwarz inequality, and the equality follows from Assumption 3.1 (v) and

(A.2). The second term of T_2 satisfies

$$\begin{aligned} T_2 &\leq \sqrt{M}\left(E_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}|g(X,\theta)|^\eta\right]\right)^{1/\eta}\left(P_0^{(1)}\{X\notin\mathcal{X}_n\}\right)^{(\eta-1)/\eta} \\ &\leq C\sqrt{M}\left(m_n^{-\eta}E_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}|g(X,\theta)|^\eta\right]\right)^{(\eta-1)/\eta} = o(1), \end{aligned} \tag{B.2}$$

where the first inequality follows from Hlder inequality, the second inequality follows from Markov inequality, and the equality follows from Assumption 3.1 (v) and (A.2). Combining these results, we obtain the uniform convergence $\sup_{\theta\in\Theta}|E_{Q_n^{(M)}}[\phi_n(B,\theta)]-E_{P_0^{(M)}}[\phi(B,\theta)]|\to 0$. Thus by the triangle inequality,

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

The conclusion is obtained from Lemma B.6 (i) and Assumption 3.1 (iii).

Lemma B.2. *Suppose that Assumption 3.1 holds. Then, for each $r>0$ and each sequence $Q_n\in\mathcal{B}_n(r)$,*

$$\sqrt{n}(\bar{T}_{Q_n^{(M)}}-\theta_0) = -\sqrt{n}(M\Sigma)^{-1}\int\Lambda_n dQ_n^{(M)}+o(1). \tag{B.3}$$

Proof. The proof is based on Rieder (1994, proofs of Theorems 6.3.4 and 6.4.5). Pick any $r>0$ and $Q_n\in\mathcal{B}_n(r)$. In this proof, we omit the upperscript "(M)" and denote $Q_n=Q_n^{(M)}$, $\bar{P}_{\theta_0,Q_n}=\bar{P}^{(M)}_{\theta_0,Q_n^{(M)}}$, so on. Observe that

$$\begin{aligned} &\left\|dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}(\bar{T}_{Q_n}-\theta_0)'\Lambda_n dQ_n^{1/2}\right\|^2 \\ =& \left\|dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\Lambda_n dQ_n^{1/2}\right\|^2+\left\|\frac{1}{2}(\bar{T}_{Q_n}-\theta_0-\psi_{n,Q_n})'\Lambda_n dQ_n^{1/2}\right\|^2 \\ &+\left\{\int\left(dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\Lambda_n dQ_n^{1/2}\right)\Lambda_n' dQ_n^{1/2}\right\}(\bar{T}_{Q_n}-\theta_0-\psi_{n,Q_n}) \\ =& \left\|dQ_n^{1/2}-d\bar{P}_{\theta_0,Q_n}^{1/2}+\frac{1}{2}\psi_{n,Q_n}'\Lambda_n dQ_n^{1/2}\right\|^2+\left\|\frac{1}{2}(\bar{T}_{Q_n}-\theta_0-\psi_{n,Q_n})'\Lambda_n dQ_n^{1/2}\right\|^2, \end{aligned} \tag{B.4}$$

where the second equality follows from

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

From the triangle inequality, the left hand side of (B.4) satisfies

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

where the second inequality follows from Lemma B.3 (i), the third inequality follows from the definition of $\bar{T}_{Q_n}$, the fourth inequality follows from the triangle inequality, and the equality follows from Lemma B.3 (ii). Thus, from (B.4),

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

This implies

$$\begin{aligned}
& o(\sqrt{M}|\bar{T}_{Q_n} - \theta_0|) + o(\sqrt{M}|\psi_{n,Q_n}|) + o(\sqrt{M/n}) \\
\geq \; & \sqrt{\frac{1}{4}(\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n})' \int \Lambda_n \Lambda_n' dQ_n (\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n})} \\
\geq \; & C\sqrt{M}|\bar{T}_{Q_n} - \theta_0 - \psi_{n,Q_n}|, \qquad \text{(B.5)}
\end{aligned}$$

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

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

$$\begin{aligned}
\psi_{n,Q_n} \;=\; & -2\{E_{Q_n}[\Lambda_n\Lambda_n']^{-1} - M^{-1}\Sigma^{-1}\} \int \Lambda_n (dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2}) dQ_n^{1/2} \\
& -2M^{-1}\Sigma^{-1} \int \Lambda_n (dQ_n^{1/2} - d\bar{P}_{\theta_0,Q_n}^{1/2}) dQ_n^{1/2}. \qquad \text{(B.6)}
\end{aligned}$$

Note that from the convex duality of partially finite programming (Borwein and Lewis, 1993), the Radon-Nikodym derivative can be written as

$$\frac{d\bar{P}^{(M)}_{\theta,Q^{(M)}}}{dQ^{(M)}} = \frac{1}{\{1+\gamma_n(\theta,Q)'\phi_n(b,\theta)\}^2}, \tag{B.7}$$

