PAPER • OPEN ACCESS

Asymptotic for Lasso Estimator in High-Dimensional Repeated Measurements Model

To cite this article: Naser Oda Jassim and Andul Hussein Saber Al-Mouel 2021 *J. Phys.: Conf. Ser.* **1818** 012039

Asymptotic for Lasso Estimator in High-Dimensional Repeated Measurements Model

Naser Oda Jassim[1], Andul Hussein Saber Al-Mouel[2]

Department of Mathematics, College of Education, for Pure Science, University of Basrah, Basrah, Iraq

Emails: [1]noj1972.khalef@gmail.com , [2]abdulhusseinsaber@yahoo.com

Abstract. In this paper, we propose high-dimensional repeated measurements model and using the Bridge estimator as penalized method that minimizes the residual sum of squares plus penalty term $\sum|\theta_j|^\gamma$. After that, under appropriate conditions, we discuss the consistency and asymptotic behavior of lasso estimator when $\gamma = 1$ as especial case and also study the consistency and limiting distribution of the Bridge estimators when $\gamma < 1$ and $\gamma > 1$. Moreover, we discuss the asymptotic of estimators by using small parameter and local asymptotic. In other words, we discuss the asymptotic behavior in a triangular array of observations.

1.Introduction

High-dimensional statistical problem can be considered as results of the large amount of data gathered today such as spectra, biomedical data, financial data, images which are described by hundreds or thousands of attributes. The relationship between the unknown parameters denoted by k which are to be estimated and the sample size denoted by n reflects the type of data whether it is high-dimensional or low-dimensional. In the sense that, when the number of parameters is larger than sample size then we have "high-dimensional" data. On the other hand, when the number of parameter is less than sample size then we have "low-dimensional" data.

Modelling high-dimensional data is challenging because of some reasons**.** One of these mainly is ordinary least squares estimator is not unique and will heavily over-fit the data. In this case, if this method has such problem so it is possible to use totally different method, for example penalized least squares methods. These methods are common and suitable to treat with the high-dimensional data when the number of explanatory variables is larger than the sample size. The penalized least squares method is used to overcome the computational problems in the high-dimensional data and it is also improved the prediction accuracy by making estimator and variable selection simultaneously. It is based on the principle that minimizing the sum of squares error with some constraints on the parameters. It can be obtained the penalized least squares estimators by minimizing the objective function which contains two parts a loss function and a penalty function. The best penalty function gives estimator that describes by three properties: unbiasedness, sparsity and continuity. It is also the estimator from the ideal aspect must have oracle properties which are consistency and asymptotic normal. One of these penalized least squares methods is called "lasso", proposed by Tibshirani (1996). This method is considered as an especial case when $\gamma = 1$ in a penalized regression method called "Bridge regression" proposed by Fran and Friedman (1993), as in the following form

$$\hat{\beta} = arg\min_{\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - X'_{ij}\beta_j\right)^2 + \frac{\lambda_n}{N}\sum_{j=1}^{k}\left|\beta_j\right|^{\gamma}\right\},$$

where λ is penalty parameter and $\frac{\lambda_n}{n}\sum_{j=1}^{k}\left|\beta_j\right|^{\gamma}$ is penalty function. The lasso method is based on the idea that minimizing the residual sum of squares plus the sum of absolute value of coefficients. In Section 2, we define our model with its some properties. In section 3, we assume that some important assumptions on the design matrix and also investigate the convexity of lasso estimator with some important definitions. In section 4, we discuss the asymptotic consistency of Bridge estimators with lasso estimator as especial case when $\gamma = 1$. Finally, in section 5, we discuss the finite sample behavior by considering " small true parameters".

2. Setting Model

Consider the response random variables $Y_{1t}, \dots, Y_{Nt}$ are generated by a repeated measurements linear model at time $t, t = 1, \dots, T$

$$Y_{it} = \mu + X'_{j(it)}\,\beta_j + v_i + \epsilon_{it}, \qquad (1)$$

for observed explanatory variables $\{x_{j(it)}\}$, unknown fixed parameters $\mu\ and\ \beta_j \in \mathbb{R}^k$, unknown errors ϵ_{it} with $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and a random effect v_i with $v_i \sim N(0, \sigma_v^2)$. The model in (1) can be rewritten as

$$Y_{it} = \mu + \sum_{j=1}^{k}\beta_j X_{j(it)} + \omega_{it}, \qquad (2)$$

Where $\omega_{it} = v_i + \epsilon_{it}$ with $\omega_{it} \sim N(0, \sigma_\omega^2)$, $\sigma_\omega^2 = \sigma_v^2 + \sigma_\epsilon^2$.

Assuming that the explanatory variables are centered to have mean 0 and put $\mu = \bar{Y}$ in this case, it can be change Y_{it} in (2) by $Y_{it} - \bar{Y}$ and concentrate on estimating β. Âgain assuming that $\bar{Y} = 0$. The model in (2) can be rewritten as

$$Y_{it} = G'_{j(it)}\theta_j + \omega_{it.} \qquad (3)$$

In matrix notation the model in (3) can be rewritten as

$$Y = G\theta + \omega\,, \qquad (4)$$

where, $Y = [Y_{11}, \dots, Y_{1T}, Y_{21}, \dots, Y_{N1}, \dots, Y_{NT}]'$ has length NT,
$G = [e\,, X]$, $e = [1, \dots 1]'$ has length NT,
$X = [X_1, \dots, X_N]'$ is a $NT \times K$ design matrix of fixed effects,
$\theta = \left[\,\mu, \beta_1, \beta_{2,}, \dots, \beta_k\right]'$ has length $k + 1$, and
$\omega = [\omega_{11}, \dots, \omega_{1T}, \omega_{21}, \dots, \omega_{N1}, \dots, \omega_{NT}]'$ has length NT.

From the model in (4), we have $Y \sim N_{NT}(G\theta, \Sigma)$, where
$E(\omega\omega') = E[(Y - G\theta)(Y - G\theta)']$
$= I_N \otimes (\sigma_\varepsilon^2 I_T + \sigma_v^2 ee')$, where $\otimes$ is Konecker product
$= \sigma_\varepsilon^2(I_N \otimes I_T) + \sigma_v^2(I_N \otimes ee')$,
replace $I_T\ by\ (E_T + J_T)$ and ee' by TJ_T, where
$J_T = \frac{1}{T}ee'$ and $E_T = I_T - J_T$, then
$\Sigma = \sigma_\varepsilon^2[I_N \otimes (E_T + J_T)] + \sigma_v^2(I_N \otimes TJ_T)$
$= \sigma_\varepsilon^2(I_N \otimes E_T) + \sigma_\varepsilon^2(I_N \otimes J_T) + T\sigma_v^2(I_N \otimes J_T)$,
by collecting terms with the same matrices, we obtain

$\Sigma = \sigma_\varepsilon^2 (I_N \otimes E_T) + (\sigma_\varepsilon^2 + T\sigma_v^2)(I_N \otimes J_T)$
$= \sigma_\varepsilon^2 Q + \sigma_1^2 P,$
where $\sigma_1^2 = (\sigma_\varepsilon^2 + T\sigma_v^2)$ and $\Sigma^{-1} = \frac{Q}{\sigma_\varepsilon^2} + \frac{P}{\sigma_1^2}$,
$\rightarrow |\Sigma| = (\sigma_\varepsilon^2)^{N(T-1)}(\sigma_1^2)^N$.

