High Dimensional Generalized Empirical Likelihood for Moment Restrictions with Dependent Data

Chang, Jinyuan and Chen, Song Xi and Chen, Xiaohong

University of Melbourne, Peking University, Yale University

2014

High Dimensional Generalized Empirical Likelihood for Moment Restrictions with Dependent Data

Jinyuan Chang* Song Xi Chen† Xiaohong Chen‡

The University of Melbourne Peking University Yale University

First version: May 2013 This version: October 2014

Abstract

This paper considers the maximum generalized empirical likelihood (GEL) estimation and inference on parameters identified by high dimensional moment restrictions with weakly dependent data when the dimensions of the moment restrictions and the parameters diverge along with the sample size. The consistency with rates and the asymptotic normality of the GEL estimator are obtained by properly restricting the growth rates of the dimensions of the parameters and the moment restrictions, as well as the degree of data dependence. It is shown that even in the high dimensional time series setting, the GEL ratio can still behave like a chi-square random variable asymptotically. A consistent test for the over-identification is proposed. A penalized GEL method is also provided for estimation under sparsity setting.

JEL classification: C14; C30; C40
Key words: Generalized empirical likelihood; High dimensionality; Penalized likelihood; Variable selection; Over-identification test; Weak dependence.

1 Introduction

In economic, financial and statistical applications, econometric models defined with a growing number of parameters and moment restrictions are increasingly employed. Vector autoregressive models, dynamic asset pricing models, dynamic panel data models and high dimensional dynamic factor models are specific examples; see, e.g., Bai and Ng (2002), Stock and Watson (2010) and Fan and Liao (2014). Due to the desire to better capture large scale dynamic fundamental relations, these models with large number of unknown parameters of interest are typically used to for time series data of high dimension due to a large number of variables (relative to the sample size).

The unconditional moment restriction models are the inferential settings of the Generalized Method of Moment (GMM) of Hansen (1982), which is perhaps the most popular econometric method for semi-parametric statistical inference. There are two dimensions that play essential roles in this method: the dimension of the moment restrictions and the dimension of the unknown parameters of interest. When both dimensions are fixed and finite, there is a huge established literature on inferential procedures, which include but not restrict to Rothenberg (1973) for the minimum distance, Hansen (1982) and Hansen and

*Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, 3010, Australia
†Corresponding author. Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. Tel.: +86 10 62760736; fax: +86 10 62760736. E-mail address: csx@gsm.pku.edu.cn
‡Department of Economics, Yale University, New Haven, CT, 06520, U.S.A.

Singleton (1982) for the GMM, Owen (1988), Qin and Lawless (1994) and Kitamura (1997) for the empirical likelihood (EL), Smith (1997), Newey and Smith (2004) and Anatolyev (2005) for the generalized empirical likelihood (GEL). Among these methods, some members of the GEL (especially the EL) have the attractive properties of the Wilks' theorem (Owen, 1988, 1990; Qin and Lawless, 1994), Bartlett correction (Chen and Cui, 2006, 2007), and a smaller second order bias (Newey and Smith, 2004; Anatolyev, 2005). See Owen (2001), Kitamura (2007) and Chen and Van Keilegom (2009) for reviews.

This paper investigates high dimensional GEL estimation and testing for weakly dependent observations when the dimensions of both the moment restrictions and the unknown parameters of interest may grow with the sample size n. Let p and r denote the dimension of the unknown parameters and the number of moment restrictions, respectively. When $r \geq p$, we investigate the impacts of p and r on the consistency, the rate of convergence and the asymptotic normality of the GEL estimator, the limiting behavior of the GEL ratio statistics as well as the overidentification test. To accommodate the potential serial dependence in the estimating functions induced by the original time series data, the blocking technique is employed. This paper establishes the consistency (with rate) and the asymptotic normality of the GEL estimator under either (fixed) finite or diverging block size M under some suitable restrictions on r, p, M and n. It is demonstrated that in general the blocking technique with a diverging block size delivers the estimation efficiency. We also discuss the impact of the smallest eigenvalue of the covariance matrix of the averaged estimating function on the consistency and the asymptotic normality of the GEL estimator. We show that, even in high dimensional nonlinear time series setting (with diverging M), the GEL ratio still behaves like a chi-square random variable asymptotically, which echoes a similar result by Fan, Zhang and Zhang (2001) for nonparametric regression with iid data. A GEL based overidentification specification test is also presented for high dimensional time series models, which extends that of Donald, Imbens and Newey (2003) for iid data from increasing dimension of moments (r) but fixed finite dimension of parameters (p) to both dimensions are allowed to diverge (as long as $r - p > 0$). Finally, when the parameter space is sparse, a penalized GEL method is proposed to allow for $p > r$, and is shown to attain the oracle property in the selection consistency as well as the asymptotic normality of the estimated non-zero parameters.

There are some studies on the EL and its related methods under high dimensionality of both the moment restrictions and the parameters of interest. Chen, Peng and Qin (2009) and Hjort, McKeague and Van Keilegom (2009) evaluated the EL ratio statistic for the mean under high dimensional setting. Tang and Leng (2010) and Leng and Tang (2012) evaluated a penalized EL when the underlying parameter is sparse in the context of the mean parameters and the general estimating equations, respectively. Fan and Liao (2014) considered penalized GMM estimation under high dimensionality and sparsity assumption. These papers assume independent data. Recently, by allowing for dependent data but losing the self-standardization property of the EL, Lahiri and Mukhopadhyay (2012) proposed a modified EL method by adding a penalty term to the original EL criterion for estimating the high-dimensional mean parameters with $r = p > n$. Lahiri and Mukhopadhyay (2012) did not implement data blocking in their modified (penalized) EL for the means despite the moment equations are serially dependent. The EL ratio statistic based on their modified EL method is no longer asymptotically pivotal. As a result, any inference based on this modified EL has to use data blocking or other HAC long-run variance estimation. The rationale in our paper is to preserve the attractive self-standardization property of the GEL in high dimensional time series setting; doing so makes our allowed dimensionality smaller than that in Lahiri and Mukhopadhyay (2012) but maintains simple GEL inference.

The rest of the paper is organized as follows. Section 2 introduces the high dimensional model framework and the basic regularity conditions. Sections 3 and 4 establish the consistency, the rate of convergence and the asymptotic normality of the GEL estimator. Sections 5 and 6 derive the asymptotic properties of the GEL ratio statistic and the overidentification specification test respectively. Section 7 presents a penalized GEL approach for parameter estimation and variable selection when the unknown parameter is sparse. Section 8 reports some simulation results and Section 9 briefly concludes. Technical lemmas and all the proofs are given in Appendix.

2 Preliminaries

2.1 Empirical Likelihood and Generalization

Let $\{X_t\}_{t=1}^n$ be a sample of size n from an $\mathbb{R}^d-$valued strictly stationary stochastic process, where d denotes the dimension of X_t, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ be a p-dimensional parameter taking values in a parameter space Θ. Consider a sequence of r-dimensional estimating equation

$$g(X_t, \boldsymbol{\theta}) = (g_1(X_t, \boldsymbol{\theta}), \ldots, g_r(X_t, \boldsymbol{\theta}))'$$

for $r \geq p$. The model information regarding the data and the parameter is summarized by moment restrictions

$$E\{g(X_t, \boldsymbol{\theta}_0)\} = \mathbf{0} \tag{1}$$

where $\boldsymbol{\theta}_0 \in \Theta$ is the true parameter. As argued in Hjort, McKeague and Van Keilegom (2009), the moment restrictions (1) can be viewed as a triangular array where $r, d, X_t, \boldsymbol{\theta}$ and $g(x, \boldsymbol{\theta})$ may all depend on the sample size n. We will explicitly allow r and/or p grow with n while considering inference for $\boldsymbol{\theta}_0$ identified by (1). Although there is often a connection between d and r which is dictated by the context of an econometrical or statistical analysis, the theoretical results established in this paper are written directly on the growth rates of r and p relative to n. Hence, we will not impose explicit conditions on d which can be either growing or fixed. Certainly, when d diverges, it would indirectly affect the underlying assumptions made in Section 2.3, for instance the moment condition and the rate of the mixing coefficients.

We assume the dependence in the time series $\{X_t\}$ satisfies the α-mixing condition (Doukhan, 1994). Specifically, let $\mathscr{F}_u^v = \sigma(X_t : u \leq t \leq v)$ be the σ-field generated by the data from a time u to a time v for $v \geq u$. Then, the α-mixing coefficients are defined as

$$\alpha_X(k) = \sup_d \sup_{A \in \mathscr{F}_{-\infty}^0, B \in \mathscr{F}_k^\infty} |P(A \cap B) - P(A)P(B)| \quad \text{for each } k \geq 1.$$

The α-mixing condition means that $\alpha_X(k) \to 0$ as $k \to \infty$. When $\{X_t\}$ are independent, $\alpha_X(k) = 0$ for all $k \geq 1$.

We employ the blocking technique (Hall, 1985; Carlstein, 1986; Künsch, 1989) to preserve the dependence among the underlying data. Let M and L be two integers denoting the block length and separation between adjacent blocks, respectively. Then, the total number of blocks is $Q = \lfloor (n - M)/L \rfloor + 1$, where $\lfloor \cdot \rfloor$ is the integer truncation operator. For each $q = 1, \ldots, Q$, the q-th data block $B_q = (X_{(q-1)L+1}, \ldots, X_{(q-1)L+M})$. The average of the estimating equation over the q-th block is

$$\phi_M(B_q, \boldsymbol{\theta}) = \frac{1}{M} \sum_{m=1}^{M} g(X_{(q-1)L+m}, \boldsymbol{\theta}). \tag{2}$$

Clearly, $E\{\phi_M(B_q, \boldsymbol{\theta}_0)\} = \mathbf{0}$. For any n and $\boldsymbol{\theta} \in \Theta$, $\{\phi_M(B_q, \boldsymbol{\theta})\}_{q=1}^Q$ is a new stationary sequence. The blockwise EL (Kitamura, 1997) is defined as

$$\mathcal{L}(\boldsymbol{\theta}) = \sup\left\{ \prod_{q=1}^{Q} \pi_q \,\middle|\, \pi_q > 0, \sum_{q=1}^{Q} \pi_q = 1, \sum_{q=1}^{Q} \pi_q \phi_M(B_q, \boldsymbol{\theta}) = \mathbf{0} \right\}. \tag{3}$$

Employing the routine optimization procedure for the blockwise EL leads to

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{q=1}^{Q} \left\{ \frac{1}{Q} \frac{1}{1 + \widehat{\lambda}(\boldsymbol{\theta})' \phi_M(B_q, \boldsymbol{\theta})} \right\}, \tag{4}$$

where $\widehat{\lambda}(\boldsymbol{\theta})$ is a stationary point of the function $q(\lambda) = -\sum_{q=1}^{Q} \log\{1 + \lambda' \phi_M(B_q, \boldsymbol{\theta})\}$.

The EL estimator for $\boldsymbol{\theta}_0$ is $\widehat{\boldsymbol{\theta}}_{EL} = \arg\max_{\boldsymbol{\theta} \in \Theta} \log\ \mathcal{L}(\boldsymbol{\theta})$. The maximization in (3) can be carried out more efficiently by solving the corresponding dual problem, which implies that $\widehat{\boldsymbol{\theta}}_{EL}$ can be obtained as

$$\widehat{\boldsymbol{\theta}}_{EL} = \arg\min_{\boldsymbol{\theta} \in \Theta} \max_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta})} \sum_{q=1}^{Q} \log\{1 + \lambda' \phi_M(B_q, \boldsymbol{\theta})\}, \tag{5}$$

where $\widehat{\Lambda}_n(\boldsymbol{\theta}) = \{\lambda \in \mathbb{R}^r : \lambda' \phi_M(B_q, \boldsymbol{\theta}) \in \mathcal{V}, q = 1, \ldots, Q\}$ for any $\boldsymbol{\theta} \in \Theta$ and $\mathcal{V}$ is an open interval containing zero.

The link function $\log(1 + v)$ in (5) can be replaced by a general concave function $\rho(v)$ (Smith, 1997). The domain of $\rho(\cdot)$ contains 0 as an interior point, and $\rho(\cdot)$ satisfies $\rho_v(0) \neq 0$ and $\rho_{vv}(0) < 0$ where $\rho_v(v) = \partial\rho(v)/\partial v$ and $\rho_{vv}(v) = \partial^2\rho(v)/\partial v^2$. The GEL estimator (Smith, 1997; Newey and Smith, 2004) is

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \Theta} \max_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta})} \sum_{q=1}^{Q} \rho(\lambda' \phi_M(B_q, \boldsymbol{\theta})), \tag{6}$$

which includes the EL estimator $\widehat{\boldsymbol{\theta}}_{EL}$ of Owen (1988), the exponential tilting (ET) estimator of Kitamura and Stutzer (1997) and Imbens, Spady and Jonson (1998) (with $\rho(v) = -\exp(v)$), the continuous updating (CU) GMM estimator of Hansen, Heaton and Yaron (1996) (with a quadratic $\rho(v)$), and many others as special cases. Define

$$\widehat{S}_n(\boldsymbol{\theta}, \lambda) = \frac{1}{Q} \sum_{q=1}^{Q} \rho(\lambda' \phi_q(\boldsymbol{\theta})). \tag{7}$$

Then $\widehat{\boldsymbol{\theta}}_n$ and its Lagrange multiplier $\widehat{\lambda}$ satisfy the score equation

$$\nabla_\lambda \widehat{S}_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\lambda}) = \mathbf{0}.$$

By the implicit function theorem [Theorem 9.28 of Rudin (1976)], for all $\boldsymbol{\theta}$ in a $\|\cdot\|_2$-neighborhood of $\widehat{\boldsymbol{\theta}}_n$, there is a $\widehat{\lambda}(\boldsymbol{\theta})$ such that $\nabla_\lambda \widehat{S}_n(\boldsymbol{\theta}, \widehat{\lambda}(\boldsymbol{\theta})) = \mathbf{0}$ and $\widehat{\lambda}(\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$. By the concavity of $\widehat{S}_n(\boldsymbol{\theta}, \lambda)$ with respect to λ, $\widehat{S}_n(\boldsymbol{\theta}, \widehat{\lambda}(\boldsymbol{\theta})) = \max_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta})} \widehat{S}_n(\boldsymbol{\theta}, \lambda)$. From the envelope theorem,

$$\mathbf{0} = \nabla_{\boldsymbol{\theta}} \widehat{S}_n(\boldsymbol{\theta}, \widehat{\lambda}(\boldsymbol{\theta}))\big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n} = \frac{1}{Q} \sum_{q=1}^{Q} \rho_v(\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)' \phi_q(\widehat{\boldsymbol{\theta}}_n)) \{\nabla_{\boldsymbol{\theta}} \phi_q(\widehat{\boldsymbol{\theta}}_n)\}' \widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n). \tag{8}$$

The role of the block size M played in the consistency and the asymptotic normality of the GEL estimator $\widehat{\boldsymbol{\theta}}_n$ will be discussed in Sections 3 and 4, respectively.

2.2 Examples

We illustrate the model setting of high dimensional moment restrictions framework through three examples.

Example 1 (High dimensional means): Suppose $\{X_t\}_{t=1}^n$ is a stationary sequence of observations, where $X_t \in \mathbb{R}^d$ and $\boldsymbol{\theta}_0 = E(X_t)$. For high dimensional data, d diverges and $g(X_t, \boldsymbol{\theta}) = X_t - \boldsymbol{\theta}$ constitutes the simplest high dimensional moment equation, which implies the dimension of observation d, the number of moment restrictions r and the number of parameters p all are the same. Under this setting and for independent data, Chen, Peng and Qin (2009) and Hjort, McKeague and Van Keilegom (2009) considered the asymptotic normality of the EL ratio, that mirrors the Wilks' theorem for finite dimensional case.

This framework can be used in other inference problems. For instance checking if two univariate stationary time series $\{Y_t\}$ and $\{Z_t\}$ have identical marginal distribution. Let

$$f_Y(s) = E(e^{\mathbf{i}sY_t}) \text{ and } f_Z(s) = E(e^{\mathbf{i}sZ_t})$$

denote the characteristic functions of the two series, respectively. Suppose all the moments of Y_t and Z_t exist, then the characteristic functions can be expressed as

$$f_Y(s) = 1 + \sum_{k=1}^{\infty} \frac{(\mathbf{i}s)^k}{k!} E(Y_t^k) \text{ and } f_Z(s) = 1 + \sum_{k=1}^{\infty} \frac{(\mathbf{i}s)^k}{k!} E(Z_t^k).$$

Let $X_t = (Y_t, Z_t)$ and $g(X_t, \boldsymbol{\theta}) = (a_1(Y_t - Z_t - \theta_1), \ldots, a_r(Y_t^r - Z_t^r - \theta_r))'$ for some nonzero constants $a_1, \ldots, a_r$. Here θ_l measures $E(Y_t^l) - E(Z_t^l)$ for $l = 1, \ldots, r$, and the a_i's are used to account for the potential diverging moments case, i.e, either $E(Y_t^l)$ or $E(Z_t^l)$ may diverge as $l \to \infty$. Then, the test for whether Y_t and Z_t having the same marginal distribution can be conducted by testing if $\boldsymbol{\theta}_0 = \mathbf{0}$ via the growing dimensional moment restrictions $E\{g(X_t, \boldsymbol{\theta}_0)\} = \mathbf{0}$ by letting $r \to \infty$.

Example 2 (Time series regression): We assume a structural model for s-dimensional time series Y_t which involve unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ of interest as well as time innovations with unknown distributional form. Specifically, assume

$$h(Y_t, \ldots, Y_{t-m}; \boldsymbol{\theta}_0) = \boldsymbol{\varepsilon}_t \in \mathbb{R}^r \tag{9}$$

where $m \geq 1$ is some constant. In this model, we can view $X_t = (Y_t', \ldots, Y_{t-m}')' \in \mathbb{R}^d$ with $d = sm$ and $g(X_t, \boldsymbol{\theta}) = h(Y_t, \ldots, Y_{t-m}; \boldsymbol{\theta})$. If $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, it implies

$$E\{g(X_t, \boldsymbol{\theta}_0)\} = \mathbf{0}.$$

For conventional vector autoregressive models

$$Y_t = \mathbf{A}_1 Y_{t-1} + \cdots + \mathbf{A}_m Y_{t-m} + \eta_t \tag{10}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_m$ are some coefficient matrices needed to be estimated and η_t is the white noise series. This model is the special case of (9) with

$$h(Y_t, \ldots, Y_{t-m}; \boldsymbol{\theta}_0) = (Y_t - \mathbf{A}_1 Y_{t-1} - \cdots - \mathbf{A}_m Y_{t-m}) \otimes (Y_t', \ldots, Y_{t-m}')'.$$

In modern high dimensional time series analysis, we always assume the dimensionality of Y_t is large in relation to sample size, i.e., $s \to \infty$ as $n \to \infty$. Under such background, the numbers of estimating equation and unknown parameters are both $s^2 m$. If we replace $(Y_t', \ldots, Y_{t-m}')'$ by $(Y_t', \ldots, Y_{t-m-l}')'$ for some fixed $l \geq 1$, the model will be over-identified. The phenomenon of over-parametrization in such model is well known (Lütkepohl, 2006). Davis, Zhang and Zheng (2012) considered the estimation of (10) under the sparsity assumption on $\mathbf{A}_i$'s. Under the sparsity, the penalized method proposed in Section 7 can be applied. Some other models share the form (9) can be found in Section 3.1 of Nordman and Lahiri (2013).

Example 3 (Conditional moment restrictions): Let $\{X_t = (Y_t', Z_t')'\}_{t=1}^n$ be a set of observations, and $\rho(y, z, \boldsymbol{\theta})$ be a known J-dimensional vector of generalized residual function. The parameter $\boldsymbol{\theta}_0$ is uniquely defined via the following conditional moment restrictions

$$E\{\rho(Y_t, Z_t, \boldsymbol{\theta}_0)|Y_t\} = \mathbf{0} \ \text{ almost surely.} \tag{11}$$

By different choices of the functional forms of the generalized residual function $\rho(y, z, \boldsymbol{\theta})$, the conditional moment restrictions (11) include many existing models in statistics and econometrics as special cases. The popular generalized linear models are special cases of (11). To appreciate this point, let $\mu(y) = E(Z|Y = y)$ and $h(\mu(y)) = y'\boldsymbol{\theta}_0$ for an increasing link function $h(\cdot)$. Then the generalized linear models are special cases of (11) with $\rho(y, z, \boldsymbol{\theta}_0) = z - h^{-1}(y'\boldsymbol{\theta}_0)$.

Let $q^K(y) = (q_{1K}(y), \dots, q_{KK}(y))'$ denote a $K \times 1$ vector of known basis functions that can approximate any square integrable functions of Y well as $K \to \infty$, such as polynomial splines, B-splines, power series, Fourier series, wavelets, Hermite polynomials and others; see, e.g., Ai and Chen (2003) and Donald, Imbens and Newey (2003). Then, (11) implies

$$E\{\rho(Y_t, Z_t, \boldsymbol{\theta}_0) \otimes q^K(Y_t)\} = \mathbf{0}. \tag{12}$$

Moreover, the unknown parameter $\boldsymbol{\theta}_0$ is a solution to this set of increasing dimensional ($r = JK$) unconditional moment restrictions (12). The dimension K will increase with n to guarantee the consistency of the estimator for $\boldsymbol{\theta}_0$ and its asymptotic efficiency. Define $g(X_t, \boldsymbol{\theta}) = \rho(Y_t, Z_t, \boldsymbol{\theta}) \otimes q^K(Y_t)$, then (12) is a special case of (1). The number of moment restrictions $r = JK$ increases as K does. For this model with iid data, Donald, Imbens and Newey (2003) apply the GEL method to the increasing number of the unconditional moment restrictions (12) to obtain efficient estimation for finite fixed dimensional $\boldsymbol{\theta}_0$. They find that the diverging rate of the moment restrictions $r = JK$ depends on the choice of the basis functions $q^K(y)$. For example, if $q^K(y)$ is a spline basis then $r = JK$ could grow at the rate of $K = o(n^{1/3})$.

2.3 Notations and Technical Conditions

Throughout the paper, we use Cs, with different subscripts, to denote positive finite constants which does not depend on the sample size n. For a matrix A, we use $\|A\|_F$ and $\|A\|_2$ to denote its Frobenius-norm and operator-norm respectively, i.e., $\|A\|_F = \{\mathrm{tr}\,(A'A)\}^{1/2}$ and $\|A\|_2 = \{\lambda_{\max}\,(A'A)\}^{1/2}$. If a is a vector, $\|a\|_2$ denotes its L_2-norm. Without causing much confusion, we denote the i-th component of $g(x, \boldsymbol{\theta})$ by $g_i(x, \boldsymbol{\theta})$; and simplify $g(X_t, \boldsymbol{\theta})$ and $\phi_M(B_q, \boldsymbol{\theta})$ by $g_t(\boldsymbol{\theta})$ and $\phi_q(\boldsymbol{\theta})$, respectively, where $\phi_M(B_q, \boldsymbol{\theta})$ is defined in (2). Furthermore, we use $g_{t,j}(\boldsymbol{\theta})$ and $\phi_{q,j}(\boldsymbol{\theta})$ to denote the j-th component of $g_t(\boldsymbol{\theta})$ and $\phi_q(\boldsymbol{\theta})$ respectively. Let $\bar{g}(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n g_t(\boldsymbol{\theta})$ and $\bar{\phi}(\boldsymbol{\theta}) = Q^{-1}\sum_{q=1}^Q \phi_q(\boldsymbol{\theta})$. Additionally, define

$$V_M = \mathrm{Var}\{M^{1/2}\phi_q(\boldsymbol{\theta}_0)\} \ \text{ and } \ V_n = \mathrm{Var}\{n^{1/2}\bar{g}(\boldsymbol{\theta}_0)\}$$

which are the covariance of the averaged estimating functions over a block and the entire sample respectively. Clearly $V_M = V_n$ if $M = n$. The following regularity conditions are needed in our analysis.

(A.1) (i) $\{X_t\}$ is strictly stationary and there exists $\gamma > 2$ such that $\sum_{k=1}^\infty k\alpha_X(k)^{1-2/\gamma} < \infty$; (ii) $M \geq L$ and $M/L \to c \geq 1$; (iii) $E\{g_t(\boldsymbol{\theta}_0)\} = \mathbf{0}$ and there are positive functions $\Delta_1(r, p)$ and $\Delta_2(\varepsilon)$ such that for any $\varepsilon > 0$,

$$\inf_{\{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\geq\varepsilon\}} \|E\{g_t(\boldsymbol{\theta})\}\|_2 \geq \Delta_1(r, p)\Delta_2(\varepsilon) > 0,$$

where $\liminf_{r,p\to\infty} \Delta_1(r, p) > 0$; (iv) $\sup_{\boldsymbol{\theta}\in\Theta} \|\bar{g}(\boldsymbol{\theta}) - E\{g_t(\boldsymbol{\theta})\}\|_2 = o_p\{\Delta_1(r, p)\}$.

