Regression Coefficient and Autoregressive Order Shrinkage and Selection via Lasso

(Appendix E: Proof of Theorem 4)

Hansheng Wang
Guanghua School of Management, Peking University
hansheng@gsm.pku.edu.cn

Guodong Li
Department of Statistics and Actuarial Science, University of Hong Kong
ligd@hkusua.hku.hk

and Chih-Ling Tsai
Graduate School of Management, University of California - Davis
cltsai@ucdavis.edu

Step 1: To facilitate the derivations of proof, let $\lambda_{\min}\{M\}$ denote the smallest eigenvalue of the symmetric matrix $M = \{m_{ij}\}$. Furthermore, define the norm of the matrix $A_{a\times b} = \{a_{ij}\}$ as $\|A_{a\times b}\|_{\mathcal{M}} = \sum_{ij} |a_{ij}|$. Then, for matrices A and $B_{b\times c} = \{b_{ij}\}$, it can be verified that

$$\begin{aligned} \|A_{a\times b} \cdot B_{b\times c}\| &= \sum_{i=1}^{a}\sum_{j=1}^{c}\left|\sum_{k=1}^{b} a_{ik}b_{kj}\right| \le \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{k=1}^{b} \|A_{a\times b}\|_{\mathcal{M}} \cdot \|B_{b\times c}\|_{\mathcal{M}} \\ &\le \|A_{a\times b}\| \cdot \|B_{b\times c}\| \cdot (abc). \end{aligned} \tag{1}$$

Moreover, for any two vectors $v_1 = (v_{11}, \cdots, v_{1a})'$ and $v_2 = (v_{21}, \cdots, v_{2b})'$, we have

$$\begin{aligned} \left|v_1' A_{a\times b} v_2\right| &= \left|\sum_{i=1}^{a}\sum_{j=1}^{b} a_{ij}v_{1i}v_{2j}\right| \le \sum_{i=1}^{a}\sum_{j=1}^{b} |a_{ij}| \cdot |v_{1i}| \cdot |v_{2j}| \\ &\le \sum_{i=1}^{a}\sum_{j=1}^{b} \|A_{a\times b}\|_{\mathcal{M}} \cdot \|v_1\| \cdot \|v_2\| \\ &= ab\|A_{a\times b}\|_{\mathcal{M}} \cdot \|v_1\| \cdot \|v_2\|. \end{aligned} \tag{2}$$

Step 2: To prove this theorem, it suffices to show that for any $\omega \notin \mathcal{N}_0$, there is a sufficiently large integer n_ω such that for any $n > n_\omega$, we have

$$\inf_{\theta\in\mathcal{B}_\delta} \lambda_{\min}\left\{n^{-1}\ddot{L}_n(\theta)\right\} > \epsilon, \tag{3}$$

where $\ddot{L}_n(\theta) = \partial^2 L_n(\theta)/(\partial\theta\partial\theta')$ is the Hessian matrix, and $\epsilon > 0$.

Applying the strong law of large numbers, we obtain that

$$n^{-1}\ddot{L}_n(\theta^0) \to_{a.s} \mathrm{E}\left\{n^{-1}\ddot{L}_n(\theta^0)\right\} \doteq \Sigma_0,$$

where $\rightarrow_{a.s}$ denote the almost sure convergence and Σ_0 is a positive definite matrix. Next, set $\epsilon = 0.1\lambda_{\min}\{\Sigma_0\}$. Then there is a probability null set $\mathcal{N}_{0a}$ and a sufficiently large integer $n_{\omega a}$ such that for any $\omega \notin \mathcal{N}_{0a}$ and $n > n_{\omega a}$, we have