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

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

where

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

Thus, if $E_Q[\phi_n(B,\theta)\phi_n(B,\theta)']$ is invertible, $\gamma_n(\theta, Q)$ is written as

$$\begin{aligned} \gamma_n(\theta,Q) &= \frac{1}{2}E_Q[\phi_n(B,\theta)\phi_n(B,\theta)']^{-1}E_Q[\phi_n(B,\theta)] \\ &\quad + \frac{1}{2}E_Q[\phi_n(B,\theta)\phi_n(B,\theta)']^{-1}E_Q[\varrho_n(B,\theta,Q)\phi_n(B,\theta)]. \end{aligned} \tag{B.8}$$

The second term of (B.6) is

$$\begin{aligned} & -2M^{-1}\Sigma^{-1}\int \Lambda_n(dQ_n^{1/2} - d\bar{P}^{1/2}_{\theta_0,Q_n})dQ_n^{1/2} \\ = & -2\Sigma^{-1}M^{-1/2}G'\Omega^{-1}E_{Q_n}[\phi_n(B,\theta_0)\phi_n(B,\theta_0)']\gamma_n(\theta_0,Q_n) \\ & +2\Sigma^{-1}M^{-1/2}G'\Omega^{-1}\left(\int \frac{\gamma_n(\theta_0,Q_n)'\phi_n(b,\theta_0)}{1+\gamma_n(\theta_0,Q_n)'\phi_n(b,\theta_0)}\phi_n(b,\theta_0)\phi_n(b,\theta_0)'dQ_n\right)\gamma_n(\theta_0,Q_n) \\ = & -\Sigma^{-1}M^{-1/2}G'\Omega^{-1}\left\{E_{Q_n}[\phi_n(B,\theta_0)] + \frac{1}{2}E_{Q_n}[\varrho_n(B,\theta_0,Q_n)\phi_n(B,\theta_0)]\right\} + o(n^{-1/2}) \\ = & -M^{-1}\Sigma^{-1}\int \Lambda_n dQ_n + o(n^{-1/2}), \end{aligned} \tag{B.9}$$

where the first equality follows from (B.7), the second equality follows from (B.8) and Lemma B.5, and the third equality follows from Lemma B.5. Similarly, the first term of (B.6) is $o(n^{-1/2})$. Therefore,

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

and $|\psi_{n,Q_n}| = O(n^{-1/2})$ from Lemma B.5. Then from (B.5),

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

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

Lemma B.3. *Suppose that Assumption 3.1 holds. Then for each $r > 0$ and each sequence $Q_n \in \mathcal{B}_n(r)$,*

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

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

Proof of (i). In this proof, we omit the upperscript "(M)" and denote $Q_n = Q_n^{(M)}$, $\bar{P}_{\theta_0,Q_n} = \bar{P}^{(M)}_{\theta_0,Q_n^{(M)}}$, so on. Denote $t_n = \bar{T}_{Q_n} - \theta_0$. Pick any $r > 0$ and $Q_n \in \mathcal{B}_n(r)$. From (B.7),

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

For T_2, Lemmas B.5 and B.6 imply

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

Thus, we focus on T_1. From (B.8),

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

Lemmas B.5 and B.6 imply that $T_{12} = o(\sqrt{M/n})$ and $T_{13} = o(\sqrt{M/n})$. Thus, we focus on T_{11}. Taylor expansions of $\phi_n(b, \bar{T}_{Q_n})$ around $\bar{T}_{Q_n} = \theta_0$ yield

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

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

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

Lemma B.4. *Suppose that Assumption 3.1 holds. Then, for each* $\xi \in \mathbb{R}^p$, $|E_{P_0^{(M)}}[\phi_n(B, \theta_0)]| = o(\sqrt{M/n})$, $|E_{P_0^{(M)}}[\phi_n(B, \theta_n)]| = O(\sqrt{M/n})$, $|E_{P_0^{(M)}}[\phi_n(B, \theta_n)\phi_n(B, \theta_n)'] - V| = o(1)$, *and* $|E_{P_0^{(M)}}[\partial \phi_n(B, \theta_n)/\partial \theta'] - \sqrt{M} G| = o(\sqrt{M})$.

Proof of the first statement. The same argument as in (B.2) yields $|E_{P_0^{(M)}}[\phi_n(B, \theta_0)]| = O(\sqrt{M} m_n^{1-\eta})$. The conclusion follows by (A.2).

