THE OPTIMAL STEIN ESTIMATION RELATIVE TO CONCENTRATION PROBABILITY – A LARGE SAMPLE APPROXIMATION

Raman Pant[1]
Dept. of Statistics, Mahatma Gandhi Kashi Vidyapith
Varanasi, India

ABSTRACT
General family of Stein rule estimators is considered in linear regression model. The large sample approximation of its sampling distribution is derived. Approximations of concentration probability of the estimators around the true value are evaluated. Optimal selection of the biasing scalar is discussed.

KEYWORDS : Linear regression model, Stein rule estimator, large sample approximation, Sampling distribution, concentration probability

MSC: 62F10

RESUMEN
La familia general de los estimadores reglados de Stein es considerada en un modelo de regresión lineal. La aproximación para muestras grandes de su distribución muestral es derivada. Aproximaciones de la probabilidad de concentración de los estimadores alrededor del verdadero valor son evaluados. La selección optimal del escalar de insesgamiento es discutido.

PALABRAS CLAVE : modelo de regresión lineal, Estimado de la regla de Stein, aproximación para muestras grandes, distribución muestral, probabilidad de concentración.

1. INTRODUCTION

Stein's counter to the well established overall dominance of the classical least squares (CLS) estimator for estimating the coefficient vector in a linear regression model has been a well debated matter in literature. The use of Stein-rule estimators in real data problems has also received considerable attention in the literature, e.g., Knight et al.(1992,1993,1993a), Bao and Wan(2007), Adkins(1995) and Wan et al.(2003).
Based on Stein's philosophy, several shrinkage estimators for the coefficient vector have been proposed in the literature; see, Judge et al.(1985)] for a detailed account. Most of the studies concerning these estimators have been done regarding the optimal choice of these characterizing scalars in order to establish their superiority over the classical least squares (CLS) estimators, e.g., Shalabh et al.(2009), Srivastava and Upadhyaya(1997),Pant and Manoj(2012) and Ullah and Ullah(1978), under the quadratic loss set up. Rao (1981) suggested to employ proper measure of proximity that is which is based on concentration of the estimate around the true unknown parameter they aim to estimate and to judge the performance of an estimator which is more intrinsic in nature. The two well known measures of concentration of estimators are Pitman Nearness and the Probability of Concentration. The Pitman Nearness criterion suffers from certain basic drawbacks e.g. lack of transitivity. Detail discussions on the merits of the two criterion of concentration see Robert et al.(1993).

[1] *email* : ramanpant@gmail.com

Performance of various shrinkage estimators for the coefficient vector of the linear regression model are judged with the Concentration Probability criterion. For this purpose the sampling distribution of these estimators is required. The exact expressions for the probability density functions of these non linear shrinkage estimators are fairly complicated and as such it becomes difficult to evaluate the expressions for the probability of concentration there from. Therefore the large sample approximation for the probability density function of these non linear shrinkage estimators is derived and the concentration probability around parameter is evaluated there from. Comparisons on concentration probabilities have been done and dominance conditions are derived for various prominent shrinkage estimators. The plan of the paper is as follows. In section 2 of the paper, we describe the model and estimators, while in section 3 we present the large sample approximation of probability density function of the proposed general class and derived the large sample approximation of concentration probability of proposed general class as well as of the classical least square estimator. Finally, in section 4, we investigate the optimal choices for the characterizing scalars for the relative dominance of these estimators over each other.

2. THE MODEL AND ESTIMATORS

Let us postulate a linear regression model

$$y = x\beta + u \tag{2.1}$$

Where y is a $T \times 1$ vector of observations on the variable to be explained, X is a $T \times p$ full column rank matrix of observation on explanatory variables, β is a $p \times 1$ vector of regression coefficients being estimated, and u is a $T \times 1$ vector of disturbances which are assumed to be small and normally distributed with mean vector 0 and the variance covariance matrix as $\sigma^2 I_T$.

The classical least square (CLS) estimator $\widehat{\beta}_0$ which is the best linear unbiased estimator of β, is given by

$$\hat{\beta}_0 = (X'X)^{-1} X'y \tag{2.2}$$

with variance covariance matrix as $\sigma^2 (X'X)^{-1}$.