Now θ can be estimated by minimizing the penalized least squares (LS),

i.e $\hat{\theta}_N(\lambda_N) = \underset{\theta \in R^{k+1}}{\operatorname{argmin}} \left\{\frac{1}{N}\sum_{i=1}^{N}(Y_{it} - G'_{it}\ \hat{\theta}_j)^2 + \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|^\gamma\right\}$ (5)

for given penalty parameter λ_N denotes the estimator $\hat{\theta}_N$ and $\gamma > 0$. Such estimator contains two especial cases, the bridge estimator in (5) becomes ridge estimator when $\gamma = 2$ and lasso estimator when $\gamma = 1$. Also if λ_N is sufficiently large and $\gamma \le 1$, then the estimators in the penalty function $\lambda_N \sum_{j=1}^{k}|\hat{\theta}_j|^\gamma$ is exactly 0. In fact, model selection methods that penalize by some nonzero parameters can be imposed as limiting cases of bridge estimation as $\gamma \rightarrow 0$ since $\lim_{\gamma \rightarrow 0} \sum_{j=1}^{k}|\theta_j|^\gamma = \sum_{j=1}^{k} I(\theta_j \neq 0)$. In fact when $\lambda_N = 0$, so the terminology in (5) becomes $\sum_{i=1}^{N}(Y_{it} - G'_{it}\ \theta)^2$ which corresponds to ordinary least squares estimator and this estimator denotes by $\hat{\theta}_N^{(0)}$. This means that (5) becomes

$$\hat{\theta}_N^{(0)} = \underset{\theta \in R^{k+1}}{\operatorname{argmin}} \sum_{i=1}^{N}(Y_{it} - G'_{it}\ \hat{\theta}_j)^2.$$

3. The Assumptions on the Design Matrix *G*

For the design matrix proposes the following regularity conditions

i. $\lim_{N\rightarrow\infty} A_N = \lim_{N\rightarrow\infty} \frac{1}{N}\sum_{i=1}^{N} G_{it}G'_{it} \rightarrow A$, (6)

where A is a nonnegative definite matrix .
ii. The random error ω_{it} are independent and identical distributed with mean 0 and has a continuous, probability density function (p.d.f) f in a neighborhood of 0. If the matrix A in (7) is nonsingular then the parametrization of the model in (1) is unique. Define the following class
$\Omega = \{\xi : \xi = \theta + \Phi\ where\ A_N\Phi = 0\}$, where θ satisfies the relation in (4).

Under conditions (i) and (ii) and assuming that A is nonsingular then we have that the least squares estimator is consistent and that

$$\sqrt{N}\ (\hat{\theta}_N^{(0)} - \theta) \underset{d}{\rightarrow} N(0, \sigma_\omega^2 A^{-1}), \text{ where } \sigma_\omega^2 = \sigma_v^2 + \sigma_\epsilon^2 .$$

Lemma 1. Under assumption (A_1) on design matrix we have that the regularity condition

$$\frac{1}{N}\max_{1\le i\le N} G'_{it}G_{it} \rightarrow 0 \text{ , as } N \rightarrow \infty,\ t = 1, \dots, T. \quad (7)$$

Proof. Consider $\varepsilon > 0$ and $R = tr(A)$. We have $A_N \rightarrow A$ and A is nonsingular and non-negative definite, this implies $tr(A_N) = \frac{1}{N}\sum_{i=1}^{N} G'_{it}G_{it} \rightarrow R > 0$.
Therefore there exists an $N^*(\varepsilon)$ such that
$\left|\frac{1}{N}\sum_{i=1}^{N} G'_{it}G_{it} - R\right| \le \varepsilon$, $N \ge N^*$.
$-\varepsilon \le \frac{1}{N}\sum_{i=1}^{N} G'_{it}XG_{it} - R \le \varepsilon$
$R - \varepsilon \le \frac{1}{N}\sum_{i=1}^{N} G'_{it}G_{it} \le R + \varepsilon$. We have
$0 \le \frac{1}{N}\max_{N^*\le i\le N} G'_{it}G_{it}$
$= \frac{1}{N}\max_{N^*\le i\le N}\left(\sum_{r=1}^{i} G'_{it}G_{it} - \sum_{r=1}^{i-1} G'_{it}G_{it}\right)$

$\leq \frac{1}{N} \max_{N^* \leq i \leq N} \big((R+\varepsilon)i - (R-\varepsilon)(i-1)\big)$
$= \max_{N^* \leq i \leq N} \frac{R+2i\varepsilon-\varepsilon}{N}$
$\leq \max_{N^* \leq i \leq N} \frac{R+2i\varepsilon}{N} \leq 3\varepsilon$
For sufficiently large N and letting $N \to \infty$, since
$0 \leq \lim_{N\to\infty} sup \frac{1}{N} \max_{N^* \leq i \leq N} G'_{it} G_{it} \leq 3\varepsilon,$

Whereas ε>0 is arbitrary. The prove is complete.

Proposition 1. The Lasso estimator $\hat{\theta}_N$ when γ =1 in equation (5) is convex.
Proof . The Objective function in (5) when γ =1 can be written as

$\hat{\theta}_N = f(\hat{\theta}) + g(\hat{\theta})$, where
$f(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\big(y_{it} - G'_{it}\hat{\theta}_j\big)^2$ and $g(\hat{\theta}) = \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|$, $t = 1, \dots, T$.

Since $\hat{\theta} \in R^k$ then the domain of both functions f and g is convex.
To prove that $f(\hat{\theta})$ is convex, we must show that $\frac{1}{N}G'G$ is nonnegative definite matrix for any $k \times 1$ vector. In matrix notation, f can be written as

$$f(\hat{\theta}) = \frac{1}{N}\big(Y - G\hat{\theta}\big)'\big(Y - G\hat{\theta}\big)$$
$$= \frac{1}{N}\Big((\omega + G\theta) - G\hat{\theta}\Big)'\Big((\omega + G\theta) - G\hat{\theta}\Big)$$
$$= \frac{1}{N}\big(\omega + G\theta - G\hat{\theta}\big)'\big(\omega + G\theta - G\hat{\theta}\big)$$
$$= \frac{1}{N}\big(\omega + G(\theta - \hat{\theta})\big)'\big(\omega + G(\theta - \hat{\theta})\big)$$
$$= \frac{1}{N}\omega'\omega + \frac{1}{N}\omega'G\big(\theta - \hat{\theta}\big) + \frac{1}{N}\big(\theta - \hat{\theta}\big)'G'\omega + \big(\theta - \hat{\theta}\big)'G'G\big(\theta - \hat{\theta}\big)$$
$$= \frac{1}{N}\omega'\omega + \frac{1}{N}\omega'G\big(\theta - \hat{\theta}\big) + \frac{1}{N}\big(\theta - \hat{\theta}\big)'G'\omega + \big(\theta - \hat{\theta}\big)'G'G\big(\theta - \hat{\theta}\big)$$

By letting $N \to \infty$, note the following results
$\frac{1}{N}\omega'\omega \xrightarrow{p} E(\omega'\omega) = \sigma_\omega^2$ (by the law of large number)
$E\left[\frac{1}{N}\sum_{i=1}^{N}\omega_i\right] = \frac{1}{N}\sum_{i=1}^{N}E[\omega_i] = 0$, therefore
$2E\left[\frac{1}{N}\sum_{i=1}^{N}\omega_i\right]G'\big(\theta - \hat{\theta}\big) = 0$ (by the law of large number)

This implies
$f(\hat{\theta}) = \sigma_\omega^2 + \frac{1}{N}\big(\theta - \hat{\theta}\big)'G'G\big(\theta - \hat{\theta}\big)$
$= \sigma_\omega^2 + \big(\theta - \hat{\theta}\big)'A\big(\theta - \hat{\theta}\big)$
Since $\big(\theta - \hat{\theta}\big) \geq 0$ for all true θ and estimator $\hat{\theta}$, therefore
$\big(\theta - \hat{\theta}\big)'A\big(\theta - \hat{\theta}\big) = \big\|G(\widehat{\theta - \theta})\big\|_2^2 \geq 0.$

Thus A is nonnegative definite matrix and this implies that $f(\hat{\theta})$ is convex.
For any $\hat{\theta}_1, \hat{\theta}_2$ and any α∈(0 , 1) , γ =1 in (5) let
$\hat{\theta} = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2$, then
$g(\hat{\theta}) = \frac{\lambda_N}{N}\big\|\alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2\big\|^\gamma$
$\leq \frac{\lambda_N}{N}\big\|\alpha\hat{\theta}\big\|^\gamma + \frac{\lambda_N}{N}\big\|(1-\alpha)\hat{\theta}_2\big\|^\gamma$
$= \frac{\lambda_N}{N}\alpha\big\|\hat{\theta}\big\|^\gamma + \frac{\lambda_N}{N}(1-\alpha)\big\|\hat{\theta}_2\big\|^\gamma$

$= \alpha g(\hat{\theta}_1) + (1-\alpha) g(\hat{\theta}_2)$.
Hence $g(\hat{\theta})$ is convex.
Since $f(\hat{\theta})$ and $g(\hat{\theta})$ are both convex , therefore
$\boldsymbol{\mathcal{L}_N(\hat{\theta}) = f(\hat{\theta}) + g(\hat{\theta})}$ **is also convex.** ■

Definition 1. (Convergence in Probability) [11]. A sequence of random variables $X_1, X_2, \ldots$ is said to converges in probability to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1 \text{ or equivalently, } \lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0.$$

Definition 2. (Convergence in Distribution) [11]. A sequence of random variables $X_1, X_2, \ldots, X_n$ is said to converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ at all points } x \text{ where } F_X(x) \text{ is continuous.}$$

Definition 3 [17]. A sequence of random variables $X_1, X_2, \ldots$ is said to be convergent to a constant c in probability, denoted $X_n \xrightarrow{p} c$ if for any given $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - c| \geq \epsilon) = 0.$$

Definition 4 [17]. A sequence of random variables $\{X_n\}$, $n = 1, 2, \ldots$, is of smaller order than a sequence random variables $\{a_n\}, n = 1, 2, \ldots,$ if

$\lim_{n\to\infty} \left\{\frac{X_n}{a_n}\right\} = 0$ in which case we write $X_n = o(a_n)$.

