Supplemental Appendix for "Select the Valid and Relevant Moments: An Information-Based LASSO for GMM with Many Moments"

Xu Cheng* Zhipeng Liao†

This Version: October, 2014

This supplemental appendix provides some auxiliary materials for "Select the Valid and Relevant Moments: An Information-Based LASSO for GMM with Many Moments"(cited as CL (2012) in this appendix). Section 1 presents and discusses two matrix algebra theorems that are used in the proof of CL (2012) . Section 2 shows that the shrinkage estimator of θ_o is as efficient as the oracle estimator asymptotically. The pointwise and uniform convergence rates of the empirical information measure are established in Section 3. Some additional simulation results are in Section 4.

The notations in this appendix are consistent with those in CL (2012). Throughout the appendix, C denotes some generic finite positive constant; $\|\cdot\|$ denotes the Euclidean norm; A' denotes the transpose of a matrix A; $\rho_{\max}(A)$ and $\rho_{\min}(A)$ denote the largest and smallest eigenvalues of a matrix A, respectively; for any square matrix A, $A \geq 0$ means that A is a positive semi-definite matrix; for any positive integers k_1 and k_2, I_{k_1} denotes the $k_1 \times k_1$ identity matrix and $\mathbf{0}_{k_1\times k_2}$ denotes the $k_1 \times k_2$ zero matrix; $A \equiv B$ means that A is defined as B; $a_n = o_p(b_n)$ means that for any constants $\epsilon_1, \epsilon_2 > 0$, there is $\Pr(|a_n/b_n| \geq \epsilon_1) < \epsilon_2$ eventually; $a_n = O_p(b_n)$ means that for any $\epsilon > 0$, there is a finite constant C_ϵ such that $\Pr(|a_n/b_n| \geq C_\epsilon) < \epsilon$ eventually; for any two sequences a_n and b_n, we use $a_n \lesssim b_n$ to denote that $a_n \leq Cb_n$ where C is some fixed finite positive constant; "$\rightarrow_p$" and "$\rightarrow_d$" denote convergence in probability and convergence in distribution, respectively; and w.p.a.1 abbreviates with probability approaching 1.

*Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, USA. Email: xucheng@sas.upenn.edu

†Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu

1 Two Useful Theorems

For any symmetric $k \times k$ real matrix A, we write its eigenvalues in decreasing order:

$$\rho_1(A) \geq \rho_2(A) \geq \cdots \geq \rho_k(A). \tag{1.1}$$

By definition, $\rho_1(A) = \rho_{\max}(A)$ and $\rho_k(A) = \rho_{\min}(A)$. Let $\|\cdot\|_s$ denote the operator norm:

$$\|A\|_s = \sup_{\|x\| \leq 1} \|Ax\| \tag{1.2}$$

for any real matrix A. The following theorem is useful for theoretical results in CL (2012).

Theorem 1.1 (Weyl's Eigenvalue Perturbation Theorem) *Let A and B are $k \times k$ symmetric real matrices. Then*

$$\max_{1 \leq j \leq k} \left|\rho_j(A) - \rho_j(B)\right| \leq \|A - B\|_s .$$

Theorem 1.1 is a simplified version of Corollary III.2.6 in Bhatia (1997), which allows A and B to be Hermitian matrices. Because $\|A\|_s \leq \|A\|$ for any finite dimensional matrix A, this theorem directly implies that

$$\max_{1 \leq j \leq k} \left|\rho_j(A) - \rho_j(B)\right| \leq \|A - B\| \tag{1.3}$$

for any $k \times k$ symmetric real matrices A and B.

Theorem 1.2 (Aronszajn's Inequality) *Let C be an $k \times k$ symmetric real matrix partitioned as*

$$C = \begin{bmatrix} A & X \\ X' & B \end{bmatrix},$$

where A is a $k_0 \times k_0$ matrix. Let the eigenvalues of A, B, and C be $\rho_1(A) \geq \cdots \geq \rho_{k_0}(A)$, $\rho_1(B) \geq \cdots \geq \rho_{k-k_0}(B)$, and $\rho_1(C) \geq \cdots \geq \rho_k(C)$, respectively. Then,

$$\rho_{i+j-1}(C) + \rho_k(C) \leq \rho_i(A) + \rho_j(B)$$

for all i, j with $i + j - 1 \leq k$.

Theorem 1.2 is a simplified version of Theorem III.2.9 in Bhatia (1997), which allows A to be a Hermitian matrix. Let $i = j = 1$ and C be a positive definite matrix, then we can use Aronszajn's Inequality to show

$$\rho_{\max}(C) = \rho_1(C) \leq \rho_1(C) + \rho_k(C) \leq \rho_1(A) + \rho_1(B) = \rho_{\max}(A) + \rho_{\max}(B), \tag{1.4}$$

which implies that the largest eigenvalue of C is bounded from above by the sum of the largest eigenvalues of A and B. The inequality (1.4) is used in the linear IV example of CL (2012).

2 Variance of the P-GMM Estimator

In this section, we prove the claims in Remark 3.5 of CL (2012) and demonstrate the shrinkage estimator is as efficient as the oracle estimator asymptotically. Using the Cauchy-Schwarz inequality, we deduce that

$$\begin{aligned} & \left|\gamma'_{n,d_\theta}\left[\Sigma_n - \left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\right]\gamma_{n,d_\theta}\right| \\ = & \left|\gamma'_{n,d_\theta}\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\Gamma'_\alpha W_n(\Omega_n - W_n^{-1})W_n\Gamma_\alpha\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\gamma_{n,d_\theta}\right| \\ \leq & \left\|\gamma'_{n,d_\theta}\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\Gamma'_\alpha W_n\right\|^2 \left\|W_n^{-1} - \Omega_n\right\| \end{aligned} \tag{2.1}$$

where $\gamma'_{n,d_\theta} = (\gamma'_{d_\theta}, \mathbf{0}'_{d_B})$ and $\gamma_{d_\theta} \in R^{d_\theta}$. Using $\rho_{\max}(\Gamma'_\alpha\Gamma_\alpha) \leq C$, which is shown in the proof of Theorem 3.3 of CL (2012), Assumptions 3.1(iii) and 3.6(iii) in CL (2012), we have

$$C^{-1} \lesssim \rho_{\min}\left(\Gamma'_\alpha W_n \Gamma_\alpha\right) \leq \rho_{\max}\left(\Gamma'_\alpha W_n \Gamma_\alpha\right) \lesssim C \tag{2.2}$$

w.p.a.1. Hence, we deduce that

$$\begin{aligned} & \left\|\gamma'_{n,d_\theta}\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\Gamma'_\alpha W_n\right\|^2 \\ = & \gamma'_{n,d_\theta}\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\Gamma'_\alpha W_n^2\Gamma_\alpha\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\gamma_{n,d_\theta} \\ \lesssim & \gamma'_{n,d_\theta}\gamma_{n,d_\theta} = \gamma'_{d_\theta}\gamma_{d_\theta}, \end{aligned} \tag{2.3}$$

which together with (2.1) and $\left\|W_n^{-1} - \Omega_n\right\| = o_p(1)$ implies that

$$\left|\gamma'_{n,d_\theta}\left[\Sigma_n - \left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\right]\gamma_{n,d_\theta}\right| = o_p(1). \tag{2.4}$$

Moreover, using similar arguments, we deduce that

$$\begin{aligned} & \left|\gamma'_{n,d_\theta}\left[\left(\Gamma'_\alpha \Omega_n^{-1}\Gamma_\alpha\right)^{-1} - \left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\right]\gamma_{n,d_\theta}\right| \\ = & \left|\gamma'_{n,d_\theta}\left(\Gamma'_\alpha \Omega_n^{-1}\Gamma_\alpha\right)^{-1}\Gamma'_\alpha\left[W_n - \Omega_n^{-1}\right]\Gamma_\alpha\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\gamma_{n,d_\theta}\right| \\ \leq & \left\|\gamma'_{n,d_\theta}\left(\Gamma'_\alpha \Omega_n^{-1}\Gamma_\alpha\right)^{-1}\Gamma'_\alpha\right\| \left\|\gamma'_{n,d_\theta}\left(\Gamma'_\alpha W_n \Gamma_\alpha\right)^{-1}\Gamma'_\alpha\right\| \left\|W_n^{-1} - \Omega_n\right\| = o_p(1). \end{aligned} \tag{2.5}$$

Hence, when W_n is an asymptotically efficient weighting matrix, the variance of $\sqrt{n}\gamma'_{d_\theta}(\widehat{\theta}_n - \theta_0)$ can be approximated by $\gamma'_{n,d_\theta}\left(\Gamma'_\alpha \Omega_n^{-1}\Gamma_\alpha\right)^{-1}\gamma_{n,d_\theta}$.