(A.2) (i) $\boldsymbol{\theta}_0 \in \mathrm{int}(\Theta)$ and Θ contains a small $\|\cdot\|_2$-neighborhood of $\boldsymbol{\theta}_0$ in which $g(x, \boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}$ for any $x \in \mathcal{X}$, the domain of X_t, and

$$\left|\frac{\partial g_i(x,\boldsymbol{\theta})}{\partial \theta_j}\right| \le T_{n,ij}(x) \quad (i = 1,\ldots,r; j = 1,\ldots,p)$$

for some functions $T_{n,ij}(x)$ with $E\{T_{n,ij}^2(X_t)\} \le C$ for any i, j; (ii) $\sup_{\boldsymbol{\theta}\in\Theta} \|g(x,\boldsymbol{\theta})\|_2 \le r^{1/2} B_n(x)$, where $E\{B_n^{\gamma}(X_t)\} \le C$ for γ given in (A.1)(i); (iii) $E\{|g_{t,j}(\boldsymbol{\theta}_0)|^{2\gamma}\} \le C$ for all $j = 1,\ldots,r$; (iv) the eigenvalues of $[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta})\}]'[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta})\}]$ in a $\|\cdot\|_2$-neighborhood of $\boldsymbol{\theta}_0$ are uniformly bounded away from zero and infinity; $\sup_{\boldsymbol{\theta}\in\Theta} \lambda_{\max}\{n^{-1}\sum_{t=1}^n g_t(\boldsymbol{\theta})g_t(\boldsymbol{\theta})'\} \le C$ with probability approaching to 1.

(A.3) In a $\|\cdot\|_2$-neighborhood of $\boldsymbol{\theta}_0$, $g(x,\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ for any $x \in \mathcal{X}$, and for some functions $K_{n,ijk}(x)$ with $E\{K_{n,ijk}^2(X_t)\} \le C$ for any i, j, k,

$$\left|\frac{\partial^2 g_i(x,\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k}\right| \le K_{n,ijk}(x) \quad (i = 1,\ldots,r; j,k = 1,\ldots,p).$$

Condition (A.1)(i) specifies the rate of decay for the mixing coefficients via a tuning parameter γ as commonly assumed in the analysis of weakly dependent data. When the data are independent, $\alpha_X(k) = 0$ for all $k \ge 1$ and this condition is automatically satisfied for any $\gamma > 2$. Kitamura (1997) assumed $\sum_{k=1}^{\infty} \alpha_X(k)^{1-2/\gamma} < \infty$ for fixed finite dimensional EL, which implies $M^{-1}\sum_{k=1}^{M} k\alpha_X(k)^{1-2/\gamma} \to 0$ by Kronecker's lemma. In the current high dimensional setting, we need stronger condition on the mixing coefficients in order to control remainder terms when analyzing the asymptotic properties of the GEL estimator and the GEL ratio. If $\{X_t\}$ is exponentially strong mixing (Fan and Yao, 2003) so that $\alpha_X(k) \sim \varrho^k$ for some $\varrho \in (0,1)$, then (A.1)(i) is automatically valid for any $\gamma > 2$. (A.1)(ii) imposes a condition regarding the two blocking quantities M and L, which is commonly assumed in the works of block bootstrap and blockwise EL. (A.1)(iii) is the population identification condition for the case of diverging parameter space. A similar assumption can be found in Chen (2007) and Chen and Pouzo (2012). The last part of (A.1) is an extension of the uniform convergence. If p is fixed, under the assumption of the compactness of Θ and some other regularity conditions, following Newey (1991), $\sup_{\boldsymbol{\theta}\in\Theta} \|\bar{g}(\boldsymbol{\theta}) - E\{g_t(\boldsymbol{\theta})\}\|_2 = o_p(1)$ which is a special case of (A.1)(iv) with $\Delta_1(r,p)$ being a constant.

As conditions (A.1)(iii) and (iv) are rather abstractive, we illustrate them via the examples given in Section 2.2. For Example 1, we can choose $\Delta_1(r,p) = 1$ and $\Delta_2(\varepsilon) = \varepsilon$. For the conditional moment restrictions model (Example 3), a common assumption in the literature is that for any $a(Y_t)$ with $E\{a^2(Y_t)\} < \infty$ there exists a $K \times 1$ vector γ_K such that $E[\{a(Y_t) - \gamma_K' q^K(Y_t)\}^2] \to 0$ as $K \to \infty$. For any $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \ge \varepsilon\}$, let $\Gamma_K(\boldsymbol{\theta})$ satisfy $E[\|E\{\rho(Y_t, Z_t, \boldsymbol{\theta})|Y_t\} - \Gamma_K(\boldsymbol{\theta})q^K(Y_t)\|_2^2] \to 0$ as $K \to \infty$. If $\sup_y \|E\{\rho(Y_t, Z_t, \boldsymbol{\theta})|Y_t = y\} - \Gamma_K(\boldsymbol{\theta})q^K(y)\|_2 = O(K^{-\lambda})$ for some $\lambda > 1/2$, then

$$\begin{aligned}\|E\{g_t(\boldsymbol{\theta})\}\|_2 &\ge \big\|E[\{\Gamma_K(\boldsymbol{\theta})q^K(Y_t)\} \otimes q^K(Y_t)]\big\|_2 - \big\|E[\{\rho(Y_t, Z_t, \boldsymbol{\theta}) - \Gamma_K(\boldsymbol{\theta})q^K(Y_t)\} \otimes q^K(Y_t)]\big\|_2 \\ &\ge \lambda_{\min}\big(E\{q^K(Y_t)q^K(Y_t)'\}\big)\|\Gamma_K(\boldsymbol{\theta})\|_F - O(K^{-\lambda})\big(\mathrm{tr}[E\{q^K(Y_t)q^K(Y_t)'\}]\big)^{1/2}.\end{aligned}$$

Under the assumption that the eigenvalues of $E\{q^K(Y_t)q^K(Y_t)'\}$ are uniformly bounded away from zero and infinity, we have

$$\begin{aligned}\inf_{\{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\ge\varepsilon\}} \|E\{g_t(\boldsymbol{\theta})\}\|_2 &\ge C\bigg[\inf_{\{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\ge\varepsilon\}} \|\Gamma_K(\boldsymbol{\theta})\|_F - K^{1/2-\lambda}\bigg] \\ &\ge C\bigg[\inf_{\{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\ge\varepsilon\}} E\big[\big\|E\{\rho(Y_t, Z_t, \boldsymbol{\theta})|Y_t\}\big\|_2\big] - K^{1/2-\lambda}\bigg].\end{aligned}$$

Hence, as $\boldsymbol{\theta}_0$ is the unique root of $E\{\rho(Y_t, Z_t, \boldsymbol{\theta})|Y_t\} = \mathbf{0}$, (A.1)(iii) holds provided that the lower bound in the above inequality is greater than or equal to $\Delta_1(r,p)\Delta_2(\varepsilon)$. In addition, if the generalized residual function $\rho(y, z, \boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}$. Then,

$$\big\|E\{\rho(Y_t, Z_t, \boldsymbol{\theta})|Y_t\}\big\|_2 \geq \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \lambda_{\min}^{1/2}\big(E\big[\{\nabla_{\boldsymbol{\theta}}\rho(Y_t, Z_t, \boldsymbol{\theta}^*)\}'\{\nabla_{\boldsymbol{\theta}}\rho(Y_t, Z_t, \boldsymbol{\theta}^*)\}|Y_t\big]\big)$$

where $\boldsymbol{\theta}^*$ is on the line joining $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$. If the eigenvalues of $E[\{\nabla_{\boldsymbol{\theta}}\rho(Y_t, Z_t, \boldsymbol{\theta})\}'\{\nabla_{\boldsymbol{\theta}}\rho(Y_t, Z_t, \boldsymbol{\theta})\}|Y_t]$ are uniformly bounded away from zero, $\Delta_1(r,p)$ and $\Delta_2(\varepsilon)$ can be chosen as some constant C and ε, respectively.

Condition (A.2)(i) assumes that the first derivatives of $g_i(x, \boldsymbol{\theta})$ near $\boldsymbol{\theta}_0$ are uniformly bounded by some functions which have bounded second moments. (A.2)(ii) generalizes the moment conditions on $g(x, \boldsymbol{\theta})$ for fixed dimensional case (Qin and Lawless, 1994; Kitamura, 1997; Newey and Smith, 2004). More generally, we can replace the factor $r^{1/2}$ by some function $\zeta(r) > 0$. We let $\zeta(r) = r^{1/2}$ to simplify the presentation. (A.2)(iii) is the moment assumption on each $g_{t,j}(\boldsymbol{\theta}_0)$. The first part of (A.2)(iv) is an extension of that assumed in the EL or the GEL in the fixed dimensional case (Qin and Lawless, 1994; Kitamura, 1997; Newey and Smith, 2004). The second one of (A.2)(iv) is to bound $\sup_{\boldsymbol{\theta}\in\Theta} \lambda_{\max}\{Q^{-1}\sum_{q=1}^{Q} \phi_q(\boldsymbol{\theta})\phi_q(\boldsymbol{\theta})\}$. Our proofs in Appendix can be easily extended to allow for $\sup_{\boldsymbol{\theta}\in\Theta} \lambda_{\max}\{Q^{-1}\sum_{q=1}^{Q} \phi_q(\boldsymbol{\theta})\phi_q(\boldsymbol{\theta})'\}$ diverging in probability. Note that we do not assume the eigenvalues of V_M or V_n being bounded away from zero and infinity, but rather leave it open for specific treatments in Sections 3 and 4 for the consistency and the asymptotic normality of the GEL estimator. Our subsequent analysis shows that, to obtain the main results of the paper, $\lambda_{\min}(V_M)$ is allowed to decay to zero at certain rates by properly restricting the diverging rates of r and p. Condition (A.3) ensures the second derivatives of $g_i(x, \boldsymbol{\theta})$ near $\boldsymbol{\theta}_0$ are uniformly bounded by functions which have bounded second moments.

3 Consistency and Convergence Rates

To study the consistency of the GEL estimator $\widehat{\boldsymbol{\theta}}_n$ defined by (6), we need the following conditions regarding the dimensionality r, the block size M and the sample size n:

$$r^2 M^{2-2/\gamma} n^{2/\gamma - 1} = o(1) \quad \text{and} \quad r^2 M^3 n^{-1} = o(1). \tag{13}$$

Theorem 1. *Assume conditions (A.1), (A.2) and that the eigenvalues of V_M are uniformly bounded away from zero and infinity. Then, if (13) holds, $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 \xrightarrow{p} 0$. If in addition, $r^2pM^2n^{-1} = o(1)$, then $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(r^{1/2}n^{-1/2})$ and $\|\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)\|_2 = O_p(r^{1/2}Mn^{-1/2})$.*

This theorem provides the consistency of the GEL estimator $\widehat{\boldsymbol{\theta}}_n$ for both independent and dependent data when the blocking size M is either finite or diverging. For independent data, $V_M = E\{g_1(\boldsymbol{\theta}_0)g_1(\boldsymbol{\theta}_0)'\}$ for any $M \geq 1$. Thus, to make r have a faster diverging rate, we select the block size $M = 1$. For dependent data,

$$V_M = E\{g_1(\boldsymbol{\theta}_0)g_1(\boldsymbol{\theta}_0)'\} + \sum_{k=1}^{M-1}\left(1 - \frac{k}{M}\right)\big[E\{g_1(\boldsymbol{\theta}_0)g_{1+k}(\boldsymbol{\theta}_0)'\} + E\{g_{1+k}(\boldsymbol{\theta}_0)g_1(\boldsymbol{\theta}_0)'\}\big].$$

However, if $\{g_t(\boldsymbol{\theta}_0)\}_{t=1}^n$ is a martingale difference sequence, then $V_M \equiv E\{g_1(\boldsymbol{\theta}_0)g_1(\boldsymbol{\theta}_0)'\}$ for any $M \geq 1$ and $M = 1$ should be used to make r have a faster diverging rate. Furthermore, if the eigenvalues of V_M are uniformly bounded away from zero and infinity for some fixed M, (13) is simplified to

$$r^2 n^{2/\gamma - 1} = o(1).$$

Here γ determines the number of moments of the estimating equation as specified in (A.2)(ii) and (A.2)(iii). Then, $r = o(n^{1/2-1/\gamma})$ ensures the consistency of the GEL estimator $\widehat{\boldsymbol{\theta}}_n$. For large enough γ, r will be made close to $o(n^{1/2})$, which is the best rate we can established.

Theorem 1 encompasses the existing consistency results for the GEL estimator in the literature. Indeed, if r is fixed and the data are independent, Theorem 1 implies that both $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2$ and $\|\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)\|_2$ are $O_p(n^{-1/2})$, which are the same as the rates obtained in Qin and Lawless (1994) for the EL and Newey and Smith (2004) for the GEL. If r is fixed but the data are dependent, Theorem 1 means that $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(n^{-1/2})$ and $\|\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)\|_2 = O_p(Mn^{-1/2})$, which coincides with the result of Kitamura (1997) for the EL estimator. If r is diverging and the data are independent, both $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2$ and $\|\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)\|_2$ are $O_p(r^{1/2}n^{-1/2})$, which retain the results in Donald, Imbens and Newey (2003) and Leng and Tang (2012).

The following is an extension of Theorem 1 by allowing V_M to be asymptotically singular, namely $\lambda_{\min}(V_M) \to 0$ as $r \to \infty$, with M being either fixed or diverging.

Corollary 1. *Assume conditions (A.1), (A.2), and that $\lambda_{\min}(V_M) \asymp r^{-\iota_1}$ for some $\iota_1 > 0$ and $\lambda_{\max}(V_M)$ is uniformly bounded away from zero and infinity. Then $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(r^{(1+\iota_1)/2}n^{-1/2})$ and $\|\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)\|_2 = O_p(r^{(1+3\iota_1)/2}Mn^{-1/2})$ provided that $r^{2+3\iota_1}M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$, $r^{2+2\iota_1}M^3n^{-1} = o(1)$, $r^{2+3\iota_1}pMn^{-1} = o(1)$ and $r^{2+\iota_1}pM^2n^{-1} = o(1)$.*

This corollary shows that when the smallest eigenvalue of V_M is not bounded away from zero, the convergence rates for $\widehat{\boldsymbol{\theta}}_n$ and the Lagrange multiplier $\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)$ become slower. Theorem 1 can be viewed as a special case of Corollary 1 with $\iota_1 = 0$.

The convergence rate of $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2$ attained in Theorem 1 is dictated by r, the number of the moment restrictions, rather than by p, the dimension of $\boldsymbol{\theta}$. Under slightly stronger conditions the next proposition improves the convergence rate to $O_p(p^{1/2}n^{-1/2})$.

Proposition 1. *Under conditions (A.1)-(A.3), assume that the eigenvalues of V_M and V_n are uniformly bounded away from zero and infinity. Then $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(p^{1/2}n^{-1/2})$ provided that $r^3M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$, $r^3M^3n^{-1} = o(1)$, $r^3pMn^{-1} = o(1)$ and $r^3p^2n^{-1} = o(1)$.*

4 Asymptotic Normality

We now turn to the asymptotic normality of the GEL estimator $\widehat{\boldsymbol{\theta}}_n$. We are in particularly interested in the effect of the block size M on the estimation efficiency. Based on the consistency of $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)$ given in Theorem 1, expanding $\nabla_\lambda \widehat{S}_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)) = \mathbf{0}$ for $\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)$ around $\lambda = \mathbf{0}$ gives

$$\mathbf{0} = \frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\phi_q(\widehat{\boldsymbol{\theta}}_n) + \frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}(\widetilde{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n), \tag{14}$$

where $\widetilde{\lambda}$ is on the line joining $\mathbf{0}$ and $\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)$. From (8) and (14), it yields

$$\left[\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)'\phi_q(\widehat{\boldsymbol{\theta}}_n))\{\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)\}'\right]\left\{\frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}(\widetilde{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'\right\}^{-1}\bar{\phi}(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}. \tag{15}$$

Based on (15), we can establish the following proposition which is the starting point in our study of the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$.

Proposition 2. *Under conditions (A.1)-(A.3), assume that the eigenvalues of V_M and V_n are uniformly bounded away from zero and infinity. If $r^2pM^2n^{-1} = o(1)$ and (13) holds, then for any vector $\boldsymbol{\alpha}_n \in \mathbb{R}^p$*

with unit L_2-norm,

$$\begin{aligned}
&\sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\
=& - \sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}\bar{g}(\boldsymbol{\theta}_0)\\
&+ O_p(r^{3/2}p^{1/2}M^{1/2}n^{-1/2}) + O_p(r^{3/2}pn^{-1/2}) + O_p(r^{3/2}M^{1-1/\gamma}n^{1/\gamma-1/2}) + O_p(r^{3/2}M^{3/2}n^{-1/2}).
\end{aligned}$$

Proposition 2 covers both the finite and diverging M cases. When M is diverging, $\|V_M - V_n\|_2 \to 0$ provided that $r = o(M)$. We can replace $V_M^{-1}V_nV_M^{-1}$ on the left-hand side of above asymptotic expansion by V_n^{-1} via adding an extra high order term on the right-hand side of the asymptotic expansion. Let

$$\boldsymbol{\beta}_n = -V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}\boldsymbol{\alpha}_n \tag{16}$$

and $U_{n,t} = n^{-1/2}\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)$ for $t = 1, \ldots, n$. From Proposition 2, a major point of interest is under what conditions $\sum_{t=1}^n U_{n,t}$ is asymptotically normal.

Let us first consider the easier case where the observations $\{X_t\}_{t=1}^n$ are independent. From Lindeberg-Feller theorem (Durrett, 2010), to attain the asymptotic normality of $\sum_{t=1}^n U_{n,t}$, it suffices to verify the following two conditions,

$$\text{(i)} \ \sum_{t=1}^n E(U_{n,t}^2) \to 1 \ \text{ and } \ \text{(ii)} \ \sum_{t=1}^n E\{U_{n,t}^2 1_{(|U_{n,t}|>\varepsilon)}\} \to 0 \ \text{ as } \ n \to \infty \ \text{ for any } \ \varepsilon > 0.$$

Actually, $V_n = V_M = E\{g_t(\boldsymbol{\theta}_0)g_t(\boldsymbol{\theta}_0)'\}$ in this case. Hence, $\sum_{t=1}^n E(U_{n,t}^2) = 1$. Note that $\|\boldsymbol{\beta}_n\|_2 \le \lambda_{\min}^{-1/2}(V_n)$ which is uniformly bounded away from infinity if $\lambda_{\min}(V_n)$ is uniformly bounded away from zero. Hence, by (A.2)(ii),

$$(n^{1/2}\varepsilon)^{\gamma-2}E[|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|^2 1_{\{|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|>n^{1/2}\varepsilon\}}] \le E\{|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|^\gamma\} \le Cr^{\gamma/2},$$

which implies that part (ii) holds if $rn^{2/\gamma-1} = o(1)$. Therefore, $\sum_{t=1}^n U_{n,t} \xrightarrow{d} N(0,1)$ provided that $rn^{2/\gamma-1} = o(1)$ for any selection of $\boldsymbol{\alpha}_n \in \mathbb{R}^p$ with unit L_2-norm.

For dependent data, we need to assume $\sup_n E\{|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|^{2+v}\} < \infty$ for some $v > 0$, namely $|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|$ has a higher than two uniformly bounded moment. This is required in the central limit theorem for dependent processes as carried out in Peligrad and Utev (1997) and Francq and Zakoïan (2005). It is used to guarantee the limit of $\mathrm{Var}\{n^{-1/2}\sum_{t=1}^n \boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)\}$ can be well defined as $n \to \infty$. More specifically, notice that

$$\mathrm{Var}\bigg\{\frac{1}{n^{1/2}}\sum_{t=1}^n \boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)\bigg\} = E\{|\boldsymbol{\beta}_n' g_1(\boldsymbol{\theta}_0)|^2\} + 2\sum_{k=1}^{n-1}\bigg(1 - \frac{k}{n}\bigg)E\{\boldsymbol{\beta}_n' g_1(\boldsymbol{\theta}_0)g_{1+k}(\boldsymbol{\theta}_0)'\boldsymbol{\beta}_n\},$$

to define the limit of above sum of series, we need that $\mathrm{Var}\{n^{-1/2}\sum_{t=1}^n \boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)\}$ is absolutely convergent, i.e.,

$$\lim_{n\to\infty}\bigg[E\{|\boldsymbol{\beta}_n' g_1(\boldsymbol{\theta}_0)|^2\} + 2\sum_{k=1}^{n-1}\bigg(1 - \frac{k}{n}\bigg)|E\{\boldsymbol{\beta}_n' g_1(\boldsymbol{\theta}_0)g_{1+k}(\boldsymbol{\theta}_0)'\boldsymbol{\beta}_n\}|\bigg] < \infty.$$

By Davydov inequality (Davydov, 1968; Rio, 1993), the absolute convergence of $\mathrm{Var}\{n^{-1/2}\sum_{t=1}^n \boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)\}$ will hold by requiring $\sup_n E\{|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|^{2+v}\} < \infty$ for some suitable v.

For high dimensional moment equation $g(x, \theta)$ with diverging r, we need

$$\sup_n E\{|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|^\gamma\} < \infty \tag{17}$$

for $\boldsymbol{\beta}_n$ defined via (16) and $\gamma > 2$ specified in (A.1)(i). A sufficient condition for (17) is to restrict

$$\boldsymbol{\beta}_n \in \mathcal{D}(K) := \bigg\{(v_1, v_2, \ldots) \in \mathbb{R}^\infty : \sum_{k=1}^{\infty} |v_k| \le K\bigg\},$$

where K is a given finite constant. To appreciate this, write $\boldsymbol{\beta}_n = (\beta_{n,1}, \ldots, \beta_{n,r})'$ and let $\kappa_r = \sum_{j=1}^r |\beta_{n,j}|$. Then,

$$E\{|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|^\gamma\} = \kappa_r^\gamma E\bigg\{\bigg|\sum_{j=1}^r \frac{|\beta_{n,j}|}{\kappa_r}\mathrm{sign}(\beta_{n,j}) g_{t,j}(\boldsymbol{\theta}_0)\bigg|^\gamma\bigg\} \le K^\gamma C,$$

where the last step is based on the Jensen's inequality and (A.2)(iii). If $\sum_{j=1}^r |\beta_{n,j}| \to \infty$ as $n \to \infty$, we can construct a counter-example such that $\sup_n E\{|\boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0)|^{2+v}\} \to \infty$ for any $v > 0$. The following theorem establishes the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$.

Theorem 2. *Under conditions (A.1)-(A.3), assume that the eigenvalues of V_M and V_n are uniformly bounded away from zero and infinity. For dependent data, if*

$$r^3 M^{2-2/\gamma} n^{2/\gamma-1} = o(1),\ \ r^3M^3n^{-1} = o(1),\ \ r^3pMn^{-1} = o(1) \ \ \textit{and} \ \ r^3p^2n^{-1} = o(1), \tag{18}$$

then for any $\boldsymbol{\alpha}_n \in \mathbb{R}^p$ with unit L_2-norm such that (17) holds,

$$\sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

converges to $N(0,1)$ as $n \to \infty$.

For finite block size M, the above asymptotic distribution holds provided that

$$r^3 n^{2/\gamma-1} = o(1) \quad \text{and} \quad r^3p^2n^{-1} = o(1).$$

Since

$$\begin{aligned}
&([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1} ([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]) \\
&\quad \times ([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1} \ge ([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_n^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1},
\end{aligned}$$

the GEL estimator is asymptotically efficient if $\|V_M - V_n\|_2 \to 0$, which implies $V_M^{-1}V_nV_M^{-1}$ is asymptotically equivalent to V_n^{-1}. This means that if $\{g_t(\boldsymbol{\theta}_0)\}_{t=1}^n$ is a martingale difference sequence, as $V_M = V_n = E\{g_1(\boldsymbol{\theta}_0)g_1(\boldsymbol{\theta}_0)'\}$ for any $M \ge 1$, selecting $M = 1$ will lead to the efficient GEL estimation. In a general case where the nature of the dependence in the estimating function is unknown, letting $M \to \infty$ at some suitable diverging rate, so that (18) is satisfied, will lead to the efficient estimation. Specifically, as

$$\|V_M - V_n\|_2 \le CrM^{-1}\sum_{k=1}^{M} k\alpha_X(k)^{1-2/\gamma},$$

under (A.1)(i) and (A.2)(ii), choosing $M \to \infty$ such that $r = o(M)$ produces the asymptotically efficient GEL estimator $\widehat{\boldsymbol{\theta}}_n$. According to (18) and $r = o(M)$, the divergence rate in M is $M = O(n^{\{(\gamma-2)/(5\gamma-2)\}\wedge 1/6})$ while $r = o(n^{\{(\gamma-2)/(5\gamma-2)\}\wedge 1/6})$, regardless p being fixed or diverging. Under such setting, the best growth rate for r is $r = o(n^{1/6})$ when $\gamma \ge 10$. In comparison with the case of finite M, while letting $M \to \infty$ can guarantee the efficiency, it does slows the divergence of r.

If the smallest eigenvalues of V_M and V_n decay to zero as $r \to \infty$, we assume $\lambda_{\min}(V_M) \asymp r^{-\iota_1}$ and $\lambda_{\min}(V_n) \asymp r^{-\iota_2}$ for some positive ι_1 and ι_2. Based on Corollary 1, by repeating the proof of Proposition 2 in Appendix, it can be shown that the leading term in the asymptotic expansion of Proposition 2 remains while the four remainder terms become

$$O_p(r^{(3+6\iota_1+\iota_2)/2}p^{1/2}M^{1/2}n^{-1/2}) + O_p(r^{(3+4\iota_1+\iota_2)/2}pn^{-1/2}) + O_p(r^{(3+5\iota_1+\iota_2)/2}M^{1-1/\gamma}n^{1/\gamma-1/2}) + O_p(r^{(3+5\iota_1+\iota_2)/2}M^{3/2}n^{-1/2})$$

provided that the conditions governing r, p, M and n assumed in Corollary 1 hold. By the central limit theorem established in Francq and Zakoïan (2005), the leading term in the asymptotic expansion of Proposition 2 converges to $N(0,1)$ regardless $\lambda_{\min}(V_M) \to 0$ and $\lambda_{\min}(V_n) \to 0$ or not. Hence, the asymptotic normality of the GEL estimator $\widehat{\boldsymbol{\theta}}_n$ is valid free of the statue of the eigenvalues of V_M and V_n. The difference is that when $\lambda_{\min}(V_M) \to 0$ and $\lambda_{\min}(V_n) \to 0$, the growth rate of r and/or M are reduced.