Definition 5 [17]. A sequence of random variables $\{X_n\}$, $n = 1, 2, \ldots,$ is said to be bounded upon order $\{a_n\}$, $n = 1, 2, \ldots,$ if there exists a real number $K < \infty$ such that $\frac{|X_n|}{a_n} \leq K$ for all n. In this case, we write $X_n = O(a_n)$.

Definition 6 (Consistency). An estimator $\hat{\theta}^{(N)}$ is said to be consistent for the parameter θ if $\lim_{N\to\infty} \left(\left|\hat{\theta}^{(N)} - \theta\right| \geq \epsilon\right) = 0, \quad \forall\, \epsilon > 0.$

Definition 7. (Kronecker product). An product of two matrices $A \in M^{p,q}$ and $B \in M^{r,s}$ in the form

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1q}B \\ a_{21}B & a_{22}B & \cdots & a_{2q}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}B & a_{p2}B & \cdots & a_{pq}B \end{bmatrix}$$

Is called kronecker product

4. Asymptotic Consistency of the Lasso Estimator

In this part, it will be discussed the consistent of lasso estimator as especial case when $\gamma = 1$ in bridge estimator. Assume that the Grim matrix A defined in (6) is nonsingular. Define the following random function

$$\mathcal{L}_N(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_{it} - G_{it}'\hat{\theta}\right)^2 + \frac{\lambda_N}{N}\sum_{j=1}^{k}\left|\hat{\theta}_j\right|^{\gamma}, \lambda_N > 0, \gamma > 0, t = 1, \ldots, T \qquad (8)$$

Which is minimized at $\hat{\theta} = \hat{\theta}_N$. Assuming that the asymptotic behavior of the bridge estimator can be determined by studying the asymptotic behavior of (8). The following theorem explains that $\hat{\theta}_N$ is a consistent estimator of θ provided that $\lambda_N = o(N)$.

Theorem 1. The lasso estimators
$\hat{\theta}_N = argmin\mathcal{L}_N(\hat{\theta}) = argmin\frac{1}{N}\sum_{i=1}^{N}(y_{it} - G'_{it}\hat{\theta})^2 + \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|^\gamma, \lambda_N > 0, t = 1, \dots, T$
when $\gamma = 1$ is consistent if the following assumptions is satisfied

1. $The\ matrix\ A$ defined in (7) is nonsingular.
2. $\frac{\lambda_N}{N} \to \lambda_0 \geq 0$ i.e. $\lambda_N = o(N)$.
3. $\hat{\theta}_N \xrightarrow{p} argmin\ (\mathcal{L})$, where $\mathcal{L}(\hat{\theta}) = (\hat{\theta} - \theta)'A((\hat{\theta} - \theta) + \lambda_0 \sum_{j=1}^{k}|\hat{\theta}_j|$

Proof. To prove that the lasso estimator $\hat{\theta}_N$ is consistent, we must show that $\mathcal{L}_N(\hat{\theta})$ defined in (8) converges in Probability to $\mathcal{L}(\hat{\theta}) + \sigma_\omega^2$. In other word, we need to show that,

$$\sup_{\hat{\theta}\in K}|\mathcal{L}_N(\hat{\theta}) - \mathcal{L}(\hat{\theta}) - \sigma_\omega^2| \xrightarrow{p} 0 \tag{9}$$

for any compact set K and that

$$\hat{\theta}_N = O_p(1). \tag{10}$$

To show this, it can be translated $\mathcal{L}_N$ in (8) to matrix notation as,

$$\mathcal{L}_N(\hat{\theta}) = \frac{1}{N}(Y - G\hat{\theta})'(Y - G\hat{\theta}) + \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|^\gamma$$
$$= \frac{1}{N}(\omega + G\theta) - G\hat{\theta})'((\omega + G\theta) - G\theta) + \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|^\gamma$$
$$= \frac{1}{N}(\omega + G(\theta - \hat{\theta}))'(\omega + G(\theta - \hat{\theta})) + \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|^\gamma$$
$$= \frac{1}{N}\left[\omega'\omega + 2\omega G(\theta - \hat{\theta}) + (\theta - \hat{\theta})'G'G(\theta - \hat{\theta})\right] + \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|^\gamma$$

Now, letting $N \to \infty$ and $\gamma = 1$ we have the following facts:

$var(\omega) = E(\omega'\omega) + E(\omega)'E(\omega) = E(\omega'\omega) = \sigma_\omega^2$.
$\frac{1}{N}\omega'\omega \xrightarrow{p} E[\omega'\omega] = \sigma_\omega^2$ (by the law of the large number),
$\frac{1}{N}G'G \to A$,
$E\left[\frac{1}{N}\sum_{i=1}^{N}\omega_{it}\right] = \frac{1}{N}\sum_{i=1}^{N}E[\omega_{it}] = 0$,
$\frac{2}{N}\omega'G(\theta - \hat{\theta}) \to 2E\left[\frac{1}{N}\sum_{i=1}^{N}\omega_{it}\right]G'(\theta - \hat{\theta}) = 0$ (by the law of the large numbers)
$\frac{\lambda_N}{N} \to \lambda_0 = 0$
Therefore,
$\mathcal{L}_N(\hat{\theta}) \xrightarrow{p} \sigma_\omega^2 + (\theta - \hat{\theta})'G'G(\theta - \hat{\theta}) + \lambda_0\sum_{j=1}^{k}|\hat{\theta}_j| = \mathcal{L}(\hat{\theta}) + \sigma_\omega^2$

Since $\mathcal{L}_N(\hat{\theta})$ pointwise convergence in probability to $\mathcal{L}(\hat{\theta})$, we conclude that,
$\sup_{\hat{\beta}\in K}|\mathcal{L}_N(\hat{\theta}) - \mathcal{L}(\hat{\theta}) - \sigma_\omega^2| \xrightarrow{p} 0$, for any compact K and that
$\hat{\theta}_N = O_p(1)$.

Therefore we have that For $\gamma \geq 1$, $\mathcal{L}_N$ is convex and
$\hat{\theta}_N = argmin\ (\mathcal{L}_N) \xrightarrow{p} \theta = argmin\ (\mathcal{L})$.
Therefore lasso estimator $\hat{\theta}_N$ is consistent when $\gamma = 1$. ■

Remark 1. For $\gamma < 1$, $\mathcal{L}_N$ will not be convex but the formula in (9) can be easy concluded it by using the same way above. To investigate the formula in (10), we have that
$\mathcal{L}_N(\hat{\theta}) \geq \frac{1}{N}\sum_{i=1}^{N}(Y_{it} - G'_{it}\hat{\theta}_j)^2 = \mathcal{L}_N^0(\hat{\theta})$ for all $\hat{\theta}$.

We note that $argmin(\mathcal{L}_N^0) = O_p(1)$, it follows that $argmin(\mathcal{L}_N) = O_p(1)$.
This implies that $\hat{\theta}_N$ are consistent when $\gamma < 1$. In spite of $\lambda_N = o(N)$ is sufficient for consistency of estimator $\hat{\theta}_N$, it will be chosen λ_N smaller order than $\sqrt{N}$ to get $\sqrt{N} - consistency$ of the ridge estimator. At the same time, if λ_N is chosen less order than $\sqrt{N}$ then $\sqrt{N}\left(\hat{\theta}_N - \theta\right)$ will be the same asymptotic distribution as $\sqrt{N}(\hat{\theta}_N^0 - \theta)$. Therefore the rate of growth of λ_N depends on whether $\gamma \geq 1$ or $\gamma < 1$ to get interesting asymptotic distribution. In the following theorem, $\sqrt{N} - consistency$ of lasso and bridge estimators will be investigated when $\lambda_N = O(\sqrt{N})$ for $\gamma \geq 1$.