Recall that

$$\Gamma'_\alpha \equiv \begin{bmatrix} \left(\frac{\partial \mathbb{E}[g_{S+A}(Z,\theta_o)]}{\partial \theta'}\right)' & \left(\frac{\partial \mathbb{E}[g_B(Z,\theta_o)]}{\partial \theta'}\right)' \\ \mathbf{0}_{d_B\times(k_0-d_A)} & -I_{d_B} \end{bmatrix} \tag{2.6}$$

and

$$\Omega_n \equiv \mathrm{Var}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n}\begin{pmatrix} g_S(Z_i,\theta_o) \\ g_A(Z_i,\theta_o) \\ g_B(Z_i,\theta_o)\end{pmatrix}\right] = \begin{bmatrix} \Omega_{S+A,n} & \Omega_{AB,n} \\ \Omega_{BA,n} & \Omega_{B,n}\end{bmatrix}, \tag{2.7}$$

where $\Omega_{S+A,n}$ denotes the leading $(k_0+d_A)\times(k_0+d_A)$ submatrix of Ω_n, $\Omega_{B,n}$ denotes the last $d_B\times d_B$ submatrix of Ω_n, $\Omega_{AB,n}$ and $\Omega_{BA,n}$ are defined accordingly. Let $\Omega^{11}_{S+A,n}$ and $\Omega^{22}_{B,n}$ denote the leading $(k_0+d_A)\times(k_0+d_A)$ and last $d_B\times d_B$ submatrices of Ω_n^{-1} respectively, that is

$$\Omega^{11}_{S+A,n} = \left(\Omega_{S+A,n} - \Omega_{AB,n}\Omega^{-1}_{B,n}\Omega_{BA,n}\right)^{-1} \text{ and } \Omega^{22}_{B,n} = \left(\Omega_{B,n} - \Omega_{BA,n}\Omega^{-1}_{S+A,n}\Omega_{AB,n}\right)^{-1}. \tag{2.8}$$

Then we can write

$$\Omega_n^{-1} = \begin{bmatrix} \Omega^{11}_{S+A,n} & -\Omega^{-1}_{S+A,n}\Omega_{AB,n}\Omega^{22}_{B,n} \\ -\Omega^{-1}_{B,n}\Omega_{BA,n}\Omega^{11}_{S+A,n} & \Omega^{22}_{B,n}\end{bmatrix} = \begin{bmatrix} \Omega^{11}_{S+A,n} & \Omega^{12}_{AB,n} \\ \Omega^{21}_{BA,n} & \Omega^{22}_{B,n}\end{bmatrix}. \tag{2.9}$$

Hence, we have

$$\Gamma'_\alpha\Omega_n^{-1}\Gamma_\alpha = \begin{bmatrix} M_n^{11} & M_n^{12} \\ M_n^{21} & M_n^{22}\end{bmatrix} \tag{2.10}$$

where $M_n^{22} = \Omega^{22}_{B,n}$,

$$\begin{aligned} M_n^{11} &= \left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial\theta'}\right)'\Omega^{11}_{S+A,n}\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial\theta'}\right) \\ &\quad + \left(\frac{\partial \mathbb{E}\left[g_B(Z,\theta_o)\right]}{\partial\theta'}\right)'\Omega^{21}_{BA,n}\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial\theta'}\right) \\ &\quad + \left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial\theta'}\right)'\Omega^{12}_{AB,n}\left(\frac{\partial \mathbb{E}\left[g_B(Z,\theta_o)\right]}{\partial\theta'}\right) \\ &\quad + \left(\frac{\partial \mathbb{E}\left[g_B(Z,\theta_o)\right]}{\partial\theta'}\right)'\Omega^{22}_{B,n}\left(\frac{\partial \mathbb{E}\left[g_B(Z,\theta_o)\right]}{\partial\theta'}\right), \\ M_n^{12} &= -\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial\theta'}\right)'\Omega^{12}_{AB,n} - \left(\frac{\partial \mathbb{E}\left[g_B(Z,\theta_o)\right]}{\partial\theta'}\right)'\Omega^{22}_{B,n} = M_n^{21\prime}. \end{aligned} \tag{2.11}$$

Let $\Sigma^*_{\theta,n}$ denote the leading $d_\theta\times d_\theta$ submatrix of $\left(\Gamma'_\alpha\Omega_n^{-1}\Gamma_\alpha\right)^{-1}$. Using the inverse formula of partitioned matrix, we have

$$\Sigma^*_{\theta,n} = \left[M_n^{11} - M_n^{12}(M_n^{22})^{-1}M_n^{21}\right]^{-1}, \tag{2.12}$$

where

$$\begin{aligned} M_n^{12}(M_n^{22})^{-1}M_n^{21} &= \left[\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right)' \Omega^{12}_{AB,n}(\Omega^{22}_{B,n})^{-1} + \left(\frac{\partial \mathbb{E}\left[g_{B}(Z,\theta_o)\right]}{\partial \theta'}\right)'\right] \\ &\quad \times \left[\Omega^{21}_{BA,n}\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right) + \Omega^{22}_{B,n}\left(\frac{\partial \mathbb{E}\left[g_{B}(Z,\theta_o)\right]}{\partial \theta'}\right)\right] \\ &= \left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right)' \Omega^{12}_{AB,n}(\Omega^{22}_{B,n})^{-1}\Omega^{21}_{BA,n}\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right) \\ &\quad + \left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right)' \Omega^{12}_{AB,n}\left(\frac{\partial \mathbb{E}\left[g_{B}(Z,\theta_o)\right]}{\partial \theta'}\right) \\ &\quad + \left(\frac{\partial \mathbb{E}\left[g_{B}(Z,\theta_o)\right]}{\partial \theta'}\right)' \Omega^{21}_{BA,n}\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right) \\ &\quad + \left(\frac{\partial \mathbb{E}\left[g_{B}(Z,\theta_o)\right]}{\partial \theta'}\right)' \Omega^{22}_{B,n}\left(\frac{\partial \mathbb{E}\left[g_{B}(Z,\theta_o)\right]}{\partial \theta'}\right). \end{aligned} \tag{2.13}$$

Hence,

$$\begin{aligned} &M_n^{11} - M_n^{12}(M_n^{22})^{-1}M_n^{21} \\ = &\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right)' \left[\Omega^{11}_{S+A,n} - \Omega^{12}_{AB,n}(\Omega^{22}_{B,n})^{-1}\Omega^{21}_{BA,n}\right]\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right). \end{aligned} \tag{2.14}$$