Proof of the second statement. Pick any $\xi \in \mathbb{R}^p$. By the triangle inequality and (B.2), we obtain

$$\begin{aligned}
|E_{P_0^{(M)}}[\phi_n(B, \theta_n)]| &\leq \sqrt{M} |E_{P_0^{(1)}}[g(X, \theta_n)]| + o(\sqrt{M/n}) \\
&\leq C \sqrt{\frac{M}{n}} E_{P_0^{(1)}}\left[\sup_{\theta \in \mathcal{U}} \left| \frac{\partial g(X, \theta)}{\partial \theta'} \right| \right] + o(\sqrt{M/n}) = O(\sqrt{M/n}),
\end{aligned}$$

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

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

$$
\begin{aligned}
& |E_{P_0^{(M)}}[\phi_n(B,\theta_n)\phi_n(B,\theta_n)'] - V| \\
\leq\ & |E_{P_0^{(M)}}[\phi_n(B,\theta_n)\phi_n(B,\theta_n)'] - E_{P_0^{(M)}}[\phi(B,\theta_n)\phi(B,\theta_n)']| \\
& + |E_{P_0^{(M)}}[\phi(B,\theta_n)\phi(B,\theta_n)'] - E_{P_0^{(M)}}[\phi(B,\theta_0)\phi(B,\theta_0)']| \\
& + |E_{P_0^{(M)}}[\phi(B,\theta_0)\phi(B,\theta_0)'] - V| \\
\leq\ & \frac{1}{M}\sum_{j,l=1}^{M} |E_{P_0^{(M)}}[g(X_j,\theta_n)g(X_l,\theta_n)'\mathbb{I}\{X_j \notin \mathcal{X}_n \text{or } X_l \notin \mathcal{X}_n\}]| + o(1) \\
\leq\ & CMm_n^{-\eta/2}\sqrt{E_{P_0^{(1)}}[|g(X,\theta_n)|^4]} + o(1), \qquad \text{(B.10)}
\end{aligned}
$$

where the second inequality follows from the continuity of $g(x,\theta)$ at θ_0 and the definition of V, and the third inequality follows from Cauchy-Schwarz and Markov inequalities. The conclusion follows by Assumption 3.1 and (A.2).

Proof of the fourth statement. Pick any $\xi \in \mathbb{R}^p$. By the triangle inequality

$$
\begin{aligned}
& |E_{P_0^{(M)}}[\partial\phi_n(B,\theta_n)/\partial\theta'] - E_{P_0^{(M)}}[\partial\phi(B,\theta_0)/\partial\theta']| \\
\leq\ & |E_{P_0^{(M)}}[\partial\phi_n(B,\theta_n)/\partial\theta' - \partial\phi(B,\theta_n)/\partial\theta']| + \sqrt{M}|E_{P_0^{(1)}}[\partial g(X,\theta_n)/\partial\theta' - \partial g(X,\theta_0)/\partial\theta']| \\
\leq\ & \sqrt{M}\left(E_{P_0^{(1)}}\left[\left|\frac{\partial g(X,\theta_n)}{\partial\theta'}\right|^{\eta}\right]\right)^{1/\eta}\left(m_n^{-\eta}E_{P_0^{(1)}}\left[\sup_{\theta\in\Theta}|g(X,\theta)|^{\eta}\right]\right)^{(\eta-1)/\eta} + o(\sqrt{M}) = o(\sqrt{M}),
\end{aligned}
$$

where the second inequality follows from the triangle, Cauchy-Schwarz, and Markov inequalities and the continuity of $\partial g(x,\theta)/\partial\theta'$ at θ_0, and the equality follows from Assumption 3.1 (v) and (A.2).

Lemma B.5. *Suppose that Assumption 3.1 holds. Then for each $r > 0$ and each sequence $Q_n \in \mathcal{B}_n(r)$,*

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

Proof of (i). Proof of the first statement. Pick any $r > 0$ and any sequence $Q_n \in \mathcal{B}_n(r)$. By the triangle inequality and Lemma B.4,

$$
\begin{aligned}
|E_{Q_n^{(M)}}[\phi_n(B,\theta_0)]| &\leq \frac{1}{\sqrt{M}}\sum_{l=1}^{M}\left|\int g_n(x,\theta_0)\left(\sqrt{dQ_n^{(1,l)}}-\sqrt{dP_0^{(1)}}\right)^2\right| \\
&\quad + \frac{2}{\sqrt{M}}\sum_{l=1}^{M}\left|\int g_n(x,\theta_0)\sqrt{dP_0^{(1)}}\left(\sqrt{dQ_n^{(1,l)}}-\sqrt{dP_0^{(1)}}\right)\right| + o(\sqrt{M/n}) \\
&\leq \sqrt{M}m_n\frac{r^2}{n} + 2\sqrt{M}\sqrt{E_{P_0^{(1)}}[|g(X,\theta_0)|^2]}\frac{r}{\sqrt{n}} + o(\sqrt{M/n}), \qquad \text{(B.11)}
\end{aligned}
$$

where the second inequality follows from Cauchy-Schwarz inequality and $Q_n \in \mathcal{B}_n(r)$. Thus, the conclusion is obtained by Assumption 3.1 (v) and (A.2).