Theorem 2. Given the above assumptions on design matrix (i) and (ii), if $\frac{\lambda_N}{\sqrt{N}} \rightarrow \lambda_0 \geq 0$ and A is nonsingular and assume that $\hat{\theta}_N = argminn\mathcal{L}_N(\hat{\theta}) = argmin\frac{1}{N}\sum_{i=1}^{N}\left(y_{it} - G_{it}'\hat{\theta}\right)^2 + \frac{\lambda_N}{N}\sum_{j=1}^{k}\left|\hat{\theta}_j\right|^{\gamma}$,
$\lambda_N > 0, \gamma > 0, t = 1, \ldots, T$, then

1. The lasso estimator $\hat{\theta}_N$ when $\gamma = 1$ satisfies $\sqrt{N}\left(\hat{\theta}_N - \theta\right) \xrightarrow{d} argmin(V_1)$,
where $V_1(u) = u'Au - 2u'W + \lambda_0\sum_{j=1}^{k}\left[u_j sgn(\theta_j) I(\theta_j \neq 0) + |u_j|\, I(\theta_j = 0)\right]$.
2. The Bridge estimator $\hat{\theta}_N$ when $\gamma > 1$ satisfies $\sqrt{N}\left(\hat{\theta}_N - \theta\right) \xrightarrow{d} argmin(V_2)$,
where $V_2(u) = u'Au - 2u'W + \lambda_0\sum_{j=1}^{k} u_j sgn(\theta_j)\left|\theta_j\right|^{\gamma-1}$ and
$W \sim N(0,\ \sigma_\omega^2 A)$.

Proof.
1. If $\gamma=1$ Define a random variable V_N where,

$$V_N(u) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(w_{it} - \frac{u'G_{it}}{\sqrt{N}}\right)^2 - \omega_{it}^2\right] + \frac{\lambda_N}{N}\sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_j\right|^{\gamma}\right], \qquad (11)$$

where $u = (u_1, \ldots, u_k)' \in R^k$, $\lambda_N > 0, \gamma = 1, t = 1, \ldots, T$.
(11) can be rewritten in matrix notation as
$V_N(u) = \left(\left\|\omega - \frac{Gu}{\sqrt{N}}\right\|^2 + \lambda_N\sum_{j=1}^{k}\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|\right) - \left(\|\omega\|^2 + \lambda_N\sum_{j=1}^{k}\left|\theta_j\right|\right)$, which can be written as

$$V_N(u) = Q_N\left(\theta + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta) \qquad (12)$$

where, $Q_N(\alpha) = \|Y - G\alpha\|^2 + \lambda_N\sum_{j=1}^{k}\left|\alpha_j\right|$
Note that $\sqrt{N}\left(\hat{\theta}_N - \theta\right)$ minimizes V_N
i.e. $\sqrt{N}\left(\hat{\theta}_N - \theta\right) = argminV_N$ by noting that

$$V_N\left(\sqrt{N}\left(\hat{\theta}_N - \theta\right)\right) = Q_N\left(\theta + \left(\hat{\theta}_N - \theta\right)\right) - Q_N(\theta) = Q_N(\hat{\theta}_N) + A$$

To show that V_N converges in distribution to V_1, noting that

$$Q_N\left(\theta + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta) = \left|(Y - G\theta) - \frac{Gu}{\sqrt{N}}\right|'\left|(Y - G\theta) - \frac{Gu}{\sqrt{N}}\right| + \lambda_N\sum_{j=1}^{k}\left|\theta_j + \frac{u_j}{\sqrt{N}}\right| - Q_N(\theta)$$

$$= (Y - G\theta)'(Y - G\theta) - \omega'\frac{Gu}{\sqrt{N}} - \frac{u'G'}{\sqrt{N}}\omega + \frac{u'G'G\,u}{N} + \lambda_N\sum_{j=1}^{k}\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left(\|\omega\|^2 + \lambda_N\sum_{j=1}^{k}\left|\theta_j\right|\right)$$

$$= \|\omega\|^2 - \frac{u'G'}{\sqrt{N}}\omega - \frac{u'G'}{\sqrt{N}}\omega + \frac{u'G'G\,u}{N}\ \lambda_N\sum_{j=1}^{k}\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \|\omega\|^2 - \lambda_N\sum_{j=1}^{k}\left|\theta_j\right|$$

$$= u'\frac{G'G}{N}u' - 2u'\frac{G'\omega}{\sqrt{N}} + \lambda_N\sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right| - \left|\theta_j\right|\right]$$

By letting $N \rightarrow \infty$ we will have the following results:
$\frac{1}{N}G'G \rightarrow A$, $\frac{G'\omega}{\sqrt{N}} \xrightarrow{d} W$, $\frac{\lambda_N}{\sqrt{N}} \rightarrow \lambda_0$ and

Journal of Physics: Conference Series **1818** (2021) 012039 doi:10.1088/1742-6596/1818/1/012039

$$\lambda_N \sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right| - |\theta_j|\right] \to \lambda_0 \sum_{j=1}^{k}\left[u_j sgn(\theta_j) I(\theta_j \neq 0) + |u_j|\, I(\theta_j = 0)\right].$$

i.e. by applying these facts, we have that

$$V_N(u) = Q_N\left(\theta + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta) \xrightarrow{d} V(u)$$
$$= u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k} u_j sgn(\theta_j) I(\theta_j \neq 0) + |u_j| I(\theta_j = 0).$$

This implies that $V_N(u) \xrightarrow{d} V_1(u)$.

2. If $\gamma > 1$, note that by using same argument above for first part and

$$\lambda_N \sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - |\theta_j|^{\gamma}\right] \to \lambda_0 \sum_{j=1}^{k} u_j sgn(\theta_j) |\theta_j|^{\gamma-1}.$$

It follows that $V_N(u) \xrightarrow{d} V_2(u)$.

Since $V_N(u) \xrightarrow{d} V(u)$ with the finite-dimensional convergence and V_N is convex when $\gamma \geq 1$ and also V has a unique minimum, therefore by convexity argument follows that

$$\sqrt{N}\left(\hat{\theta}_N - \theta\right) = argmin\ V_N(u) \xrightarrow{d} argmin\, V_1(u) \text{ if } \gamma = 1 \text{ and}$$
$$\sqrt{N}\left(\hat{\theta}_N - \theta\right) = argmin\ V_N(u) \xrightarrow{d} argmin\, V_2(u) \text{ if } \gamma > 1.$$

Also we have $argmin\,(V) = A^{-1}\, W \sim N(0, \sigma_\omega^2 A^{-1})$ $when\ \lambda_0 = 0$. ■

Remark 2. In Theorem 2, it is seen that when $\lambda_0 > 0$ and $\gamma \geq 1$ then the nonzero parameters of the lasso estimator are estimated with some asymptotic bias. This bias is caused by the part $\lambda_0 \sum_{j=1}^{k}\left[u_j sgn(\theta_j) I(\theta_j \neq 0) + |u_j|\, I(\theta_j = 0)\right]$ in $V_1(u)$.

In other words, assume that $\theta_S \neq 0$ and $\theta_{S^c} = 0$, in this case, we have
$V_1(u) = u'Au - 2u'W + \lambda_0 \sum_{j\in S} u_j sgn(\theta_j) + \lambda_0 \sum_{j\in S^c} |u_j|$.

Moreover, theorem 2 shows that when $\gamma > 1$ the value of parameters that will be shrinkage towards 0 increases with amount of coefficients being estimated. Therefore, for large coefficients, the bias of their estimators for $\gamma > 1$ may be unsatisfactory large. In the following theorem, it will be showed that $\lambda_N = O(N^{\frac{\gamma}{2}})$ is necessary for $\gamma < 1$ but $\lambda_N = O(\sqrt{N})$ is sufficient.

$$V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k} |u_j|^{\gamma}\, I(\theta_j = 0).$$

Proof. Let $\hat{\theta}_N = argmin\ \mathcal{L}_N(\hat{\theta}) = argmin\left[\frac{1}{N}\sum_{i=1}^{N}(y_{it} - X_{it}'\hat{\theta})^2 + \frac{\lambda_N}{N}\sum_{j=1}^{k}|\hat{\theta}_j|^{\gamma}\right]$
where $\gamma < 1$, $\lambda_N > 0$ and $t = 1, \ldots, T$.
Define a random variable V_N as

$$V_N(u) = \sum_{i=1}^{N}\left[\left(w_{it} - \frac{u' G_{it}}{\sqrt{N}}\right)^2 - \omega_{it}^2\right] + \lambda_N \sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - |\theta_j|^{\gamma}\right],$$

where $u' \in R^k$, $\gamma < 1$, $\lambda_N > 0$ and $t = 1, \ldots, T$.
It will be proved same as theorem two but it will be added some complexity because the objective function $\mathcal{L}_N$ is nonconvex. In matrix notation the above equation is

$$V_N(u) = \left(\left\|\omega - \frac{Gu}{\sqrt{N}}\right\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma}\right) - \left(\|\omega\|^2 + \lambda_N \sum_{j=1}^{k}|\theta_j|^{\gamma}\right),$$

which can be written as

$$V_N(u) = Q_N\left(\theta + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta)$$

where, $Q_N(\alpha) = \|Y - G\alpha\|^2 + \lambda_N \sum_{j=1}^{k}|\alpha_j|^{\gamma}$