By definition,

$$\begin{aligned} &\Omega^{11}_{S+A,n} - \Omega^{12}_{AB,n}(\Omega^{22}_{B,n})^{-1}\Omega^{21}_{BA,n} \\ = &\Omega^{11}_{S+A,n} - \Omega^{-1}_{S+A,n}\Omega_{AB,n}\Omega^{22}_{B,n}(\Omega^{22}_{B,n})^{-1}\Omega^{-1}_{B,n}\Omega_{BA,n}\Omega^{11}_{S+A,n} \\ = &\Omega^{11}_{S+A,n} - \Omega^{-1}_{S+A,n}\Omega_{AB,n}\Omega^{-1}_{B,n}\Omega_{BA,n}\Omega^{11}_{S+A,n} \\ = &\Omega^{-1}_{S+A,n}\left(\Omega_{S+A,n} - \Omega_{AB,n}\Omega^{-1}_{B,n}\Omega_{BA,n}\right)\Omega^{11}_{S+A,n} = \Omega^{-1}_{S+A,n}, \end{aligned} \tag{2.15}$$

which together with (2.14) implies that

$$M_n^{11} - M_n^{12}(M_n^{22})^{-1}M_n^{21} = \left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right)' \Omega^{-1}_{S+A,n}\left(\frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}\right). \tag{2.16}$$

In CL (2012), we define that

$$\frac{\partial m_{S+A}(\theta_o)}{\partial \theta'} = \frac{\partial \mathbb{E}\left[g_{S+A}(Z,\theta_o)\right]}{\partial \theta'}. \tag{2.17}$$

Hence, using (2.12) and (2.16) we get

$$\Sigma^*_{\theta,n} = \left[\left(\frac{\partial m_{S+A}(\theta_o)}{\partial \theta'} \right)' \Omega^{-1}_{S+A,n} \left(\frac{\partial m_{S+A}(\theta_o)}{\partial \theta'} \right) \right]^{-1}. \tag{2.18}$$

This proves the claim in Remark 3.5 of of CL (2012).

3 Empirical Information Measure

In this section, we study the stochastic properties of the empirical information measure under the assumption that the data are i.i.d. Recall that

$$\begin{aligned} \dot{\Gamma}_{S,n} &\equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\partial g_S(Z_i,\dot{\theta}_n)}{\partial \theta'},\ \dot{\Omega}_{S,n} \equiv \frac{1}{n}\sum_{i=1}^{n} g_S(Z_i,\dot{\theta}_n) g_S(Z_i,\dot{\theta}_n)', \\ \dot{\Gamma}_{S+\ell,n} &\equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\partial g_{S+\ell}(Z_i,\dot{\theta}_n)}{\partial \theta'},\ \dot{\Omega}_{S+\ell,n} \equiv \frac{1}{n}\sum_{i=1}^{n} g_{S+\ell}(Z_i,\dot{\theta}_n) g_{S+\ell}(Z_i,\dot{\theta}_n)'. \end{aligned} \tag{3.1}$$

The estimators of the variance matrices V_S and $V_{S+\ell}$ are thus constructed as

$$\dot{V}_{n,S} \equiv \dot{\Gamma}'_{S,n} \dot{\Omega}^{-1}_{S,n} \dot{\Gamma}_{S,n} \text{ and } \dot{V}_{n,S+\ell} \equiv \dot{\Gamma}'_{S+\ell,n} \dot{\Omega}^{-1}_{S+\ell,n} \dot{\Gamma}_{S+\ell,n}. \tag{3.2}$$

The empirical information measure is defined as

$$\dot{\mu}_{n,\ell} = \rho_{\max}(\dot{V}_{n,S} - \dot{V}_{n,S+\ell}). \tag{3.3}$$

For any moment function $g_\ell(Z,\theta)$ and any $\theta \in \Theta$, we define $g_{\theta,\ell}(Z,\theta) = \frac{\partial g_\ell(Z,\theta)}{\partial \theta'}$. By definition, $g_{\theta,\ell}(Z,\theta)$ is a $1 \times d_\theta$ vector with the k-th element being $\frac{\partial g_\ell(Z,\theta)}{\partial \theta(k)}$, where $\theta(k)$ denotes the k-th element of θ for any $k = 1, ..., d_\theta$. We use $g_{\theta\theta,\ell}(Z,\theta)$ denote the $d_\theta \times d_\theta$ matrix whose (k,j)-th element being $\frac{\partial^2 g_\ell(Z,\theta)}{\partial \theta(k) \partial \theta(j)}$, where $\theta(k)$ and $\theta(j)$ $(k, j = 1, ..., d_\theta)$ denote the k-th and j-th elements of θ respectively.

Assumption 3.1 (i) *The preliminary estimator* $\dot{\theta}_n$ *satisfies*

$$\sqrt{n}(\dot{\theta}_n - \theta_o) = O_p(1);$$

(ii) $g_\ell(Z,\theta)$ *is twice differentiable in* θ *a.e. for any* $\ell \in D$; (iii) *for any* $\ell \in D$

$$\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}} \|g_{\theta,\ell}(Z,\theta)\|^4\right] \;\leq\; C, \tag{3.4}$$

$$\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}} \|g_{\theta\theta,\ell}(Z,\theta)\|^2\right] \;\leq\; C; \tag{3.5}$$

(iv) $\mathbb{E}\left[g_\ell^4(Z,\theta_o)\right] \leq C$ *for any* $\ell \in D$; (v) $\Omega_S(\theta_o)$ *and* $\Omega_{S+\ell}(\theta_o)$ *are finite and positive definite matrices for any* $\ell \in D$.

Lemma 3.1 *Under Assumption 3.1, we have* $\left|\dot{\mu}_{n,\ell} - \mu_{o,\ell}\right| = O_p(n^{-\frac{1}{2}})$ *for any* $\ell \in D$.

Lemma 3.1 establishes the convergence rate of individual empirical information measure. This result is sufficient for studying the properties of the penalized GMM estimator when the number k_n of moment conditions is fixed. However, when k_n is divergent, such a "pointwise" result will not be enough. We next establish a strong result which provides the convergence rate of the empirical information measure uniformly over $\ell \in D$. The following conditions are needed.

Assumption 3.2 (i) $\min_{\ell\in D}\rho_{\min}\left(\Omega_{S+\ell}\right) \geq C^{-1}$ *for all* n; (ii)

$$\max_{\ell\leq k_n}\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}} \|g_{\theta,\ell}(Z,\theta)\|^4\right] \;\leq\; C, \tag{3.6}$$

$$\max_{\ell\leq k_n}\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}} \|g_{\theta\theta,\ell}(Z,\theta)\|^2\right] \;\leq\; C; \tag{3.7}$$

(iii) $\max_{\ell\leq k_n}\mathbb{E}\left[g_\ell^4(Z,\theta_o)\right] \leq C$.

Lemma 3.2 *Under Assumptions 3.1 and 3.2, we have* $\max_{\ell\in D}\left|\dot{\mu}_{n,\ell} - \mu_{o,\ell}\right| = O_p(\sqrt{k_n/n})$.