Let us consider the following class of Stein-rule estimators for β.

$$\hat{\beta} = \left[1 - \frac{K_1 s}{\beta_0' X'X\beta_0 + K_2 s + K_3}\right].\hat{\beta}_0 \tag{2.3}$$

which is characterized by three non-stochastic scalars K_1, K_2 and K_3. Here $s = \frac{1}{T-p}(y - X\widehat{\beta}_0)'(y - X\hat{\beta}_0)$ is the residual sum of squares.

The class is fairly general as it encompasses many interesting cases. For example, setting $K_1 = 0$, we get the CLS estimator $\widehat{\beta}_0$. On the other hand, if we set $K_1 > 0$, and $K_2 = K_3 = 0$, we obtain a class of estimators

$$\widehat{\beta}_1 = \left[1 - \frac{K_1 s}{\widehat{\beta}_0' X'X\widehat{\beta}_0}\right]\widehat{\beta}_0 \tag{2.4}$$

which is a special case of the class of estimators considered by Srivastava and Upadhyaya(1997).
Similarly, by setting $K_1>0$ and $K_3=0$ in (2.3), we get another interesting class of estimators, viz.,

$$\widehat{\beta}_2 = \left[1 - \frac{K_1 s}{\widehat{\beta}_0' X'X\widehat{\beta}_0' + K_2 s}\right]\widehat{\beta}_0 \tag{2.5}$$

which reduces to the well known Double K- class estimator of Ullah and Ullah(1978) if we take $K_1 = K_1^*$ and $K_2 = 1 - K_1^*$, the properties of which were studied by Vinod(1980), Carter(1989) and Menjoge(1984).
Another interesting possibility is when we choose $K_1>0$ and $K_2=0$ in (2.3), this provides the class of estimators as

$$\hat{\beta}_3 = \left[1 - \frac{K_1 s}{\hat{\beta}_0' X'X\hat{\beta}_0 + K_3}\right]\hat{\beta}_0. \tag{2.6}$$

An approximation to the sampling distribution of $\hat{\beta}$ is derived and studies its properties under the concentration probability criterion. The optimal choices of the characterizing scalars K_1, K_2 and K_3 for the relative dominance of the constituent members of this class over each other are derived there from.

3. THE LARGE SAMPLE APPROXIMATION OF PROBABILITY DISTRIBUTION

Before presenting the large sample approximation of the probability distribution function of the class of estimator $\hat{\beta}$, let us introduce the following notations. Let us denote by

$$r = \frac{1}{\sigma}(X'X)^{1/2}(\hat{\beta} - \beta) \quad , \quad \xi(r) = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}r'r}$$
$$\alpha = (X'X)^{1/2}\beta \quad , \quad \theta = \frac{\beta'X'X\beta}{\sigma^2 T} \tag{3.1}$$

Where r denotes the estimator in its standardized form on the basis of its leading term analysis, $\xi(r)$ denotes the probability density function of a standard normal vector r in terms of and θ denotes the noncentrality parameter.

Theorem 3.1 The large sampling approximations for the probability density function of the stochastic vector variable r, up to the order $O_p(T^{-3/2})$, is given by

$$f(r) = \left[1 + e_{-1/2} + e_{-1} + e_{-3/2} + O_p(T^{-2-j})\right] \xi(r) \qquad ; \quad j \geq 0 \tag{3.2}$$

Where

$$e_{-1/2} = -\frac{k_1(\alpha'r)}{(k_2+\theta)\sigma T}$$
$$e_{-1} = \frac{K_1}{(K_2+\theta)T}\left[(p - r'r) - \frac{(K_1+4)}{2(K_2+\theta)\sigma^2 T}(\alpha'\alpha - r'\alpha'\alpha r)\right] \tag{3.3}$$

$$e_{-3/2} = \frac{K_1}{(K_2+\theta)\sigma T}\left[\frac{(\alpha'r)}{(K_2+\theta)\sigma^2 T}\left\{\sigma^2 p(\theta+1)(r'r-p-1) - \frac{2(K_2\theta - \frac{2}{T}(4\alpha'\alpha - \alpha'r)^2)}{(K_2+\theta)^3}\right\}\right.$$
$$- \frac{p(\alpha'r)}{T} + \frac{2(r'r-p-1)(\alpha'r)}{(K_2+\theta)} + \frac{K_1(r'r-p-2)(\alpha'r)}{(K_2+\theta)T}$$
$$\left. - \frac{(6K_1+1)((\alpha'r)^2 - 3\alpha'\alpha)}{6(K_2+\theta)^2\sigma^2 T^2}\right]$$