Proof of the second statement. Pick any $r > 0$ and any sequence $Q_n \in \mathcal{B}_n(r)$. From the triangle inequality,

$$
\begin{aligned}
&|E_{Q_n^{(M)}}[\phi_n(B,\theta_0)\phi_n(B,\theta_0)'] - V| \\
\leq\ & |E_{Q_n^{(M)}}[\phi_n(B,\theta_0)\phi_n(B,\theta_0)'] - E_{P_0^{(M)}}[\phi_n(B,\theta_0)\phi_n(B,\theta_0)']| \\
& + |E_{P_0^{(M)}}[\phi_n(B,\theta_0)\phi_n(B,\theta_0)'] - V| \\
\leq\ & \frac{1}{M}\sum_{j,l=1}^{M}\left|\int g_n(x_j,\theta_0)g_n(x_l,\theta_0)'\left(\sqrt{dQ_n^{j,l}}-\sqrt{dP_0^{j,l}}\right)^2\right| \\
& + \frac{2}{M}\sum_{j,l=1}^{M}\left|\int g_n(x_j,\theta_0)g_n(x_l,\theta_0)'\sqrt{dP_0^{j,l}}\left(\sqrt{dQ_n^{j,l}}-\sqrt{dP_0^{j,l}}\right)\right| + o(1) \\
\leq\ & m_n^2 M a_n^2 + 2Ma_n\sqrt{E_{P_0^{(1)}}[|g_n(X,\theta_0)|^4]} + o(1) = o(1), \qquad \text{(B.12)}
\end{aligned}
$$

where the second inequality follows from the triangle inequality and Lemma B.4, the third inequality follows from Cauchy-Schwarz inequality and $Q_n \in \mathcal{B}_n(r)$, and the equality follows from Assumption 3.1 (v) and (A.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

$$
\Gamma_n = \{\gamma \in \mathbb{R}^m : |\gamma| \leq A_n\}, \qquad \text{(B.13)}
$$

with $A_n\sqrt{M}m_n \to 0$ and $A_n\sqrt{n/M} \to \infty$ ((A.2) guarantees existence of A_n). Observe that

$$
\sup_{\gamma\in\Gamma_n, b\in\mathcal{X}_n^M, \theta\in\Theta} |\gamma'\phi_n(b,\theta)| \leq A_n\sqrt{M}m_n \to 0. \qquad \text{(B.14)}
$$

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

$$\begin{aligned} -1 &= R_n(Q_n^{(M)}, \theta_n, 0) \leq R_n(Q_n^{(M)}, \theta_n, \tilde{\gamma}) \\ &= -1 + \tilde{\gamma}' E_{Q_n^{(M)}}[\phi_n(B, \theta_n)] - \tilde{\gamma}' E_{Q_n^{(M)}}\left[\frac{\phi_n(B, \theta_n)\phi_n(B, \theta_n)'}{\{1 + \dot{\gamma}'\phi_n(B, \theta_n)\}^3}\right]\tilde{\gamma} \\ &\leq -1 + |\tilde{\gamma}||E_{Q_n^{(M)}}[\phi_n(B, \theta_n)]| - C|\tilde{\gamma}|^2, \end{aligned} \tag{B.15}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, and the second inequality follows from (B.14), Lemma B.5 (i), and Assumption 3.1 (vi). Thus, Lemma B.5 (i) implies

$$C|\tilde{\gamma}| \leq |E_{Q_n^{(M)}}[\phi_n(B, \theta_n)]| = O(\sqrt{M/n}). \tag{B.16}$$

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 $\partial R_n(Q_n^{(M)}, \theta_0, \tilde{\gamma})/\partial\gamma = 0$ for all n large enough. Since $R_n(Q_n^{(M)}, \theta_0, \gamma)$ is concave in γ for all n large enough, $\tilde{\gamma} = \arg\max_{\gamma\in\mathbb{R}^m} R_n(Q_n^{(M)}, \theta_n, \gamma)$ for all n large enough. Thus, the first statement is obtained. Also, from (B.16), the second statement is obtained. Using condition (A.2), the third statement follows from

$$\sup_{b\in\mathcal{X}_n^M} |\gamma_n(\theta_n, Q_n^{(M)})'\phi_n(b, \theta_n)| = O(Mn^{-1/2}m_n) = o(1). \tag{B.17}$$

Lemma B.6. *Suppose that Assumption 3.1 holds. Then for each $r > 0$ and each sequence $Q_n \in \mathcal{B}_n(r)$,*

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

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

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

$$\sup_{b\in\mathcal{X}_n^M, \theta\in\Theta} |\tilde{\gamma}'\phi_n(b, \theta)| \leq Mm_n/\sqrt{n} \to 0. \tag{B.18}$$

By a similar argument to (B.12),

$$\begin{aligned} & |E_{Q_n^{(M)}}[\phi_n(B, \bar{T}_{Q_n^{(M)}})\phi_n(B, \bar{T}_{Q_n^{(M)}})']| \\ \leq\ & \sum_{j=1}^{M} E_{P_0^{1,j}}\left[\sup_{\theta\in\Theta}|g_n(X_1, \theta)g_n(X_j, \theta)'|\right] + O(m_n^2 M a_n^2) + O(Ma_n) \\ \leq\ & 12\left(E_{P_0}\left[\sup_{\theta\in\Theta}|g(X, \theta)|^\eta\right]\right)^{2/\eta} \sum_{j=1}^{M} \alpha(j-1)^{1-2/\eta} + O(1), \end{aligned} \tag{B.19}$$

where the second inequality follows from Davydov (1968, Corollary) under Assumption 3.1 (i) and the condition on a_n. Thus, $|E_{Q_n^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})\phi_n(B,\bar{T}_{Q_n^{(M)}})']|$ is finite for all n large enough.