Proof of Lemma 3.1. It is clear that the matrices V_S, $V_{S+\ell}$, $\dot{V}_{n,S}$ and $\dot{V}_{n,S+\ell}$ are $d_\theta \times d_\theta$ symmetric real matrices. For the ease of notation, let $\Omega_S \equiv \Omega_S(\theta_o)$ and $\Omega_{S+\ell} \equiv \Omega_{S+\ell}(\theta_o)$. Invoking Weyl's Eigenvalue Perturbation Theorem and the triangle inequality, we obtain

$$\begin{aligned}\left|\dot{\mu}_{n,\ell} - \mu_{o,\ell}\right| &= \left|\rho_{\max}(\dot{V}_{n,S} - \dot{V}_{n,S+\ell}) - \rho_{\max}(V_{n,S} - V_{n,S+\ell})\right| \\ &\leq \left\|\dot{V}_{n,S} - V_S - (\dot{V}_{n,S+\ell} - V_{S+\ell})\right\| \\ &\leq \left\|\dot{V}_{n,S} - V_S\right\| + \left\|\dot{V}_{n,S+\ell} - V_{S+\ell}\right\| \end{aligned} \tag{3.8}$$

which implies that

$$\max_{\ell \in D} \left|\dot{\mu}_{n,\ell} - \mu_{o,\ell}\right| \leq \left\|\dot{V}_{n,S} - V_S\right\| + \max_{\ell \in D} \left\|\dot{V}_{n,S+\ell} - V_{S+\ell}\right\|. \tag{3.9}$$

The above inequality is useful to establish the convergence rate of $\dot{\mu}_{n,\ell}$.

As d_θ is a fixed integer, for the ease of notation, we assume that $d_\theta = 1$ in the rest of the proof. When $d_\theta > 1$, a complete proof can be conducted by applying the same argument element by element. By definition, we can write

$$\dot{V}_{n,S} - V_S = \left(\dot{\Gamma}_{S,n} - \Gamma_S\right)' \dot{\Omega}_{S,n}^{-1} \dot{\Gamma}_{S,n} + \Gamma_S' \left(\dot{\Omega}_{S,n}^{-1} - \Omega_S^{-1}\right) \dot{\Gamma}_{S,n} + \Gamma_S' \Omega_S^{-1} \left(\dot{\Gamma}_{S,n} - \Gamma_S\right). \tag{3.10}$$

By the triangle inequality and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \left\|\dot{V}_{n,S} - V_S\right\| \leq & \left\|\dot{\Gamma}_{S,n} - \Gamma_S\right\| \left\|\dot{\Omega}_{S,n}^{-1}\right\| \left\|\dot{\Gamma}_{S,n}\right\| \\ & + \|\Gamma_S\| \left\|\dot{\Omega}_{S,n}^{-1} - \Omega_S^{-1}\right\| \left\|\dot{\Gamma}_{S,n}\right\| \\ & + \|\Gamma_S\| \left\|\Omega_S^{-1}\right\| \left\|\dot{\Gamma}_{S,n} - \Gamma_S\right\|. \end{aligned} \tag{3.11}$$

By the mean value theorem,

$$\begin{aligned} \dot{\Gamma}_{\ell,n} - \Gamma_\ell &= \frac{1}{n} \sum_{i=1}^{n} g_{\theta,\ell}(Z, \dot{\theta}_n) - \mathbb{E}\left[g_{\theta,\ell}(Z, \theta_o)\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} g_{\theta,\ell}(Z, \theta_o) - \mathbb{E}\left[g_{\theta,\ell}(Z, \theta_o)\right] + \frac{1}{n} \sum_{i=1}^{n} g_{\theta\theta,\ell}(Z, \widetilde{\theta}_{\ell,n})(\dot{\theta}_n - \theta_o), \end{aligned} \tag{3.12}$$

where $\widetilde{\theta}_{\ell,n}$ is some value between $\dot{\theta}_n$ and θ_o. Using (3.4) and the Markov inequality,

$$\frac{1}{n} \sum_{i=1}^{n} g_{\theta,\ell}(Z, \theta_o) - \mathbb{E}\left[g_{\theta,\ell}(Z, \theta_o)\right] = O_p(n^{-\frac{1}{2}}) \tag{3.13}$$

for any $\ell \in S$. By the consistency of $\dot{\theta}_n$, we know that $\widetilde{\theta}_{\ell,n} \in \{\theta \in \Theta : \|\theta - \theta_o\| \leq \delta\}$ w.p.a.1. We next show that

$$\sup_{\{\theta \in \Theta : \|\theta - \theta_o\| \leq \delta\}} \left\| \frac{1}{n} \sum_{i=1}^{n} g_{\theta\theta,\ell}(Z, \theta) - \mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z, \theta)\right] \right\| = O_p(n^{-\frac{1}{2}}). \tag{3.14}$$

Using (3.5), the Markov inequality and the Maximum inequality (Section 4.3 of Pollard, 1989),

$$\begin{aligned}
& \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} g_{\theta\theta,\ell}(Z,\widetilde{\theta}_{\ell,n}) - \mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z,\widetilde{\theta}_{\ell,n})\right]\right| \geq M n^{-\frac{1}{2}}\right) \\
\leq\ & \Pr\left(\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(g_{\theta\theta,\ell}(Z,\theta) - \mathbb{E}\left[g_{\theta\theta,\ell}(Z,\theta)\right]\right)\right| \geq M\right) \\
\lesssim\ & \frac{\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}\left|g_{\theta\theta,\ell}(Z,\theta)\right|^2\right]}{M^2} \lesssim \frac{1}{M^2}
\end{aligned} \tag{3.15}$$

where M is any large positive constant, which proves (3.14). Hence using Assumptions 3.1(i), 3.1(iii), and (3.14), we have

$$\frac{1}{n}\sum_{i=1}^{n} g_{\theta\theta,\ell}(Z,\widetilde{\theta}_{\ell,n})(\dot{\theta}_n - \theta_o) = \mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z,\widetilde{\theta}_{\ell,n})\right](\dot{\theta}_n - \theta_o) + o_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}), \tag{3.16}$$

where $\mathbb{E}_Z[\cdot]$ denotes the expectation taking with respect to Z, which together with (3.12) and (3.13) implies that

$$\left\|\dot{\Gamma}_{\ell,n} - \Gamma_\ell\right\| = O_p(n^{-\frac{1}{2}}) \tag{3.17}$$

for any $\ell \in S$. This together with the fact that $d_S = k_0$ is a fixed integer yields

$$\left\|\dot{\Gamma}_{S,n} - \Gamma_S\right\|^2 = \sum_{\ell\in S}\left\|\dot{\Gamma}_{\ell,n} - \Gamma_\ell\right\| = O_p(n^{-\frac{1}{2}}). \tag{3.18}$$

By the definition of $\dot{\Omega}_{S,n}$ and Ω_S, we can write

$$\dot{\Omega}_{S,n} - \Omega_S = \frac{1}{n}\sum_{i=1}^{n} g_S(Z_i,\dot{\theta}_n)g_S(Z_i,\dot{\theta}_n)' - \mathbb{E}\left[g_S(Z,\theta_o)g_S(Z,\theta_o)'\right]. \tag{3.19}$$

By the mean value theorem,

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} g_\ell(Z_i,\dot{\theta}_n)g_k(Z_i,\dot{\theta}_n) - \mathbb{E}\left[g_\ell(Z,\theta_o)g_k(Z,\theta_o)\right] \\
=\ & \frac{1}{n}\sum_{i=1}^{n} g_\ell(Z_i,\theta_o)g_k(Z_i,\theta_o) - \mathbb{E}\left[g_\ell(Z,\theta_o)g_k(Z,\theta_o)\right] \\
& +(\dot{\theta}_n - \theta_o)'\frac{1}{n}\sum_{i=1}^{n} g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})g_k(Z_i,\widetilde{\theta}_{1,n}) \\
& +(\dot{\theta}_n - \theta_o)'\frac{1}{n}\sum_{i=1}^{n} g_{\theta,k}(Z_i,\widetilde{\theta}_{1,n})g_\ell(Z_i,\widetilde{\theta}_{1,n})
\end{aligned} \tag{3.20}$$

for any $\ell, k \in S$, where $\widetilde{\theta}_{1,n}$ is some value between $\dot{\theta}_n$ and θ_o. Using Assumption 3.1(iv), the Hölder's inequality and the Markov inequality,

$$\frac{1}{n}\sum_{i=1}^{n} g_\ell(Z,\theta_o)g_k(Z,\theta_o) - \mathbb{E}\left[g_\ell(Z,\theta_o)g_k(Z,\theta_o)\right] = O_p(n^{-\frac{1}{2}}). \tag{3.21}$$