The large sampling approximations for the sampling distributions of the estimators $\hat{\beta}_0, \hat{\beta}_2, \hat{\beta}_3$ and also of the least square estimator b can be obtained from (3.2) along with (3.3) by substitution of $K_2 = K_3 = 0$, $K_3 = 0$, $K_2 = 0$ *and* $K_1 = 0$ respectively. This does not disturb the sampling behavior of these estimators.
The theorem is derived in section 6.

4. CONCENTRATION PROBABILITY OF ESTIMATORS

The concentration of an estimator $\tilde{\beta}$ around the true unknown value β is defined in terms of the probability of its concentration around β. Thus the concentration probability of estimator $\tilde{\beta}$ in the neighborhood of β is given by

$$CP(\tilde{\beta}) = \Pr\left\{\left|\tilde{\beta} - \beta\right| \le \overline{m}\right\} = \Pr\left\{\left|\tilde{\beta}_j - \beta_j\right| \le m_j \;\; ; \;\; j = 1,2,\ldots\ldots p\right\} \tag{4.1}$$

Where $\overline{m} = col(m_1, m_2, \ldots\ldots m_p)$ is an arbitrarily chosen constant vector with jth element as m_j, and $\tilde{\beta}_j$ and β_j being jth elements of the estimator $\tilde{\beta}$ and the parameter vector β respectively. This gives the concentration of the estimator $\tilde{\beta}$ in the region bounded by planes $\left\{\left|\tilde{\beta}_j - \beta_j\right| \le m_j \;\; ; \;\; j = 1,2,\ldots\ldots p\right\}$ in the p-dimensional Euclidean space.

Let

$$\phi(m) = \int_{-m_p}^{m_p} \ldots\ldots \int_{-m_1}^{m_1} \xi(z) dz_1 \ldots\ldots\ldots dz_p \tag{4.2}$$

Where $\xi(z)$ is the standard multivariate normal density of variable vector z.

The concentration probability of least square estimator of b around β can be shown as

$$CP(b) = \phi(m) \tag{4.3}$$

Theorem 4.1: The large sample approximation for the concentration probability of estimator $\hat{\beta}$ around β in the region $(\left|r_j\right| \le m_j ; j = 1,2,\ldots.p)$ of the p dimensional Euclidean space, to the order $O_p(T^{\frac{-3}{2}})$, is given by

$$CP(\hat{\beta}) = \left\{1 + \frac{K_1}{(K_2 + \theta)T} trE - \frac{K_1(K_1 + 4)}{2(K_2 + \theta)^2 \sigma^2 T^2} \alpha' E \alpha\right\} \phi(m) \tag{4.4}$$

Where E is a diagonal matrix of constants with elements as $E = diag(e_1, e_2 \ldots\ldots e_p)$ where

$$e_j = \frac{m_j e^{-\frac{1}{2}m_j^2}}{\int_0^{m_j} e^{-\frac{1}{2}r_j^2} dr_j} \quad ; \quad j = 1,2,\ldots\ldots\ldots p \;.$$

5. THE OPTIMAL CHOICE OF ESTIMATOR

The concentration probability of estimators of least square estimator b and the general class of Stein-rule estimator $\hat{\beta}$ is same up to the order $O_p(T^{-\frac{1}{2}})$. However if we go up to the order $O_p(T^{-\frac{3}{2}})$ the difference in their concentration is given by

$$CP(\hat{\beta}) - CP(b) = \frac{K_1}{(K_2 + \theta)T}\left\{trE - \frac{(K_1 + 4)}{2(K_2 + \theta)\sigma^2 T} \alpha' E \alpha\right\}\phi(m) \tag{5.1}$$