Note that $\sqrt{N}\left(\hat{\theta}_N - \theta\right)$ minimizes V_N
i.e. $argminV_N = \sqrt{N}\left(\hat{\theta}_N - \theta\right)$ by noting that
$V_N\left(\sqrt{N}\left(\hat{\theta}_N - \theta\right)\right) = Q_N\left(\theta + \left(\hat{\theta}_N - \theta\right)\right) - Q_N(\theta) = Q_N + A.$

To show that V_N converges in distribution to V, noting that

$$V_N(u) = Q_N\left(\theta + \frac{u_j}{\sqrt{N}}\right) - Q_N(\beta) = (Y - G\theta)'(Y - G\theta) - \omega'\frac{Gu}{\sqrt{N}} - \frac{u'G'}{\sqrt{N}}\omega$$
$$+ \frac{u'G'G\,u}{N} + \lambda_N - \left(\|\omega\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_j\right|^{\gamma}\right)$$
$$= \|\omega\|^2 - \frac{u'G'}{\sqrt{N}}\omega - \frac{u'G'}{\sqrt{N}}\omega + \frac{u'G'G\,u}{N}\ \lambda_N \sum_{j=1}^{k}\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \|\omega\|^2 - \lambda_N \sum_{j=1}^{k}\left|\theta_j\right|^{\gamma}$$
$$= u'\frac{G'G}{N}u' - 2u'\frac{G'\omega}{\sqrt{N}} + \lambda_N \sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_j\right|^{\gamma}\right]$$

By letting $N \to \infty$ we will have the following results:

$$\frac{1}{N}G'G \to A, \quad \frac{G'\omega}{\sqrt{N}} \xrightarrow{d} W,$$

since $\lambda_N = O\left(N^{\frac{\gamma}{2}}\right) = o\left(\sqrt{N}\right)$ then $\frac{\lambda_N}{N^{\frac{\gamma}{2}}} \to \lambda_0$ and this implies that

$$\lambda_N \sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_j\right|^{\gamma}\right] \to 0 \quad \text{if } \theta_j \neq 0. \text{ Therefore}$$
$$\lambda_N \sum_{j=1}^{k}\left[\left|\theta_j + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_j\right|^{\gamma}\right] \to \lambda_0 \sum_{j=1}^{k}\left|u_j\right|^{\gamma} I\left(\theta_j = 0\right)$$

Which is uniform convergence over compact sets on u. It implies then that

$$V_N(u) \xrightarrow{d} V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k}\left|u_j\right|^{\gamma} I\left(\theta_j = 0\right),$$

which is uniform convergence on the space of functions on compact sets. Since The random variable V_N is nonconvexity due to $\gamma < 1$ hence To prove that
$argmin(V_N) = \sqrt{N}\left(\hat{\theta}_N - \theta\right) \xrightarrow{d} argmin\,(V)$, it is sufficient to show that
$argmin(V_N) = O_p(1)$.

It is noted that

$$V_N(u) \geq \sum_{i=1}^{N}\left[\left(w_{it} - \frac{u'G_{it}}{\sqrt{N}}\right)^2 - \omega_{it}^2\right] - \lambda_N \sum_{j=1}^{k}\left|\frac{u_j}{\sqrt{N}}\right|^{\gamma}$$
$$\geq \sum_{i=1}^{N}\left[\left(w_{it} - \frac{u'G_{it}}{\sqrt{N}}\right)^2 - \omega_{it}^2\right] - (\lambda_0 + \delta)\sum_{j=1}^{k}\left|u_j\right|^{\gamma}$$
$$= V_N^{(m)}(u), \text{ for all } u \text{ and } N \text{ sufficiently large.}$$

Since $\gamma < 1$, this implies that the terms $\left[\left(w_{it} - \frac{u'G_{it}}{\sqrt{N}}\right)^2 - \omega_{it}^2\right]$ is grow faster than the terms $\left|u_j\right|^{\gamma}$, hence we have that $argmin\left(V_N^{(m)}\right) = O_p(1)$.
It follows that $argmin(V_N) = O_p(1)$. Since $argmin\,(V)$ is unique and
$argmin(v) = O_p(1)$.
Therefore, $argmin(V_N) = \sqrt{N}\left(\hat{\theta}_N - \theta\right) \xrightarrow{d} argmin\,(V)$. ■

Remark 3. In the above theorem, it is seen that $\lambda_N = O\left(N^{\frac{\gamma}{2}}\right)$ is necessary for $\gamma < 1$ but $\lambda_N = O(\sqrt{N})$ is sufficient. It also indicates that if $\gamma < 1$ then it can be estimated non-zero coefficients normally with unbiased asymptotic. In the other hand, the estimates of zero coefficients are shrunk to zero with positive probability. It is cleared that theorem 3 is in contrast with theorem 2 when $\gamma \geq 1$ which indicates that the nonzero parameters are estimated with some asymptotic bias.

5. An asymptotic of Finite Sample and Small True Parameters

he Bridge estimator for $\gamma \leq 1$ has specific feature which is the ability of obtaining exact 0 parameter estimates. In this section, we discuss the possibility of obtaining this "exact 0" phenomenon in finite samples when the true parameter is close to 0 but nonzero.

To do this, it will be defined a triangular array of observations by defining

$$Y_{N(it)} = \mu + \boldsymbol{X}'_{\boldsymbol{Nj(it)}}\boldsymbol{\beta}_{\boldsymbol{Nj}} + \nu_{Ni} + \varepsilon_{N(it)}, \text{ for } i = 1, \dots, N \text{ and } t = 1, \dots, T \tag{13}$$

which can be written as

$$Y_{N(it)} = \mu + \boldsymbol{X}'_{\boldsymbol{N(it)}}\boldsymbol{\beta}_{\boldsymbol{N}} + \omega_{N(it)} \tag{14}$$

$\sigma_\omega^2 = \sigma_\nu^2 + \sigma_\varepsilon^2$ where $\omega_{N(it)} = \nu_{Ni} + \varepsilon_{N(it)}$, with $\omega_{N(it)} \sim N(0, \sigma_\omega^2)$,

(14) can be rewritten as

$$Y_{N(it)} = \boldsymbol{G}'_{\boldsymbol{N(it)}}\boldsymbol{\theta}_{\boldsymbol{N}} + \omega_{N(it)} \tag{15}$$

where $\boldsymbol{G}_{\boldsymbol{N(it)}} = [\boldsymbol{e}, \boldsymbol{X}_{\boldsymbol{N(it)}}]$, $e = [1, \dots 1]'$ and $\boldsymbol{\theta}_{\boldsymbol{N}} = [\mu, \boldsymbol{\beta}_{\boldsymbol{N}}]'$

Triangular array of $\boldsymbol{G}_{\boldsymbol{N(it)}}$ can be written as

$$\boldsymbol{G}_{\boldsymbol{N(it)}} = \begin{bmatrix} \mathbf{1} & \boldsymbol{X}_{\mathbf{1,1,1}} & & & \\ \mathbf{1} & \boldsymbol{X}_{\mathbf{2,1,2}} & \boldsymbol{X}_{\mathbf{2,2,2}} & & \\ \mathbf{1} & \boldsymbol{X}_{\mathbf{3,1,3}} & \boldsymbol{X}_{\mathbf{3,2,3}} & \boldsymbol{X}_{\mathbf{3,3,3}} & \dots \\ \dots & \dots & \dots & \dots & \\ \mathbf{1} & \boldsymbol{X}_{\boldsymbol{N},\mathbf{1},\boldsymbol{T}} & \boldsymbol{X}_{\boldsymbol{N},\mathbf{2},\boldsymbol{T}} & \dots & \boldsymbol{X}_{\boldsymbol{N},\boldsymbol{N},\boldsymbol{T}} \end{bmatrix}$$

Assuming for the design matrix $\boldsymbol{G}_{\boldsymbol{N(it)}}$ the following regularity conditions.

$$\frac{1}{N}\sum_{i=1}^{N} G_{N(it)}G'_{N(it)} \rightarrow A, \; t = 1, \dots, T \tag{16}$$

where A is positive definite matrix and

$$\frac{1}{N}\max_{1\leq i\leq N} G_{N(it)}G'_{N(it)} \rightarrow 0 \tag{17}$$

These are explicit analogues of (6) and (7).