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

$$
\begin{aligned}
&= -1+\tilde{\gamma}'E_{Q_n^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})] - \tilde{\gamma}'E_{Q_n^{(M)}}\left[\frac{\phi_n(B,\bar{T}_{Q_n^{(M)}})\phi_n(B,\bar{T}_{Q_n^{(M)}})'}{\{1+\dot{\gamma}'\phi_n(B,\bar{T}_{Q_n^{(M)}})\}^3}\right]\tilde{\gamma} \\
&\geq -1+\sqrt{M/n}|E_{Q_n^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})]| - CM/n, \qquad \text{(B.20)}
\end{aligned}
$$

for all n large enough, where $\dot{\gamma}$ is a point one the line joining 0 and $\tilde{\gamma}$, the inequality follows from (B.18) and $\tilde{\gamma}'\tilde{\gamma}=M/n$. From the definitions of $\gamma_n(\bar{T}_{Q_n^{(M)}},Q_n)$ and $\bar{T}_{Q_n^{(M)}}$,

$$
\begin{aligned}
&-1+\sqrt{M/n}|E_{Q_n^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})]| - CM/n \\
&\leq R_n(Q_n,\bar{T}_{Q_n^{(M)}},\tilde{\gamma}) \leq R_n(Q_n,\bar{T}_{Q_n^{(M)}},\gamma_n(\bar{T}_{Q_n^{(M)}},Q_n)) \leq R_n(Q_n,\theta_0,\gamma_n(\theta_0,Q_n^{(M)})), \qquad \text{(B.21)}
\end{aligned}
$$

where the first inequality follows from (B.20). From $|\gamma_n(\theta_0,Q_n^{(M)})|=O(\sqrt{M/n})$ and $|E_{Q_n^{(M)}}[\phi_n(B,\theta_0)]| = O(\sqrt{M/n})$ (by Lemma B.5), similar to (B.15) we have

(B.22)

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

The conclusion follows by (B.21) and (B.22).

Proof of the second statement. Pick any $r>0$ and any sequence $Q_n\in\mathcal{B}_n(r)$. From the triangle inequality,

$$
\begin{aligned}
&|E_{Q_n^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})\phi_n(B,\bar{T}_{Q_n^{(M)}})'] - E_{P_0^{(M)}}[\phi(B,\theta_0)\phi(B,\theta_0)']| \\
\leq\ &|E_{Q_n^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})\phi_n(B,\bar{T}_{Q_n^{(M)}})'] - E_{P_0^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})\phi_n(B,\bar{T}_{Q_n^{(M)}})']| \\
&+|E_{P_0^{(M)}}[\phi_n(B,\bar{T}_{Q_n^{(M)}})\phi_n(B,\bar{T}_{Q_n^{(M)}})' - \phi(B,\bar{T}_{Q_n^{(M)}})\phi(B,\bar{T}_{Q_n^{(M)}})']| \\
&+|E_{P_0^{(M)}}[\phi(B,\bar{T}_{Q_n^{(M)}})\phi(B,\bar{T}_{Q_n^{(M)}})'] - E_{P_0^{(M)}}[\phi(B,\theta_0)\phi(B,\theta_0)']| := T_1+T_2+T_3.
\end{aligned}
$$

By a similar argument to (B.12) combined with $\bar{T}_{Q_n^{(M)}}\to\theta_0$ (Lemma B.1 (ii)), the term T_1 satisfies

$$
T_1 \leq m_n^2 M a_n^2 + 2Ma_n\sqrt{E_{P_0^{(1)}}\left[\sup_{\theta\in\mathcal{U}}|g_n(X,\theta)|^4\right]} = o(1).
$$

By a similar argument to (B.10) combined with $\bar{T}_{Q_n^{(M)}}\to\theta_0$, the term T_2 satisfies

$$
T_2 \leq CMm_n^{-\eta/2}\sqrt{E_{P_0^{(1)}}\left[\sup_{\theta\in\mathcal{U}}|g(X,\theta)|^4\right]} = o(1).
$$