By the consistency of $\dot{\theta}_n$, we know that $\widetilde{\theta}_{1,n} \in \{\theta \in \Theta : \|\theta - \theta_o\| \leq \delta\}$ w.p.a.1. Using the Cauchy-Schwarz inequality and the triangle inequality, we get

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n} g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})g_k(Z_i,\widetilde{\theta}_{1,n})\right\| &= \left\|\frac{1}{n}\sum_{i=1}^{n} g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})g_k(Z_i,\theta_o)\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})\left[g_k(Z_i,\widetilde{\theta}_{1,n}) - g_k(Z_i,\theta_o)\right]\right\| \\
\leq \quad & \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} g_k^2(Z_i,\theta_o)} \\
&+\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,k}(Z_i,\widetilde{\theta}_{2,n})\right\|^2}\left\|\dot{\theta}_n - \theta_o\right\|
\end{aligned} \tag{3.22}$$

where $\widetilde{\theta}_{2,n}$ is some value between $\widetilde{\theta}_{1,n}$ and θ_o. By Assumption 3.1(iv) and the Markov inequality, we have

$$\frac{1}{n}\sum_{i=1}^{n} g_\ell^2(Z_i,\theta_o) = O_p(1) \text{ for all } \ell. \tag{3.23}$$

We next show that

$$\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}\left|\frac{1}{n}\sum_{i=1}^{n}\|g_{\theta,\ell}(Z_i,\theta)\|^2 - \mathbb{E}_Z\left[\|g_{\theta,\ell}(Z_i,\theta)\|^2\right]\right| = O_p(n^{-\frac{1}{2}}). \tag{3.24}$$

Using (3.4), the Markov inequality and the Maximum inequality (Section 4.3 of Pollard, 1989),

$$\begin{aligned}
&\Pr\left(\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}\left|\frac{1}{n}\sum_{i=1}^{n}|g_{\theta,\ell}(Z_i,\theta)|^2 - \mathbb{E}_Z\left[|g_{\theta,\ell}(Z_i,\theta)|^2\right]\right| \geq Mn^{-\frac{1}{2}}\right) \\
\lesssim \quad & \frac{\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}|g_{\theta,\ell}(Z,\theta)|^4\right]}{M^2} \lesssim \frac{1}{M^2}
\end{aligned} \tag{3.25}$$

which implies that (3.24) holds. Using Assumption (3.4) and the result in (3.24), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{j,n})\right|^2 \\
=\ &\frac{1}{n}\sum_{i=1}^{n}\left|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{j,n})\right|^2 - \mathbb{E}_Z\left[\left|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{j,n})\right|^2\right] + \mathbb{E}_Z\left[\left|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{j,n})\right|^2\right] \\
=\ &O_p(1) \text{ for any } \ell \text{ and any } j=1,2. \qquad (3.26)
\end{aligned}$$

Using the results in (3.22), (3.23), (3.26), and Assumption 3.1(i), we have

$$\frac{1}{n}\sum_{i=1}^{n} g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})g_k(Z_i,\widetilde{\theta}_{1,n})(\dot{\theta}_n-\theta_o) = O_p(n^{-\frac{1}{2}}). \qquad (3.27)$$

Similarly, we can show that

$$\frac{1}{n}\sum_{i=1}^{n} g_{\theta,k}(Z_i,\widetilde{\theta}_{1,n})g_\ell(Z_i,\widetilde{\theta}_{1,n})(\dot{\theta}_n-\theta_o) = O_p(n^{-\frac{1}{2}}). \qquad (3.28)$$

Combining the results in (3.19), (3.20), (3.21), (3.27), and (3.28), we have

$$\left\|\dot{\Omega}_{S,n}-\Omega_S\right\| = O_p(n^{-\frac{1}{2}}). \qquad (3.29)$$

By the Jensen's inequality and assumption (3.4),

$$\|\Gamma_S\|^2 = \sum_{\ell\in S}\|\mathbb{E}\left[g_{\theta,\ell}(Z,\theta_o)\right]\|^2 \le \sum_{\ell\in S}\mathbb{E}\left[\|g_{\theta,\ell}(Z,\theta_o)\|^2\right] < C. \qquad (3.30)$$

Now, using Assumption 3.1(iv), the results in (3.11), (3.18), (3.29), and (3.30), we deduce that

$$\left\|\dot{V}_{n,S}-V_S\right\| = O_p(n^{-\frac{1}{2}}). \qquad (3.31)$$

Similarly, we can show that $\left\|\dot{V}_{n,S+\ell}-V_{S+\ell}\right\| = O_p(n^{-\frac{1}{2}})$ for any $\ell\in D$, which together with (3.31) and the inequality in (3.9) implies that $\left|\dot{\mu}_{n,\ell}-\mu_{o,\ell}\right| = O_p(n^{-\frac{1}{2}})$ for any $\ell\in D$. ■

Proof of Lemma 3.2. To obtain the desired result, we use the inequality in (3.9). Note that we have shown $||\dot{V}_{n,S}-V_S|| = O_p(n^{-\frac{1}{2}})$ in the proof above. Let $\tau_n=\sqrt{k_n/n}$. It remains to show

$$\max_{\ell\in D}\left\|\dot{V}_{n,S+\ell}-V_{S+\ell}\right\| = O_p(\tau_n). \qquad (3.32)$$

To this end, we write the equality

$$\begin{aligned}
\dot{V}_{n,S+\ell} - V_{S+\ell} &= \left(\dot{\Gamma}_{S+\ell,n} - \Gamma_{S+\ell}\right)' \dot{\Omega}_{S+\ell,n}^{-1} \left(\dot{\Gamma}_{S+\ell,n} - \Gamma_{S+\ell}\right) \\
&\quad + \Gamma_{S+\ell}' \dot{\Omega}_{S+\ell,n}^{-1} \left(\dot{\Gamma}_{S+\ell,n} - \Gamma_{S+\ell}\right) \\
&\quad + \left(\dot{\Gamma}_{S+\ell,n} - \Gamma_{S+\ell}\right)' \dot{\Omega}_{S+\ell,n}^{-1} \Gamma_{S+\ell} \\
&\quad - \Gamma_{S+\ell}' \dot{\Omega}_{S+\ell,n}^{-1} \left(\dot{\Omega}_{S+\ell,n} - \Omega_{S+\ell}\right) \Omega_{S+\ell}^{-1} \Gamma_{S+\ell}. \qquad (3.33)
\end{aligned}$$

Below we show (i) $\max_{\ell \in D} ||\dot{\Gamma}_{S+\ell,n} - \Gamma_{S+\ell}|| = O_p(\tau_n)$; (ii) $\max_{\ell \in D} ||\dot{\Omega}_{S+\ell,n} - \Omega_{S+\ell}|| = O_p(\tau_n)$; (iii) $\min_{\ell \in D} \rho_{\min}(\dot{\Omega}_{S+\ell,n}) \geq C^{-1}$ w.p.a.1; and (iv) $\max_{\ell \in D} \|\Gamma_{S+\ell}\| \leq C$. They are sufficient to obtain $\max_{\ell \in D} ||\dot{V}_{n,S+\ell} - V_{S+\ell}|| = O_p(\tau_n)$ by the triangle inequality and the Cauchy-Schwarz inequality.