The condition will definitely hold true as long as

$$0 < K_1 < \frac{2trE}{\lambda_{\max}(E)} - 4 \tag{5.2}$$

Assuming without loss of generality $e_1 \le e_2 \le \ldots\ldots\ldots \le e_p$ the sufficient condition for dominance of the class of shrinkage estimators $\hat{\beta}$ over the classical least square estimator is

$$0 < K_1 < 2\left(\frac{\sum_{j=1}^{p} e_j}{e_p} - 2\right) \tag{5.3}$$

In particular if $m_1 = m_2 = \ldots\ldots\ldots\ldots = m_p = m_0$ the condition (5.3) reduces to

$$0 < K_1 < 2(p-2) \quad ; \quad p > 2 \tag{5.4}$$

This matches exactly with the necessary and sufficient condition of dominance of Stein rule estimator over least square estimator under the predictive risk criterion.

To explore further among the various choices of these estimators within the class of $\hat{\beta}$, we observe that the concentration of $\hat{\beta}$ up to order $O_p(T^{-\frac{3}{2}})$ does not involve the characterizing scalar K_3.

The pair of estimators $\tilde{\beta} \equiv (\hat{\beta}, \hat{\beta}_2)$ as well as $\tilde{\tilde{\beta}} \equiv (\hat{\beta}_0, \hat{\beta}_3)$ has the same concentration probability approx. up to the order $O_p(T^{-\frac{3}{2}})$. The difference in their concentration probability is given by

$$CP(\tilde{\tilde{\beta}}) - CP(\tilde{\beta}) = \frac{K_1 K_2}{(K_2+\theta)\theta T}\left\{trE - \frac{(K_1+4)(K_2+2\theta)}{2(K_2+\theta)\theta\sigma^2 T}\alpha' E\alpha\right\}\phi(m) \tag{5.5}$$

The difference will be positive if and only if

$$0 < K_1 < \frac{2trE}{\alpha' E\alpha \big/ \alpha'\alpha}\,\frac{K_2+\theta}{K_2+2\theta} - 4 \tag{5.6}$$

Sufficient condition to hold (5.6) good is

$$0 < K_1 < \frac{trE}{\lambda_{\max}(E)} - 4 \tag{5.7}$$

Which in particular case when $m_1 = m_2 = \ldots\ldots\ldots\ldots = m_p = m_0$ reduces to

$$0 < K_1 < (p-4) \quad ; \quad p > 4 \tag{5.8}$$

For selecting the best estimator in the pair $\tilde{\tilde{\beta}} \equiv (\hat{\beta}_0, \hat{\beta}_3)$, the difference in the concentration probabilities of estimators $\hat{\beta}_0$ and $\hat{\beta}_3$ to the order $O_p(T^{-2})$ comes out to be

$$CP(\hat{\beta}_0) - CP(\hat{\beta}_3) = \frac{K_1 K_3}{\theta^2\sigma^2 T^2}\left\{trE - \frac{(K_1+4)}{\theta\sigma^2 T}\alpha' E\alpha\right\}\phi(m) \tag{5.9}$$

This is positive if and only if we have

$$0 < K_1 < \frac{trE}{\alpha' E\alpha \big/ \alpha'\alpha} - 4 \quad , \quad K_3 > 0 \tag{5.10}$$

The sufficient condition for (5.10) to hold true comes out to be

$$0 < K_1 < \frac{trE}{\lambda_{\max}(E)} - 4 \quad , \quad K_3 > 0 \tag{5.11}$$

The necessary and sufficient condition when $m_1 = m_2 = ... = m_p = m_0$ comes out to be

$$0 < K_1 < (p-4) \quad ; \quad p > 4 \tag{5.12}$$

For selecting the best estimator in the pair $\tilde{\beta} \equiv (\hat{\beta}, \hat{\beta}_2)$, the difference in the concentration probabilities of estimators is

$$CP(\hat{\beta}_2) - CP(\beta) = \frac{K_1 K_3}{(K_2+\theta)^2 \sigma^2 T^2} \left\{ trE - \frac{(K_1+4)}{(K_2+\theta)\sigma^2 T} \alpha' E \alpha \right\} \phi(m) \tag{5.13}$$