To put the growth rate of r into perspectives and to highlight the impacts of data dependence, we consider the independent analogue of Theorem 2 in the following, whose proof is obtained by assigning $\alpha_X(k) = 0$ and $M = 1$ in Proposition 2.

Corollary 2. *Under conditions (A.1)-(A.3), assume that the eigenvalues of $E\{g_1(\boldsymbol{\theta}_0)g_1(\boldsymbol{\theta}_0)'\}$ are uniformly bounded away from zero and infinity. For independent data, if $r^3p^2n^{-1} = o(1)$ and $r^3n^{2/\gamma-1} = o(1)$, then for any $\boldsymbol{\alpha}_n \in \mathbb{R}^p$ with unit L_2-norm,*

$$\sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_n^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty.$$

The above corollary shows that under independence, the growth rate for r is $o(n^{1/3-2/(3\gamma)})$ if p is fixed. If γ is sufficiently large, the rate of r can be close to $o(n^{1/3})$. If p grows with r and $p/r \to y \in (0,1]$, then $r = o(n^{\{1/3-2/(3\gamma)\}\wedge 1/5})$. In particular, if $\gamma \geq 5$, $r = o(n^{1/5})$ which retains Theorem 2 in Leng and Tang (2012) for the EL estimator. Comparing the growth rates for r under the dependent and independent settings, when M is diverging, we see a slowing down in the rate under dependence from $o(n^{1/5})$ to $o(n^{1/6})$ if the best moment conditions hold.

If p, the dimension of $\boldsymbol{\theta}$, is fixed, as in a case of conditional moment restrictions in Example 3, the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$ can be attained with some ease. It can be shown that $\boldsymbol{\beta}_n$ is automatically in $\mathcal{D}(K)$ for a large enough K, which implies the condition (17) holds for any $\boldsymbol{\alpha}_n \in \mathbb{R}^p$ with unit L_2-norm. This is summarized in the following corollary.

Corollary 3. *Under conditions (A.1)-(A.3), assume that the eigenvalues of V_M and V_n are uniformly bounded away from zero and infinity. For dependent data, if p is fixed, then for any $\boldsymbol{\alpha}_n \in \mathbb{R}^p$ with unit L_2-norm,*

$$\sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

converges to $N(0,1)$ as $n \to \infty$, provided that $r^3M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$ and $r^3M^3n^{-1} = o(1)$.

This Corollary with $M = 1$ recovers that in Donald, Imbens and Newey (2003) for iid data.

5 Generalized Empirical Likelihood Ratios

The EL ratio $w_n(\boldsymbol{\theta}) = -2\log\{Q^Q\mathcal{L}(\boldsymbol{\theta})\}$ for $\mathcal{L}(\boldsymbol{\theta})$ defined in (4) plays an important role in the statistical inference. A prominent result for fixed dimensional EL is its resembling the parameter likelihood by have a limiting chi-square distribution under a wide range of situations, as demonstrated in Owen (1988), Chen and Cui (2003), Qin and Lawless (1994) and Chen and Van Keilegom (2009) for independent data, and Kitamura (1997) for dependent data.

For GEL, we define the GEL ratio as

$$w_n(\boldsymbol{\theta}) = \frac{2\rho_{vv}(0)}{\rho_v^2(0)}\bigg\{Q\rho(0) - \max_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta})}\sum_{q=1}^{Q}\rho(\lambda'\phi_q(\boldsymbol{\theta}))\bigg\} \tag{19}$$

which is the extension of the EL ratio in the GEL framework.

We consider the asymptotic distribution of the GEL ratio $w_n(\boldsymbol{\theta}_0)$ when both r and p are diverging. Under such setting, a natural form of the Wilks' theorem is

$$(2r)^{-1/2}\{w_n(\boldsymbol{\theta}_0) - r\} \xrightarrow{d} N(0,1) \quad \text{as } r \to \infty. \tag{20}$$

For the case of means where $g_t(\boldsymbol{\theta}) = X_t - \boldsymbol{\theta}$ with independent observations, Chen, Peng and Qin (2009) and Hjort, McKeague and Van Keilegom (2009) evaluated the impact of the dimensionality on the asymptotic distribution (20) for the EL ratio by providing various diverging rates for r. For parameters defined by general moment restrictions, establishing the limiting distribution of the GEL ratio is far more challenging. We need the following stronger version of (A.1)(i):

(A.1)'(i) There is some $\eta > 8$ such that $\alpha_X(k)^{1-2/\gamma} \asymp k^{-\eta}$ where γ is given in (A.2).

Condition (A.1)'(i) is used to guarantee the leading order term of (19) has the similar probabilistic behavior as the chi-square distribution. It is automatically satisfied with $\eta = \infty$ if X_t is exponentially strong mixing or independent. We also need the following conditions:

$$r^3 M^{2-2/\gamma} n^{2/\gamma-1} = o(1), \quad r^3 M^3 n^{-1} = o(1) \quad \text{and} \quad r^{3/2} M^{-1}\sum_{k=1}^{M} k\alpha_X(k)^{1-2/\gamma} = o(1). \tag{21}$$

Define

$$\xi = \frac{\eta - 8}{4\eta + 4}1_{\{8<\eta<32\}} + \frac{2}{11}1_{\{32\le\eta\le\infty\}} + 1_{\{\text{indenpendent data}\}}. \tag{22}$$

The next theorem establishes the asymptotic distribution of $w_n(\boldsymbol{\theta}_0)$.

Theorem 3. *Under conditions (A.1)'(i), (A.1)(ii) and (A.2)(iii), assume that the eigenvalues of V_n are uniformly bounded away from zero and infinity. If (21) holds and $r = o(n^{\xi})$ where ξ is defined in (22), then*

$$(2r)^{-1/2}\{w_n(\boldsymbol{\theta}_0) - r\} \xrightarrow{d} N(0,1) \quad \textit{as } r \to \infty.$$

This theorem is new for dependent data and includes some established results for independent data as special cases. For independent data, this theorem implies that the asymptotic normality of the GEL ratio is valid if $r = o(n^{1/3-2/(3\gamma)})$, which is the same as that in Hjort, McKeague and Van Keilegom (2009) for the EL ratio with independent data. Our result is more general than theirs since we allow for GEL ratio and for dependent data. For dependent data, the block size is $M = O(n^{(\gamma-2)/(4\gamma-2)})$ if $2 < \gamma < 8$ and $M = O(n^{1/5})$ otherwise, and hence the asymptotic distribution (20) holds if $r = o(n^{\delta})$ with

$$\delta = \min\left(\frac{\eta-8}{4\eta+4}1_{\{8<\eta<32\}} + \frac{2}{11}1_{\{32\le\eta\le\infty\}}, \frac{\gamma-2}{6\gamma-3}1_{\{2<\gamma<8\}} + \frac{2}{15}1_{\{\gamma\ge 8\}}\right).$$

When η and γ are sufficiently large, the best diverging rate is $r = o(n^{2/15})$ for the dependent case, which is slower than the rate of $r = o(n^{1/3-2/(3\gamma)})$ for the independence case.

6 Test for Over-identification

For moment restrictions, it is important to check on the validity of the model by testing the following hypotheses

$$H_0 : E\{g(X_t, \boldsymbol{\theta}_0)\} = \mathbf{0} \text{ for some } \boldsymbol{\theta}_0 \in \Theta \text{ v.s. } H_1 : E\{g(X_t, \boldsymbol{\theta})\} \neq \mathbf{0} \text{ for any } \boldsymbol{\theta} \in \Theta.$$

We consider testing the above hypothesis when $r > p$, namely the moment equation overly identify the parameter $\boldsymbol{\theta}$.

We formulate the test statistic as the GEL ratio $w_n(\widehat{\boldsymbol{\theta}}_n)$. For the EL ratio, it has been demonstrated in the fixed dimensional case by Qin and Lawless (1994) and Kitamura (1997) that

$$w_n(\widehat{\boldsymbol{\theta}}_{EL}) \xrightarrow{d} \chi^2_{r-p} \quad \text{as } n \to \infty$$

under H_0. This mirrors the J-test of Hansen (1982)'s GMM with fixed and finite dimensions r and p.

To formulate the GEL specification test allowing for increasing dimensions r and p, we are to study the asymptotic distribution of $w_n(\widehat{\boldsymbol{\theta}}_n)$ under H_0 first. We only need to consider its leading order $n\bar{g}(\widehat{\boldsymbol{\theta}}_n)'\{M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\}^{-1}\bar{g}(\widehat{\boldsymbol{\theta}}_n)$ as the remainder terms in the asymptotic expansion of $w_n(\widehat{\boldsymbol{\theta}}_n)$ can be shown to be of a smaller order. Since $\widehat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}_0$ under H_0, Lemma 17 in Appendix establishes the relationship between the asymptotic distributions of $n\bar{g}(\widehat{\boldsymbol{\theta}}_n)'\{M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\}^{-1}\bar{g}(\widehat{\boldsymbol{\theta}}_n)$ and $n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)$ under H_0. We need the following conditions:

$$\begin{gathered} r^3pn^{-1} = o(1), \quad pr^{-1/2} = o(1), \quad r^3M^3n^{-1} = o(1), \\ r^3M^{2-2/\gamma}n^{2/\gamma-1} = o(1) \text{ and } r^{3/2}M^{-1}\sum_{k=1}^{M} k\alpha_X(k)^{1-2/\gamma} = o(1). \end{gathered} \tag{23}$$

Compared with the conditions for the asymptotic distribution of $w_n(\boldsymbol{\theta}_0)$ in (21), the first two restrictions in (23) are the extra ones used to control the remainder terms.

Theorem 4. *Under conditions (A.1)'(i), (A.1)(ii), (A.1)(iii), (A.1)(iv), (A.2) and (A.3), assume that the eigenvalues of V_n are bounded away from zero and infinity. If (23) holds and $r = o(n^{\xi})$, where ξ is defined in (22), then*

$$\{2(r-p)\}^{-1/2}\{w_n(\widehat{\boldsymbol{\theta}}_n) - (r-p)\} \xrightarrow{d} N(0,1) \quad \text{as } r \to \infty.$$

The asymptotic normality can be used to derive the over-identification test under high dimensionality and dependence. Specifically, H_0 is rejected if

$$\{2(r-p)\}^{-1/2}\{w_n(\widehat{\boldsymbol{\theta}}_n) - (r-p)\} > z_{1-\alpha}$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of $N(0,1)$.

To show the above GEL test for over-identification is consistent, we assume that under the alternative hypothesis H_1,

$$\inf_{\boldsymbol{\theta}\in\Theta} \|E\{g(X_t, \boldsymbol{\theta})\}\|_2 \geq \varsigma. \tag{24}$$

The following theorem describes the behavior of $\{2(r-p)\}^{-1/2}\{w_n(\widehat{\boldsymbol{\theta}}_n) - (r-p)\}$ under H_1.

Theorem 5. *Under conditions (A.1)(i), (A.1)(ii), (A.1)(iv), (A.2)(ii) and (24), if there is a positive constant ϵ such that $r^2M^{1-2/\gamma}n^{2/\gamma-1}(\log n)^{\epsilon}\varsigma^{-2} = O(1)$, $r^{1/2}Mn^{-1}\varsigma^{-1} = o(1)$ and $\Delta_1(r,p)\varsigma^{-1} = O(1)$, then*

$$\{2(r-p)\}^{-1/2}\{w_n(\widehat{\boldsymbol{\theta}}_n) - (r-p)\} \xrightarrow{p} \infty \quad \text{as } r \to \infty.$$

This theorem shows that the GEL test is consistent. Unlike Theorem 4, this theorem does not require the block size $M \to \infty$ and assumes the weaker condition (A.1)(i) (instead of the more restrictive one (A.1)'(i)). From the proof given in the Appendix which follows the technique developed in Chang, Tang and Wu (2013), the test statistic $\{2(r-p)\}^{-1/2}\{w_n(\widehat{\boldsymbol{\theta}}_n) - (r-p)\}$ diverges to infinity at least at the rate of $O(r^{1/2})$ under H_1.

7 Penalized Generalized Empirical Likelihood

In high dimensional data analysis, when the dimension of parameters is large, i.e., $p \to \infty$, a more reasonable assumption is that only a subset of the parameters are nonzero. Write $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0p})' \in \mathbb{R}^p$ and define $\mathcal{A} = \{j : \theta_{0j} \neq 0\}$ with its cardinality $s = |\mathcal{A}|$. Without loss of generality, let $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)'}, \boldsymbol{\theta}^{(2)'})'$, where $\boldsymbol{\theta}^{(1)} \in \mathbb{R}^s$ and $\boldsymbol{\theta}^{(2)} \in \mathbb{R}^{p-s}$ correspond to the nonzero and zero components respectively, i.e., $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^{(1)'}, \mathbf{0}')'$. Under such sparsity, we can allow the number of parameters is larger than the number of estimating equations, i.e., $p > r$. However, we still need to assume $s \leq r$, which means that the "real" parameters can be uniquely identified by the moment restrictions (1). To carry out the statistical inference on $\boldsymbol{\theta}$ under the sparsity assumption, we add a penalty term in (6) and the penalized GEL estimator is defined as

$$\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})} = \arg\min_{\boldsymbol{\theta}\in\Theta} \max_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta})} \bigg\{ \sum_{q=1}^{Q} \rho(\lambda' \phi_M(B_q, \boldsymbol{\theta})) + Q \sum_{j=1}^{p} p_\tau(|\theta_j|) \bigg\}$$

where $p_\tau(\cdot)$ is some penalty function with a tuning parameter τ. The following conditions are imposed on the penalty function $p_\tau(\cdot)$ and the tuning parameter τ.

(A.4) $\liminf_{\tau \to 0} \liminf_{\theta \to 0+} p'_\tau(\theta)/\tau > 0$.

(A.5) There exists a positive constant C such that $\max_{j \in \mathcal{A}} p_\tau(|\theta_{0j}|) \leq C\tau$.

Conditions (A.4) and (A.5) hold for many penalty functions such as the one in Fan and Li (2001) and the minimax concave penalty of Zhang (2010). Define

$$\begin{aligned}\mathbf{S}(\boldsymbol{\theta}_0) = & \big([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\big)^{-1} \big([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\big) \\ & \times \big([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\big)^{-1}.\end{aligned}$$

We correspondingly decompose $\mathbf{S}(\boldsymbol{\theta}_0)$ as

$$\mathbf{S}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{S}_{11}(\boldsymbol{\theta}_0) & \mathbf{S}_{12}(\boldsymbol{\theta}_0) \\ \mathbf{S}_{21}(\boldsymbol{\theta}_0) & \mathbf{S}_{22}(\boldsymbol{\theta}_0) \end{pmatrix} \tag{25}$$

where $\mathbf{S}_{11}(\boldsymbol{\theta}_0)$ and $\mathbf{S}_{22}(\boldsymbol{\theta}_0)$ are $s \times s$ and $(p-s) \times (p-s)$ matrices, respectively. The following restrictions are needed

$$s\tau r^{-1} n M^{-1} = O(1) \quad \text{and} \quad \tau (r^{-1}n)^{1/2} M^{-1} \to \infty. \tag{26}$$

Write the penalized GEL estimator $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})} = (\widehat{\boldsymbol{\theta}}_n^{(1)'}, \widehat{\boldsymbol{\theta}}_n^{(2)'})'$ and define

$$\mathbf{S}_p(\boldsymbol{\theta}_0) = \mathbf{S}_{11}(\boldsymbol{\theta}_0) - \mathbf{S}_{12}(\boldsymbol{\theta}_0)\mathbf{S}_{22}^{-1}(\boldsymbol{\theta}_0)\mathbf{S}_{21}(\boldsymbol{\theta}_0).$$

The following theorem describes the basic properties of the penalized GEL estimator.

Theorem 6. *Under conditions (A.1)-(A.5), assume that the eigenvalues of V_M are uniformly bounded away from zero and infinity. If $\max_{j\in\mathcal{A}} p'_\tau(|\theta_{0j}|) = o(r^{-1/2}n^{-1/2})$, $\min_{j\in\mathcal{A}} |\theta_{0j}|/\tau \to \infty$ and (26) holds, the following results hold.*

1. *$P\{\widehat{\boldsymbol{\theta}}_n^{(2)} = \mathbf{0}\} \to 1$ as $n \to \infty$, provided that (13) holds and $r^2pM^2n^{-1} = o(1)$;*

2. *In addition, if the eigenvalues of V_n are uniformly bounded away from zero and infinity, then for any $\boldsymbol{\alpha}_n \in \mathbb{R}^s$ with unit L_2-norm, then*

$$\sqrt{n}\boldsymbol{\alpha}_n'\mathbf{S}_p^{-1/2}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}) \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty,$$

   provided that

   (a) for independent data, $r^3p^2n^{-1} = o(1)$ and $r^3n^{2/\gamma-1} = o(1)$;

   (b) for dependent data, (18) holds and $\boldsymbol{\alpha}_n$ satisfies (17) with

$$\begin{aligned}\boldsymbol{\beta}_n = & -V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}_0)\}]\big([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\big)^{-1} \\ & \times [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}^{(1)}} g_t(\boldsymbol{\theta}_0)\}]\{\mathbf{S}_{11}(\boldsymbol{\theta}_0) - \mathbf{S}_{12}(\boldsymbol{\theta}_0)\mathbf{S}_{22}^{-1}(\boldsymbol{\theta}_0)\mathbf{S}_{21}(\boldsymbol{\theta}_0)\}^{1/2}\boldsymbol{\alpha}_n.\end{aligned}$$

Similar to the consistency of GEL estimator, if the eigenvalues of $E\{g_t(\boldsymbol{\theta}_0)g_t(\boldsymbol{\theta}_0)'\}$ are uniformly bounded away from zero and infinity, result (i) still holds without blocking technique if $r^2n^{2/\gamma-1} = o(1)$ and $r^2pn^{-1} = o(1)$ are satisfied. Comparing Theorem 6 with Theorem 2 and Corollary 2, since $\mathbf{S}_p(\boldsymbol{\theta}_0) \le \mathbf{S}_{11}(\boldsymbol{\theta}_0)$, the penalized GEL estimator is more efficient in estimating the nonzero components. Leng and Tang (2012) considered the theoretical results of the penalized EL estimator for independent data by assuming $p/r \to c \in (0,1)$. Our results extend theirs to penalized GEL estimator for weakly dependent data without requiring $p/r \to c \in (0,1)$.

8 Simulation Results

In this section, we present simulation results to compare the finite sample performance of the GEL estimators with the GMM estimator in the high dimensional time series setting. Three versions of the GEL estimators were considered in the simulations: the EL, the ET and the CU estimators. We experimented two forms of the moment restrictions: one was linear, and the other was nonlinear. The penalized GEL estimator was also considered in the non-linear case.

We first conducted simulation for the linear moment restrictions with $g(X_t, \boldsymbol{\theta}) = X_t - \boldsymbol{\theta}$. The observations $\{X_t\}_{t=1}^n$ were generated according to the vector autoregressive (VAR) model of order 1: $X_t = \psi X_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim N(\mathbf{0}, \Sigma_\varepsilon)$, $\Sigma_\varepsilon = (\sigma_{i,j})_{p\times p}$, $\sigma_{i,i} = 1 - \psi^2$, $\sigma_{i,i\pm1} = 0.5(1-\psi^2)$ and $\sigma_{i,j} = 0$ for $|i-j| > 1$. The stationary distribution of X_t is $N(\mathbf{0}, \Sigma_x)$ where $\Sigma_x = (\tilde{\sigma}_{i,j})_{p\times p}$ and $\tilde{\sigma}_{i,i} = 1$, $\tilde{\sigma}_{i,i\pm1} = 0.5$ and $\tilde{\sigma}_{i,j} = 0$ for $|i-j| > 1$. In this model, $p = r$ and the true parameter $\boldsymbol{\theta}_0 = \mathbf{0} \in \mathbb{R}^p$.

The second simulation model was the generalized linear model. The covariates $\{Z_t\}_{t=1}^n$ were generated with the same VAR(1) process as the $\{X_t\}_{t=1}^n$ in the first model setting. The response variables $\{Y_t\}_{t=1}^n$ were generated from the Bernoulli distribution such that $P(Y_t = 1|Z_t) = \exp(1+Z_t'\boldsymbol{\theta}_0)/\{1+\exp(1+Z_t'\boldsymbol{\theta}_0)\}$ with the true parameter $\boldsymbol{\theta}_0 = (0.8, 0.2, 0, \ldots, 0)' \in \mathbb{R}^p$. Then

$$E\left\{Y_t - \frac{\exp(1+Z_t'\boldsymbol{\theta}_0)}{1+\exp(1+Z_t'\boldsymbol{\theta}_0)}\,\middle|\, Z_t\right\} = 0.$$

In this setting, we have nonlinear moment restrictions

$$g(X_t, \boldsymbol{\theta}) = \begin{pmatrix} Z_t \\ W_t \end{pmatrix}\left\{Y_t - \frac{\exp(1+Z_t'\boldsymbol{\theta})}{1+\exp(1+Z_t'\boldsymbol{\theta})}\right\},$$

where $W_t = (Z_{1,t}^2, \ldots, Z_{p,t}^2)'$ for $Z_t = (Z_{1,t}, \ldots, Z_{p,t})'$. This model is over-identified. We considered both non-penalized and penalized estimators under this model setting.

In both simulation models, we chose $n = 500$, 1000 and 2000, respectively. The parameter ψ in the VAR(1) process capturing the serial dependence was set to be $0.1, 0.3$ and 0.5, respectively. The dimension p was pegged to the sample size n such that $p = \lfloor cn^{2/15} \rfloor$, where $c = 10$ and 12 in the first model setting, and $c = 5$ and 6 in the second model setting, respectively. Simulations results were based on 200 repetitions. For each repetition of each model setting, we obtained the parameter estimates $\widehat{\boldsymbol{\theta}}$'s based on the four considered estimation methods: EL, GMM, ET and CU under five regimes regarding the blocking parameters L and M:

Regime (i). $L = M = 1$; Regime (ii). $M = \lfloor n^{1/5} \rfloor$ and $L = \lfloor 0.5M \rfloor$;

Regime (iii). $L = M = \lfloor n^{1/5} \rfloor$; Regime (iv). $M = \lfloor 3n^{1/5} \rfloor$ and $L = \lfloor 0.5M \rfloor$;

Regime (v). $L = M = \lfloor 3n^{1/5} \rfloor$.

Regime (i) means no blocking. Regimes (ii) and (iv) assigned the block size M to be twice of the block separation parameter L; and Regimes (iii) and (v) prescribed $M = L$. For each repetition of the second model setting, we additionally considered the parameter estimates $\widehat{\boldsymbol{\theta}}$'s based on the penalized GEL estimation methods. The penalty function $p_\tau(u)$ used in the simulation satisfied:

$$p_\tau'(u) = \tau \left\{ I(u \le \tau) + \frac{(a\tau - u)_+}{(a-1)\tau} I(u > \tau) \right\}$$

for $u > 0$, where $a = 3.7$, and $s_+ = s$ for $s > 0$ and 0 otherwise. This penalty function is given in Fan and Li (2001). We applied the method given in Leng and Tang (2012) to determine the penalty parameter τ. In each simulation replication, we calculated the L_2 distance between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ as $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = \{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\}^{1/2}$.

Tables 1 and 2 report empirical medians of the squared estimation errors for the EL, ET, CU and GMM estimators in the first simulation model with $c = 10$ and $c = 12$, respectively. And Tables 3 and 4 summarize the empirical median for the second simulation model with the extra penalized GEL estimators. We had also collected the average of the squared estimation errors, which exhibited similar patterns as the empirical median. Hence, we only report the median of squared estimation errors per the suggestion of one referee.

It is noted that the performance of each estimator at each given blocking regime was improved when the sample size was increased, which confirms the convergence of these estimators. For the second nonlinear model, we observed that the performance of three GEL estimators and their penalized analogues were improved under the blocking regimes (ii)-(v) which were bona fide blocking since $L, M > 1$. This was not that surprising since dependence was presence in both simulated models, and applying the blocking can improve the efficiency of the estimation. However, the performance of the GMM estimator were largely similar regardless of the blocking regimes used. The empirical medians of the squared estimation errors of the GMM estimator were much larger than those of the GEL estimators, which confirmed the existing research on GMM versus GEL for finite fixed dimensional settings (Newey and Smith, 2004; Anatolyev, 2005). Among the three GEL estimators, we observed that while they were largely similar under the first simulation model, the EL and the ET estimators performed better than the CU estimator for the logistic regression model. This might be due to the multivariate asymmetry in the moment conditions, which makes the bias term of the CU estimator more pronounced, as shown in Newey and Smith (2004) and Anatolyev (2005). We note that the estimation efficiency among the GEL estimator with respect to the different regimes of the blocking width selection was largely comparable to each other for the simple mean models. However, in the case of the generalized linear model, the regimes (iv) and (v), with the block width $M = \lfloor 3n^{1/5} \rfloor$, led to the best performance. We also observed that under the second model setting where the parameter is sparse, the penalized GEL estimators were much more efficient than their non-penalized counterparts, which confirmed our Theorem 6.