First, by the triangle inequality, we get

$$\left\|\dot{\Gamma}_{S+\ell,n} - \Gamma_{S+\ell}\right\| \leq \left\|\Gamma_{S+\ell,n}(\dot{\theta}_n) - \Gamma_{S+\ell,n}(\theta_o)\right\| + \left\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\right\|. \qquad (3.34)$$

Recall that by definition, $\dot{\Gamma}_{S+\ell,n} = \Gamma_{S+\ell,n}(\dot{\theta}_n)$, where the subscript n indicates sample average, and $\Gamma_{S+\ell} = \Gamma_{S+\ell}(\theta_o)$, which is the population version evaluated at the true value. Using the Bonferroni inequality and the Markov inequality, we deduce that

$$\begin{aligned}
& \Pr\left(\max_{\ell \in D} \|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\| \geq M\tau_n\right) \\
\leq & \sum_{\ell \in D} \Pr\left(\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\| \geq M\tau_n\right) \\
\leq & \sum_{\ell \in D} \frac{\mathbb{E}\left[\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\|^2\right]}{M^2\tau_n^2} \qquad (3.35)
\end{aligned}$$

where M is any positive constant. As $k_n = k_0 + d_D$ and $\Gamma_{S+\ell,n}(\theta_o)$ and $\Gamma_{S+\ell}(\theta_o)$ only have (k_0+1) columns, we obtain

$$\begin{aligned}
& \sum_{\ell \in D} \frac{\mathbb{E}\left[\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\|^2\right]}{M^2\tau_n^2} \\
= & \sum_{\ell \in D} \frac{n\mathbb{E}\left[\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\|^2\right]}{M^2 k_n} \\
\leq & \frac{\max_{\ell \in D} \mathbb{E}\left[n\left\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\right\|^2\right]}{M^2} \\
\leq & \frac{(k_0+1)\max_{\ell \leq k_n} \mathbb{E}\left[\|g_{\theta,\ell}(Z,\theta_o)\|^2\right]}{M^2} \qquad (3.36)
\end{aligned}$$

where the first inequality holds by $d_D \leq k_n$ and the second inequality holds because the observations are i.i.d. and $n\mathbb{E}[\,\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\|^2]$ is bounded by $(k_0+1)\max_{\ell\leq k_n}\mathbb{E}[\,\|g_{\theta,\ell}(Z,\theta_o)\|^2]$. Combining the results in (3.35), (3.36), and assumption (3.6) yields

$$\Pr\left(\max_{\ell\in D}\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\| \geq M\tau_n\right) \lesssim \frac{1}{M}. \tag{3.37}$$

As M can be sufficiently large, we get

$$\max_{\ell\in D}\|\Gamma_{S+\ell,n}(\theta_o) - \Gamma_{S+\ell}(\theta_o)\| = O_p(\tau_n). \tag{3.38}$$

As d_θ is a fixed integer, for the ease of notation, we assume that $d_\theta = 1$ in the rest of the proof. When $d_\theta > 1$, a complete proof can be conducted by applying the same argument element by element. To show $\max_{\ell\in D}\left\|\Gamma_{S+\ell,n}(\dot{\theta}_n) - \Gamma_{S+\ell,n}(\theta_o)\right\| = O_p(\tau_n)$, we note that by the mean value theorem,

$$\frac{1}{n}\sum_{i=1}^n g_{\theta,\ell}(Z_i,\dot{\theta}_n) - \frac{1}{n}\sum_{i=1}^n g_{\theta,\ell}(Z_i,\theta_o) = \frac{1}{n}\sum_{i=1}^n g_{\theta\theta,\ell}(Z_i,\widetilde{\theta}_n)(\dot{\theta}_n - \theta_o) \tag{3.39}$$

for any ℓ, where $\widetilde{\theta}_n$ is some value between $\dot{\theta}_n$ and θ_o. We can write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n g_{\theta\theta,\ell}(Z_i,\widetilde{\theta}_n)(\dot{\theta}_n - \theta_o) \\
=\ &\left[\frac{1}{n}\sum_{i=1}^n g_{\theta\theta,\ell}(Z_i,\widetilde{\theta}_n) - \mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z,\widetilde{\theta}_n)\right]\right](\dot{\theta}_n - \theta_o) \\
&+\mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z,\widetilde{\theta}_n)\right](\dot{\theta}_n - \theta_o).
\end{aligned} \tag{3.40}$$

By the Cauchy-Schwarz inequality, the Jensen's inequality, and Assumptions 3.1(i), and (3.7),

$$\begin{aligned}
&\max_{\ell\leq k_n}\left\|\mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z,\widetilde{\theta}_n)\right](\dot{\theta}_n - \theta_o)\right\| \\
\leq\ &\left\|\dot{\theta}_n - \theta_o\right\|\max_{\ell\leq k_n}\mathbb{E}_Z\left[\left\|g_{\theta\theta,\ell}(Z,\widetilde{\theta}_n)\right\|\right] \\
\leq\ &\left\|\dot{\theta}_n - \theta_o\right\|\max_{\ell\leq k_n}\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}\|g_{\theta\theta,\ell}(Z,\theta)\|\right] = O_p(n^{-\frac{1}{2}}).
\end{aligned} \tag{3.41}$$

We next show that

$$\max_{\ell\leq k_n}\left[\frac{1}{n}\sum_{i=1}^n g_{\theta\theta,\ell}(Z_i,\widetilde{\theta}_n) - \mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z,\widetilde{\theta}_n)\right]\right] = O_p(\tau_n). \tag{3.42}$$

Using (3.7), the Bonferroni inequality, the Maximum inequality (Section 4.3 of Pollard, 1989) and the Markov inequality,

$$\begin{aligned}
&\Pr\left(\max_{\ell\leq k_n}\left|\frac{1}{n}\sum_{i=1}^{n} g_{\theta\theta,\ell}(Z_i,\widetilde{\theta}_n)-\mathbb{E}_Z\left[g_{\theta\theta,\ell}(Z,\widetilde{\theta}_n)\right]\right|\geq M\tau_n\right)\\
\leq\ & \sum_{\ell\leq k_n}\Pr\left(\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}\left|\frac{1}{n}\sum_{i=1}^{n} g_{\theta\theta,\ell}(Z_i,\theta)-\mathbb{E}\left[g_{\theta\theta,\ell}(Z,\theta)\right]\right|\geq M\tau_n\right)\\
\leq\ & \sum_{\ell\leq k_n}\frac{\mathbb{E}\left[\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|\leq\delta\}}\left|g_{\theta\theta,\ell}(Z,\theta)\right|^2\right]}{M^2k_n}\lesssim\frac{1}{M^2} \qquad (3.43)
\end{aligned}$$

which shows (3.42). Using Assumptions 3.1(i), the results in (3.40), (3.41), and (3.42), we get

$$\max_{\ell\leq k_n}\left\|\frac{1}{n}\sum_{i=1}^{n} g_{\theta\theta,\ell}(Z_i,\widetilde{\theta}_n)(\dot{\theta}_n-\theta_o)\right\|=O_p(\tau_n). \qquad (3.44)$$

Combining the results in (3.39) and (3.44) yields

$$\begin{aligned}
&\max_{\ell\in D}\left\|\Gamma_{S+\ell,n}(\dot{\theta}_n)-\Gamma_{S+\ell,n}(\theta_o)\right\|\\
\leq\ & (k_0+1)\max_{\ell\leq k_n}\left\|\frac{1}{n}\sum_{i=1}^{n} g_{\theta\theta,\ell}(Z_i,\widetilde{\theta}_n)(\dot{\theta}_n-\theta_o)\right\|=O_p(\tau_n), \qquad (3.45)
\end{aligned}$$

which together with (3.38) further implies that

$$\max_{\ell\in D}\left\|\dot{\Gamma}_{S+\ell,n}-\Gamma_{S+\ell}\right\|=O_p(\tau_n). \qquad (3.46)$$