The term T_3 satisfies

$$\begin{aligned} T_3 &\leq \sum_{j=1}^{M} |E_{P_0^{j.1}}[g(X_j, \bar{T}_{Q_n^{(M)}})g(X, \bar{T}_{Q_n^{(M)}})' - g(X_j, \theta_0)g(X_1, \theta_0)']| \\ &\leq \sum_{j=1}^{M} E_{P_0^{j.1}}\left[\sup_{\theta\in\mathcal{U}}\left|\frac{\partial g(X_j,\theta)}{\partial\theta'}\right|\left|\frac{\partial g(X_1,\theta)}{\partial\theta'}\right|\right]|\bar{T}_{Q_n^{(M)}} - \theta_0|^2 \\ &\quad +2\sum_{j=1}^{M} E_{P_0^{j.1}}\left[\sup_{\theta\in\mathcal{U}}\left|\frac{\partial g(X_j,\theta)}{\partial\theta'}\right||g(X_1,\theta)|\right]|\bar{T}_{Q_n^{(M)}} - \theta_0|, \end{aligned}$$

where the first inequality follows from the triangle inequality and strict stationarity of P_0, and the second inequality follows from a Taylor expansion around $\bar{T}_{Q_n^{(M)}} = \theta_0$. By a similar argument in (B.19) using Davydov (1968, Corollary) and $\bar{T}_{Q_n^{(M)}} \to \theta_0$ (Lemma B.1 (ii)), we have $T_3 = o(1)$ and the conclusion follows.

Proof of the third statement. Pick any $r > 0$ and any sequence $Q_n \in \mathcal{B}_n(r)$. From the triangle inequality,

$$\begin{aligned} & |E_{Q_n^{(M)}}[\partial\phi_n(B, \bar{T}_{Q_n^{(M)}})/\partial\theta'] - E_{P_0^{(M)}}[\partial\phi(B, \theta_0)/\partial\theta']| \\ \leq\ & |E_{Q_n^{(M)}}[\partial\phi_n(B, \bar{T}_{Q_n^{(M)}})/\partial\theta'] - E_{P_0^{(M)}}[\partial\phi_n(B, \bar{T}_{Q_n^{(M)}})/\partial\theta']| \\ & +|E_{P_0^{(M)}}[\partial\phi_n(B, \bar{T}_{Q_n^{(M)}})/\partial\theta' - \partial\phi(B, \bar{T}_{Q_n^{(M)}})/\partial\theta']| \\ & +|E_{P_0^{(M)}}[\partial\phi(B, \bar{T}_{Q_n^{(M)}})/\partial\theta'] - E_{P_0^{(M)}}[\partial\phi(B, \theta_0)/\partial\theta']| := T_1 + T_2 + T_3. \end{aligned}$$

By the triangle inequality, the term T_1 satisfies

$$\begin{aligned} T_1 &\leq \sqrt{M}\int |\partial g_n(x, \bar{T}_{Q_n^{(M)}})/\partial\theta'|\left(\sqrt{dQ_n^{(1)}} - \sqrt{dP_0^{(1)}}\right)^2 \\ &\quad +2\sqrt{M}\int |\partial g_n(x, \bar{T}_{Q_n^{(M)}})/\partial\theta'|\sqrt{dP_0^{(1)}}\left(\sqrt{dQ_n^{(1)}} - \sqrt{dP_0^{(1)}}\right) \\ &\leq o(\sqrt{M/n}) + 2r\sqrt{\frac{M}{n}}\sqrt{E_{P_0^{(1)}}\left[\sup_{\theta\in\mathcal{U}}|\partial g_n(x,\theta)/\partial\theta'|^2\right]}, \end{aligned}$$

where the second inequality follows from Cauchy-Schwartz inequality, $Q_n \in \mathcal{B}_n(r)$, and Assumption 3.1 (v). Thus, by (A.2), $T_1 = o(1)$. By Cauchy-Schwarz inequality, the term T_2 satisfies

$$\begin{aligned} & |E_{P_0^{(M)}}[\partial\phi_n(B, \bar{T}_{Q_n^{(M)}})/\partial\theta' - \partial\phi(B, \bar{T}_{Q_n^{(M)}})/\partial\theta']| \\ \leq\ & \sqrt{M}\sqrt{E_{P_0^{(1)}}\left[\sup_{\theta\in\mathcal{U}}|\partial g(X,\theta)/\partial\theta'|^2\right]}\sqrt{P_0^{(1)}\{X \notin \mathcal{X}_n\}} = O(\sqrt{M}m_n^{-\eta/2}), \end{aligned}$$

where the equality follows from Markov inequality. By $m_n \to \infty$, we have $T_2 = o(\sqrt{M})$. Also the term T_3 is $o(\sqrt{M})$ 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 same as for Lemma B.5 (ii) except using Lemma B.6 (i) instead of Lemma B.5 (i).