9 Conclusion

In this paper, we have investigated the asymptotic properties of the GEL estimator, the GEL ratio statistic and the over-identification specification test for high dimensional moment restriction models with increasing number of parameters and weakly dependent data. We have also investigated a penalized GEL approach that is designed for the high dimensional sparse parameter situation with $p > r$, although the true but unknown number of non-zero parameters is not larger than r. We establish the oracle property of the penalized GEL estimator. Both theoretical and simulation studies find the penalization leads to efficiency gain for the GEL estimators even for dependent data.

We establish the consistency and the asymptotic normality of the high-dimensional GEL and the penalized GEL estimators allowing for fixed block size M for time series data. However, when the unconditional moment functions $\{g(X_t, \boldsymbol{\theta}_0)\}_{t=1}^n$ are autocorrelated, the simple limiting distributions of the GEL ratio statistic and the over-identification specification test are established when the block size M diverges with the sample size n. How to practically select M is a quite challenging problem. As indicated in Hall, Horowitz and Jing (1995) and Lahiri (2003), although there has been much research in determining the order of magnitude of M, there is in general a lack of research for selecting the tuning parameter, the coefficient of M for general nonlinear time series models. The simulation study reported in Section 8 shows that $M = \lfloor 3n^{1/5} \rfloor$ led to satisfactory performance. Instead of blocking, one could also perform local smoothing of the unconditional moment functions $\{g(X_t, \boldsymbol{\theta}_0)\}_{t=1}^n$ to reduce temproal dependence (Smith, 1997; Anatolyev, 2005; Kitamura, 2007), which introduces an alternative tuning parameter, however. We leave it to future research about the performance of this local smoothing GEL approach for high dimensinal time series models.

Acknowledgements

The authors are grateful to the co-editor Jianqing Fan, an Associate Editor, two anonymous referees and Qiwei Yao for constructive comments and suggestions. This work is based on part of Jinyuan Chang's PhD thesis at Peking University. We thank Jean-Michel Zakoïan for sharing a longer version of his paper, and Chenlei Leng and Cheng Yong Tang for sharing their R code. Chang and S.X. Chen were supported by National Natural Science Foundation of China Grants 11131002 and G0113, Center for Statistical Science and LMEQF at Peking University.

Appendix

Throughout the Appendix, C denotes a generic positive finite constant that may be different in different uses. For any $q = 1, \ldots, Q$ and $k = 1, \ldots, M$, let $\beta_1(q,k) = \#\{j < q : X_{(q-1)L+k} \in B_j\}$ and $\beta_2(q,k) = \#\{j > q : X_{(q-1)L+k} \in B_j\}$. These two quantities denote the times of the k-th element of the q-th block occurs in the blocks before and after the q-th block, respectively. Let $\bar{g}(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n g_t(\boldsymbol{\theta})$, $\bar{\phi}(\boldsymbol{\theta}) = Q^{-1}\sum_{q=1}^Q \phi_q(\boldsymbol{\theta})$, $V_M = \text{Var}\{M^{1/2}\phi_q(\boldsymbol{\theta}_0)\}$, $\widehat{\Omega}(\boldsymbol{\theta}) = Q^{-1}\sum_{q=1}^Q \phi_q(\boldsymbol{\theta})\phi_q(\boldsymbol{\theta})'$ and $\Omega(\boldsymbol{\theta}) = E\{\phi_q(\boldsymbol{\theta})\phi_q(\boldsymbol{\theta})'\}$.

Some Lemmas I

The lemmas proposed in this subsection are used to prove Theorem 1.

Lemma 1. $\beta_1(q,k) = (q-1) \wedge \lfloor (M-k)/L \rfloor$ *and* $\beta_2(q,k) = (Q-q) \wedge \lfloor (k-1)/L \rfloor$.

PROOF: For $t = (q-1)L+k$, suppose $X_t \in B_{\bar{q}}$ where $\bar{q} < q$. Then there exists a positive integer $\bar{k} \in [1, M]$ such that $(q-1)L+k = (\bar{q}-1)L+\bar{k}$. It means $\bar{q} = q-(\bar{k}-k)/L$. From this, we can get $\bar{k} = k+iL$ for some $i \in \{1, \ldots, q-1\}$. Note that $\bar{k} \in [1, M]$, then $i \le \lfloor (M-k)/L \rfloor$. Hence, $\beta_1(q,k) = (q-1) \wedge \lfloor (M-k)/L \rfloor$. By the same argument, $\beta_2(q,k) = (Q-q) \wedge \lfloor (k-1)/L \rfloor$. □

Lemma 2. *Under conditions (A.1)(ii) and (A.2)(ii),* $\sup_{\boldsymbol{\theta} \in \Theta} \|\bar{\phi}(\boldsymbol{\theta}) - \bar{g}(\boldsymbol{\theta})\|_2 = O_p(r^{1/2} M n^{-1})$.

PROOF: By Jensen's inequality,

$$E\bigg\{\sup_{\boldsymbol{\theta} \in \Theta} \|\bar{\phi}(\boldsymbol{\theta}) - \bar{g}(\boldsymbol{\theta})\|_2\bigg\} \le \frac{1}{MQ}\bigg\{n-(Q-1)L-M+\sum_{\beta_1(q,k)=0} \beta_2(q,k) + n - MQ\bigg\} \cdot E\bigg\{\sup_{\boldsymbol{\theta} \in \Theta} \|g_t(\boldsymbol{\theta})\|_2\bigg\}.$$

From Lemma 1 and (A.1)(ii), $\sum_{\beta_1(q,k)=0} \beta_2(q,k) \le (Q-1)(M-L)$ for sufficiently large n. Noting that $Q = \lfloor (n-M)/L \rfloor + 1$, then for sufficiently large n

$$E\bigg\{\sup_{\boldsymbol{\theta} \in \Theta} \|\bar{\phi}(\boldsymbol{\theta}) - \bar{g}(\boldsymbol{\theta})\|_2\bigg\} \le 2LM^{-1}Q^{-1} \cdot E\bigg\{\sup_{\boldsymbol{\theta} \in \Theta} \|g_t(\boldsymbol{\theta})\|_2\bigg\}.$$

Hence, (A.1)(ii) and (A.2)(ii) lead to the conclusion. □

Lemma 3. *Under conditions (A.1)(i) and (A.2)(iii),* $\|\widehat{\Omega}(\boldsymbol{\theta}_0) - \Omega(\boldsymbol{\theta}_0)\|_F = O_p(rM^{1/2}n^{-1/2})$.

PROOF: Note that

$$\begin{aligned} E\{\|\widehat{\Omega}(\boldsymbol{\theta}_0) - \Omega(\boldsymbol{\theta}_0)\|_F^2\} &= Q^{-1}E(\mathrm{tr}\{[\phi_q(\boldsymbol{\theta}_0)\phi_q(\boldsymbol{\theta}_0)' - \Omega(\boldsymbol{\theta}_0)]^2\}) \\ &\quad + Q^{-2}\sum_{q_1 \ne q_2} E(\mathrm{tr}\{[\phi_{q_1}(\boldsymbol{\theta}_0)\phi_{q_1}(\boldsymbol{\theta}_0)' - \Omega(\boldsymbol{\theta}_0)][\phi_{q_2}(\boldsymbol{\theta}_0)\phi_{q_2}(\boldsymbol{\theta}_0)' - \Omega(\boldsymbol{\theta}_0)]\}) \\ &=: A_1 + A_2. \end{aligned}$$

As $A_1 \le Q^{-1}E\{\|\phi_q(\boldsymbol{\theta}_0)\|_2^4\}$, by Jensen's inequality and (A.2)(iii), $A_1 = O\left(r^2Mn^{-1}\right)$. At the same time,

$$A_2 = Q^{-2}\sum_{u,v=1}^{r}\sum_{q_1 \ne q_2} E\{[\phi_{q_1,u}(\boldsymbol{\theta}_0)\phi_{q_1,v}(\boldsymbol{\theta}_0) - \Omega_{u,v}(\boldsymbol{\theta}_0)][\phi_{q_2,v}(\boldsymbol{\theta}_0)\phi_{q_2,u}(\boldsymbol{\theta}_0) - \Omega_{v,u}(\boldsymbol{\theta}_0)]\},$$

where $\Omega_{u,v}(\boldsymbol{\theta}_0)$ denotes the (u,v)-element of $\Omega(\boldsymbol{\theta}_0)$. By Davydov inequality and (A.2)(iii), $|A_2| \le Cr^2Q^{-2}\sum_{q_1 \ne q_2} \alpha_\phi\{|q_1 - q_2|\}^{1-2/\gamma}$. Hence, by (A.1)(i), $A_2 = O(r^2Mn^{-1})$. From Markov inequality, $\|\widehat{\Omega}(\boldsymbol{\theta}_0) - \Omega(\boldsymbol{\theta}_0)\|_F = O_p(rM^{1/2}n^{-1/2})$. □

Lemma 4. *Under conditions (A.1)(ii), (A.2)(ii) and (A.2)(iv), then* $\sup_{\boldsymbol{\theta} \in \Theta} \lambda_{\max}\{\widehat{\Omega}(\boldsymbol{\theta})\} = O_p(1)$ *provided that* $rMn^{-1} = o(1)$.

PROOF: Using the same approach as in the proof of Lemma 2,

$$\sup_{\boldsymbol{\theta} \in \Theta}\sup_{\|x\|_2=1}\bigg\{\bigg|\frac{1}{MQ}\sum_{q=1}^{Q}\sum_{t \in B_q} x'g_t(\boldsymbol{\theta})g_t(\boldsymbol{\theta})'x - \frac{1}{n}\sum_{t=1}^{n} x'g_t(\boldsymbol{\theta})g_t(\boldsymbol{\theta})'x\bigg|\bigg\} = O_p(rMn^{-1}).$$

By Jensen's inequality, for any $\|x\|_2 = 1$,

$$\frac{1}{Q}\sum_{q=1}^{Q} x'\phi_q(\boldsymbol{\theta})\phi_q(\boldsymbol{\theta})'x \le \frac{1}{MQ}\sum_{q=1}^{Q}\sum_{t \in B_q} x'g_t(\boldsymbol{\theta})g_t(\boldsymbol{\theta})'x.$$

Then $\sup_{\boldsymbol{\theta} \in \Theta} \lambda_{\max}\{\widehat{\Omega}(\boldsymbol{\theta})\} \le \sup_{\boldsymbol{\theta} \in \Theta} \lambda_{\max}\{n^{-1}\sum_{i=1}^{n} g_t(\boldsymbol{\theta})g_t(\boldsymbol{\theta})'\} + o_p(1)$. The result can be implied by (A.2)(iv). □

Lemma 5. *Under condition (A.2)(ii), define* $\delta_n = o(r^{-1/2}Q^{-1/\gamma})$ *and* $\Lambda_n = \{\lambda \in \mathbb{R}^r : \|\lambda\|_2 \le \delta_n\}$*, we have* $\sup_{1\le q\le Q, \boldsymbol{\theta}\in\Theta, \lambda\in\Lambda_n} |\lambda'\phi_q(\boldsymbol{\theta})| \xrightarrow{p} 0$*. Also w.p.a.1,* $\Lambda_n \subset \widehat{\Lambda}_n(\boldsymbol{\theta})$ *for all* $\boldsymbol{\theta} \in \Theta$*.*

Proof: From (A.2)(ii) and Markov inequality, $\sup_{1\le q\le Q,\boldsymbol{\theta}\in\Theta} \|\phi_q(\boldsymbol{\theta})\|_2 = O_p(r^{1/2}Q^{1/\gamma})$. Then,

$$\sup_{1\le q\le Q,\boldsymbol{\theta}\in\Theta,\lambda\in\Lambda_n} |\lambda'\phi_q(\boldsymbol{\theta})| \le \delta_n \cdot \sup_{1\le q\le Q,\boldsymbol{\theta}\in\Theta} \|\phi_q(\boldsymbol{\theta})\|_2 \xrightarrow{p} 0.$$

It also implies w.p.a.1 $\lambda'\phi_q(\boldsymbol{\theta}) \in \mathcal{V}$ for all $\boldsymbol{\theta}\in\Theta$ and $\|\lambda\|_2 \le \delta_n$. □

Lemma 6. *Under conditions (A.1)(i), (A.2)(i) and (A.2)(iii), assume that* $\lambda_{\max}(V_M)$ *is uniformly bounded away from infinity. If* $r^2M^3n^{-1} = o(1)$*,* $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 = O_p(\tau_n)$ *and* $rpM\tau_n^2 = o(1)$*, then* $\|\widehat{\Omega}(\boldsymbol{\theta}) - \widehat{\Omega}(\boldsymbol{\theta}_0)\|_2 = O_p(r^{1/2}p^{1/2}M^{-1/2}\tau_n)$*.*

Proof: Choose $x \in \mathbb{R}^r$ with unit L_2-norm such that $\lambda_{\max}\{\widehat{\Omega}(\boldsymbol{\theta}) - \widehat{\Omega}(\boldsymbol{\theta}_0)\} = x'\{\widehat{\Omega}(\boldsymbol{\theta}) - \widehat{\Omega}(\boldsymbol{\theta}_0)\}x$. Then,

$$\begin{aligned}
&|\lambda_{\max}\{\widehat{\Omega}(\boldsymbol{\theta}) - \widehat{\Omega}(\boldsymbol{\theta}_0)\}| \\
&\quad\le \frac{1}{Q}\sum_{q=1}^{Q} |x'\phi_q(\boldsymbol{\theta}) - x'\phi_q(\boldsymbol{\theta}_0)| \cdot \|\phi_q(\boldsymbol{\theta}) - \phi_q(\boldsymbol{\theta}_0)\|_2 + \frac{2}{Q}\sum_{q=1}^{Q} |x'\phi_q(\boldsymbol{\theta}_0)| \cdot \|\phi_q(\boldsymbol{\theta}) - \phi_q(\boldsymbol{\theta}_0)\|_2 \\
&\quad\le \frac{1}{Q}\sum_{q=1}^{Q} \|\phi_q(\boldsymbol{\theta}) - \phi_q(\boldsymbol{\theta}_0)\|_2^2 + 2[\lambda_{\max}\{\widehat{\Omega}(\boldsymbol{\theta}_0)\}]^{1/2}\bigg\{\frac{1}{Q}\sum_{q=1}^{Q} \|\phi_q(\boldsymbol{\theta}) - \phi_q(\boldsymbol{\theta}_0)\|_2^2\bigg\}^{1/2}.
\end{aligned}$$

Note that $r^2M^3n^{-1} = o(1)$, by Lemmas 3 and $\lambda_{\max}(V_M)$ is uniformly bounded away from infinity, $\lambda_{\max}\{\widehat{\Omega}(\boldsymbol{\theta}_0)\} = O_p(M^{-1})$. From (A.2)(i), $Q^{-1}\sum_{q=1}^{Q} \|\phi_q(\boldsymbol{\theta}) - \phi_q(\boldsymbol{\theta}_0)\|_2^2 = rp \cdot O_p(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2^2)$. If $rpM\tau_n^2 = o(1)$, then $|\lambda_{\max}\{\widehat{\Omega}(\boldsymbol{\theta}) - \widehat{\Omega}(\boldsymbol{\theta}_0)\}| = O_p(r^{1/2}p^{1/2}M^{-1/2}\tau_n)$. Using the same argument, $|\lambda_{\min}\{\widehat{\Omega}(\boldsymbol{\theta}) - \widehat{\Omega}(\boldsymbol{\theta}_0)\}| = O_p(r^{1/2}p^{1/2}M^{-1/2}\tau_n)$. This completes the proof. □

Lemma 7. *Under conditions (A.1)(i), (A.1)(ii), (A.2)(i), (A.2)(ii) and (A.2)(iii), assume that the eigenvalues of* V_M *are uniformly bounded away from zero and infinity. If* $r^2M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$*,* $r^2M^3n^{-1} = o(1)$*,* $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p(\tau_n)$*,* $rpM\tau_n^2 = o(1)$ *and* $\|\bar{g}(\widetilde{\boldsymbol{\theta}})\|_2 = O_p(r^{1/2}n^{-1/2})$*, then* $\widehat{\lambda}(\widetilde{\boldsymbol{\theta}}) = \arg\max_{\lambda\in\widehat{\Lambda}_n(\widetilde{\boldsymbol{\theta}})} \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \lambda)$ *exists w.p.a.1,* $\sup_{\lambda\in\widehat{\Lambda}_n(\widetilde{\boldsymbol{\theta}})} \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \lambda) = \rho(0) + O_p(rMn^{-1})$ *and* $\|\widehat{\lambda}(\widetilde{\boldsymbol{\theta}})\|_2 = O_p(r^{1/2}Mn^{-1/2})$ *where* $\widehat{S}_n(\boldsymbol{\theta}, \lambda)$ *is defined in (7).*

Proof: Pick $\delta_n = o(r^{-1/2}Q^{-1/\gamma})$ and $r^{1/2}Mn^{-1/2} = o(\delta_n)$, which is guaranteed by $r^2M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$. From Lemma 2 and Triangle inequality, then $\|\bar{\phi}(\widetilde{\boldsymbol{\theta}})\|_2 \le \|\bar{g}(\widetilde{\boldsymbol{\theta}})\|_2 + O_p(r^{1/2}Mn^{-1})$ which implies $\|\bar{\phi}(\widetilde{\boldsymbol{\theta}})\|_2 = O_p(r^{1/2}n^{-1/2})$. Let $\bar{\lambda} = \arg\max_{\lambda\in\Lambda_n} \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \lambda)$, where Λ_n is defined in Lemma 5. By Lemmas 3, 5 and 6, noting $\rho_{vv}(0) < 0$,

$$\begin{aligned}
\rho(0) = \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, 0) \le \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \bar{\lambda}) &= \rho(0) + \rho_v(0)\bar{\lambda}'\bar{\phi}(\widetilde{\boldsymbol{\theta}}) + \frac{1}{2}\bar{\lambda}'\bigg\{\frac{1}{Q}\sum_{q=1}^{Q} \rho_{vv}(\dot{\lambda}'\phi_q(\widetilde{\boldsymbol{\theta}}))\phi_q(\widetilde{\boldsymbol{\theta}})\phi_q(\widetilde{\boldsymbol{\theta}})'\bigg\}\bar{\lambda} \\
&\le \rho(0) + |\rho_v(0)| \cdot \|\bar{\lambda}\|_2 \cdot \|\bar{\phi}(\widetilde{\boldsymbol{\theta}})\|_2 - C\|\bar{\lambda}\|_2^2 \cdot \{M^{-1} + o_p(M^{-1})\}.
\end{aligned}$$

where $\dot{\lambda}$ lies on the jointing line between $\mathbf{0}$ and $\bar{\lambda}$. Hence, $\|\bar{\lambda}\|_2 \le C \cdot M \cdot \|\bar{\phi}(\widetilde{\boldsymbol{\theta}})\|_2 \cdot \{1 + o_p(1)\} = O_p(r^{1/2}Mn^{-1/2}) = o_p(\delta_n)$. Thus $\bar{\lambda} \in \mathrm{int}(\Lambda_n)$ w.p.a.1. Since $\Lambda_n \subset \widehat{\Lambda}_n(\widetilde{\boldsymbol{\theta}})$ w.p.a.1, $\widehat{\lambda}(\widetilde{\boldsymbol{\theta}}) = \bar{\lambda}$ w.p.a.1 by the concavity of $\widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \lambda)$ and $\widehat{\Lambda}_n(\widetilde{\boldsymbol{\theta}})$. Then,

$$\widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \widehat{\lambda}(\widetilde{\boldsymbol{\theta}})) \le \rho(0) + |\rho_v(0)| \cdot \|\widehat{\lambda}(\widetilde{\boldsymbol{\theta}})\|_2 \cdot \|\bar{\phi}(\widetilde{\boldsymbol{\theta}})\|_2 - C \cdot M^{-1} \cdot \|\widehat{\lambda}(\widetilde{\boldsymbol{\theta}})\|_2^2 \cdot \{1 + o_p(1)\}$$

leads to $\sup_{\lambda\in\widehat{\Lambda}_n(\widetilde{\boldsymbol{\theta}})} \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \lambda) = \rho(0) + O_p(rMn^{-1})$. □

Proof of Theorem 1

Choose $\delta_n = o(r^{-1/2}Q^{-1/\gamma})$ and $r^{1/2}Mn^{-1/2} = o(\delta_n)$. Let $\bar{\lambda} = \text{sign}\{\rho_v(0)\} \cdot \delta_n \bar{\phi}(\widehat{\boldsymbol{\theta}}_n)/\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2$, then $\bar{\lambda} \in \Lambda_n$. By Taylor expansion, Lemmas 4 and 5, noting $\rho_{vv}(0) < 0$,

$$\begin{aligned}\widehat{S}_n(\widehat{\boldsymbol{\theta}}_n, \bar{\lambda}) &= \rho(0) + \rho_v(0)\bar{\lambda}'\bar{\phi}(\widehat{\boldsymbol{\theta}}_n) + \frac{1}{2}\bar{\lambda}'\bigg\{\frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}(\dot{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'\bigg\}\bar{\lambda} \\ &\geq \rho(0) + |\rho_v(0)| \cdot \delta_n \cdot \|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2 - C \cdot O_p(1) \cdot \|\bar{\lambda}\|_2^2.\end{aligned}$$

Meanwhile, by the same way in the proof of Lemma 3, $\|\bar{g}(\boldsymbol{\theta}_0) - E\{g_t(\boldsymbol{\theta}_0)\}\|_2 = O_p(r^{1/2}n^{-1/2})$. Since $E\{g_t(\boldsymbol{\theta}_0)\} = \mathbf{0}$, $\|\bar{g}(\boldsymbol{\theta}_0)\|_2 = O_p(r^{1/2}n^{-1/2})$. Then, from Lemma 7,

$$\widehat{S}_n(\widehat{\boldsymbol{\theta}}_n, \bar{\lambda}) \leq \sup_{\lambda \in \widehat{\Lambda}_n(\widehat{\boldsymbol{\theta}}_n)} \widehat{S}_n(\widehat{\boldsymbol{\theta}}_n, \lambda) \leq \sup_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta}_0)} \widehat{S}_n(\boldsymbol{\theta}_0, \lambda) = \rho(0) + O_p(rMn^{-1}).$$

Hence, $\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2 = O_p(\delta_n)$. Consider any $\varepsilon_n \to 0$ and let $\tilde{\lambda} = \text{sign}\{\rho_v(0)\} \cdot \varepsilon_n \bar{\phi}(\widehat{\boldsymbol{\theta}}_n)$, then $\|\tilde{\lambda}\|_2 = o_p(\delta_n)$. Using the same way above, we can obtain

$$|\rho_v(0)| \cdot \varepsilon_n \cdot \|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2^2 - C \cdot O_p(1) \cdot \varepsilon_n^2 \cdot \|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2^2 = O_p(rMn^{-1}).$$

Then, $\varepsilon_n\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2^2 = O_p(rMn^{-1})$. Thus, $\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2^2 = O_p(rMn^{-1})$. From Lemma 2, $\|\bar{g}(\widehat{\boldsymbol{\theta}}_n)\|_2 = O_p(r^{1/2}M^{1/2}n^{-1/2})$. If $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2$ does not converge to zero in probability, then there exists a subsequence $\{(n_*, M_*, r_*, p_*)\}$ such that $\|\widehat{\boldsymbol{\theta}}_{n_*} - \boldsymbol{\theta}_0\|_2 \geq \varepsilon$ a.s. for some positive constant ε. By (A.1)(iv), $\|E\{g_t(\widehat{\boldsymbol{\theta}}_{n_*})\}\|_2 = o_p\{\Delta_1(r_*, p_*)\} + O_p(r_*^{1/2}M_*^{1/2}n_*^{-1/2})$. On the other hand, from (A.1)(iii), $\|E\{g_t(\widehat{\boldsymbol{\theta}}_{n_*})\}\|_2 \geq \Delta_1(r_*, p_*)\Delta_2(\varepsilon)$. As $\liminf_{r,p\to\infty} \Delta_1(r, p) > 0$, it is a contradiction. Hence, $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 \xrightarrow{p} 0$. By (A.2)(iv), $\|\bar{g}(\widehat{\boldsymbol{\theta}}_n) - \bar{g}(\boldsymbol{\theta}_0)\|_2 \geq C\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2$ w.p.a.1. Then, $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(r^{1/2}M^{1/2}n^{-1/2})$. In addition, if $r^2pM^2n^{-1} = o(1)$, from Lemmas 3 and 6, $\lambda_{\max}\{\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\} \leq CM^{-1}$ w.p.a.1. By repeating the above arguments, we can obtain $\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\|_2 = O_p(r^{1/2}n^{-1/2})$ and $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(r^{1/2}n^{-1/2})$. From Lemma 7, $\|\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)\|_2 = O_p(r^{1/2}Mn^{-1/2})$. Therefore, we complete the proof of Theorem 1. □

Proof of Corollary 1

To construct Corollary 1, we need the analogue of Lemma 7 listed below.

Lemma 8. *Under conditions (A.1)(i), (A.1)(ii), (A.2)(i), (A.2)(ii) and (A.2)(iii), assume that* $\lambda_{\min}(V_M) \asymp r^{-\iota_1}$ *for some* $\iota_1 > 0$.