$$\lambda_{\min}\Big\{n^{-1}\ddot{L}_n(\theta^0)\Big\} > 2\epsilon. \tag{4}$$

Moreover,

$$\begin{aligned}\lambda_{\min}\{n^{-1}\ddot{L}_n(\theta)\} &= \inf_{\tilde{\theta}:\|\tilde{\theta}\|=1.0} \tilde{\theta}'\Big\{n^{-1}\ddot{L}_n(\theta)\Big\}\tilde{\theta} \\ &\geq \inf_{\tilde{\theta}:\|\tilde{\theta}\|=1.0} \tilde{\theta}'\Big\{n^{-1}\ddot{L}_n(\theta^0)\Big\}\tilde{\theta} + \inf_{\tilde{\theta}:\|\tilde{\theta}\|=1.0} \tilde{\theta}'\Big\{n^{-1}\ddot{L}_n(\theta) - n^{-1}\ddot{L}_n(\theta^0)\Big\}\tilde{\theta} \\ &\geq \lambda_{\min}\Big\{n^{-1}\ddot{L}_n(\theta^0)\Big\} + \inf_{\tilde{\theta}:\|\tilde{\theta}\|=1.0} \tilde{\theta}'\Big\{n^{-1}\ddot{L}_n(\theta) - n^{-1}\ddot{L}_n(\theta^0)\Big\}\tilde{\theta} \\ &\geq \lambda_{\min}\Big\{n^{-1}\ddot{L}_n(\theta^0)\Big\} - p^2\|n^{-1}\ddot{L}_n(\theta) - n^{-1}\ddot{L}_n(\theta^0)\|_{\mathcal{M}}^2, \end{aligned} \tag{5}$$

where the last inequality is due to (2).

Finally we need to show that there is a probability null set $\mathcal{N}_{0b}$ and a sufficiently large number $n_{\omega b}$ such that for any $\omega \notin /\mathcal{N}_{0b}$ and $n > n_{\omega b}$, we have

$$\sup_{\theta\in\mathcal{B}_\delta} \|n^{-1}\ddot{L}_n(\theta) - n^{-1}\ddot{L}_n(\theta^0)\|_{\mathcal{M}} < \sqrt{\epsilon}/p. \tag{6}$$

The result of (6) together with (4) and (5) implies that (3) holds by setting the probability null set $\mathcal{N}_0 = \mathcal{N}_{0a} \cup \mathcal{N}_{0b}$ and $n_\omega = \max\{n_{\omega a}, n_{\omega,b}\}$. This completes the proof.

Step 3: In the last step, we show that (6) holds, which is implied by the following three equations:

$$\sup_{\theta\in\mathcal{B}_\delta}\left\|\frac{1}{n}\frac{\partial^2 L_n(\theta)}{\partial\beta\partial\beta'} - \frac{1}{n}\frac{\partial L^2(\theta^0)}{\partial\beta\partial\beta'}\right\|_{\mathcal{M}} < \frac{\sqrt{\epsilon}}{4p} \quad a.s., \tag{7}$$

$$\sup_{\theta\in\mathcal{B}_\delta}\left\|\frac{1}{n}\frac{\partial^2 L_n(\theta)}{\partial\beta\partial\phi'} - \frac{1}{n}\frac{\partial L^2(\theta^0)}{\partial\beta\partial\phi'}\right\|_{\mathcal{M}} < \frac{\sqrt{\epsilon}}{4p} \quad a.s., \tag{8}$$

$$\sup_{\theta\in\mathcal{B}_\delta}\left\|\frac{1}{n}\frac{\partial^2 L_n(\theta)}{\partial\phi\partial\phi'} - \frac{1}{n}\frac{\partial L^2(\theta^0)}{\partial\phi\partial\phi'}\right\|_{\mathcal{M}} < \frac{\sqrt{\epsilon}}{4p} \quad a.s.. \tag{9}$$

Because the proofs of (7), (8), and (9) are very similar, we only provide the detailed proof for (7). Let $\phi_0 = -1$, $\tilde{\phi} = (\phi_0, \phi')'$, $\tilde{\phi}^0 = (\phi_0, \phi^{0\prime})'$ $Y_t = (y_t, y_{t-1}, \cdots, y_{t-q})'$, and $X_t = (x_t, x_{t-1}, \cdots, x_{t-q})'$. Then, we have

$$n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\beta\partial\beta'} = \frac{1}{n}\sum_{t=q+1}^{n_0}\Big[X_t'\tilde{\phi}\tilde{\phi}'X_t\Big].$$