Second, by the triangle inequality, we get

$$\left\|\dot{\Omega}_{S+\ell,n}-\Omega_{S+\ell}\right\|\leq\left\|\Omega_{S+\ell,n}(\dot{\theta}_n)-\Omega_{S+\ell,n}(\theta_o)\right\|+\left\|\Omega_{S+\ell,n}(\theta_o)-\Omega_{S+\ell}(\theta_o)\right\|. \qquad (3.47)$$

By Assumption 3.2(iii) and similar arguments used in showing (3.38), we have

$$\max_{\ell\in D}\left\|\Omega_{S+\ell,n}(\theta_o)-\Omega_{S+\ell}(\theta_o)\right\|=O_p(\tau_n). \qquad (3.48)$$

Next we show $\max_{\ell\in D}\left\|\Omega_{S+\ell,n}(\dot{\theta}_n)-\Omega_{S+\ell,n}(\theta_o)\right\|=O_p(\tau_n)$. By the mean value theorem, for any ℓ, $j\leq k_n$, there is

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}g_\ell(Z_i,\dot{\theta}_n)g_j(Z_i,\dot{\theta}_n)-\frac{1}{n}\sum_{i=1}^{n}g_\ell(Z_i,\theta_o)g_j(Z_i,\theta_o)\\
=\;&\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\widetilde{\theta}_n)(\dot{\theta}_n-\theta_o)\\
&+\frac{1}{n}\sum_{i=1}^{n}g_{\theta,j}(Z_i,\widetilde{\theta}_n)g_\ell(Z_i,\widetilde{\theta}_n)(\dot{\theta}_n-\theta_o).
\end{aligned}\tag{3.49}$$

Using the mean value theorem and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\widetilde{\theta}_n)-\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\theta_o)\right\|\\
=\;&\left\|\frac{1}{n}\sum_{i=1}^{n}g_{\theta,j}(Z_i,\widetilde{\theta}_n)g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})\right\|\left\|\widetilde{\theta}_n-\theta_o\right\|\\
\leq\;&\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,j}(Z_i,\widetilde{\theta}_n)\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})\right\|^2}\left\|\widetilde{\theta}_n-\theta_o\right\|.
\end{aligned}\tag{3.50}$$

Using similar arguments in showing (3.42), but replacing assumption (3.7) with assumption (3.6), we have

$$\max_{j\leq k_n}\left[\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,j}(Z_i,\widetilde{\theta}_n)\right\|^2-\mathbb{E}_Z\left[\left\|g_{\theta,j}(Z_i,\widetilde{\theta}_n)\right\|^2\right]\right]=O_p(\tau_n)\tag{3.51}$$

which together with Assumptions 3.1(i) and (3.6) further implies that

$$\begin{aligned}
&\max_{j\leq k_n}\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,j}(Z_i,\widetilde{\theta}_n)\right\|^2\left\|\widetilde{\theta}_n-\theta_o\right\|\\
=\;&\max_{j\leq k_n}\left|\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,j}(Z_i,\widetilde{\theta}_n)\right\|^2-\mathbb{E}_Z\left[\left\|g_{\theta,j}(Z_i,\widetilde{\theta}_n)\right\|^2\right]\right|\left\|\widetilde{\theta}_n-\theta_o\right\|\\
&+\max_{j\leq k_n}\mathbb{E}_Z\left[\left\|g_{\theta,j}(Z_i,\widetilde{\theta}_n)\right\|^2\right]\left\|\widetilde{\theta}_n-\theta_o\right\|\\
=\;&O_p(\tau_n).
\end{aligned}\tag{3.52}$$

By the same arguments, we have

$$\max_{\ell\leq k_n}\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_{1,n})\right\|^2\left\|\widetilde{\theta}_n-\theta_o\right\|=O_p(\tau_n).\tag{3.53}$$

Combining the results in (3.50), (3.52), and (3.53), we have

$$\max_{\ell,j\leq k_n}\left\|\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\widetilde{\theta}_n)-\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\theta_o)\right\|=O_p(\tau_n). \tag{3.54}$$

By the Cauchy-Schwarz inequality,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\theta_o)\right\|\leq\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}g_j^2(Z_i,\theta_o)}. \tag{3.55}$$

By Assumption 3.2(iii), the Bonferroni inequality and the Markov inequality,

$$\max_{j\leq k_n}\frac{1}{n}\sum_{i=1}^{n}g_j^2(Z_i,\theta_o)=O_p(1). \tag{3.56}$$

Using the result in (3.51) and assumption (3.6), we have

$$\begin{aligned}\max_{\ell\leq k_n}\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)\right\|^2 &\leq \max_{\ell\leq k_n}\left[\frac{1}{n}\sum_{i=1}^{n}\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)\right\|^2-\mathbb{E}_Z\left[\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)\right\|^2\right]\right]\\ &\quad+\max_{\ell\leq k_n}\mathbb{E}_Z\left[\left\|g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)\right\|^2\right]\\ &= O_p(1)\end{aligned} \tag{3.57}$$

which together with (3.55) and (3.56) implies that

$$\max_{\ell,j\leq k_n}\left\|\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\theta_o)\right\|=O_p(1). \tag{3.58}$$

Under Assumption 3.1(i) and (3.58),

$$\max_{\ell,j\leq k_n}\left\|\frac{1}{n}\sum_{i=1}^{n}g_{\theta,\ell}(Z_i,\widetilde{\theta}_n)g_j(Z_i,\theta_o)\right\|\left\|\dot{\theta}_n-\theta_o\right\|=O_p(\tau_n). \tag{3.59}$$

By the same arguments,

$$\max_{\ell,j\leq k_n}\left\|\frac{1}{n}\sum_{i=1}^{n}g_{\theta,j}(Z_i,\widetilde{\theta}_n)g_\ell(Z_i,\theta_o)\right\|\left\|\dot{\theta}_n-\theta_o\right\|=O_p(\tau_n). \tag{3.60}$$

Combining the results in (3.49), (3.59) and (3.60), we have

$$\max_{\ell,j\leq k_n}\left|\frac{1}{n}\sum_{i=1}^{n}g_\ell(Z_i,\dot{\theta}_n)g_j(Z_i,\dot{\theta}_n)-\frac{1}{n}\sum_{i=1}^{n}g_\ell(Z_i,\theta_o)g_j(Z_i,\theta_o)\right|=O_p(\tau_n) \tag{3.61}$$

which together with the fact that $\Omega_{S+\ell,n}(\theta)$ only has $(k_0+1)^2$ many components implies that

$$\max_{\ell\in D}\left\|\Omega_{S+\ell,n}(\dot{\theta}_n)-\Omega_{S+\ell,n}(\theta_o)\right\|=O_p(\tau_n). \tag{3.62}$$

This combined with (3.47) and (3.48) further yields

$$\max_{\ell\in D}\left\|\dot{\Omega}_{S+\ell,n}-\Omega_{S+\ell}\right\|=O_p(\tau_n). \tag{3.63}$$

Using the Weyl's eigenvalue perturbation theorem, we have

$$\left|\rho_{\min}\left(\dot{\Omega}_{S+\ell,n}\right)-\rho_{\min}\left(\Omega_{S+\ell}\right)\right|\leq\left\|\dot{\Omega}_{S+\ell,n}-\Omega_{S+\ell}\right\|\leq\max_{\ell\in D}\left\|\dot{\Omega}_{S+\ell,n}-\Omega_{S+\ell}\right\| \tag{3.64}$$

which implies that

$$\rho_{\min}\left(\dot{\Omega}_{S+\ell,n}\right)\geq\rho_{\min}\left(\Omega_{S+\ell}\right)-\max_{\ell\in D}\left\|\dot{\Omega}_{S+\ell,n}-\Omega_{S+\ell}\right\|=\rho_{\min}\left(\Omega_{S+\ell}\right)-O_p(\tau_n). \tag{3.65}$$