(a). *If* $r^{2+3\iota_1}M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$, $r^{2+2\iota_1}M^3n^{-1} = o(1)$, $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p(\tau_n)$, $r^{1+2\iota_1}pM\tau_n^2 = o(1)$ *and* $\|\bar{g}(\widetilde{\boldsymbol{\theta}})\|_2 = O_p(r^{(1+\iota_1)/2}n^{-1/2})$, *then* $\widehat{\lambda}(\widetilde{\boldsymbol{\theta}}) = \arg\max_{\lambda \in \widehat{\Lambda}_n(\widetilde{\boldsymbol{\theta}})} \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \lambda)$ *exists w.p.a.1,* $\sup_{\lambda \in \widehat{\Lambda}_n(\widetilde{\boldsymbol{\theta}})} \widehat{S}_n(\widetilde{\boldsymbol{\theta}}, \lambda) = \rho(0) + O_p(r^{1+2\iota_1}Mn^{-1})$ *and* $\|\widehat{\lambda}(\widetilde{\boldsymbol{\theta}})\|_2 = O_p(r^{(1+3\iota_1)/2}Mn^{-1/2})$ *where* $\widehat{S}_n(\boldsymbol{\theta}, \lambda)$ *is defined in (7).*

(b). *If* $r^{2+2\iota_1}M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$ *and* $r^{2+2\iota_1}M^3n^{-1} = o(1)$, *then* $\widehat{\lambda}(\boldsymbol{\theta}_0) = \arg\max_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta}_0)} \widehat{S}_n(\boldsymbol{\theta}_0, \lambda)$ *exists w.p.a.1,* $\sup_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta}_0)} \widehat{S}_n(\boldsymbol{\theta}_0, \lambda) = \rho(0) + O_p(r^{1+\iota_1}Mn^{-1})$.

Proof: We first prove (a). Pick $\delta_n = o(r^{-1/2}Q^{-1/\gamma})$ and $r^{(1+3\iota_1)/2}Mn^{-1/2} = o(\delta_n)$, which is guaranteed by $r^{2+3\iota_1}M^{2-2/\gamma}n^{2/\gamma-1} = o(1)$. From Lemma 2 and Triangle inequality, then $\|\bar{\phi}(\widetilde{\boldsymbol{\theta}})\|_2 \leq \|\bar{g}(\widetilde{\boldsymbol{\theta}})\|_2 +$

$O_p(r^{1/2}Mn^{-1})$ which implies $\|\bar\phi(\widetilde{\boldsymbol\theta})\|_2 = O_p(r^{(1+\iota_1)/2}n^{-1/2})$. Let $\bar\lambda = \arg\max_{\lambda\in\Lambda_n}\widehat S_n(\widetilde{\boldsymbol\theta},\lambda)$, where Λ_n is defined in Lemma 5. By Lemmas 3, 5 and 6, noting $\rho_{vv}(0) < 0$,

$$\rho(0) = \widehat S_n(\widetilde{\boldsymbol\theta},0) \le \widehat S_n(\widetilde{\boldsymbol\theta},\bar\lambda) = \rho(0) + \rho_v(0)\bar\lambda'\bar\phi(\widetilde{\boldsymbol\theta}) + \frac{1}{2}\bar\lambda'\bigg\{\frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}(\dot\lambda'\phi_q(\widetilde{\boldsymbol\theta}))\phi_q(\widetilde{\boldsymbol\theta})\phi_q(\widetilde{\boldsymbol\theta})'\bigg\}\bar\lambda$$
$$\le \rho(0) + |\rho_v(0)|\cdot\|\bar\lambda\|_2\cdot\|\bar\phi(\widetilde{\boldsymbol\theta})\|_2 - C\|\bar\lambda\|_2^2\cdot\{M^{-1}r^{-\iota_1} + o_p(M^{-1}r^{-\iota_1})\}.$$

where $\dot\lambda$ lies on the jointing line between $\mathbf{0}$ and $\bar\lambda$. Therefore, $\|\bar\lambda\|_2 \le C\cdot Mr^{\iota_1}\cdot\|\bar\phi(\widetilde{\boldsymbol\theta})\|_2\cdot\{1+o_p(1)\} = O_p(r^{(1+3\iota_1)/2}Mn^{-1/2}) = o_p(\delta_n)$. Thus $\bar\lambda\in\mathrm{int}(\Lambda_n)$ w.p.a.1. Since $\Lambda_n\subset\widehat\Lambda_n(\widetilde{\boldsymbol\theta})$ w.p.a.1, $\widehat\lambda(\widetilde{\boldsymbol\theta}) = \bar\lambda$ w.p.a.1 by the concavity of $\widehat S_n(\widetilde{\boldsymbol\theta},\lambda)$ and $\widehat\Lambda_n(\widetilde{\boldsymbol\theta})$. Then,

$$\widehat S_n(\widetilde{\boldsymbol\theta},\widehat\lambda(\widetilde{\boldsymbol\theta})) \le \rho(0) + |\rho_v(0)|\cdot\|\widehat\lambda(\widetilde{\boldsymbol\theta})\|_2\cdot\|\bar\phi(\widetilde{\boldsymbol\theta})\|_2 - C\cdot M^{-1}r^{-\iota_1}\cdot\|\widehat\lambda(\widetilde{\boldsymbol\theta})\|_2^2\cdot\{1+o_p(1)\}$$

leads to $\sup_{\lambda\in\widehat\Lambda_n(\widetilde{\boldsymbol\theta})}\widehat S_n(\widetilde{\boldsymbol\theta},\lambda) = \rho(0) + O_p(r^{1+2\iota_1}Mn^{-1})$. The proof of (b) is similar to that of (a). □

Here, we begin to prove Corollary 1. Choose $\delta_n = o(r^{-1/2}Q^{-1/\gamma})$ and $r^{(1+3\iota_1)/2}Mn^{-1/2} = o(\delta_n)$. Let $\bar\lambda = \mathrm{sign}\{\rho_v(0)\}\cdot\delta_n\bar\phi(\widehat{\boldsymbol\theta}_n)/\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2$, then $\bar\lambda\in\Lambda_n$. By Taylor expansion, Lemmas 4 and 5, noting $\rho_{vv}(0) < 0$,

$$\widehat S_n(\widehat{\boldsymbol\theta}_n,\bar\lambda) = \rho(0) + \rho_v(0)\bar\lambda'\bar\phi(\widehat{\boldsymbol\theta}_n) + \frac{1}{2}\bar\lambda'\bigg\{\frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}(\dot\lambda'\phi_q(\widehat{\boldsymbol\theta}_n))\phi_q(\widehat{\boldsymbol\theta}_n)\phi_q(\widehat{\boldsymbol\theta}_n)'\bigg\}\bar\lambda$$
$$\ge \rho(0) + |\rho_v(0)|\cdot\delta_n\cdot\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2 - C\cdot O_p(1)\cdot\|\bar\lambda\|_2^2.$$

Meanwhile, by the same way in the proof of Lemma 3, $\|\bar g(\boldsymbol\theta_0) - E\{g_t(\boldsymbol\theta_0)\}\|_2 = O_p(r^{1/2}n^{-1/2})$. Since $E\{g_t(\boldsymbol\theta_0)\} = \mathbf{0}$, $\|\bar g(\boldsymbol\theta_0)\|_2 = O_p(r^{1/2}n^{-1/2})$. Then, from Lemma 8(b),

$$\widehat S_n(\widehat{\boldsymbol\theta}_n,\bar\lambda) \le \sup_{\lambda\in\widehat\Lambda_n(\widehat{\boldsymbol\theta}_n)}\widehat S_n(\widehat{\boldsymbol\theta}_n,\lambda) \le \sup_{\lambda\in\widehat\Lambda_n(\boldsymbol\theta_0)}\widehat S_n(\boldsymbol\theta_0,\lambda) = \rho(0) + O_p(r^{1+\iota_1}Mn^{-1}).$$

Hence, $\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2 = O_p(\delta_n)$. Consider any $\varepsilon_n\to 0$ and let $\tilde\lambda = \mathrm{sign}\{\rho_v(0)\}\cdot\varepsilon_n\bar\phi(\widehat{\boldsymbol\theta}_n)$, then $\|\tilde\lambda\|_2 = o_p(\delta_n)$. Using the same way above, we can obtain

$$|\rho_v(0)|\cdot\varepsilon_n\cdot\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2^2 - C\cdot O_p(1)\cdot\varepsilon_n^2\cdot\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2^2 = O_p(r^{1+\iota_1}Mn^{-1}).$$

Then, $\varepsilon_n\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2^2 = O_p(r^{1+\iota_1}Mn^{-1})$. Thus, $\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2^2 = O_p(r^{1+\iota_1}Mn^{-1})$. From Lemma 2, $\|\bar g(\widehat{\boldsymbol\theta}_n)\|_2 = O_p(r^{(1+\iota_1)/2}M^{1/2}n^{-1/2})$. If $\|\widehat{\boldsymbol\theta}_n - \boldsymbol\theta_0\|_2$ does not converge to zero in probability, then there exists a subsequence $\{(n_*,M_*,r_*,p_*)\}$ such that $\|\widehat{\boldsymbol\theta}_{n_*} - \boldsymbol\theta_0\|_2 \ge \varepsilon$ a.s. for some positive constant ε. By (A.1)(iv), $\|E\{g_t(\widehat{\boldsymbol\theta}_{n_*})\}\|_2 = o_p\{\Delta_1(r_*,p_*)\} + O_p(r_*^{(1+\iota_1)/2}M_*^{1/2}n_*^{-1/2})$. On the other hand, from (A.1)(iii), $\|E\{g_t(\widehat{\boldsymbol\theta}_{n_*})\}\|_2 \ge \Delta_1(r_*,p_*)\Delta_2(\varepsilon)$. As $\liminf_{r,p\to\infty}\Delta_1(r,p) > 0$, it is a contradiction. Hence, $\|\widehat{\boldsymbol\theta}_n - \boldsymbol\theta_0\|_2 \xrightarrow{p} 0$. By (A.2)(iv), $\|\bar g(\widehat{\boldsymbol\theta}_n) - \bar g(\boldsymbol\theta_0)\|_2 \ge C\|\widehat{\boldsymbol\theta}_n - \boldsymbol\theta_0\|_2$ w.p.a.1. Then, $\|\widehat{\boldsymbol\theta}_n - \boldsymbol\theta_0\|_2 = O_p(r^{(1+\iota_1)/2}M^{1/2}n^{-1/2})$. In addition, if $r^{2+\iota_1}pM^2n^{-1} = o(1)$, from Lemmas 3 and 6, $\lambda_{\max}\{\widehat\Omega(\widehat{\boldsymbol\theta}_n)\} \le CM^{-1}$ w.p.a.1. By repeating the above arguments, we can obtain $\|\bar\phi(\widehat{\boldsymbol\theta}_n)\|_2 = O_p(r^{(1+\iota_1)/2}n^{-1/2})$ and $\|\widehat{\boldsymbol\theta}_n - \boldsymbol\theta_0\|_2 = O_p(r^{(1+\iota_1)/2}n^{-1/2})$. From Lemma 8, $\|\widehat\lambda(\widehat{\boldsymbol\theta}_n)\|_2 = O_p(r^{(1+3\iota_1)/2}Mn^{-1/2})$. Therefore, we complete the proof of Corollary 1. □

Some Lemmas II

The lemmas proposed in this subsection are used to establish Proposition 1, Proposition 2 and Theorem 2. The proof of Proposition 1 is based on the asymptotic expansion given in Proposition 2, so we will first construct the proof of Proposition 2 later.

Lemma 9. *Under conditions (A.1)-(A.2), assume that $\lambda_{\max}(V_M)$ is uniformly bounded away from infinity. If $r^2M^{2-2/\gamma}n^{2/\gamma-1}=o(1)$, $r^2pM^2n^{-1}=o(1)$ and $r^2M^3n^{-1}=o(1)$, then for any $x\in\mathbb{R}^p$, $y,z\in\mathbb{R}^r$,*

$$\left\|\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)'\phi_q(\widehat{\boldsymbol{\theta}}_n))\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)x-\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)x\right\|_2=O_p(rp^{1/2}M^{1/2}n^{-1/2})\cdot\|x\|_2,$$

$$\left|\frac{M}{Q}\sum_{q=1}^{Q}y'\rho_{vv}(\widetilde{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'z-\frac{M}{Q}\sum_{q=1}^{Q}\rho_{vv}(0)y'\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'z\right|$$
$$=O_p(rM^{1-1/\gamma}n^{1/\gamma-1/2})\cdot\|y\|_2\cdot\|z\|_2,$$

where $\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)$ and $\widetilde{\lambda}$ are defined in (15).

Proof: From Theorem 1, both $\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)$ and $\widetilde{\lambda}$ are $O_p(r^{1/2}Mn^{-1/2})=o_p(\delta_n)$ where δ_n is defined in Lemma 5. By Taylor expansion and Cauchy-Schwarz inequality,

$$\left\|\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)'\phi_q(\widehat{\boldsymbol{\theta}}_n))\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)x-\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)x\right\|_2^2$$
$$\leq\left[\frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}^2(\dot{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\{\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)'\phi_q(\widehat{\boldsymbol{\theta}}_n)\}^2\right]\left[\frac{1}{Q}\sum_{q=1}^{Q}x'\{\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)\}'\{\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)\}x\right]$$

where $\dot{\lambda}$ lies on the jointing line between $\mathbf{0}$ and $\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)$. From Lemma 5 and $\lambda_{\max}\{\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\}=O_p(M^{-1})$ which is provided by Lemmas 3 and 6, we obtain

$$\sum_{q=1}^{Q}\rho_{vv}^2(\dot{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\big\{\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)'\phi_q(\widehat{\boldsymbol{\theta}}_n)\big\}^2\leq C\sum_{q=1}^{Q}\big\{\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n)'\phi_q(\widehat{\boldsymbol{\theta}}_n)\big\}^2\cdot\{1+o_p(1)\}=O_p(rMn^{-1}).$$

On the other hand,

$$\frac{1}{Q}\sum_{q=1}^{Q}x'\big\{\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)\big\}'\big\{\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n)\big\}x\leq\frac{1}{MQ}\sum_{q=1}^{Q}\sum_{t\in B_q}\big\|\nabla_{\boldsymbol{\theta}}g_t(\widehat{\boldsymbol{\theta}}_n)\cdot x\big\|_2^2=O_p(rp)\cdot\|x\|_2^2.$$

Using the same argument, we can obtain the other result. □

Lemma 10. *Under conditions (A.1)(i), (A.1)(ii) and (A.3), then $\|\nabla_{\boldsymbol{\theta}}\bar{\phi}(\boldsymbol{\theta})-\nabla_{\boldsymbol{\theta}}\bar{\phi}(\boldsymbol{\theta}^*)\|_F=O_p(r^{1/2}p\cdot\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2)$ for any $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$ in a neighborhood of $\boldsymbol{\theta}_0$, and $\|\nabla_{\boldsymbol{\theta}}\bar{\phi}(\boldsymbol{\theta}_0)-E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}\|_F=O_p(r^{1/2}p^{1/2}n^{-1/2})$ provided that $M=o(n^{1/2})$.*

Proof: Using Taylor expansion and noting (A.3), the first conclusion holds. Using the same method in the proof of Lemma 3, $\|\nabla_{\boldsymbol{\theta}}\bar{g}(\boldsymbol{\theta}_0)-E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}\|_F=O_p(r^{1/2}p^{1/2}n^{-1/2})$. By the same way in the proof of Lemma 2, $\|\nabla_{\boldsymbol{\theta}}\bar{\phi}(\boldsymbol{\theta}_0)-\nabla_{\boldsymbol{\theta}}\bar{g}(\boldsymbol{\theta}_0)\|_F=O_p(r^{1/2}p^{1/2}Mn^{-1})$. Hence, by Triangle inequality, we can obtain $\|\nabla_{\boldsymbol{\theta}}\bar{\phi}(\boldsymbol{\theta}_0)-E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}\|_F=O_p(r^{1/2}p^{1/2}n^{-1/2})$. □

Proof of Proposition 2

Define

$$\boldsymbol{\beta}=([E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}])^{-1/2}\boldsymbol{\alpha}_n,$$

then

$$\begin{aligned}\|E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}\cdot\boldsymbol{\beta}\|_2^2 &= \boldsymbol{\alpha}_n'(U'U)^{-1/2}U'V_n^{-1/2}V_M^2V_n^{-1/2}U(U'U)^{-1/2}\boldsymbol{\alpha}_n\\ &\leq \lambda_{\max}(V_n^{-1/2}V_M^2V_n^{-1/2})\cdot\|U(U'U)^{-1/2}\boldsymbol{\alpha}_n\|_2^2 = \lambda_{\max}^2(V_M)\lambda_{\min}^{-1}(V_n),\end{aligned}$$

where $U = V_n^{1/2}V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]$. Therefore, $\|E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}\cdot\boldsymbol{\beta}\|_2 = O(1)$. Meanwhile,

$$\begin{aligned}\|\boldsymbol{\beta}\|_2^2 &\leq \lambda_{\min}^{-1}([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])\\ &\leq \lambda_{\max}^2(V_M)\lambda_{\min}^{-1}([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])\lambda_{\min}^{-1}(V_n).\end{aligned}$$

Hence, $\|\boldsymbol{\beta}\|_2 \leq C$. From Lemma 5,

$$\frac{M}{Q}\sum_{q=1}^{Q}\rho_{vv}(\widetilde{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)' = \rho_{vv}(0)M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\cdot\{1+o_p(1)\}.$$

Noting Lemmas 3 and 6, we know the eigenvalues of $M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)$ are uniformly bounded away from zero and infinity w.p.a.1. Hence, the eigenvalues of $MQ^{-1}\sum_{q=1}^{Q}\rho_{vv}(\widetilde{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'$ are uniformly bounded away from zero and infinity w.p.a.1. By Lemma 9 and (15),

$$\boldsymbol{\beta}'\{\nabla_{\boldsymbol{\theta}}\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\}'\bigg\{\frac{M}{Q}\sum_{q=1}^{Q}\rho_{vv}(\widetilde{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n))\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'\bigg\}^{-1}\bar{\phi}(\widehat{\boldsymbol{\theta}}_n) = O_p(r^{3/2}p^{1/2}M^{1/2}n^{-1}).$$

From Lemmas 10 and 9,

$$\begin{aligned}&\boldsymbol{\beta}'[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'\{M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\}^{-1}\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\\ &= O_p(r^{3/2}p^{1/2}M^{1/2}n^{-1}) + O_p(r^{3/2}M^{1-1/\gamma}n^{1/\gamma-1}) + O_p(r^{3/2}pn^{-1}).\end{aligned}$$

Note that Lemmas 3 and 6,

$$\begin{aligned}&\boldsymbol{\beta}'[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)\\ &= O_p(r^{3/2}p^{1/2}M^{1/2}n^{-1}) + O_p(r^{3/2}M^{1-1/\gamma}n^{1/\gamma-1}) + O_p(r^{3/2}M^{3/2}n^{-1}) + O_p(r^{3/2}pn^{-1}).\end{aligned}$$

Expanding $\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, by Lemmas 10 and 2,

$$\begin{aligned}&\boldsymbol{\beta}'[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\ &= -\boldsymbol{\beta}'[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}\bar{g}(\boldsymbol{\theta}_0) + O_p(r^{3/2}p^{1/2}M^{1/2}n^{-1}) + O_p(r^{3/2}M^{1-1/\gamma}n^{1/\gamma-1})\\ &\quad + O_p(r^{3/2}pn^{-1}) + O_p(r^{3/2}M^{3/2}n^{-1}).\end{aligned}$$

Hence, we obtain Proposition 2. □

Proof of Proposition 1

From Proposition 2, if we pick $\boldsymbol{\alpha}_n = (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)/\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2$, we can obtain that

$$\begin{aligned}&\sqrt{n}\cdot\lambda_{\max}^{-1/2}\{[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\}\\ &\qquad\qquad\cdot\lambda_{\min}\{[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\}\cdot\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2\\ &= O_p\{\|\sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}V_nV_M^{-1}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]'V_M^{-1}\bar{g}(\boldsymbol{\theta}_0)\|_2\}\\ &\quad + O_p(r^{3/2}p^{1/2}M^{1/2}n^{-1/2}) + O_p(r^{3/2}pn^{-1/2}) + O_p(r^{3/2}M^{1-1/\gamma}n^{1/\gamma-1/2}) + O_p(r^{3/2}M^{3/2}n^{-1/2}).\end{aligned}$$

Note that

$$\lambda_{\max}^{-1/2}\{[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\}\lambda_{\min}\{[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]\} > C,$$

which is assumed in (A.2)(iv) and the eigenvalues of V_M and V_n are uniformly bounded away from zero and infinity. Therefore, by

$$E(\sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} \bar{g}(\boldsymbol{\theta}_0)) = p,$$

we complete the proof of Proposition 1. □

Proof of Theorem 2

From Proposition 2, it is only need to show

$$S_n := -\sqrt{n}\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} \bar{g}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0,1).$$

Let

$$x_{n,t} = -\boldsymbol{\alpha}_n'([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2}[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} \bar{g}(\boldsymbol{\theta}_0) =: \boldsymbol{\beta}_n' g_t(\boldsymbol{\theta}_0),$$

then $S_n = n^{-1/2}\sum_{t=1}^n x_{n,t}$. As restriction (17) holds, $\sup_n \sup_{1\le t\le n} E(|x_{n,t}|^\gamma) < \infty$. On the other hand, $\mathrm{Var}(S_n) = 1$. Note that (A.1)(i), then by the central limit theorem proposed in Francq and Zakoïan (2005), we have $S_n \xrightarrow{d} N(0,1)$. □

Some Lemmas III

To prove Theorem 3, we employ the blocking technique by splitting the observations to big blocks of length h and small blocks of length b. Suppose that $\tilde{B}_i = (X_{(i-1)(h+b)+1}, \ldots, X_{i(h+b)}) = (\tilde{B}_{i1}, \tilde{B}_{i2})$, where $\tilde{B}_{i1} = (X_{(i-1)(h+b)+1}, \ldots, X_{(i-1)(h+b)+h})$, $\tilde{B}_{i2} = (X_{(i-1)(h+b)+(h+1)}, \ldots, X_{i(h+b)})$ and $b < h$. Then $n = T(h+b) + m$, where $m < h + b$. Later, we will discuss the selection of b and h. By a similar argument to those in finding the order of $\|\bar{g}(\boldsymbol{\theta}_0)\|_2$ and the proof of Lemma 2, we can obtain $\|\bar{g}(\boldsymbol{\theta}_0) - (Th)^{-1}\sum_{i=1}^T \sum_{t\in\tilde{B}_{i1}} g_t(\boldsymbol{\theta}_0)\|_2 = O_p(r^{1/2}T^{1/2}b^{1/2}n^{-1})$. Furthermore, define $\widetilde{V}_n = \mathrm{Var}\{h^{-1/2}\sum_{t=1}^h g_t(\boldsymbol{\theta}_0)\}$, $Z_{T,i} = h^{-1/2}\widetilde{V}_n^{-1/2}\sum_{t\in\tilde{B}_{i1}} g_t(\boldsymbol{\theta}_0)$ and $G_{T,k} = \sum_{i=1}^k Z_{T,i}$, then $E(Z_{T,i}) = \mathbf{0}$ and $E(Z_{T,i}Z_{T,i}') = I_r$. It can be shown that $|x'(V_n - \widetilde{V}_n)y| \le Crh^{-1}\sum_{k=1}^h k\alpha_X(k)^{1-2/\gamma}\cdot\|x\|_2\|y\|_2$. Define $\mathscr{G}_{T,0} = \{\varnothing, \Omega\}$, $\mathscr{G}_{T,k} = \sigma(Z_{T,1}, \ldots, Z_{T,k})$, $k = 1, \ldots, T$, and $S_{T,k} = T^{-1}(2r)^{-1/2}\{(\sum_{i=1}^k Z_{T,i}')(\sum_{i=1}^k Z_{T,i}) - kr\}$. $E_{T,k}(\cdot)$ denote the conditional expectation given $\mathscr{G}_{T,k}$. Let $D_{T,k} = S_{T,k} - S_{T,k-1} = T^{-1}(2r)^{-1/2}(2Z_{T,k}'G_{T,k-1} + \|Z_{T,k}\|_2^2 - r)$. The following lemmas are used to establish Theorem 3.

Lemma 11. *Under conditions (A.1)(i) and (A.2)(iii), assume that the eigenvalues of V_n are uniformly bounded away from zero and infinity. Then*

$$E\{(\|Z_{T,k}\|_2^2 - r)^2\} \le Cr^2h^2,$$

$$E\{|E_{T,k-1}(Z_{T,k}'G_{T,k-1})|\} \le Crk^{1/2}\left\{\sum_{l=1}^h \alpha_X(b+l)^{1-2/\gamma}\right\}^{1/2},$$

$$E\{E_{T,k-1}^2(\|Z_{T,k}\|_2^2 - r)\} \le Cr^2h^2\alpha_X(b+1)^{1-2/\gamma},$$

and for any $i \ne j$,

$$|E\{(\|Z_{T,i}\|_2^2 - r)(\|Z_{T,j}\|_2^2 - r)\}| \le Cr^2h^2\alpha_X\{(b+h)|i-j| - h + 1\}^{1-2/\gamma},$$

provided that $rh^{-1}\sum_{k=1}^h k\alpha_X(k)^{1-2/\gamma} = o(1)$.