As a result,

$$\left\|n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\beta\partial\beta'} - n^{-1}\frac{\partial^2 L_n(\theta^0)}{\partial\beta\partial\beta'}\right\|_{\mathcal{M}} \leq R_1 + R_2, \tag{10}$$

where $R_1 = n^{-1}\sum_{t=q+1}^{n}\left\|X_t'\tilde{\phi}(\tilde{\phi}-\tilde{\phi}^0)'X_t\right\|$ and $R_2 = n^{-1}\sum_{t=q+1}^{n}\left\|X_t'(\tilde{\phi}-\tilde{\phi}^0)\tilde{\phi}^{0\prime}X_t\right\|$. Applying the inequality (1), we have

$$\begin{aligned} R_1 &\le n^{-1}\sum_{t=q+1}^{n}\|X_t'\tilde{\phi}\|_{\mathcal{M}}\cdot\|(\tilde{\phi}-\tilde{\phi}^0)'X_t\|_{\mathcal{M}}\cdot p^2 \\ &\le n^{-1}\sum_{t=q+1}^{n}\|X_t'\|_{\mathcal{M}}\|\tilde{\phi}\|_{\mathcal{M}}\cdot\|(\tilde{\phi}-\tilde{\phi}^0)\|_{\mathcal{M}}\|X_t\|_{\mathcal{M}}\cdot p^4. \end{aligned}$$

Because $\|\tilde{\phi}\|_{\mathcal{M}} = \sum_{j=0}^{q}|\tilde{\phi}_j| \le q\|\tilde{\phi}\|$, the right hand side of the above inequality can be further bounded by

$$\begin{aligned} &\le n^{-1}\sum_{t=q+1}^{n}\|X_t'\|_{\mathcal{M}}^2(q\|\tilde{\phi}\|)\cdot(q\|(\tilde{\phi}-\tilde{\phi}^0)\|)\cdot p^4 \\ &\le n^{-1}\sum_{t=q+1}^{n}\|X_t'\|_{\mathcal{M}}^2\Big\{\|\tilde{\phi}^0\|+\|\tilde{\phi}^0-\tilde{\phi}\|\Big\}\cdot\|\tilde{\phi}-\tilde{\phi}^0\|\cdot q^2p^4. \end{aligned}$$

By the definition of $\mathcal{B}_\delta$, we know that for any $\theta \in \mathcal{B}_\delta$, we have $\|\tilde{\phi}-\tilde{\phi}^0\| \le \|\theta-\theta^0\| \le \delta$. Hence, the right hand side of the above inequality is bounded by

$$\left(n^{-1}\sum_{t=q+1}^{n}\|X_t'\|_{\mathcal{M}}^2\right)\|\tilde{\phi}^0\|q^2p^4\delta + \left(n^{-1}\sum_{t=q+1}^{n}\|X_t'\|_{\mathcal{M}}^2\right)q^2p^4\delta^2. \tag{11}$$

Applying the strong law of the large number, we know that $n^{-1}\sum_{t=q+1}^{n}\|X_t'\|_{\mathcal{M}}^2 \to_{a.s} \mathrm{E}\|X_t'\|^2 < \infty$. Moreover, $\|\tilde{\phi}^0\|^2$ and (p,q) are constants and do not change with respect to the choice of $\theta \in \mathcal{B}_\delta$. This allows us to choose a sufficiently small δ such that

$$\mathrm{E}\Big(\|X_t'\|_{\mathcal{M}}^2\Big)\times\|\tilde{\phi}^0\|q^2p^4\delta + \mathrm{E}\Big(\|X_t'\|_{\mathcal{M}}^2\Big)\times q^2p^4\delta^2 < \frac{0.5\sqrt{\epsilon}}{4p}.$$

Hence, the quantity (11) is bounded by $0.5\sqrt{\epsilon}/(4p)$ almost surely. Because the quantity (11) is independent of the choice of $\theta \in \mathcal{B}_\delta$, we have that $\sup_{\theta\in\mathcal{B}_\delta} R_1 < 0.5\sqrt{\epsilon}/(4p)$ almost surely. Analogously, we can show that $\sup_{\theta\in\mathcal{B}_\delta} R_2 < 0.5\sqrt{\epsilon}/(4p)$. Consequently, both upper bound findings together with (10) imply (7). Similar techniques can be used to establish (8) and (9), which completes the proof of (6).