Thus the estimator $\hat{\beta}_2$ will have better concentration around β than that of β if and only if

$$0 < K_1 < \frac{trE}{\alpha' E \alpha \big/ \alpha'\alpha} \frac{(K_2+\theta)}{\theta} - 4 \tag{5.14}$$

The sufficient condition to hold (5.14) good is

$$0 < K_1 < \frac{trE}{\lambda_{\max}(E)} - 4 \quad , \quad K_3 > 0 \tag{5.15}$$

For $m_1 = m_2 = ... = m_p = m_0$ it reduces to

$$0 < K_1 < (p-4) \quad ; \quad p > 4 \tag{5.16}$$

Thus we summarize that if the characterizing scalar K_1 is chosen as $0 < K_1 < 2(p-2)$ with $K_2 > 0$, $K_3 > 0$ the estimator $\hat{\beta}$ will definitely superior to the classical least square estimator b. Further, if K_1 is chosen $0 < K_1 < (p-4)$ the estimator $\hat{\beta}_0$ will give the best performance.

7. PROOF OF THE RESULTS

Observing that

$$s - \sigma^2 = \lambda_{-\frac{1}{2}} + \lambda_{-1} \tag{6.1}$$

$$where \quad \lambda_{-1/2} = \varepsilon$$

$$\lambda_{-1} = \frac{p\sigma^2 - u'P_X u}{T}$$

$$with \quad P_X = X(X'X)^{-1}X' \quad and$$

$$E(\varepsilon) = 0$$

$$E(\varepsilon^2) = \frac{2\sigma^4}{T} \tag{6.2}$$

and negative suffixes in λ denote the order of terms in probability. Thus,

$$r = \frac{1}{\sigma}(X'X)^{\frac{1}{2}}(\hat{\beta} - \beta) \cong r_0 + r_{-\frac{1}{2}} + r_{-1} + r_{-\frac{3}{2}} + O_p(T^{-\frac{3}{2}-j}) \quad ; \quad j \geq 0$$

Where $r_0 = \dfrac{1}{\sigma}(X'X)^{-\frac{1}{2}} X'u = z$

$$r_{-\frac{1}{2}} = -\frac{K_1}{(K_2+\theta)\sigma T}(X'X)^{\frac{1}{2}}\beta$$

$$r_{-1} = -\frac{K_1}{(K_2+\theta)T}z - \frac{K_1}{(K_2+\theta)^2\sigma^3 T}(\theta\varepsilon - \frac{2u'X\beta}{T})(X'X)^{\frac{1}{2}}\beta$$

$$r_{-\frac{3}{2}} = \frac{K_1}{(K_2+\theta)^2\sigma^3}\left[\frac{\left\{K_3+(1+\theta)z'z\sigma^2\right\}}{T^2} + \frac{1}{(K_2+\theta)\sigma^2 T}\left\{K_2\varepsilon + \frac{2\sigma z'\alpha}{T}\right\}\left\{\theta\varepsilon - \frac{2\sigma z'\alpha}{T}\right\}\right]\alpha$$
$$- \frac{K_1(p-2)\alpha}{(K_2+\theta)\sigma T^2} - \frac{K_1}{(K_2+\theta)^2\sigma^2 T}\left\{\theta\varepsilon z - \frac{2\sigma z'\alpha}{T}z\right\}$$

The characteristic function of the random vector r is defined as

$$\psi_r(h) = E(e^{ih'r})$$
$$= E\left[\exp ih'\left\{r_0 + r_{-\frac{1}{2}} + r_{-1} + r_{-\frac{3}{2}} + O_p(T^{-\frac{3}{2}-j})\right\}\right] \quad ; \quad j \geq 0$$