PROOF: As $rh^{-1}\sum_{k=1}^{h} k\alpha_X(k)^{1-2/\gamma}=o(1)$, $\lambda_{\max}(\widetilde{V}_n^{-1})\le C$. Then, $E(\|Z_{T,k}\|_2^4)\le Ch^{-2}E\{\|\sum_{t=1}^{h} g_t(\boldsymbol{\theta}_0)\|_2^4\}$. By Triangle and Jensen's inequalities, $E\{\|\sum_{t=1}^{h} g_t(\boldsymbol{\theta}_0)\|_2^4\}\le Cr^2h^4$. Hence, $E\{(\|Z_{T,k}\|_2^2-r)^2\}\le Cr^2h^2$. By Cauchy-Schwarz inequality,

$$E\{|E_{T,k-1}(Z'_{T,k}G_{T,k-1})|\}\le\{E(\|G_{T,k-1}\|_2^2)\}^{1/2}[E\{\|E_{T,k-1}(Z_{T,k})\|_2^2\}]^{1/2}.$$

Using the same method in the proof of Lemma 3, $E(\|G_{T,k-1}\|_2^2)\le Crk$. On the other hand, note that $\|E_{T,k-1}(Z_{T,k})\|_2^2\le C\sum_{t\in\tilde{B}_{k1}}\|E_{T,k-1}\{g_t(\boldsymbol{\theta}_0)\}\|_2^2$. Hence,

$$\begin{aligned}
E\{\|E_{T,k-1}(Z_{T,k})\|_2^2\} &\le C\sum_{t\in\tilde{B}_{k1}}\sum_{j=1}^{r}E[E_{T,k-1}^2\{g_{tj}(\boldsymbol{\theta}_0)\}]\\
&\le C\sum_{t\in\tilde{B}_{k1}}\sum_{j=1}^{r}[E\{|g_{tj}(\boldsymbol{\theta}_0)|^\gamma\}]^{2/\gamma}\alpha_X\{t+b-(k-1)(h+b)\}^{1-2/\gamma} \qquad (27)\\
&= Cr\sum_{l=1}^{h}\alpha_X(b+l)^{1-2/\gamma}.
\end{aligned}$$

This is based on the fact that if $E(X)=0$, then $(E[\{E(X|\mathcal{F})\}^2])^{1/2}=\sup\{E(XY):Y\in\mathcal{F},E(Y^2)=1\}$ for any σ-field $\mathcal{F}$ (details can be found in Durrett (2010)), and Davydov inequality. Then, $E\{|E_{T,k-1}(Z'_{T,k}G_{T,k-1})|\}\le Crk^{1/2}\{\sum_{l=1}^{h}\alpha_X(b+l)^{1-2/\gamma}\}^{1/2}$. Using the same argument above, we can obtain $E(\|Z_{T,k}\|_2^{2\gamma})\le Cr^\gamma h^\gamma$. Then, by the same argument of (27),

$$E\{E_{T,k-1}^2(\|Z_{T,k}\|_2^2-r)\}\le Cr^2h^2\alpha_X(b+1)^{1-2/\gamma}. \qquad (28)$$

For any $i\ne j$, by Davydov inequality,

$$|E\{(\|Z_{T,i}\|_2^2-r)(\|Z_{T,j}\|_2^2-r)\}|\le C\{E(|\|Z_{T,i}\|_2^2-r|^\gamma)\}^{2/\gamma}\alpha_X\{(b+h)|i-j|-h+1\}^{1-2/\gamma}.$$

Hence, we complete the proof of this lemma. $\square$

Lemma 12. *Under conditions (A.1)(i) and (A.2)(iii), assume that the eigenvalues of V_n are uniformly bounded away from zero and infinity. Then*

$$E\{\|E_{T,k-1}(Z_{T,k}Z'_{T,k})-I_r\|_F^2\}\le Cr^2h^2\alpha_X(b+1)^{1-2/\gamma},$$
$$E[\{G'_{T,k-1}E_{T,k-1}(Z_{T,k})\}^2]\le Cr^2h^2k^2\alpha_X(b+1)^{1/2-1/\gamma}.$$

provided that $rh^{-1}\sum_{k=1}^{h} k\alpha_X(k)^{1-2/\gamma}=o(1)$.

PROOF: Note that

$$\begin{aligned}
&E\{\|E_{T,k-1}(Z_{T,k}Z'_{T,k})-I_r\|_F^2\}\\
&\le CE\left(\left\|E_{T,k-1}\left\{\left[h^{-1/2}\sum_{t\in\tilde{B}_{k1}}g_t(\boldsymbol{\theta}_0)\right]\left[h^{-1/2}\sum_{t\in\tilde{B}_{k1}}g_t(\boldsymbol{\theta}_0)\right]'\right\}-\widetilde{V}_n\right\|_F^2\right)\\
&\le C\sum_{u,v=1}^{r}E\left(\left|E_{T,k-1}\left\{\left[h^{-1/2}\sum_{t\in\tilde{B}_{k1}}g_{tu}(\boldsymbol{\theta}_0)\right]\left[h^{-1/2}\sum_{t\in\tilde{B}_{k1}}g_{tv}(\boldsymbol{\theta}_0)\right]\right\}-\widetilde{V}_n(u,v)\right|^2\right)\\
&\le Cr^2h^2\alpha_X(b+1)^{1-2/\gamma},
\end{aligned}$$

where $\widetilde{V}_n(u,v)$ denotes the (u,v)-element of $\widetilde{V}_n$. The last step is similar to (28). Then we obtain the first conclusion. As

$$\begin{aligned}\|E_{T,k-1}(Z_{T,k})\|_2^4 &\le \{E_{T,k-1}(Z_{T,k})\}'\{E_{T,k-1}(Z_{T,k}Z'_{T,k})\}\{E_{T,k-1}(Z_{T,k})\}\\ &\le \|E_{T,k-1}(Z_{T,k})\|_2^2\|E_{T,k-1}(Z_{T,k}Z'_{T,k})-I_r\|_F+\|E_{T,k-1}(Z_{T,k})\|_2^2,\end{aligned}$$

then by Cauchy-Schwarz inequality,

$$\begin{aligned}&E\{\|E_{T,k-1}(Z_{T,k})\|_2^4\}\\ &\le [E\{\|E_{T,k-1}(Z_{T,k})\|_2^4\}]^{1/2}[E\{\|E_{T,k-1}(Z_{T,k}Z'_{T,k})-I_r\|_F^2\}]^{1/2}+E\{\|E_{T,k-1}(Z_{T,k})\|_2^2\}.\end{aligned}$$

Let $U=[E\{\|E_{T,k-1}(Z_{T,k})\|_2^4\}]^{1/2}$, from (27) and the first result in this lemma,

$$U^2\le CUrh\alpha_X(b+1)^{1/2-1/\gamma}+Crh\alpha_X(b+1)^{1-2/\gamma}.$$

Then, $U\le Crh\alpha_X(b+1)^{1/2-1/\gamma}$. Hence, $E\{\|E_{T,k-1}(Z_{T,k})\|_2^4\}\le Cr^2h^2\alpha_X(b+1)^{1-2/\gamma}$. Also,

$$E[\{G'_{T,k-1}E_{T,k-1}(Z_{T,k})\}^2]\le\{E(\|G_{T,k-1}\|_2^4)\}^{1/2}[E\{\|E_{T,k-1}(Z_{T,k})\|_2^4\}]^{1/2}\le Cr^2h^2k^2\alpha_X(b+1)^{1/2-1/\gamma}.$$

We complete the proof. □

Lemma 13. *Under conditions (A.1)(i) and (A.2)(iii), assume that the eigenvalues of V_n are uniformly bounded away from zero and infinity. Then $r^{-1}T^{-2}\sum_{j=2}^T(T-j)G'_{T,j-1}Z_{T,j}=o_p(1)$ provided that $rh^{-1}\sum_{k=1}^h k\alpha_X(k)^{1-2/\gamma}=o(1)$ and $n^2h^2\alpha_X(b+1)^{1-2/\gamma}=o(1)$.*

Proof: Note that

$$\begin{aligned}&E\bigg(\bigg[\frac{1}{rT^2}\sum_{j=2}^T(T-j)\{G'_{T,j-1}Z_{T,j}-E_{T,j-1}(G'_{T,j-1}Z_{T,j})\}\bigg]^2\bigg)\\ &\qquad=\frac{1}{r^2T^4}\sum_{j=2}^T(T-j)^2E[\{G'_{T,j-1}Z_{T,j}-E_{T,j-1}(G'_{T,j-1}Z_{T,j})\}^2].\end{aligned}$$

By the first result of Lemma 12,

$$\begin{aligned}E\{(G'_{T,j-1}Z_{T,j})^2\}&=E(\|G_{T,j-1}\|_2^2)+E[G'_{T,j-1}\{E_{T,j-1}(Z_{T,j}Z'_{T,j})-I_r\}G_{T,j-1}]\\ &\le Crj+\{E(\|G_{T,j-1}\|_2^4)\}^{1/2}[E\{\|E_{T,j-1}(Z_{T,j}Z'_{T,j})-I_r\|_F^2\}]^{1/2}\\ &\le Crj+Cr^2h^2j^2\alpha_X(b+1)^{1/2-1/\gamma}.\end{aligned}$$

Using Cauchy-Schwarz inequality and the fact $E\{|E_{T,j-1}(G'_{T,j-1}Z_{T,j})|^2\}\le E\{(G'_{T,j-1}Z_{T,j})^2\}$,

$$E\bigg[\frac{1}{r^2T^4}\sum_{j=2}^T(T-j)^2\{G'_{T,j-1}Z_{T,j}-E_{T,j-1}(G'_{T,j-1}Z_{T,j})\}^2\bigg]\le Cr^{-1}+Cnh\alpha_X(b+1)^{1/2-1/\gamma}\to 0.$$

Then, $r^{-1}T^{-2}\sum_{j=2}^T(T-j)\{G'_{T,j-1}Z_{T,j}-E_{T,j-1}(G'_{T,j-1}Z_{T,j})\}=o_p(1)$. From Lemma 11, we have $r^{-1}T^{-2}\sum_{j=2}^T(T-j)E_{T,j-1}(G'_{T,j-1}Z_{T,j})=o_p(1)$. Hence, we complete the proof. □

Lemma 14. *Under conditions (A.1)(i) and (A.2)(iii), assume that the eigenvalues of V_n are uniformly bounded away from zero and infinity. If $rh^{-1}\sum_{k=1}^h k\alpha_X(k)^{1-2/\gamma}=o(1)$, $rT\sum_{l=1}^h\alpha_X(b+l)^{1-2/\gamma}=o(1)$ and $rh^2\alpha_X(b+1)^{1-2/\gamma}=o(1)$, then $\sum_{k=1}^T E_{T,k-1}(D_{T,k})=o_p(1)$.*

Proof: Note that

$$\sum_{k=1}^{T} E_{T,k-1}(D_{T,k}) = \frac{2}{(2r)^{1/2}T}\sum_{k=1}^{T} E_{T,k-1}(Z'_{T,k}G_{T,k-1}) + \frac{1}{(2r)^{1/2}T}\sum_{k=1}^{T} E_{T,k-1}(\|Z_{T,k}\|_2^2 - r)$$
$$=: I_1 + I_2.$$

From Lemma 11,

$$E(|I_1|) \le \frac{2}{(2r)^{1/2}T}\sum_{k=1}^{T} E\{|E_{T,k-1}(Z'_{T,k}G_{T,k-1})|\} \le Cr^{1/2}T^{1/2}\bigg\{\sum_{l=1}^{h}\alpha_X(b+l)^{1-2/\gamma}\bigg\}^{1/2} \to 0$$

and

$$E(|I_2|) \le \frac{1}{(2r)^{1/2}T}\sum_{k=1}^{T}[E\{E^2_{T,k-1}(\|Z_{T,k}\|_2^2 - r)\}]^{1/2} \le Cr^{1/2}h\alpha_X(b+1)^{1/2-1/\gamma} \to 0.$$

Then, we complete the proof of this lemma. □

Lemma 15. *Under conditions (A.1)(i) and (A.2)(iii), assume that the eigenvalues of V_n are uniformly bounded away from zero and infinity. If $rh^{-1}\sum_{k=1}^{h} k\alpha_X(k)^{1-2/\gamma} = o(1)$, $r^2n^2h^2\alpha_X(b+1)^{1-2/\gamma} = o(1)$ and $rh^3n^{-1} = o(1)$, then $S_{T,T} \xrightarrow{d} N(0,1)$.*

Proof: We will use the martingale central limit theorem to show $S_{T,T} \xrightarrow{d} N(0,1)$. Note that

$$S_{T,T} = \sum_{k=1}^{T} D_{T,k} = \sum_{k=1}^{T}\{D_{T,k} - E_{T,k-1}(D_{T,k})\} + \sum_{k=1}^{T} E_{T,k-1}(D_{T,k}).$$

The first part on the right hand of above equation are the sum of a sequence of martingale difference with respect to $\{\mathscr{G}_{T,k}\}_{k=0}^{T}$. From Lemma 14, $S_{T,T} = \sum_{k=1}^{T}\{D_{T,k} - E_{T,k-1}(D_{T,k})\} + o_p(1)$.

By the martingale central limit theorem (Billingsley, 1995), in order to show the conclusion, it is sufficient to show that, letting $\sigma^2_{T,k} = E_{T,k-1}[\{D_{T,k} - E_{T,k-1}(D_{T,k})\}^2]$, as $T \to \infty$,

$$V_{T,T} := \sum_{k=1}^{T}\sigma^2_{T,k} \xrightarrow{p} 1 \quad \text{and} \quad \sum_{k=1}^{T} E\{D_{T,k} - E_{T,k-1}(D_{T,k})\}^4 \to 0.$$

For the first part,

$$\begin{aligned}V_{T,T} = &\frac{2}{rT^2}\sum_{k=1}^{T}(G'_{T,k-1}[E_{T,k-1}(Z_{T,k}Z'_{T,k}) - \{E_{T,k-1}(Z_{T,k})\}\{E_{T,k-1}(Z'_{T,k})\}]G_{T,k-1})\\ &+ \frac{2}{rT^2}\sum_{k=1}^{T} G'_{T,k-1}[E_{T,k-1}\{Z_{T,k}(\|Z_{T,k}\|_2^2 - r)\} - E_{T,k-1}(Z_{T,k})\cdot E_{T,k-1}(\|Z_{T,k}\|_2^2 - r)]\\ &+ \frac{1}{2rT^2}\sum_{k=1}^{T}[E_{T,k-1}\{(\|Z_{T,k}\|_2^2 - r)^2\} - E^2_{T,k-1}(\|Z_{T,k}\|_2^2 - r)]\\ =:& I_1 + I_2 + I_3.\end{aligned}$$

We will show that $I_1 \xrightarrow{p} 1, I_2 \xrightarrow{p} 0$ and $I_3 \xrightarrow{p} 0$.

Note that $0 \le I_3 \le (2r)^{-1}T^{-2}\sum_{k=1}^{T} E_{T,k-1}\{(\|Z_{T,k}\|_2^2 - r)^2\}$ and $E\{(\|Z_{T,k}\|_2^2 - r)^2\} \le Cr^2h^2$, then $I_3 \xrightarrow{p} 0$. Using Cauchy-Schwarz, Triangle and Jensen's inequalities,

$$
\begin{aligned}
&|E_{T,k-1}\{G'_{T,k-1}Z_k(\|Z_{T,k}\|_2^2 - r)\}| \\
&\quad \le \{E_{T,k-1}(G'_{T,k-1}Z_{T,k}Z'_{T,k}G_{T,k-1})\}^{1/2}[E_{T,k-1}\{(\|Z_{T,k}\|_2^2 - r)^2\}]^{1/2} \\
&\quad \le \{\|G_{T,k-1}\|_2^2 + \|G_{T,k-1}\|_2^2 \cdot \|E_{T,k-1}(Z_{T,k}Z'_{T,k}) - I_r\|_F\}^{1/2}[E_{T,k-1}\{(\|Z_{T,k}\|_2^2 - r)^2\}]^{1/2} \\
&\quad \le C\|G_{T,k-1}\|_2[E_{T,k-1}\{(\|Z_{T,k}\|_2^2 - r)^2\}]^{1/2}\{1 + \|E_{T,k-1}(Z_{T,k}Z'_{T,k}) - I_r\|_F^{1/2}\}.
\end{aligned}
$$

Then, by Cauchy-Schwarz inequality and Lemmas 11 and 12,

$$
\begin{aligned}
&E[|E_{T,k-1}\{G'_{T,k-1}Z_k(\|Z_{T,k}\|_2^2 - r)\}|] \\
&\qquad \le C\{E(\|G_{T,k-1}\|_2^2)\}^{1/2}[E\{(\|Z_{T,k}\|_2^2 - r)^2\}]^{1/2} \\
&\qquad\quad + C\{E(\|G_{T,k-1}\|_2^2)\}^{1/2}[E\{(\|Z_{T,k}\|_2^2 - r)^4\}]^{1/4}[E\{\|E_{T,k-1}(Z_{T,k}Z'_{T,k}) - I_r\|_F^2\}]^{1/4} \\
&\qquad \le Cr^{3/2}hk^{1/2} + Cr^2h^{3/2}k^{1/2}\alpha_X(b+1)^{1/4-1/(2\gamma)}.
\end{aligned}
$$

Hence, $r^{-1}T^{-2}\sum_{k=1}^{T} E[|G'_{T,k-1}E_{T,k-1}\{Z_{T,k}(\|Z_{T,k}\|_2^2 - r)\}|] \to 0$. By the same argument, we can obtain $r^{-1}T^{-2}\sum_{k=1}^{T} E\{|E_{T,k-1}(G'_{T,k-1}Z_{T,k}) \cdot E_{T,k-1}(\|Z_{T,k}\|_2^2 - r)|\} \to 0$. Thus, $I_2 \xrightarrow{p} 0$. Note that

$$
\begin{aligned}
I_1 = &\frac{2}{rT^2}\sum_{k=1}^{T}\|G_{T,k-1}\|_2^2 \\
&+ \frac{2}{rT^2}\sum_{k=1}^{T} G'_{T,k-1}\{E_{T,k-1}(Z_{T,k}Z'_{T,k}) - E_{T,k-1}(Z_{T,k}) \cdot E_{T,k-1}(Z'_{T,k}) - I_r\}G_{T,k-1}.
\end{aligned}
$$

By Triangle inequality,

$$
\begin{aligned}
&\left|\frac{2}{rT^2}\sum_{k=1}^{T} G'_{T,k-1}\{E_{T,k-1}(Z_{T,k}Z'_{T,k}) - E_{T,k-1}(Z_{T,k}) \cdot E_{T,k-1}(Z'_{T,k}) - I_r\}G_{T,k-1}\right| \\
&\qquad \le \frac{2}{rT^2}\sum_{k=1}^{T}\|G_{T,k-1}\|_2^2\|E_{T,k-1}(Z_{T,k}Z'_{T,k}) - I_r\|_F + \frac{2}{rT^2}\sum_{k=1}^{T}\{G'_{T,k-1}E_{T,k-1}(Z_{T,k})\}^2.
\end{aligned}
$$

By Cauchy-Schwarz inequality and Lemma 12,

$$
\begin{aligned}
&E\{\|G_{T,k-1}\|_2^2\|E_{T,k-1}(Z_{T,k}Z'_{T,k}) - I_r\|_F\} \\
&\le \{E(\|G_{T,k-1}\|_2^4)\}^{1/2}[E\{\|E_{T,k-1}(Z_{T,k}Z'_{T,k}) - I_r\|_F^2\}]^{1/2} \\
&\le Cr^2h^2k^2\alpha_X(b+1)^{1/2-1/\gamma}.
\end{aligned}
$$

Then,

$$
E\left\{\frac{2}{rT^2}\sum_{k=1}^{T}\|G_{T,k-1}\|_2^2\|E_{T,k-1}(Z_{T,k}Z'_{T,k}) - E_r\|_F\right\} \le Crnh\alpha_X(b+1)^{1/2-1/\gamma} \to 0.
$$

On the other hand, by Lemma 12,

$$
E\left[\frac{2}{rT^2}\sum_{k=1}^{T}\{G'_{T,k-1}E_{T,k-1}(Z_{T,k})\}^2\right] \le Crnh\alpha_X(b+1)^{1/2-1/\gamma} \to 0.
$$

Then, $I_1 = 2r^{-1}T^{-2}\sum_{k=1}^T \|G_{T,k-1}\|_2^2 + o_p(1)$.

From Lemma 13,

$$\frac{2}{rT^2}\sum_{k=1}^T \|G_{T,k-1}\|_2^2 = \frac{2}{rT^2}\sum_{i=1}^T (T-i)\|Z_{T,i}\|_2^2 + \frac{4}{rT^2}\sum_{j=2}^T (T-j)G'_{T,j-1}Z_{T,j}$$
$$= \frac{2}{rT^2}\sum_{i=1}^T (T-i)\|Z_{T,i}\|_2^2 + o_p(1).$$

In order to prove $I_1 \xrightarrow{p} 1$, it is only need to show $2r^{-1}T^{-2}\sum_{i=1}^T (T-i)(\|Z_{T,i}\|_2^2 - r) \xrightarrow{p} 0$. Note that

$$E\bigg\{\frac{2}{rT^2}\sum_{i=1}^T (T-i)(\|Z_{T,i}\|_2^2 - r)\bigg\} = 0,$$

it is sufficient to show

$$\frac{4}{r^2T^4}\bigg[\sum_{i=1}^T (T-i)^2 E\{(\|Z_{T,i}\|_2^2 - r)^2\} + \sum_{i\neq j}(T-i)(T-j)E\{(\|Z_{T,i}\|_2^2 - r)(\|Z_{T,j}\|_2^2 - r)\}\bigg] \to 0,$$

which can be derived from Lemma 11. Hence, $I_1 \xrightarrow{p} 1$.

For the second part, we only need to prove $\sum_{k=1}^T E(D_{T,k}^4) \to 0$. Note that

$$D_{T,k}^4 \leq Cr^{-2}T^{-4}\{(G'_{T,k-1}Z_{T,k})^4 + (\|Z_{T,k}\|_2^2 - r)^4\}$$

and

$$(G'_{T,k-1}Z_{T,k})^4 = \sum_{i_1,\ldots,i_4=1}^{k-1}\sum_{j_1,\ldots,j_4=1}^{r} Z_{T,i_1,j_1}Z_{T,i_2,j_2}Z_{T,i_3,j_3}Z_{T,i_4,j_4}Z_{T,k,j_1}Z_{T,k,j_2}Z_{T,k,j_3}Z_{T,k,j_4},$$

where $Z_{T,i,j}$ denotes the jth component of $Z_{T,i}$. By the same way of the Lemma 15 in Francq and Zakoïan (2007), $r^{-2}h^{-2}E\{(G'_{T,k-1}Z_{T,k})^4\} \leq Ck^2$. Then, $\sum_{k=1}^T E(D_{T,k}^4) \leq Ch^2T^{-1} + Cr^2h^4T^{-3} \to 0$. Hence, we complete the proof. $\square$

Lemma 16. *Under conditions (A.1)(i) and (A.2)(iii), assume that the eigenvalues of V_n are uniformly bounded away from zero and infinity. Then $(2r)^{-1/2}\{n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0) - r\} \xrightarrow{d} N(0,1)$ provided that $r^{3/2}h^{-1}\sum_{k=1}^h k\alpha_X(k)^{1-2/\gamma} = o(1)$, $rbh^{-1} = o(1)$, $r^2n^2h^2\alpha_X(b+1)^{1-2/\gamma} = o(1)$ and $rh^3n^{-1} = o(1)$.*

Proof: Note that

$$(2r)^{-1/2}\{n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0) - r\}$$
$$= \frac{n}{Th}S_{T,T} + O_p\bigg\{r^{3/2}h^{-1}\sum_{k=1}^h k\alpha_X(k)^{1-2/\gamma}\bigg\} + O_p(r^{1/2}b^{1/2}h^{-1/2}).$$

Then, by Lemma 15, we have $(2r)^{-1/2}\{n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0) - r\} \xrightarrow{d} N(0,1)$. $\square$