Consider $\boldsymbol{\theta}_{\boldsymbol{N}} = \boldsymbol{\theta} + \frac{z}{\sqrt{N}}$ and define

$$\hat{\theta}_N = argmin\left[\sum_{i=1}^{N}\left(Y_{N(it)} - \boldsymbol{\theta}'\boldsymbol{G}_{\boldsymbol{N(it)}}\right)^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_j\right|^{\gamma}\right] \tag{18}$$

this formulation achieves the asymptotic properties of Bridge estimator when one or more of true parameters are near to 0 but nonzero.

Theorem 4. Consider $Y_{N(it)} = \boldsymbol{G}'_{\boldsymbol{N(it)}}\boldsymbol{\theta}_{\boldsymbol{N}} + \omega_{N(it)}$ for $i = 1, \dots, N$, $t = 1, \dots, T$ with $\theta_N = \theta + \frac{z}{\sqrt{N}}$ and assume that the regularity conditions (16) and (17). Let

$\hat{\theta}_N = argmin\sum_{i=1}^{N}\left(Y_{N(it)} - \boldsymbol{\theta}'\boldsymbol{G}_{\boldsymbol{N(it)}}\right)^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_j\right|^{\gamma}$ for some $\gamma > 1$.

1. If $\frac{\lambda_N}{\sqrt{N}} \rightarrow \lambda_0$ then $\sqrt{N}\left(\hat{\theta}_N - \theta_N\right) \overset{d}{\rightarrow} argmin\,(v)$, where

$V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k} u_j sgn(\theta_j)\left|\theta_j\right|^{\gamma-1}$.

2. $\sqrt{N}\left(\hat{\theta}_N - \theta_N\right) \overset{d}{\rightarrow} argmin\,(v)$ if $\theta = 0$ and $\frac{\lambda_N}{N^{\frac{\gamma}{2}}} \rightarrow \lambda_0 \geq 0$, where

$V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k}\left|u_j + z_j\right|^{\gamma}$.

Proof.

1. Define random variable

$$V_N(u) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\omega_{N(it)} - \frac{u'G'_{N(it)}}{N}\right)^2 - \omega_{N(it)}^2\right] + \frac{\lambda_N}{N}\sum_{j=1}^{k}\left[\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_{Nj}\right|^{\gamma}\right],$$

where $u' \in R^k$, $\lambda_N > 0, \gamma > 1$, $\mathrm{T} = 1, \dots, T$.

Above equation can be written as

$$V_N(u) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\omega_{N(it)} - \frac{u' G'_{N(it)}}{N}\right)^2 + \frac{\lambda_N}{N}\sum_{j=1}^{k}\left|\theta_{Nj} + \frac{u_j}{N}\right|^{\gamma}\right] - \left[\frac{\omega^2_{N(it)}}{N} + \sum_{j=1}^{k}\left|\theta_{Nj}\right|^{\gamma}\right] \quad (19)$$

It can be rewritten (19) in matrix notation as

$V_N(u) = \left[\left\|\omega_N - \frac{G_N u}{\sqrt{N}}\right\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma}\right] - \left[\|\omega\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj}\right|^{\gamma}\right]$, which can be written as $V_N(u) = Q_N\left(\theta_N + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta_N)$

Where $Q_N(\theta_N) = \|Y_N - G_N\theta_N\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj}\right|^{\gamma}$

It is cleared that $\sqrt{N}\left(\hat{\theta}_N - \theta_N\right)$ minimizes V_N.

i.e. $argmin\, V_N = \sqrt{N}\left(\hat{\theta}_N - \theta_N\right)$ by noting that

$$V_N\left(\sqrt{N}\left(\hat{\theta}_N - \theta_N\right)\right) = Q_N\left(\theta_N + \left(\hat{\theta}_N - \theta_N\right)\right) - Q_N(\theta_N) = Q_N\left(\hat{\theta}_N\right) - Q_N(\theta_N) = Q_N\left(\hat{\theta}_N\right) + A$$

and conclude that the objective function Q_N is similar to the Lasso objective function, minimized at $\hat{\theta}_N$ when $\gamma = 1$. To show that V_N converges in distribution to V, noting that

$$Q_N\left(\theta_N + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta_N) = \left|(Y_N - G_N\theta_N) - \frac{Gu}{\sqrt{N}}\right|' \left|(Y_N - G_N\theta_N) - \frac{G_N u}{\sqrt{N}}\right|$$
$$+\lambda_N \sum_{j=1}^{k}\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - Q_N(\theta)$$
$$= (Y_N - G_N\theta_N)'(Y_N - G_N\theta_N) - \omega_N' \frac{G_N u}{\sqrt{N}} - \frac{u' G'_N}{\sqrt{N}}\omega + \frac{u' G'_N G_N\, u}{\sqrt{N}}$$
$$+ \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left(\|\omega_N\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj}\right|^{\gamma}\right)$$
$$= \|\omega_N\|^2 - \frac{u' G'_N}{\sqrt{N}}\omega - \frac{u' G'_N}{\sqrt{N}}\omega_N + \frac{u' G'_N G_N\, u}{N}\ \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \|\omega_N\|^2 - \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj}\right|^{\gamma}$$
$$= u' \frac{G'_N G_N}{N} u - 2u' \frac{G'_N \omega}{\sqrt{N}} + \lambda_N \sum_{j=1}^{k}\left[\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_{Nj}\right|^{\gamma}\right] \quad (20)$$

By letting $N \to \infty$ we will have the following results:

$\frac{1}{N} G'_N G_N \to A$

$\frac{G'_N \omega_N}{\sqrt{N}} \xrightarrow{d} W \sim N(0, \sigma^2_\omega\, A)$

$\frac{\lambda_N}{\sqrt{N}} \to \lambda_0$

$\sum_{j=1}^{k}\left[\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_{Nj}\right|^{\gamma}\right] \to \sum_{j=1}^{k} u_j sgn(\theta_{Nj})\left|\theta_{Nj}\right|^{\gamma-1}$ if $\gamma > 1$.

i.e. by combining these facts, we have that

$$V_N(u) = Q_N\left(\theta_N + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta_N) \xrightarrow{d} V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k} u_j sgn(\theta_{Nj})\left|\theta_{Nj}\right|^{\gamma-1} \text{ if } \gamma > 1$$

$i.e.\ V_N(u) \xrightarrow{d} V(u)$.

This implies that

$$argmin\,(V_N) = \sqrt{N}\left(\hat{\theta}_N - \theta_N\right) \xrightarrow{d} argmin\,(V).$$

2. If $\beta = 0$ then we have that $\beta_N = \frac{z}{\sqrt{N}}$ and (20) becomes

$$Q_N\left(\theta_N + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta_N) = u' \frac{G'_N G_N}{N} u - 2u' \frac{G'_N \omega_N}{\sqrt{N}} + \lambda_N \sum_{j=1}^{k}\left[\left|\frac{z_j}{\sqrt{N}} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\frac{z_j}{\sqrt{N}}\right|^{\gamma}\right]$$

By letting $N \to \infty$ and $\gamma > 1$ we have that the following results

$\frac{\lambda_N}{N^{\frac{\gamma}{2}}} \to \lambda_0 \geq 0$, and this implies that

$\lambda_N \sum_{j=1}^{k}\left[\left|\frac{z_j}{\sqrt{N}}+\frac{u_j}{\sqrt{N}}\right|^{\gamma}-\left|\frac{z_j}{\sqrt{N}}\right|^{\gamma}\right] \to \lambda_0 \sum_{j=1}^{k}\left|u_j+z_j\right|^{\gamma}$

Combining above results, we have that

$V(u)=u'Au-2u'W+\lambda_0\sum_{j=1}^{k}\left|u_j+z_j\right|^{\gamma}$.

Hence $V_N(u)\overset{d}{\to}V(u)$

Therefore

$argmin\,(V_N)=\sqrt{N}\left(\hat{\theta}_N-\theta_N\right)=\sqrt{N}\left(\hat{\theta}_N-\frac{z}{\sqrt{N}}\right)\overset{d}{\to} argmin\,(V)$. ■

Remark 4. Part 1 of theorem 4 proposes that the asymptotic bias indicated in theorem 2 is still continue if one or more parameters are large. While part 2 proposes that the usefulness of using penalty with $\gamma > 1$ is restricted to situation where all coefficients are smaller than sample size N. The following corollary explains that when $\gamma = 2$ then all coefficients will be less than sample size N .