On expansion and evaluating the respective expectations the approximation for the characteristic function of random vector variable r, up to the order $O_p(T^{-\frac{3}{2}-j})$ as

$$\psi_r(h) = \left\{1 + \psi_{-\frac{1}{2}} + \psi_{-1} + \psi_{-\frac{3}{2}} + O_p(T^{-\frac{3}{2}-j})\right\}\exp(-\frac{1}{2}h'h) \quad ; \; j \geq 0$$

Employing this approximation for the characteristic function in the inversion theorem to get the large sample approximations for the probability density function $f(r)$

$$f(r) = \frac{1}{(2\pi)^p}\int_{-\infty}^{\infty}..........\int_{-\infty}^{\infty}\exp(-ih'r)\ \psi_h(r)\,dh$$

For evaluating the approximation for the concentration probability of estimator $\hat{\beta}$ to be close to β we apply

$$CP(\hat{\beta}) = \int_{-m_p}^{m_p}........\int_{-m_1}^{m_1} f(r)dr_1 dr_2.........dr_p$$

Evaluating this multiple integral we get the result derived in the theorem.

RECEIVED: MARCH, 2017
REVISED: JULY, 2017

REFERENCES

[1]. ADKINS, L.C. and EELLS, J.B. (1995): Improved estimators of energy models, **Energy Econ.** 17, 15-25.

[2]. BAO, H. and .WAN, A.T.K (2007): Improved estimators of hedonic housing price models, **J. Real Estate Res.** 29, 267-304.

[3]. CARTER, R.A.L. (1989): Improved Stein-rule Estimator for Regression Problems, **Journal of econometrics**, 17, 113-123.

[4]. JUDGE, G.J., GRIFFITHS, W.E., HILL, R.C., LUTKEPOHL, H. and LEE, T.C. (1985): ***The Theory and Practice of Econometrics***, John Wiley, NY.

[5]. KNIGHT, J.R., HILL, R.C. and SIRMANS, C.F. (1992):Biased prediction of housing values,, **J. Amer. Real Estate Urban Econ. Assoc.** 20, 427-456.

[6]. KNIGHT, J.R., HILL, R.C. and SIRMANS, C.F. (1993): Estimation of hedonic price models using non-sample information: A Monto Carlo study, **J. Urban Econ.**34, 319-346.

[7]. KNIGHT, J.R., HILL, R.C. and SIRMANS, C.F. (1993a): Stein rule estimation in real state appraisal, **The Appraisal Journal** 15, 539-544.

[8]. MENJOGE, S.S. (1984): On double K-Class Estimators of Coefficients in Linear Regression, **Economic Letters**, 15, 295-300.

[9]. PANT, R. and MANOJ KUMAR (2012): The Optimal Stein Estimation in Linear Models, **Investigations in Mathematical Sciences**, Vol 2, No. 2, 2012, 171-184.
[10]. RAO, C.R. (1981): **Some comments on the minimum Mean Square as a Criterion of Estimation, Statistics and Related Topic**s edited by M.Csorgo, D.A.Dawson, J.N.K.Rao, and A.K.M.E.Saleh, Amesterdam, North Holland, 123-147.
[11]. ROBERT, C.P., HWANG, J.T. and STRAWDERMAN, W.E. (1993): Is Pitman Closeness a Reasonable Criterion?, **Journal of American Statistical Association**, 88, 57–76.
[12]. SHALABH, TOUTENBERG, H. and HEUMANN, C. (2009): Stein-rule estimation under an extended balanced loss function, **Journal of Statistical Computation and Simulation**,10, 1259-1273.
[13]. SRIVASTAVA, V.K. and UPADHYAYA, S. (1977): Properties of Stein-Like Estimators in the Regression Model when Disturbances are Small, **Journal of Statistical Research**, 11, 5-12.
[14]. ULLAH, A. and ULLAH, S. (1978): Double K-Class Estimators of Coefficients in Linear Regression, **Econometrica**, 44, 705-722.
[15]. VINOD, H.D. (1980): Improved Stein-Rule Estimator for Regression Problem, **Journal of Econometrics**, 12, 143-150.
[16]. WAN, A.T.K., CHATURVEDI, A. and ZOU, G.H. (2003): Unbiased estimation of the MSE matrices of improved estimators in linear regression, **J. Appl. Stat.** 30, 173-189.