Lemma B.7. *Suppose that Assumption 3.1 holds. Then, for each $r > 0$ and each sequence $Q_n \in \mathcal{B}_n(r)$, $\bar{T}_{P_n^{(M)}} \to_p \theta_0$ under Q_n.*

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

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

The first term is $o_p(1)$ from a UWLLN. The second and third terms are $o(1)$ by similar arguments in (B.1) and (B.2), respectively. Therefore, $\sup_{\theta\in\Theta} |E_{P_n^{(M)}}[\phi_n(B,\theta)] - E_{P_0^{(M)}}[\phi(B,\theta)]| = o_p(1)$. Since the first statement of Lemma B.9 (i) implies $|E_{P_n^{(M)}}[\phi_n(B,\bar{T}_{P_n^{(M)}})]| \xrightarrow{p} 0$, we obtain

$$|E_{P_0^{(M)}}[\phi(B,\bar{T}_{P_n^{(M)}})]| \le |E_{P_0^{(M)}}[\phi(B,\bar{T}_{P_n^{(M)}})] - E_{P_n^{(M)}}[\phi_n(B,\bar{T}_{P_n^{(M)}})]| + |E_{P_n^{(M)}}[\phi_n(B,\bar{T}_{P_n^{(M)}})]| \xrightarrow{p} 0.$$

The conclusion follows from Assumption 3.1 (iii).

Lemma B.8. *Suppose that Assumption 3.1 holds. Then for each $r > 0$ and each sequence $Q_n \in \mathcal{B}_n(r)$, the followings hold under Q_n:*

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

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

$$|E_{P_n^{(M)}}[\phi_n(B,\theta_0)]| \leq \left|E_{P_n^{(M)}}[\phi_n(B,\theta_0)] - \frac{1}{n_B}\sum_{j=1}^{n_B} E_{Q_n^{(M,(j-1)L+1)}}[\phi_n(B,\theta_0)]\right| + \max_{1\leq j\leq n_B} |E_{Q_n^{(M,(j-1)L+1)}}[\phi_n(B,\theta_0)]|. \tag{B.23}$$

The first term is $O_p(\sqrt{M/n})$ by the CLT. The second term is $O(\sqrt{M/n})$ by a similar argument in (B.11). Therefore, the conclusion follows.

Proof of the second statement. By the triangle inequality,

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

where the first term is $o(1)$ by a similar argument in (B.12), and the second term is $o_p(1)$ by UWLLN.

Proof of (ii). The proof is exactly as for Lemma B.5 (ii) except using Lemma B.8 (i) instead of Lemma B.5 (i).

Lemma B.9. *For each $r > 0$ and each sequence $Q_n \in \mathcal{B}_n(r)$, the followings hold under Q_n:*

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

Proof of (i). By UWLLN,

$$\sup_{\theta\in\Theta}\left|E_{P_n^{(M)}}[\phi_n(B,\theta)\phi_n(B,\theta)'] - \frac{1}{n_B}\sum_{j=1}^{n_B} E_{Q_n^{(M,(j-1)L+1)}}[\phi_n(B,\theta)\phi_n(B,\theta)']\right| \xrightarrow{p} 0, \tag{B.24}$$

By applying the same argument in (B.19), we have $\sup_{\theta\in\Theta} |E_{P_n^{(M)}}[\phi_n(B,\bar{T}_{P_n^{(M)}})\phi_n(B,\bar{T}_{P_n^{(M)}})']| = O(1)$. From here the proof of the first statement is the same as for the first statement of Lemma B.6 (i) except using Lemma B.8 instead of Lemma B.5.

The second statement follows from (B.24) and Lemma B.6 (i). The third statement of the lemma follows from continuity $\partial\phi_n(b,\theta)/\partial\theta'$ at θ_0, Lemmas B.6 (i) and B.7.

Proof of (ii). The proof is similar to the proof of Lemma B.5 (ii) except using Lemma B.9 (i) instead of Lemma B.5 (i).

Lemma B.10. *Suppose that Assumption 3.1 holds. Then for each $r > 0$ and each sequence $Q_n \in \mathcal{B}_n(r)$,*

$$\sqrt{n}(\bar{T}_{P_n^{(M)}} - \theta_0) = -\sqrt{n}(M\Sigma)^{-1} \int \Lambda_n dP_n^{(M)} + o_p(1) \quad \textit{under } Q_n, \tag{B.25}$$

$$\sqrt{n}\left(\bar{T}_{P_n^{(M)}} - \frac{1}{n_B}\sum_{j=1}^{n_B} \bar{T}_{Q_n^{(M,(j-1)L+1)}}\right) \xrightarrow{d} N(0, \Sigma^{-1}) \quad \textit{under } Q_n, \tag{B.26}$$

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

Proof. The proof of (B.25) is similar to that of Lemma B.2. Replace $Q_n^{(M)}$ with $P_n^{(M)}$ and use Lemmas B.8 and B.9 instead of Lemmas B.5 and B.6.

Now we prove (B.26). Lemma B.2 shows that for any $Q_n \in \mathcal{B}_n(r)$ and for any block j,

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

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 (B.25) one obtains

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

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

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

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 B.5 (i), and Davydov (1968, Corollary), and the third equality follows from Definition 3.1 (ii) and (iii).