Corollary 1. Consider $Y_{N(it)} = \boldsymbol{G}'_{\boldsymbol{N(it)}}\boldsymbol{\theta_N} + \omega_{N(it)}$ for $i = 1,\dots,N,\ t = 1,\dots,T$ with $\theta_N = \theta + \frac{z}{\sqrt{N}}$ and assume that the regularity conditions (16) and (17). Let $\hat{\theta}_N = argmin\sum_{i=1}^{N}\left(Y_{N(it)}-\boldsymbol{\theta}'\boldsymbol{G}_{\boldsymbol{N(it)}}\right)^2+\lambda_N\sum_{j=1}^{k}\left|\theta_j\right|^{\gamma}$.

if $\theta = 0$, $\frac{\lambda_N}{N^{\frac{\gamma}{2}}} \to \lambda_0 \geq 0$ and $\sqrt{N}\left(\hat{\theta}_N-\theta_N\right)\overset{d}{\to} argmin\,(v)$ where

$V(u)=u'Au-2u'W+\lambda_0\sum_{j=1}^{k}\left|u_j+z_j\right|^{\gamma}$ then for ridge estimation $\gamma = 2$ we have that

$\sqrt{\mathrm{N}}\left(\hat{\theta}_{\mathrm{N}}-\frac{\mathrm{z}}{\sqrt{\mathrm{N}}}\right)\overset{\mathrm{d}}{\to}(\mathrm{A}+\lambda_0\mathrm{I})^{-1}(\mathrm{W}-\lambda_0\mathrm{z})$

$\sim N[(-\lambda_0(A+\lambda_0 I)^{-1}z),(\sigma_\omega^2(A+\lambda_0 I)^{-1}A(A+\lambda_0 I)^{-1})]$

Proof. Since $\sqrt{N}\left(\hat{\theta}_N-\frac{z}{\sqrt{N}}\right)\overset{d}{\to} argmin\,V(u)$

$and\ \ argmin\,V(u)$ represents derivative of V with respect to u.

i.e. $argmin\,V(u)=\frac{\partial v}{\partial u}\left[u'Au-2u'W+\lambda_0\sum_{j=1}^{k}\left|u_j+z_j\right|^{2}.\right]=0$

$0=2Au-2W+2\lambda_0(u+z)$

$0=Au-W+\lambda_0 u+\lambda_0 z$

$0=(A+\lambda_0 I)u-W+\lambda_0 z$

$(A+\lambda_0 I)u=(W-\lambda_0 z)$

$u=(A+\lambda_0 I)^{-1}(W-\lambda_0 z)$

Also $E(u)=E[(A+\lambda_0 I)^{-1}(W-\lambda_0 z)]$

$=(A+\lambda_0 I)^{-1}E(W+\lambda_0 z)$

$=-\lambda_0(A+\lambda_0 I)^{-1}z.$

$Var(u)=var[(A+\lambda_0 I)^{-1}(W-\lambda_0 z)]$

$=[\sigma_\omega^2\,(A+\lambda_0 I)^{-1}A(A+\lambda_0 I)^{-1}]$

Hence

$\sqrt{N}\left(\hat{\theta}_N-\frac{z}{\sqrt{N}}\right)\overset{d}{\to}(A+\lambda_0 I)^{-1}(W-\lambda_0 z)$

$\sim N[(-\lambda_0(A+\lambda_0 I)^{-1}z),(\sigma_\omega^2(A+\lambda_0 I)^{-1}A(A+\lambda_0 I)^{-1})]$.

Remark 5. Corollary 1 explains that by select λ_0 reasonably, we will have the mean square error $G'\hat{\theta}_N$ smaller than that of $G'\hat{\theta}_N^0$.

Theorem 5. Assume that $Y_{N(it)} = \boldsymbol{G}'_{\boldsymbol{N(it)}}\boldsymbol{\theta_N} + \omega_{N(it)}$ and

$\hat{\theta}_N = argmin\ \mathcal{L}_N\left(\hat{\beta}\right)=argmin\sum_{i=1}^{N}\left(Y_{N(it)}-\boldsymbol{\theta}'\boldsymbol{G}_{\boldsymbol{N(it)}}\right)^2+\lambda_N\sum_{j=1}^{k}\left|\theta_j\right|^{\gamma}$ for some $\gamma \leq 1$ and $\frac{\lambda_N}{N^{\frac{\gamma}{2}}} \to \lambda_0 \geq 0$ with $\theta_N = \theta + \frac{z}{\sqrt{N}}$. Also suppose that the regularity conditions (16) and (17) are satisfied, then

1. The lasso estimator $\hat{\theta}_N$ when $\gamma = 1$ satisfies $\sqrt{N}\left(\hat{\theta}_N-\theta\right)\overset{d}{\to} argminV$.

where $V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k}[u_j sgn(\theta_j)\, I(\theta_j \neq 0) + |u_j + z_j|\, I(\theta_j = 0)]$.
2. The Bridge estimator $\hat{\theta}_N$ when $\gamma < 1$ Satisfies $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} argminV,$ where
$V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k}|u_j + z_j|^{\gamma}\, I(\theta_j = 0)$.

Proof.
1. Define a random variable V_N as
$$V_N(u) = \sum_{i=1}^{N}\left[\left(w_{N(it)} - \frac{u'G_{N(it)}}{\sqrt{N}}\right)^2 - \omega_{N(it)}^2\right] + \lambda_N \sum_{j=1}^{k}\left[\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_{Nj}\right|^{\gamma}\right],$$
where $u' \in R^k$, $\gamma = 1$, $\lambda_N > 0$ and $t = 1, \ldots, T$.
It will be proved same as theorem 2 and 3 but will be add some changes. In matrix notation the above equation is
$$V_N(u) = \left(\left\|\omega_N - \frac{G_N u}{\sqrt{N}}\right\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|\right) - \left(\|\omega_N\|^2 + \lambda_N \sum_{j=1}^{k}|\theta_{Nj}|\right),$$ which can be

written as
$$V_N(u) = Q_N\left(\theta_N + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta_N)$$
where, $Q_N(\alpha) = \|Y - G\alpha\|^2 + \lambda_N \sum_{j=1}^{k}|\alpha_j|^{\gamma}$
Note that $\sqrt{N}\,(\hat{\theta}_N - \theta_N)$ minimizes V_N
i.e. $argminV_N = \sqrt{N}\,(\hat{\theta}_N - \theta_N)$ by noting that
$$V_N\left(\sqrt{N}\,(\hat{\theta}_N - \theta_N)\right) = Q_N\left(\theta_N + (\hat{\theta}_N - \theta_N)\right) - Q_N(\theta_N) = Q_N(\hat{\theta}_N) + A$$

and observe that the objective function Q_N is similar to the Lasso objective function, minimized at $\hat{\theta}_N$. To show that V_N converges in distribution to V, noting that
$$V_N(u) = Q_N\left(\theta_N + \frac{u_j}{\sqrt{N}}\right) - Q_N(\theta_N) = (Y_N - G_N\theta_N)'(Y_N - G_N\theta_N) - \omega_N' \frac{G_N u}{\sqrt{N}} - \frac{u'G_N'}{\sqrt{N}}\omega + \frac{u'G_N'G\,u}{\sqrt{N}} + \lambda_N\left(\|\omega_N\|^2 + \lambda_N \sum_{j=1}^{k}\left|\theta_{N\,j}\right|\right)$$
$$= \|\omega_N\|^2 - \frac{u'G_N'}{\sqrt{N}}\omega_N - \frac{u'G_N'}{\sqrt{N}}\omega_N + \frac{u'G_N'G_N\,u}{\sqrt{N}} + \lambda_N \sum_{j=1}^{k}\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \|\omega_N\|^2 - \lambda_N \sum_{j=1}^{k}|\theta_{Nj}|^{\gamma}$$
$$= u'\frac{G_N'G_N}{\sqrt{N}}u' - 2u'\frac{G_N'\omega_N}{\sqrt{N}} + \lambda_N \sum_{j=1}^{k}\left[\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - \left|\theta_{Nj}\right|^{\gamma}\right]$$

By letting $N \to \infty$ with $\theta_N = \theta + \frac{z}{j}$ we will have the following results:
$\frac{1}{N}G_N'G \to A, \quad \frac{G_N'\omega}{\sqrt{N}} \xrightarrow{d} W,$
since $\lambda_N = O\left(N^{\frac{\gamma}{2}}\right) = o(\sqrt{N})$ then $\frac{\lambda_N}{N^{\frac{\gamma}{2}}} \to \lambda_0$, it follows that
$$\lambda_N \sum_{j=1}^{k}\left[\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right| - |\theta_{Nj}|\right] \to \lambda_0 \sum_{j=1}^{k}[u_j sgn(\theta_j) I(\theta_j \neq 0) + |u_j + z_j|\, I(\theta_{Nj} = 0)].$$