Proof of Theorem 3

Let $\widehat{\lambda}(\boldsymbol{\theta}_0) = \arg\max_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta}_0)}\sum_{q=1}^{Q}\rho(\lambda'\phi_q(\boldsymbol{\theta}_0))$. From Lemma 2, $\|\bar{\phi}(\boldsymbol{\theta}_0)-\bar{g}(\boldsymbol{\theta}_0)\|_2 = O_p(r^{1/2}Mn^{-1})$. Hence, $\|\bar{\phi}(\boldsymbol{\theta}_0)\|_2 = O_p(r^{1/2}n^{-1/2})$. Then, by Lemma 7, $\|\widehat{\lambda}(\boldsymbol{\theta}_0)\|_2 = O_p(r^{1/2}Mn^{-1/2})$. Meanwhile, $\sup_{1\le q\le Q}|\widehat{\lambda}(\boldsymbol{\theta}_0)'\phi_q(\boldsymbol{\theta}_0)| = o_p(1)$. Expanding $\max_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta}_0)}\sum_{q=1}^{Q}\rho(\lambda'\phi_q(\boldsymbol{\theta}_0))$ around $\lambda=\mathbf{0}$,

$$\max_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta}_0)}\sum_{q=1}^{Q}\rho(\lambda'\phi_q(\boldsymbol{\theta}_0)) = \sum_{q=1}^{Q}\bigg[\rho(0)+\rho_v(0)\widehat{\lambda}(\boldsymbol{\theta}_0)'\phi_q(\boldsymbol{\theta}_0)+\frac{1}{2}\rho_{vv}(\dot{\lambda}'\phi_q(\boldsymbol{\theta}_0))\{\widehat{\lambda}(\boldsymbol{\theta}_0)'\phi_q(\boldsymbol{\theta}_0)\}^2\bigg]$$

where $\dot{\lambda}$ lies on the line joining $\widehat{\lambda}(\boldsymbol{\theta}_0)$ and $\mathbf{0}$. On the other hand, from $\nabla_\lambda\widehat{S}_n(\boldsymbol{\theta}_0,\widehat{\lambda}(\boldsymbol{\theta}_0))=\mathbf{0}$, we have

$$\widehat{\lambda}(\boldsymbol{\theta}_0) = -\bigg\{\frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}(\ddot{\lambda}'\phi_q(\boldsymbol{\theta}_0))\phi_q(\boldsymbol{\theta}_0)\phi_q(\boldsymbol{\theta}_0)'\bigg\}^{-1}\bigg\{\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\phi_q(\boldsymbol{\theta}_0)\bigg\}$$

for some $\ddot{\lambda}$ lies on the line joining $\widehat{\lambda}(\boldsymbol{\theta}_0)$ and $\mathbf{0}$. Hence,

$$\max_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta}_0)}\sum_{q=1}^{Q}\rho(\lambda'\phi_q(\boldsymbol{\theta}_0)) = Q\rho(0)-Q\rho_v^2(0)\bar{\phi}(\boldsymbol{\theta}_0)'\ddot{\Omega}^{-1}\bar{\phi}(\boldsymbol{\theta}_0)+\frac{1}{2}Q\rho_v^2(0)\bar{\phi}(\boldsymbol{\theta}_0)'\ddot{\Omega}^{-1}\dot{\Omega}\ddot{\Omega}^{-1}\bar{\phi}(\boldsymbol{\theta}_0)$$

where $\dot{\Omega} = Q^{-1}\sum_{q=1}^{Q}\rho_{vv}(\dot{\lambda}\phi_q(\boldsymbol{\theta}_0))\phi_q(\boldsymbol{\theta}_0)\phi_q(\boldsymbol{\theta}_0)'$ and $\ddot{\Omega} = Q^{-1}\sum_{q=1}^{Q}\rho_{vv}(\ddot{\lambda}\phi_q(\boldsymbol{\theta}_0))\phi_q(\boldsymbol{\theta}_0)\phi_q(\boldsymbol{\theta}_0)'$. Then, the generalized empirical likelihood ratio can be written as

$$\begin{aligned}w_n(\boldsymbol{\theta}_0) &= 2Q\rho_{vv}(0)\bar{\phi}(\boldsymbol{\theta}_0)'\ddot{\Omega}^{-1}\bar{\phi}(\boldsymbol{\theta}_0)-Q\rho_{vv}(0)\bar{\phi}(\boldsymbol{\theta}_0)'\ddot{\Omega}^{-1}\dot{\Omega}\ddot{\Omega}^{-1}\bar{\phi}(\boldsymbol{\theta}_0)\\ &= Q\bar{\phi}(\boldsymbol{\theta}_0)'\widehat{\Omega}^{-1}(\boldsymbol{\theta}_0)\bar{\phi}(\boldsymbol{\theta}_0)+Q\bar{\phi}(\boldsymbol{\theta}_0)'\{2\rho_{vv}(0)\ddot{\Omega}^{-1}-\widehat{\Omega}^{-1}(\boldsymbol{\theta}_0)-\rho_{vv}(0)\ddot{\Omega}^{-1}\dot{\Omega}\ddot{\Omega}^{-1}\}\bar{\phi}(\boldsymbol{\theta}_0).\end{aligned}$$

By the same argument of Lemma 9,

$$\|\dot{\Omega}-\rho_{vv}(0)\widehat{\Omega}(\boldsymbol{\theta}_0)\|_2 = O_p(rM^{-1/\gamma}n^{1/\gamma-1/2}) = \|\ddot{\Omega}-\rho_{vv}(0)\widehat{\Omega}(\boldsymbol{\theta}_0)\|_2.$$

From Lemmas 3 and $\|M\Omega(\boldsymbol{\theta}_0)-V_n\|_2 = O(rM^{-1})$, we know the eigenvalues of $M\widehat{\Omega}(\boldsymbol{\theta}_0)$ are uniformly bounded away from zero and infinity. Hence,

$$w_n(\boldsymbol{\theta}_0) = Q\bar{\phi}(\boldsymbol{\theta}_0)'\widehat{\Omega}^{-1}(\boldsymbol{\theta}_0)\bar{\phi}(\boldsymbol{\theta}_0)+O_p(r^2M^{1-1/\gamma}n^{1/\gamma-1/2}).$$

By Lemmas 2 and 3, we have

$$\begin{aligned}(2r)^{-1/2}\{w_n(\boldsymbol{\theta}_0)-r\} &= (2r)^{-1/2}\{n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)-r\}+O_p(r^{5/2}M^{2-2/\gamma}n^{2/\gamma-1})\\ &\quad+O_p(r^{3/2}M^{1-1/\gamma}n^{1/\gamma-1/2})+O_p(r^{3/2}M^{3/2}n^{-1/2})\\ &\quad+O_p\bigg\{r^{3/2}M^{-1}\sum_{k=1}^{M}k\alpha_X(k)^{1-2/\gamma}\bigg\}.\end{aligned}\tag{29}$$

The key step is to show $(2r)^{-1/2}\{n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)-r\}\xrightarrow{d}N(0,1)$. In the independent case, the requirements in Lemma 16 can be simplified as $rbh^{-1}=o(1)$ and $rh^3n^{-1}=o(1)$. We can pick $b=0$ and $h=1$, then $r=o(n)$. In this case, we can regard $\eta=\infty$. In the dependent case with $\eta<\infty$, suppose $b\asymp n^{\kappa_1}$ and $h\asymp n^{\kappa_2}$, where $0<\kappa_1<\kappa_2<1$. Note that (A.1)'(i), the requirements in Lemma 16 turn to

$$r=o(n^{2\kappa_2/3}),\quad r=o(n^{\kappa_2-\kappa_1}),\quad r=o(n^{(\eta\kappa_1-2-2\kappa_2)/2})\ \text{ and }\ r=o(n^{1-3\kappa_2}),$$

where $\eta\kappa_1-2-2\kappa_2>0$ and $1-3\kappa_2>0$. In the following, we will consider the selection of (κ_1,κ_2) to satisfy these inequalities. From $2\kappa_2+2-\eta\kappa_1<0$, $3\kappa_2-1<0$ and $\kappa_1<\kappa_2$, we can get $\frac{2\kappa_2+2}{\eta}<\kappa_1<\kappa_2<\frac{1}{3}$. In order to guarantee there exists the solution for above inequalities in $(0,1)^2$, it is necessary to require $\eta>8$. If $8<\eta<\infty$,

$$\begin{aligned}\xi &:= \sup_{\substack{\frac{2}{\eta-2}<\kappa_2<\frac{1}{3}\\ \frac{2\kappa_2+2}{\eta}<\kappa_1<\kappa_2}} \min\left(\frac{2\kappa_2}{3},\kappa_2-\kappa_1,\frac{\eta\kappa_1-2-2\kappa_2}{2},1-3\kappa_2\right)\\ &= \sup_{\frac{2}{\eta-2}<\kappa_2<\frac{1}{3}} \min\left\{\frac{2\kappa_2}{3},\frac{(\eta-2)\kappa_2-2}{\eta+2},1-3\kappa_2\right\}\\ &= \frac{\eta-8}{4\eta+4}1_{(8<\eta<32)}+\frac{2}{11}1_{(32\le\eta<\infty)}.\end{aligned}$$

In the dependent case with $\eta=\infty$ where X_t is exponentially strong mixing. The requirements in Lemma 16 turn to $r^{3/2}h^{-1}=o(1)$, $rbh^{-1}=o(1)$ and $rh^3n^{-1}=o(1)$. Then,

$$r=o(n^{2\kappa_2/3}),\quad r=o(n^{\kappa_2-\kappa_1})\ \text{ and }\ r=o(n^{1-3\kappa_2}).$$

In this setting,

$$\xi := \sup_{\substack{0<\kappa_2<\frac{1}{3}\\ 0<\kappa_1<\kappa_2}} \min\left(\frac{2\kappa_2}{3},\kappa_2-\kappa_1,1-3\kappa_2\right)=\frac{2}{11}.$$

Define

$$\xi=\frac{\eta-8}{4\eta+4}1_{\{8<\eta<32\}}+\frac{2}{11}1_{\{32\le\eta\le\infty\}}+1_{\{\text{indenpendent data}\}}.$$

Hence, if $r=o(n^{\xi})$, then $(2r)^{-1/2}\{n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)-r\}\xrightarrow{d}N(0,1)$. If (21) holds, the other terms in (29) are $o_p(1)$. We complete the proof of Theorem 3. □

Proof of Theorem 4

In order to establish Theorem 4, we need the following lemma.

Lemma 17. *For any $\widetilde{\boldsymbol{\theta}}\in\Theta$ and $r\times r$ matrix $\widehat{V}_n$ such that $\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2=O_p(p^{1/2}n^{-1/2})$ and $\|\widehat{V}_n-V_n\|_2=o_p(r^{-1/2})$, if $\|\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})-E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}\|_2=o_p(p^{-1/2})$ for any $\dot{\boldsymbol{\theta}}$ with $\|\dot{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2\le\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2$, and the eigenvalues of $[E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}]'[E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}]$ and V_n are uniformly bounded away from zero and infinity, then $(2r)^{-1/2}\{n\bar{g}(\widetilde{\boldsymbol{\theta}})'\widehat{V}_n^{-1}\bar{g}(\widetilde{\boldsymbol{\theta}})-n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)\}\xrightarrow{p}0$ provided that $p=o(r^{1/2})$.*

Proof: Note that

$$\begin{aligned}&(2r)^{-1/2}|n\bar{g}(\widetilde{\boldsymbol{\theta}})'\widehat{V}_n^{-1}\bar{g}(\widetilde{\boldsymbol{\theta}})-n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)|\\ &=(2r)^{-1/2}|n\bar{g}(\widetilde{\boldsymbol{\theta}})'\widehat{V}_n^{-1}\bar{g}(\widetilde{\boldsymbol{\theta}})-n\bar{g}(\boldsymbol{\theta}_0)'\widehat{V}_n^{-1}\bar{g}(\boldsymbol{\theta}_0)|+(2r)^{-1}|n\bar{g}(\boldsymbol{\theta}_0)'\widehat{V}_n^{-1}\bar{g}(\boldsymbol{\theta}_0)-n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)|\\ &=: I_1+I_2.\end{aligned}$$

We only need to show $I_1\xrightarrow{p}0$ and $I_2\xrightarrow{p}0$.

For I_1, by Taylor expansion, $\bar{g}(\widetilde{\boldsymbol{\theta}})=\bar{g}(\boldsymbol{\theta}_0)+\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})\cdot(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$. Then,

$$I_1\le(2r)^{-1/2}|2n(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\{\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})\}'\widehat{V}_n^{-1}\bar{g}(\boldsymbol{\theta}_0)|+(2r)^{-1/2}|n(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\{\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})\}'\widehat{V}_n^{-1}\{\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})\}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)|.$$

As the eigenvalues of V_n are uniformly bounded away from zero and infinity, and $\|\widehat{V}_n - V_n\|_2 = o_p(r^{-1/2})$, then the eigenvalues of $\widehat{V}_n$ are uniformly bounded away from zero and infinity w.p.a.1. Hence,

$$\begin{aligned}
&\|\{\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})\}'\widehat{V}_n^{-1} - [E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}]'V_n^{-1}\|_2 \\
&\le \|(\{\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})\}' - [E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}])\widehat{V}_n^{-1}\|_2 + \|[E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}]'(\widehat{V}_n^{-1} - V_n^{-1})\|_2 \\
&= o_p(p^{-1/2}) + o_p(r^{-1/2}) = o_p(p^{-1/2}).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
E(\|[E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}]'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)\|_2^2) &= E\{\mathrm{tr}(V_n^{-1}[E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}][E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}]'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0)\bar{g}(\boldsymbol{\theta}_0))'\} \\
&= n^{-1}\mathrm{tr}([E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}]'V_n^{-1}[E\{\nabla_{\boldsymbol{\theta}}g(\boldsymbol{\theta}_0)\}]) \\
&\le Cpn^{-1}.
\end{aligned}$$

Then,

$$\|\{\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}})\}'\widehat{V}_n^{-1}\bar{g}(\boldsymbol{\theta}_0)\|_2 = O_p(p^{1/2}n^{-1/2}) + o_p(r^{1/2}p^{-1/2}n^{-1/2}).$$

Therefore,

$$I_1 \le O_p(pr^{-1/2}) + o_p(1) \xrightarrow{p} 0$$

provided that $p = o(r^{1/2})$.

For I_2,

$$I_2 = (2r)^{-1/2}|n\bar{g}(\boldsymbol{\theta}_0)'(\widehat{V}_n^{-1} - V_n^{-1})\bar{g}(\boldsymbol{\theta}_0)| = O(r^{-1/2}n)o_p(r^{-1/2})O_p(rn^{-1}) = o_p(1).$$

Hence, we complete the proof of this lemma. $\square$

Remark: This lemma is similar to the Lemma 6.1 of Donald, Imbens and Newey (2003). However, we work on the operator-norm in establishing the consistency results, whereas Donald, Imbens and Newey (2003) employed the Frobenius-norm. The matrix $\widehat{V}_n$ and $\widetilde{\boldsymbol{\theta}}$ are the consistency estimators of V_n and $\boldsymbol{\theta}_0$ respectively.

Here, we begin to establish Theorem 4. From Proposition 1, we know $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(p^{1/2}n^{-1/2})$. By the same argument of the proof of Theorem 3, we have

$$w_n(\widehat{\boldsymbol{\theta}}_n) = Q\bar{\phi}(\widehat{\boldsymbol{\theta}}_n)'\widehat{\Omega}^{-1}(\widehat{\boldsymbol{\theta}}_n)\bar{\phi}(\widehat{\boldsymbol{\theta}}_n) + O_p(r^2M^{1-1/\gamma}n^{1/\gamma-1/2}).$$

Note Lemma 2,

$$w_n(\widehat{\boldsymbol{\theta}}_n) = n\bar{g}(\widehat{\boldsymbol{\theta}}_n)'\{M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\}^{-1}\bar{g}(\widehat{\boldsymbol{\theta}}_n) + O_p(r^2M^{1-1/\gamma}n^{1/\gamma-1/2}) + O_p(rMn^{-1/2}) + O_p(rM^2n^{-1}).$$

By Lemmas 3 and 6, it yields that

$$\begin{aligned}
\big\|M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n) - V_n\big\|_2 &= O_p(r^{1/2}pM^{1/2}n^{-1/2} + rM^{3/2}n^{-1/2}) + O_p\bigg\{rM^{-1}\sum_{k=1}^{M}k\alpha_X(k)^{1-2/\gamma}\bigg\} \\
&= o_p(r^{-1/2}).
\end{aligned}$$

Noting Lemma 10, for any $\dot{\boldsymbol{\theta}}$ such that $\|\dot{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 \le \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = O_p(p^{1/2}n^{-1/2})$,

$$\|\nabla_{\boldsymbol{\theta}}\bar{g}(\dot{\boldsymbol{\theta}}) - E\{\nabla_{\boldsymbol{\theta}}g_t(\boldsymbol{\theta}_0)\}\|_F = O_p(r^{1/2}p^{3/2}n^{-1/2}) = o_p(p^{-1/2}).$$

By Lemma 17, we can get $n\bar{g}(\widehat{\boldsymbol{\theta}}_n)'\{M\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_n)\}^{-1}\bar{g}(\widehat{\boldsymbol{\theta}}_n) - n\bar{g}(\boldsymbol{\theta}_0)'V_n^{-1}\bar{g}(\boldsymbol{\theta}_0) = o_p(r^{1/2})$. Then, by Lemma 16, we complete the proof of Theorem 4. $\square$

Proof of Theorem 5

We only need to prove that for some $c > 1$, $P\{w_n(\widehat{\boldsymbol{\theta}}_n) > cr\} \to 1$. To prove this, we use the technique for the proof of Theorem 1 in Chang, Tang and Wu (2013). Let

$$\widetilde{\lambda} = \frac{-\rho_v(0)}{2\rho_{vv}(0)Q^\omega} \frac{e}{\max_{1\le q\le Q} \|\phi_q(\widehat{\boldsymbol{\theta}}_n)\|_2}$$

where e is a r-dimensional vector with unit L_2-norm, and $\omega > 0$ will be determined later. Then, $\widetilde{\lambda} \in \widehat{\Lambda}_n(\widehat{\boldsymbol{\theta}}_n)$ when Q is sufficiently large. Note that $\rho_{vv}(0) < 0$, by Taylor expansion, we have

$$\begin{aligned} w_n(\widehat{\boldsymbol{\theta}}_n) &= \frac{2\rho_{vv}(0)}{\rho_v^2(0)} \bigg\{ Q\rho(0) - \max_{\lambda\in\widehat{\Lambda}_n(\widehat{\boldsymbol{\theta}}_n)} \sum_{q=1}^Q \rho(\lambda'\phi_q(\widehat{\boldsymbol{\theta}}_n)) \bigg\} \\ &\ge \frac{1}{Q^\omega} \sum_{q=1}^Q \frac{e'\phi_q(\widehat{\boldsymbol{\theta}}_n)}{\max_{1\le q\le Q} \|\phi_q(\widehat{\boldsymbol{\theta}}_n)\|_2} \\ &\qquad\qquad - \frac{1}{4Q^{2\omega}\rho_{vv}(0)} \sum_{q=1}^Q \frac{\rho_{vv}(\dot{\lambda}\phi_q(\widehat{\boldsymbol{\theta}}_n))e'\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'e}{\max_{1\le q\le Q} \|\phi_q(\widehat{\boldsymbol{\theta}}_n)\|_2^2} \end{aligned}$$

where $\dot{\lambda}$ lies on the jointing line between $\widetilde{\lambda}$ and $\mathbf{0}$. By the definition of $\widetilde{\lambda}$, we have

$$\frac{1}{4Q^{2\omega}\rho_{vv}(0)} \sum_{q=1}^Q \frac{\rho_{vv}(\dot{\lambda}\phi_q(\widehat{\boldsymbol{\theta}}_n))e'\phi_q(\widehat{\boldsymbol{\theta}}_n)\phi_q(\widehat{\boldsymbol{\theta}}_n)'e}{\max_{1\le q\le Q} \|\phi_q(\widehat{\boldsymbol{\theta}}_n)\|_2^2} \le \frac{1}{2}Q^{1-2\omega} \quad \text{w.p.a. 1.}$$

Hence, for any $c > 1$,

$$P\{w_n(\widehat{\boldsymbol{\theta}}_n) \le cr\} \le P\bigg\{ \sum_{q=1}^Q \frac{e'\phi_q(\widehat{\boldsymbol{\theta}}_n)}{\max_{1\le q\le Q} \|\phi_q(\widehat{\boldsymbol{\theta}}_n)\|_2} \le crQ^\omega + \frac{1}{2}Q^{1-\omega} \bigg\} + o(1).$$

From (A.2)(ii), we have $\|\phi_q(\widehat{\boldsymbol{\theta}}_n)\|_2 \le r^{1/2}M^{-1}\sum_{t\in B_q} B_n(X_t)$. Then, by Markov inequality,

$$P\bigg\{ \max_{1\le q\le Q} \|\phi_q(\widehat{\boldsymbol{\theta}}_n)\|_2 > (cK)^{-1}r^{1/2}Q^{1/\gamma}(\log Q)^{\epsilon/2} \bigg\} \to 0$$

for each fixed $K > 0$, which implies that

$$P\{w_n(\widehat{\boldsymbol{\theta}}_n) \le cr\} \le P\bigg\{ \sum_{q=1}^Q e'\phi_q(\widehat{\boldsymbol{\theta}}_n) \le K^{-1}(rQ^\omega + Q^{1-\omega})r^{1/2}Q^{1/\gamma}(\log Q)^{\epsilon/2} \bigg\} + o(1).$$

Let $rQ^\omega = Q^{1-\omega}$, i.e., $Q^\omega = Q^{1/2}r^{-1/2}$, then

$$P\{w_n(\widehat{\boldsymbol{\theta}}_n) \le cr\} \le P\bigg\{ \sum_{q=1}^Q e'\phi_q(\widehat{\boldsymbol{\theta}}_n) \le 2K^{-1}rQ^{1/\gamma+1/2}(\log Q)^{\epsilon/2} \bigg\} + o(1).$$

On the other hand, by Lemma 2 and (A.1)(iv),

$$\sum_{q=1}^Q e'\phi_q(\widehat{\boldsymbol{\theta}}_n) = Qe'\bar{g}(\widehat{\boldsymbol{\theta}}_n) + O_p(r^{1/2}) = Qe'E\{g_t(\widehat{\boldsymbol{\theta}}_n)\} + O_p(r^{1/2}) + o_p\{Q\Delta_1(r,p)\}.$$

Select $e = E\{g_t(\widehat{\boldsymbol{\theta}}_n)\}/\|E\{g_t(\widehat{\boldsymbol{\theta}}_n)\}\|_2$. Then,

$$\begin{aligned}
&P\{w_n(\widehat{\boldsymbol{\theta}}_n) \leq cr\} \\
&\leq P\big[\|E\{g_t(\widehat{\boldsymbol{\theta}}_n)\}\|_2 \leq 2K^{-1}rQ^{1/\gamma-1/2}(\log Q)^{\epsilon/2} + O_p(r^{1/2}Q^{-1}) + o_p\{\Delta_1(r,p)\}\big] + o(1) \\
&\leq P\big[\varsigma \leq 2K^{-1}rQ^{1/\gamma-1/2}(\log Q)^{\epsilon/2} + O_p(r^{1/2}Q^{-1}) + o_p\{\Delta_1(r,p)\}\big] + o(1).
\end{aligned}$$

As $r^2M^{1-2/\gamma}n^{2/\gamma-1}(\log n)^{\epsilon}\varsigma^{-2} = O(1)$, $r^{1/2}Mn^{-1}\varsigma^{-1} = o(1)$ and $\Delta_1(r,p)\varsigma^{-1} = O(1)$, we can choose sufficiently large K to guarantee

$$P\big[\varsigma \leq 2K^{-1}rQ^{1/\gamma-1/2}(\log Q)^{\epsilon/2} + O_p(r^{1/2}Q^{-1}) + o_p\{\Delta_1(r,p)\}\big] \to 0,$$

which leads to $P\{w_n(\widehat{\boldsymbol{\theta}}_n) \leq cr\} \to 0$ for any $c > 1$. Hence, we complete the proof. □

Proof of Theorem 6

Let

$$\widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta}, \lambda) = \frac{1}{Q}\sum_{q=1}^{Q}\rho(\lambda'\phi_q(\boldsymbol{\theta})) + \sum_{j=1}^{p}p_\tau(|\theta_j|) \;\text{ for any } \boldsymbol{\theta} \in \Theta \text{ and } \lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta}).$$

Then,

$$\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})} = \arg\min_{\boldsymbol{\theta}\in\Theta}\max_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta})}\widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta}, \lambda) \;\text{ and }\; \widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta}\max_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta})}\widehat{S}_n(\boldsymbol{\theta}, \lambda).$$

The following lemma will be used to construct Theorem 6.