Combining the above inequality with Assumption 3.2(i), we have

$$\min_{\ell\in D}\rho_{\min}\left(\dot{\Omega}_{S+\ell,n}\right)\geq\min_{\ell\in D}\rho_{\min}\left(\Omega_{S+\ell}\right)-O_p(\tau_n)\geq\frac{1}{2C}\text{ w.p.a.1.} \tag{3.66}$$

Under (3.6) and the fact that k_0+1 is a fixed integer, we deduce that

$$\begin{aligned}\max_{\ell\in D}\left\|\Gamma_{S+\ell}\right\|^2 &= \max_{\ell\in D}\sum_{j\in S\cup\{\ell\}}\left\|\mathbb{E}\left[g_{\theta,j}(Z,\theta_o)\right]\right\|^2\\ &\leq (k_0+1)\max_{j\leq k_n}\left\|\mathbb{E}\left[g_{\theta,j}(Z,\theta_o)\right]\right\|^2\leq C. \end{aligned}\tag{3.67}$$

This completes the proof. ■

4 Some Extra Simulation Results

Table S-1. Finite Sample Bias (BS), Standard Deviations (SD) and RMSEs (RE) with $K = 10$

	Automatic Estimate						Conservative GMM Estimate					
	n=250			n=2500			n=250			n=2500		
	BS	SD	RE	BS	SD	RE	BS	SD	RE	BS	SD	RE
(.1 .2)	.0032	.0825	.0825	.0001	.0248	.0248	-.0035	.2604	.2604	.0007	.0786	.0786
(.1 .5)	.0028	.0851	.0852	.0001	.0248	.0248	-.0035	.2604	.2604	.0007	.0786	.0786
(.3 .2)	.0043	.0816	.0817	.0002	.0232	.0232	-.0012	.1609	.1609	.0003	.0501	.0501
(.3 .5)	.0043	.0816	.0817	.0002	.0232	.0232	-.0012	.1609	.1609	.0003	.0501	.0501
	Pooled GMM Estimate						Aggressive GMM Estimate					
	n=250			n=2500			n=250			n=2500		
	BS	SD	RE	BS	SD	RE	BS	SD	RE	BS	SD	RE
(.1 .2)	.0057	.0788	.0790	.0005	.0246	.0246	.1077	.0787	.1334	.1034	.0245	.1063
(.1 .5)	.0057	.0788	.0790	.0005	.0246	.0246	.1394	.0788	.1601	.1364	.0244	.1386
(.3 .2)	.0049	.0734	.0736	.0004	.0230	.0230	.0939	.0734	.1192	.0902	.0229	.0930
(.3 .5)	.0049	.0734	.0736	.0004	.0230	.0230	.1216	.0734	.1421	.1190	.0228	.1212
	Post-Shrinkage GMM Estimate						Oracle GMM Estimate					
	n=250			n=2500			n=250			n=2500		
	BS	SD	RE	BS	SD	RE	BS	SD	RE	BS	SD	RE
(.1 .2)	.0033	.0814	.0815	.0001	.0247	.0247	.0020	.0790	.0790	.0001	.0246	.0246
(.1 .5)	.0029	.0839	.0840	.0001	.0247	.0247	.0020	.0790	.0790	.0001	.0246	.0246
(.3 .2)	.0057	.0896	.0898	.0003	.0236	.0236	.0017	.0735	.0735	.0001	.0230	.0230
(.3 .5)	.0057	.0897	.0898	.0003	.0236	.0236	.0017	.0735	.0735	.0001	.0230	.0230

(i) The "automatic" estimation is obtained simultaneously with moment selection. (ii) The "conservative" estimation uses Z_S. (iii) The "pooled" estimation uses all valid IVs, including Z_S, Z_A, and Z_{B_0}. (iv) The "aggressive" estimation uses all available IVs, including invalid ones. (v) The "post-shrinkage" estimation uses Z_S plus IVs selected by the shrinkage procedure. (vi) The "oracle" estimation uses Z_S and Z_A.

Table S-2. Finite Sample Bias (BS), Standard Deviations (SD) and RMSEs (RE) with $K = 30$

	Automatic Estimate						Conservative GMM Estimate					
	n=250			n=2500			n=250			n=2500		
	BS	SD	RE	BS	SD	RE	BS	SD	RE	BS	SD	RE
(.1 .2)	.0044	.0807	.0808	-.0005	.0249	.0249	-.0046	.2618	.2618	-.0005	.0771	.0771
(.1 .5)	.0041	.0808	.0809	-.0005	.0249	.0249	-.0046	.2618	.2618	-.0005	.0771	.0771
(.3 .2)	.0045	.0794	.0796	-.0004	.0232	.0232	-.0021	.1627	.1628	-.0000	.0491	.0491
(.3 .5)	.0045	.0794	.0796	-.0004	.0232	.0232	-.0021	.1627	.1628	-.0000	.0491	.0491
	Pooled GMM Estimate						Aggressive GMM Estimate					
	n=250			n=2500			n=250			n=2500		
	BS	SD	RE	BS	SD	RE	BS	SD	RE	BS	SD	RE
(.1 .2)	.0147	.0767	.0781	.0009	.0247	.0248	.1792	.0765	.1949	.1752	.0249	.1770
(.1 .5)	.0147	.0767	.0781	.0009	.0247	.0248	.1924	.0765	.2071	.1902	.0249	.1918
(.3 .2)	.0128	.0714	.0726	.0008	.0230	.0230	.1567	.0714	.1722	.1530	.0231	.1547
(.3 .5)	.0128	.0714	.0726	.0008	.0230	.0230	.1682	.0714	.1828	.1661	.0231	.1677
	Post-Shrinkage GMM Estimate						Oracle GMM Estimate					
	n=250			n=2500			n=250			n=2500		
	BS	SD	RE	BS	SD	RE	BS	SD	RE	BS	SD	RE
(.1 .2)	.0038	.0792	.0793	-.0004	.0248	.0248	.0017	.0771	.0771	-.0004	.0248	.0248
(.1 .5)	.0039	.0796	.0797	-.0004	.0248	.0248	.0017	.0771	.0771	-.0004	.0248	.0248
(.3 .2)	.0067	.0885	.0888	-.0003	.0234	.0234	.0015	.0718	.0718	-.0003	.0230	.0230
(.3 .5)	.0067	.0885	.0888	-.0003	.0234	.0234	.0015	.0718	.0718	-.0003	.0230	.0230

(i) The "automatic" estimation is obtained simultaneously with moment selection. (ii) The "conservative" estimation uses Z_S. (iii) The "pooled" estimation uses all valid IVs, including Z_S, Z_A, and Z_{B_0}. (iv) The "aggressive" estimation uses all available IVs, including invalid ones. (v) The "post-shrinkage" estimation uses Z_S plus IVs selected by the shrinkage procedure. (vi) The "oracle" estimation uses Z_S and Z_A.

References

Bhatia, R. (1997): Matrix Analysis. New York: Springer.

Cheng, X. and Liao, Z. (2012): "Select the Valid and Relevant Moments: An Information-Based LASSO for GMM with Many Moments", Working Paper, University of Pennsylvania and University of California Los Angeles.

Pollard, D. (1989): "Asymptotics via Empirical Processes," *Statistical Science*, 4(4), 341-354.