Hence $V_N(u) \xrightarrow{d} V(u) = u'Au - 2u'W + \lambda_0 \sum_{j=1}^{k}[u_j sgn(\theta_j)\, I(\theta_j \neq 0) + |u_j + z_j|\, I(\theta_j = 0)]$. Therefore,
$argmin(V_N) = \sqrt{N}\,(\hat{\theta}_N - \theta_N) \xrightarrow{d} argmin\,(V).$
2. Since The random variable V_N is nonconvexity due to $\gamma < 1$, There are some added complexities to the second part of random variable V_N. Since since $\lambda_N = O\left(N^{\frac{\gamma}{2}}\right) = o(\sqrt{N})$ then $\frac{\lambda_N}{N^{\frac{\gamma}{2}}} \to \lambda_0$, this implies that
$$\lambda_N \sum_{j=1}^{k}\left[\left|\theta_{Nj} + \frac{u_j}{\sqrt{N}}\right|^{\gamma} - |\theta_{Nj}|^{\gamma}\right] \to 0 \quad \text{if } \theta_{Nj} \neq 0.$$ Therefore

$\lambda_N \sum_{j=1}^{k}\left[\left|\theta_{Nj}+\frac{u_j}{\sqrt{N}}\right|^{\gamma}-\left|\theta_{Nj}\right|^{\gamma}\right] \to \lambda_0 \sum_{j=1}^{k}\left|u_j+z_j\right|^{\gamma} I\left(\theta_{Nj}=0\right)$, where $\theta_N=\theta+\frac{z}{j}$
Which is uniform convergence over compact sets on u. It implies then that
$V_N(u) \overset{d}{\to} V(u)=u'Au-2u'W+\lambda_0\sum_{j=1}^{k}\left|u_j+z_j\right|^{\gamma} I\left(\theta_{Nj}=0\right)$ which is uniform convergence on the space of functions on compact sets.
To prove that $argmin(V_N)=\sqrt{N}\left(\hat{\theta}_N-\theta_N\right)\overset{d}{\to} argmin\,(V)$, it is sufficient to show that $argmin(V_N)=O_p(1)$. It is noted that

$$V_N(u) \geq \sum_{i=1}^{N}\left[\left(w_{N(it)}-\frac{u'G_{N(it)}}{\sqrt{N}}\right)^2-\omega_{N(it)}^2\right]-\lambda_N\sum_{j=1}^{k}\left|\frac{u_j}{\sqrt{N}}+\frac{z_j}{\sqrt{N}}\right|^{\gamma}$$

$$\geq \sum_{i=1}^{N}\left[\left(w_{N(it)}-\frac{u'G_{N(it)}}{\sqrt{N}}\right)^2-\omega_{N(it)}^2\right]-(\lambda_0+\delta)\sum_{j=1}^{k}\left|u_j+z_j\right|^{\gamma}$$

$=V_N^{(m)}(u)$, for all u and N sufficiently large.

Since $\gamma < 1$, this implies that the terms $\left[\left(w_{N(it)}-\frac{u'G_{N(it)}}{\sqrt{N}}\right)^2-\omega_{N(it)}^2\right]$ is grow faster than the terms $\left|u_j+z_j\right|^{\gamma}$, hence we have that $argmin\,\left(V_N^{(m)}\right)=O_p(1)$.
It follows that $argmin(V_N)=O_p(1)$. Since $argmin\,(V)$ is unique and $argmin(v)=O_p(1)$.
Therefore, $argmin(V_N)=\sqrt{N}\left(\hat{\theta}_N-\theta_N\right)\overset{d}{\to} argmin\,(V)$. ■

Remark 6. In contrast with the theorem 4 when $\gamma \leq 1$, the small parameters may be estimated exact 0 in finite sample even when large parameters are present.

6. Conclusion

In this paper, the discussion of the study has introduced the high-dimensional repeated measurements model. It has studied the asymptotic behavior of bridge estimator and lasso estimator as especial case of it with put some assumption on design matrix for the two topics. In first topic which is mentioned in section 4, we discuss the asymptotic in general with large parameters. It can be concluded that the consistency of bridge estimator will be sufficient for two cases. In the first case the sufficient consistency of bridge estimator is occurred when tuning parameter λ_N is asymptotically smaller than N for all value of shrinking parameter γ.

While in the second case, the sufficient consistency of bridge estimator is occurred when λ_N is asymptotically bounded by $\sqrt{N}$ and the value of shrinking parameters are greater than or equal one. But in this case, the bias of their estimators may be unsatisfactory large when γ is greater than one while the magnitude of shrinking parameters towards zero increases. Furthermore, we conclude that the necessary condition for consistency the bridge estimator is occurred when λ_N is asymptotically bounded by $N^{\frac{\gamma}{2}}$ and the value of shrinking parameter γ is less than one. Moreover, the nonzero parameters can be estimated at the usual rate without any asymptotic bias while shrinking the estimates of zero parameters to zero with positive probability.

In second topic which is mentioned in section 5, the study showed that the asymptotic of bridge estimator in small parameters when a triangular array of observations are assumed with simple change on the parameters. In this case, we conclude that the same results above about the consistency of bridge estimators are obtained when λ_N is asymptotically smaller than $\sqrt{N}$ and the value of shrinking parameters are greater than one. On the other hand, if one or more of the parameters is large then the asymptotic bias suggested by first topic would still continue. Furthermore, we conclude that even when large parameters are existence, the small parameters can be estimated as exactly zero in finite samples when λ_N is asymptotically smaller than $N^{\frac{\gamma}{2}}$ and the value of shrinking parameter γ is less than or equal one.

References

[1] Bühlmann P and De Geer S V 2011 Statistics for High Dimensional data (Springer-Verlag Berlin Heidelberg).
[2] Charles S D 2002 Statistical Methods for the Analysis of Repeated measurements (Springer-Verlag New York, Inc).
[3] Chatterjee A (2011) Strong Consistency of Lasso Estimator Indian Journal of Statistics **73** 55-78.
[4] Chatterjee A and Lahiri S N 2010 Asymptotic Properties of the Residual Bootstrap of Lasso Estimator Proceeding of the American Mathematical Society **138** 4497-4509.
[5] Chatterjee S 2014 Assumptionless Consistency of the Lasso **1303** 5817v5.
[6] Lahmann E L 1999 Elements of Large Sample Theory (Springer-Verlag New York, Inc).
[7] FU W J 1998 Penalized Regressions: The Bridge Versus the Lasso Journal of Computational and Graphical Statistics **7** 397-416.
[8] Gao X and Huang J 2010 Asymptotic Analysis of High−dimensional LAD Regression with Lasso Statistics Sinica **20** 1484−1506.
[9] Huang J, Horowttz J L and MA S 2008 Asymptotic Properties of Bridge Estimators in Sparse High-dimensional Regression Model The Annals of Statistics **28** 1356-1378.
[10] Hogg R, Mckean J and Craig A 2005 Introduction to Mathematical Statistics (Pearson Education Inc).
[11] Knight K 1998 Limiting Distributions For L_1 Regression Estimators Under General Conditions The Annals Of Statistics **26** 755-770.
[12] Knight K and Fu W 2000 Asymptotic For Lasso-Type Estimator The Annals Of Statistics **28** 1356-1378.
[13] Meinshausen N and Yu B 2009 Lasso-Type Recovery of Sparse Representations for High-Dimensional Data Institute of Mathematics Statistics **37** 246-270.
[14] Pendse G 2011 A Tutorial on the Lasso and the Shooting Algorithm P.A.I.N. Group Imaging and Analysis Group Mclean Hospital Harvard Medical School.
[15] Shao J and Li Q 2015 Regularizing Lasso: Aconsistent Variable Selection Method Statistica Sinica **25** 975-992.
[16] Vonesh E F and Chinchilli M (1997) Linear and Nonlinear Models for the Analysis of Repeated Measurements (Mancel Dakker, Inc., New York).
[17] Vershynin R 2018 High−dimensional Probability (Roman Vershynin).
[18] Zhang C and Huang J 2008 The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression The Annals of Statistics **36** 1567-1594.
[19] Zhao P and Yu B 2006 On Model Selection Consistency of Lasso Journal of Machine Learning Research **7** 2541-2563.