Lemma 18. *Under conditions (A.1), (A.2) and (A.5), assume that the eigenvalues of V_M are uniformly bounded away from zero and infinity. If (13) holds, $r^2pM^2n^{-1} = o(1)$ and $s\tau r^{-1}M^{-1}n = O(1)$, then $\|\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})} - \boldsymbol{\theta}_0\|_2 = O_p(r^{1/2}n^{-1/2})$.*

Proof: Choose $\delta_n = o(r^{-1/2}Q^{-1/\gamma})$ and $r^{1/2}Mn^{-1/2} = o(\delta_n)$. Let $\bar{\lambda} = \mathrm{sign}\{\rho_v(0)\}\delta_n\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})/\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\|_2$, then $\bar{\lambda} \in \Lambda_n$ where Λ_n is defined in Lemma 5. By Taylor expansion, Lemmas 4 and 5, noting $\rho_{vv}(0) < 0$, we have

$$\begin{aligned}
\widehat{S}_n(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}, \bar{\lambda}) &= \rho(0) + \rho_v(0)\bar{\lambda}'\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}) + \frac{1}{2}\bar{\lambda}'\bigg\{\frac{1}{Q}\sum_{q=1}^{Q}\rho_{vv}(\dot{\lambda}'\phi_q(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}))\phi_q(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\phi_q(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})'\bigg\}\bar{\lambda} \\
&\geq \rho(0) + |\rho_v(0)|\delta_n\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\|_2 - C\|\bar{\lambda}\|_2^2 \cdot O_p(1).
\end{aligned}$$

On the other hand,

$$\widehat{S}_n^{(\mathrm{pe})}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}, \bar{\lambda}) \leq \sup_{\lambda\in\widehat{\Lambda}_n(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})}\widehat{S}_n^{(\mathrm{pe})}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}, \lambda) \leq \sup_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta}_0)}\widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta}_0, \lambda).$$

By Lemma 7 and (A.5), as $sr^{-1}\tau n = O(1)$,

$$\begin{aligned}
\sup_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta}_0)}\widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta}_0, \lambda) &= \sup_{\lambda\in\widehat{\Lambda}_n(\boldsymbol{\theta}_0)}\widehat{S}_n(\boldsymbol{\theta}_0, \lambda) + \sum_{j=1}^{p}p_\tau(|\theta_{0j}|) \\
&= \rho(0) + O_p(rMn^{-1} + s\tau) = \rho(0) + O_p(rMn^{-1}).
\end{aligned}$$

Note that $\widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta},\lambda) \geq \widehat{S}_n(\boldsymbol{\theta},\lambda)$ for any $\boldsymbol{\theta} \in \Theta$ and $\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta})$, it yields $\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\|_2 = O_p(\delta_n)$. Consider any $\varepsilon_n \to 0$ and let $\widetilde{\lambda} = \mathrm{sign}\{\rho_v(0)\}\varepsilon_n\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})$, then $\|\widetilde{\lambda}\|_2 = o_p(\delta_n)$. Using the same way above, we can obtain

$$|\rho_v(0)| \cdot \varepsilon_n \|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\|_2^2 - O_p(1)\cdot \varepsilon_n^2 \|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\|_2^2 = O_p(rMn^{-1}).$$

Then, $\varepsilon_n \|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\|_2^2 = O_p(rMn^{-1})$. Thus, $\|\bar{\phi}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\|_2 = O_p(r^{1/2}M^{1/2}n^{-1/2})$. Following the same arguments given in the proof of Theorem 1, we can obtain $\|\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})} - \boldsymbol{\theta}_0\|_2 = O_p(r^{1/2}n^{-1/2})$. □

Here, we begin to prove Theorem 6. $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}$ and its Lagrange multiplier $\widehat{\lambda}^{(\mathrm{pe})}$ satisfy the score equation

$$\mathbf{0} = \nabla_\lambda \widehat{S}_n^{(\mathrm{pe})}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}, \widehat{\lambda}^{(\mathrm{pe})}) = \nabla_\lambda \widehat{S}_n(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}, \widehat{\lambda}^{(\mathrm{pe})}).$$

By the implicit theorem (Theorem 9.28 of Rudin, 1976), for all $\boldsymbol{\theta}$ in a $\|\cdot\|_2$-neighborhood of $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}$, there is a $\widehat{\lambda}(\theta)$ such that $\nabla_\lambda \widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta}, \widehat{\lambda}(\boldsymbol{\theta})) = \mathbf{0}$ and $\widehat{\lambda}(\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$. By the concavity of $\widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta},\lambda)$ with respect to λ, $\widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta}, \widehat{\lambda}(\boldsymbol{\theta})) = \max_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta})} \widehat{S}_n(\boldsymbol{\theta},\lambda)$. From the envelope theorem,

$$\begin{aligned}\mathbf{0} &= \nabla_{\boldsymbol{\theta}} \widehat{S}_n^{(\mathrm{pe})}(\boldsymbol{\theta}, \widehat{\lambda}(\boldsymbol{\theta}))\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}} \\ &= \frac{1}{Q}\sum_{q=1}^{Q} \rho_v(\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})'\phi_q(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}))\big\{\nabla_{\boldsymbol{\theta}}\phi_q(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})})\big\}'\widehat{\lambda}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}) + \sum_{q=1}^{Q} \nabla_{\boldsymbol{\theta}} p_\tau(|\theta_j|)\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}}.\end{aligned} \tag{30}$$

For any $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 = O_p(r^{1/2}n^{-1/2})$ and $\|\bar{g}(\boldsymbol{\theta})\|_2 = O_p(r^{1/2}n^{-1/2})$, define

$$\mathbf{h}(\boldsymbol{\theta}) = \frac{1}{Q}\sum_{q=1}^{Q}\rho_v(\widehat{\lambda}(\boldsymbol{\theta})'\phi_q(\boldsymbol{\theta}))\big\{\nabla_{\boldsymbol{\theta}}\phi_q(\boldsymbol{\theta})\big\}'\widehat{\lambda}(\boldsymbol{\theta}) + \sum_{q=1}^{Q}\nabla_{\boldsymbol{\theta}} p_\tau(|\theta_j|).$$

Write $\mathbf{h}(\boldsymbol{\theta}) = (h_1(\boldsymbol{\theta}),\ldots,h_p(\boldsymbol{\theta}))'$. From Lemma 7, it yields that $\|\widehat{\lambda}(\boldsymbol{\theta})\|_2 = O_p(r^{1/2}Mn^{-1/2})$ which implies $\sup_{1\leq q\leq Q}|\widehat{\lambda}(\boldsymbol{\theta})'\phi_q(\boldsymbol{\theta})| = o_p(1)$. For each $j = 1,\ldots,p$,

$$\begin{aligned}h_j(\boldsymbol{\theta}) = \frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\widehat{\lambda}(\boldsymbol{\theta}_0)'\frac{\partial\phi_q(\boldsymbol{\theta}_0)}{\partial\theta_j} + \frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\widehat{\lambda}(\boldsymbol{\theta}_0)'\frac{\partial^2\phi_q(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\boldsymbol{\theta}'}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + p_\tau'(|\theta_j|)\mathrm{sign}(\theta_j) \\ + \text{higher order terms.}\end{aligned}$$

From (A.4), there exists a positive constant C such that $p_\tau'(|\theta_j|) \geq C\tau$. On the other hand, as $\tau(r^{-1}n)^{1/2}M^{-1} \to \infty$,

$$\max_{j\notin\mathcal{A}}\left|\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\widehat{\lambda}(\boldsymbol{\theta}_0)'\frac{\partial\phi_q(\boldsymbol{\theta}_0)}{\partial\theta_j}\right| = O_p(r^{1/2}Mn^{-1/2}) = o_p(\tau).$$

Similarly, we can show

$$\max_{j\notin\mathcal{A}}\left|\frac{1}{Q}\sum_{q=1}^{Q}\rho_v(0)\widehat{\lambda}(\boldsymbol{\theta}_0)'\frac{\partial^2\phi_q(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\boldsymbol{\theta}'}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right| = o_p(\tau).$$

Hence, $p_\tau'(|\theta_j|)\mathrm{sign}(\theta_j)$ dominates the sign of $h_j(\boldsymbol{\theta})$ uniformly for all $j\notin\mathcal{A}$. If $\widehat{\boldsymbol{\theta}}_n^{(2)} \neq \mathbf{0}$, there exists some $j\notin\mathcal{A}$ such that $\widehat{\theta}_{n,j}\neq 0$. Under our above arguments, we can find

$$P\big\{h_j(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}) \neq 0\big\} \to 1.$$

It is a contradiction. Hence, $\widehat{\boldsymbol{\theta}}_n^{(2)} = \mathbf{0}$.

Nextly, we consider the second result. From (25), it yields

$$[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1} \times [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}] = \begin{pmatrix} (\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{-1} & * \\ * & * \end{pmatrix}.$$

Let

$$[E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2} = \begin{pmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & * \end{pmatrix},$$

where $\mathbf{U}$ is a $s \times s$ symmetric matrix, then $\mathbf{U}\mathbf{U}' + \mathbf{V}\mathbf{V}' = (\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{-1}$. For any $\boldsymbol{\alpha}_n \in \mathbb{R}^s$ such that $\|\boldsymbol{\alpha}_n\|_2 = 1$, define

$$\widetilde{\boldsymbol{\alpha}}_n = \begin{pmatrix} \mathbf{U}' \\ \mathbf{V}' \end{pmatrix} (\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{1/2} \boldsymbol{\alpha}_n.$$

Then,

$$\widetilde{\boldsymbol{\alpha}}_n' \widetilde{\boldsymbol{\alpha}}_n = \boldsymbol{\alpha}_n' (\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{1/2} (\mathbf{U}\mathbf{U}' + \mathbf{V}\mathbf{V}') (\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{1/2} \boldsymbol{\alpha}_n = 1.$$

Following the same argument for Proposition 2, we know it still holds for $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})}$. Note that

$$\begin{aligned} &\widetilde{\boldsymbol{\alpha}}_n' ([E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} V_n V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}])^{-1/2} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}]' V_M^{-1} [E\{\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)\}] (\widehat{\boldsymbol{\theta}}_n^{(\mathrm{pe})} - \boldsymbol{\theta}_0) \\ &\quad = \boldsymbol{\alpha}_n' (\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{-1/2} (\widehat{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}), \end{aligned}$$

then we establish the second result following Proposition 2. □

References

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions, *Econometrica*, **71**, 1795–1843.

Anatolyev, S. (2005). GMM, GEL, serial correlation, and asymptotic bias, *Econometrica*, **73**, 983–1002.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximating factor models, *Econometrica*, **70**, 191–221.

Billingsley, P. (1995). *Probability and Measure* (3rd edition), Wiley, New York.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *The Annals of Statistics*, **14**, 1171–1179.

Chang, J., Tang, C. Y. and Wu, Y. (2013). Marginal empirical likelihood and sure independence feature screening, *The Annals of Statistics*, **41**, 2123–2148

Chen, S. X and Cui, H. (2003). An extended empirical likelihood for generalized linear models, *Statistica Sinica*, **13**, 69–81.

Chen, S. X. and Cui, H. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters, *Biometrika*, **93**, 215–220.

Chen, S. X. and Cui, H. (2007). On the second properties of empirical likelihood with moment restrictions, *Journal of Econometrics*, **141**, 492–516.

Chen, S. X., Peng, L. and Qin, Y. L. (2009). Effects of data dimension on empirical likelihood, *Biometrika*, **96**, 711–722.

Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression (with discussion), *Test*, **18**, 415–447.

Chen, X. (2007). "Large sample sieve estimation of semi-nonparametric models" in *The Handbook of Econometrics*, **6B**, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland.

Chen, X. and Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals, *Econometrica*, **80**, 277–321.

Davis, R. A., Zhang, P. and Zheng, T. (2012). Sparse vector autoregressive modelling. Available at *arXiv:1207.0520*.

Davydov, Y. A. (1968). On convergence of distributions generated by stationary stochastic processes, *Theory of Probability and its Applications*, **13**, 691–696.

Donald, S. G., Imbens, G. W. and Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions, *Journal of Econometrics*, **117**, 55–93.

Doukhan, P. (1994). *Mixing: properties and exmaples*, Springer-Verlag, Berlin.

Durrett, R. (2010). *Probability: Theory and Examples* (4th edition), Cambridge University Press.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.

Fan, J. and Liao, Y. (2014). Endogeneity in high dimensions, *The Annals of Statistics*, **42**, 872–917.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon, *The Annals of Statistics*, **29**, 153–193.

Francq, C. and Zakoïan, J. M. (2005). A central limit theorem for mixing triangular arrays of variables whose dependence is allowed to grow with the sample size, *Econometric Theory*, **21**, 1165–1171.

Francq, C. and Zakoïan, J. M. (2007). HAC estimation and strong linearity testing in weak ARMA models, *Journal of Multivariate Analysis*, **98**, 114–144.

Hall, P. (1985). Resampling a coverage pattern, *Stochastical Processes Application*, **20**, 231–246.

Hall, P., Horowitz, J. and Jing, B.-Y. (1995). On blocking rules for the bootstrap and dependent data, *Biometrika*, **82**, 561–574.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029–1054.

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

Hansen, L. P. and Singleton, K. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica*, **50**, 1269–1286.

Hjort, N. L., McKeague, I. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood, *The Annals of Statistics*, **37**, 1079–1111.

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

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

Kitamura, Y. (2007). "Empirical likelihood methods in econometrics: theory and practice" in Advances in Economics and Econometrics: Ninth World Congress of the Econometric Society , R. Blundell, W. K. Newey and T. Personn (eds.), Cambridge University Press.

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

Künsch, H. R. (1989). The jackkinife and the bootstrap for general stationary observations, *The Annals of Statistics*, **17**, 1217–1241.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*, Springer, New York.

Lahiri, S. N. and Mukhopadhyay, S. (2012). A penalized empirical likelihood method in high dimensions, *The Annals of Statistics*, **40**, 2511–2540.

Leng, C. and Tang, C. Y. (2012). Penalized empirical likelihood and growing dimensional general estimating equations, *Biometrika*, **99**, 703–716.

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity, *Econometrica*, **59**, 1161–1167.

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

Nordman, D. J. and Lahiri, S. N. (2013). A review of empirical likelihood methods for time series, *Journal of Statistical Planning and Inference*, to appear.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75**, 237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions, *The Annals of Statistics*, **18**, 90–120.

Owen, A. (2001). *Empirical Likelihood*, Chapman and Hall-CRC, New York.

Peligrad, M. and Utev, S. (1997). Central limit theorem for linear processes, *The Annals of Probability*, **25**, 443–456.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations, *The Annals of Statistics*, **22**, 300–325.

Rio, E. (1993). Covariance inequalities for strongly mixing processes, *Annales de l'Institut Henri Poincaré*, **29**, 589–597.

Rothenberg, T. J. (1973). *Efficient Estimation with a priori Information*, Yale University Press, New Haven, USA.

Rudin, W. (1976). *Principles of Mathematical Analysis*, McGraw-Hill, New York.

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

Stock, J. H. and Watson, M. W. (2010). "Dynamic factor models", *Oxford Handbook of Economic Forecasting*, M. Clements and D. Hendry (eds), Chapter 2.

Tang, C. Y. and Leng, C. (2010). Penalized high dimensional empirical likelihood, *Biometrika*, **97**, 905–920.

Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty, *The Annals of Statistics*, **38**, 894–942.

Table 1: Empirical medians of the squared estimation errors ($\times 10^2$) of the empirical likelihood (EL), the exponential tilting (ET), the continuous updating (CU) and the optimal GMM for the high dimensional mean model with $p = \lfloor 10n^{2/15} \rfloor$.

	Sample size	$n = 500$			$n = 1000$			$n = 2000$		
L, M	Method \ ψ	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
(i)	EL	0.99	1.25	1.69	0.61	0.73	1.04	0.46	0.56	0.69
	ET	0.98	1.24	1.68	0.61	0.72	1.03	0.44	0.55	0.71
	CU	0.99	1.26	1.70	0.62	0.73	1.03	0.42	0.54	0.70
	GMM	1.20	1.42	1.95	0.81	0.98	1.32	0.57	0.70	0.95
(ii)	EL	0.97	1.23	1.67	0.61	0.73	1.00	0.40	0.51	0.69
	ET	0.97	1.22	1.68	0.62	0.74	1.00	0.42	0.52	0.69
	CU	0.98	1.22	1.67	0.60	0.72	1.01	0.44	0.51	0.68
	GMM	1.20	1.41	1.97	0.82	0.99	1.32	0.58	0.72	0.97
(iii)	EL	0.93	1.21	1.63	0.59	0.71	0.99	0.40	0.51	0.69
	ET	0.95	1.20	1.63	0.62	0.72	0.99	0.41	0.50	0.69
	CU	0.95	1.21	1.64	0.58	0.70	1.00	0.41	0.51	0.67
	GMM	1.20	1.40	1.94	0.81	0.98	1.31	0.57	0.71	0.96
(iv)	EL	0.98	1.25	1.67	0.61	0.72	1.00	0.39	0.50	0.66
	ET	0.98	1.24	1.68	0.62	0.71	1.01	0.40	0.51	0.66
	CU	0.97	1.25	1.69	0.61	0.71	1.02	0.41	0.52	0.68
	GMM	1.23	1.45	1.94	0.85	1.03	1.32	0.57	0.72	0.98
(v)	EL	0.96	1.21	1.64	0.57	0.71	0.98	0.37	0.50	0.66
	ET	0.98	1.20	1.68	0.57	0.69	0.99	0.39	0.51	0.65
	CU	0.97	1.22	1.68	0.58	0.70	1.00	0.40	0.51	0.68
	GMM	1.21	1.42	1.92	0.85	1.01	1.30	0.57	0.72	0.97

Table 2: Empirical medians of the squared estimation errors ($\times 10^2$) of the empirical likelihood (EL), the exponential tilting (ET), the continuous updating (CU) and the optimal GMM for the high dimensional mean model with $p = \lfloor 12n^{2/15} \rfloor$.

	Sample size	$n = 500$			$n = 1000$			$n = 2000$		
L, M	Method \ ψ	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
(i)	EL	0.79	1.02	1.39	0.44	0.57	0.77	0.30	0.35	0.44
	ET	0.80	0.99	1.39	0.45	0.58	0.78	0.33	0.36	0.45
	CU	0.79	1.01	1.40	0.46	0.57	0.79	0.32	0.34	0.43
	GMM	1.15	1.49	2.00	0.79	1.01	1.30	0.55	0.73	0.89
(ii)	EL	0.79	0.97	1.36	0.43	0.54	0.75	0.28	0.35	0.45
	ET	0.78	0.96	1.38	0.41	0.53	0.74	0.30	0.36	0.44
	CU	0.77	0.98	1.39	0.40	0.52	0.75	0.26	0.33	0.42
	GMM	1.15	1.48	1.99	0.81	1.01	1.31	0.56	0.71	0.90
(iii)	EL	0.76	0.93	1.33	0.41	0.53	0.75	0.28	0.33	0.43
	ET	0.77	0.94	1.35	0.40	0.52	0.73	0.24	0.33	0.43
	CU	0.74	0.92	1.36	0.38	0.51	0.74	0.25	0.32	0.40
	GMM	1.14	1.47	1.99	0.80	1.01	1.29	0.54	0.70	0.89
(iv)	EL	0.77	0.98	1.35	0.41	0.52	0.71	0.26	0.32	0.42
	ET	0.78	0.98	1.36	0.41	0.53	0.72	0.27	0.33	0.44
	CU	0.79	0.97	1.37	0.40	0.51	0.71	0.28	0.36	0.43
	GMM	1.15	1.48	1.99	0.80	1.02	1.31	0.56	0.72	0.91
(v)	EL	0.75	0.95	1.34	0.40	0.51	0.69	0.23	0.32	0.40
	ET	0.77	0.96	1.35	0.38	0.50	0.71	0.26	0.30	0.42
	CU	0.76	0.97	1.36	0.39	0.49	0.70	0.26	0.32	0.41
	GMM	1.14	1.46	1.98	0.79	1.01	1.27	0.55	0.72	0.90

Table 3: Empirical medians of the squared estimation errors ($\times 10^2$) of the empirical likelihood (EL), the penalized empirical likelihood (PEL), the exponential tilting (ET), the penalized exponential tilting (PET), the continuous updating (CU), the penalized continuous updating (PCU) and the optimal GMM for the high dimensional generalized linear model with $p = \lfloor 5n^{2/15} \rfloor$.

	Sample size	$n = 500$			$n = 1000$			$n = 2000$		
L, M	Method \ ψ	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
(i)	EL	3.16	3.59	3.67	3.13	3.42	3.62	2.85	2.89	3.00
	PEL	0.95	1.05	1.10	0.84	0.93	1.05	0.74	0.78	0.89
	ET	3.05	3.19	3.38	3.03	3.10	3.20	2.70	2.83	2.92
	PET	0.84	0.95	1.00	0.82	0.89	0.95	0.67	0.80	0.83
	CU	3.77	4.01	4.35	3.73	3.94	4.16	3.18	3.30	3.46
	PCU	1.14	1.22	1.26	1.05	1.18	1.22	0.95	1.01	1.05
	GMM	6.72	7.06	7.12	6.30	6.43	6.60	5.83	5.88	5.92
(ii)	EL	3.11	3.32	3.48	3.07	3.10	3.35	2.63	2.86	2.93
	PEL	0.95	1.05	1.14	0.83	0.89	1.00	0.77	0.84	0.87
	ET	3.05	3.10	3.11	2.85	2.95	3.00	2.37	2.53	2.76
	PET	0.89	0.93	0.95	0.77	0.85	0.89	0.74	0.82	0.84
	CU	3.26	3.45	3.77	3.19	3.29	3.62	2.79	3.02	3.24
	PCU	1.00	1.10	1.14	0.95	1.00	1.07	0.84	0.95	0.98
	GMM	6.80	7.11	7.34	6.33	6.55	6.63	5.83	5.92	5.96
(iii)	EL	2.81	2.90	2.97	2.76	2.85	2.92	2.53	2.70	2.85
	PEL	0.77	0.84	0.93	0.71	0.77	0.85	0.63	0.70	0.77
	ET	2.65	2.70	2.76	2.45	2.65	2.68	2.21	2.41	2.51
	PET	0.84	0.89	0.93	0.71	0.82	0.86	0.66	0.71	0.77
	CU	3.16	3.42	3.56	3.11	3.32	3.46	2.70	2.83	3.18
	PCU	0.95	1.05	1.10	0.84	0.95	1.00	0.77	0.84	0.89
	GMM	6.76	7.06	7.23	6.28	6.50	6.60	5.81	5.86	5.90
(iv)	EL	2.24	2.35	2.43	2.19	2.28	2.39	1.97	2.12	2.26
	PEL	0.71	0.75	0.84	0.63	0.65	0.71	0.55	0.63	0.68
	ET	2.14	2.28	2.33	2.07	2.21	2.26	1.92	2.05	2.12
	PET	0.63	0.72	0.79	0.60	0.66	0.77	0.55	0.57	0.66
	CU	2.86	2.90	3.11	2.74	2.81	3.03	2.28	2.43	2.74
	PCU	0.90	0.95	1.00	0.77	0.84	0.95	0.63	0.73	0.82
	GMM	7.16	7.24	7.30	6.30	6.50	6.59	5.85	5.92	6.03
(v)	EL	2.19	2.26	2.32	2.07	2.14	2.17	1.61	1.76	1.84
	PEL	0.63	0.71	0.77	0.55	0.63	0.70	0.45	0.50	0.57
	ET	1.79	1.94	2.07	1.64	1.79	2.00	1.38	1.52	1.70
	PET	0.50	0.55	0.65	0.45	0.51	0.63	0.32	0.45	0.52
	CU	2.74	2.83	3.08	2.53	2.68	2.90	2.10	2.35	2.83
	PCU	0.84	0.87	0.95	0.71	0.80	0.84	0.55	0.68	0.80
	GMM	7.04	7.13	7.20	6.29	6.47	6.55	5.84	5.92	5.97

Table 4: Empirical medians of the squared estimation errors ($\times 10^2$) of the empirical likelihood (EL), the penalized empirical likelihood (PEL), the exponential tilting (ET), the penalized exponential tilting (PET), the continuous updating (CU), the penalized continuous updating (PCU) and the optimal GMM for the high dimensional generalized linear model with $p = \lfloor 6n^{2/15} \rfloor$.

	Sample size n	$n = 500$			$n = 1000$			$n = 2000$		
L, M	Method \ ψ	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
(i)	EL	2.63	2.75	2.82	2.48	2.60	2.76	2.31	2.42	2.54
	PEL	0.81	0.84	0.87	0.75	0.78	0.81	0.68	0.72	0.77
	ET	2.46	2.60	2.71	2.35	2.51	2.62	2.19	2.23	2.34
	PET	0.74	0.79	0.85	0.69	0.76	0.79	0.64	0.69	0.71
	CU	2.83	2.96	3.04	2.75	2.83	2.95	2.62	2.71	2.87
	PCU	0.88	0.92	0.95	0.82	0.88	0.93	0.76	0.80	0.86
	GMM	5.84	5.93	6.09	5.47	5.53	5.62	4.73	4.83	4.99
(ii)	EL	2.50	2.79	2.85	2.42	2.58	2.66	2.18	2.30	2.39
	PEL	0.73	0.83	0.86	0.66	0.78	0.80	0.56	0.66	0.71
	ET	2.35	2.44	2.59	2.28	2.37	2.45	2.15	2.23	2.32
	PET	0.65	0.73	0.79	0.62	0.67	0.75	0.57	0.61	0.66
	CU	2.79	2.82	2.94	2.61	2.74	2.81	2.58	2.67	2.73
	PCU	0.82	0.87	0.93	0.78	0.86	0.89	0.70	0.73	0.81
	GMM	5.89	5.96	6.13	5.45	5.56	5.64	4.70	4.81	5.05
(iii)	EL	2.37	2.65	2.69	2.09	2.22	2.42	1.96	2.12	2.31
	PEL	0.68	0.76	0.80	0.58	0.70	0.73	0.46	0.60	0.67
	ET	2.20	2.36	2.44	1.98	2.15	2.27	1.90	1.99	2.08
	PET	0.61	0.72	0.76	0.45	0.57	0.63	0.42	0.50	0.57
	CU	2.64	2.80	2.84	2.59	2.68	2.76	2.50	2.58	2.63
	PCU	0.76	0.84	0.89	0.74	0.79	0.82	0.65	0.73	0.77
	GMM	5.86	5.92	6.10	5.42	5.50	5.61	4.68	4.75	5.02
(iv)	EL	2.21	2.30	2.41	1.99	2.16	2.24	1.73	1.80	1.97
	PEL	0.58	0.67	0.72	0.51	0.58	0.63	0.40	0.46	0.53
	ET	1.92	2.04	2.18	1.82	1.89	1.97	1.46	1.62	1.83
	PET	0.54	0.59	0.61	0.46	0.52	0.54	0.34	0.39	0.48
	CU	2.58	2.63	2.74	2.31	2.46	2.70	2.23	2.31	2.47
	PCU	0.68	0.78	0.82	0.65	0.73	0.78	0.57	0.66	0.71
	GMM	6.04	6.10	6.16	5.60	5.64	5.78	4.70	4.79	5.01
(v)	EL	2.16	2.25	2.32	1.84	2.02	2.15	1.63	1.76	1.92
	PEL	0.56	0.60	0.67	0.47	0.52	0.61	0.37	0.42	0.52
	ET	1.90	1.96	2.06	1.71	1.76	1.84	1.34	1.43	1.59
	PET	0.54	0.58	0.61	0.43	0.46	0.51	0.30	0.32	0.36
	CU	2.53	2.58	2.66	2.24	2.44	2.61	2.17	2.21	2.29
	PCU	0.68	0.72	0.77	0.56	0.67	0.75	0.56	0.60	0.67
	GMM	5.93	6.08	6.15	5.52	5.59	5.71	4.64	4.72	4.88