References

[1] Andrews, D. W. K. (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.

[2] Andrews, D. W. K. (1987) Consistency in nonlinear econometric models: a generic uniform law of large numbers, *Econometrica*, 55, 1465-1471.

[3] Andrews, D. W. K. (1988) Robust estimation of location in a Gaussian parametric model, *Advances in Econometrics*, 7, 3-44.

[4] Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-858.

[5] Ashley, R. and D. Vaughan (1986) Measuring measurement error in economic time series, *Journal of Business and Economic Statistics*, 4, 95-103.

[6] Beran, R. (1977) Minimum Hellinger distance estimates for parametric models, *Annals of Statistics*, 5, 445-463.

[7] Beran, R. (1978) An efficient and robust adaptive estimator of location, *Annals of Statistics*, 6, 292-313.

[8] Beran, R. (1980) Asymptotic lower bounds for risk in robust estimation, *Annals of Statistics*, 8, 1252-1264.

[9] Beran, R. (1981) Efficient robust tests in parametric models, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 73-86.

[10] Beran, R. (1982) Robust estimation in models for independent non-identically distributed data, *Annals of Statistics*, 10, 415-428.

[11] Beran, R. (1984) Minimum distance procedures, in *Handbook of Statistics*, ed. by Krishnaiah, P. and P. Sen, pp. 741-754. Elsevier Science.

[12] Bickel, P. J. (1981) Quelques aspects de la statistique robuste, in *Ecole d'Et de Probabilits de Saint Flour IX 1979*, ed. by P. Hennequin, pp. 1-72. Springer.

[13] Borwein, J. M. and A. S. Lewis (1993) Partially-finite programming in L1 and the existence of maximum entropy estimates, *SIAM Journal of Optimization*, 3, 248-267.

[14] Davydov, Y. (1968) Convergence of distributions generated by stationary stochastic processes, *Theory of Probability and its Applications*, 8, 675-683.

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

[16] Hall, P. and J. L. Horowitz (1996) Bootstrap critical values for tests based on generalized-method-of-moments estimators, *Econometrica*, 64, 891-916.

[17] Hansen, L. P. (1982) Large sample properties of generalized methods of moments estimators, *Econometrica*, 50, 1029-1054.

[18] Hansen, L. P., Heaton, J. and A. Yaron (1996) Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics*, 14, 262-280.

[19] Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables, *Annals of Probability*, 12, 141-153.

[20] Imbens, G. W., Spady, R. H. and P. Johnson (1998) Information theoretic approaches to inference in moment condition models, *Econometrica*, 66, 333-357.

[21] Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes, *Annals of Statistics*, 25, 2084-2102.

[22] Kitamura, Y. (1998) Comparing misspecified dynamic econometric models using nonparametric likelihood, Working Paper, Department of Economics, University of Wisconsin.

[23] Kitamura, Y. (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.

[24] Kitamura, Y. (2007) Empirical likelihood methods in econometrics: theory and practice, in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, vol. 3, ed. by Blundell, R., Newey, W. K. and T. Persson, Cambridge University Press, Cambridge.

[25] Kitamura, Y. and T. Otsu (2005) Minimax estimation and testing for moment condition models via large deviations, Manuscript, Department of Economics, Yale University.

[26] Kitamura, Y., Otsu, T. and K. Evdokimov (2013) Robustness, infinitesimal neighborhoods, and moment restrictions, *Econometrica*, 81, 1185-1201.

[27] Kitamura, Y. and M. Stutzer (1997) An information theoretic alternative to generalized method of moments estimation, *Econometrica*, 65, 861-874.

[28] Martin, R. and V. Yohai (1986) Influence functionals for time series, *Annals of Statistics*, 14, 781-818.

[29] Newey, W. K. and R. J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 72, 219-255.

[30] Newey, W. K. and K. D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.

[31] Newey, W. K. and K. D. West (1994) Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, 61, 631–654.

[32] Owen, A. (2001) *Empirical Likelihood*, Chapman and Hall/CRC.

[33] Pollard, D. (2002) *A User's Guide to Measure Theoretic Probability*, Cambridge University Press.

[34] Pötscher, B. M. and I. R. Prucha (1989) A uniform law of large numbers for dependent and heterogeneous data processes, *Econometrica*, 57(3), 675-683.

[35] Qin, J. and J. Lawless (1994) Empirical likelihood and general estimating equations, *Annals of Statistics*, 22, 300-325.

[36] Rieder, H. (1978) A robust asymptotic testing model, *Annals of Statistics*, 6, 1080-1094.

[37] Rieder, H. (1994) *Robust Asymptotic Statistics*, Springer-Verlag.

[38] Schennach, S. M. (2007) Point estimation with exponentially tilted empirical likelihood, *Annals of Statistics*, 35, 634-672.

[39] White, H. (1982) Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-25.