The Optimal Stein Estimation relative to Concentration Probability

Raman Pant
Dept. of Statistics
Mahatma Gandhi Kashi Vidyapith
Varanasi.

Abstract

General family of Stein rule estimators is considered for general linear regression model. The approximation to the probability density function of the estimator is derived assuming the disturbances to be small. The concentration probability around the true parameter is evaluated there from. Using this concentration probability the optimal choice of biasing scalars is discussed.

Keywords: Simple linear regression model, Stein rule estimator, Small disturbance approximation, Sampling distribution, Concentration probability.

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), Wan et al.((2003)), McClatchey and VandenHul(2005), Grauer and Hakansson(2001) and Stevenson(2001).

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(1977) and Ullah and Ullah(1978). Shrinkage estimators possess an implicit optimality with respect to classical least square estimator, but the gains of these estimators should be assessed in terms of their concentration around the true unknown parameter they aim to estimate. This is important because the two approaches are contradictory in nature. The quadratic loss theme emphasizes on the deviations of estimate from the true value while the concentration theme revolves around the closeness of estimates around the parameter. The mean squared error criterion safeguards from larger deviations of the estimate rather than providing a measure of proximity of the estimates to the unknown value.

Rao(1981) suggested to employ a proper measure of proximity of estimates to the true value for judging 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. For detailed discussions on the merits of the two criterion of concentration see Robert et al (1993).

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 expression for the probability of concentration there from. Therefore the small **disturbance approximation** is used to get the approximation of probability density function of these non linear shrinkage estimators 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 small disturbance approximation of probability density function of the proposed general class and derived the 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

$$y = x\beta + u \qquad (2.1)$$

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

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

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

However, employing Stein's philosophy, we can improve upon the performance of least squares estimator by shrinking it towards the null vector. 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 $s = \frac{1}{T-p}(y - X\hat{\beta}_0)'(y - X\hat{\beta}_0)$ is the residual sum of squares.

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

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

which is a special case of the class of estimators considered by Srivastava and Upadhyaya(1977).

Similarly, by setting $K_1>0$ and $K_3=0$ in (2.3), we get another interesting class of estimators, viz.,

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

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(1981), Menjoge(1984) and Srivastava and Chaturvedi(1986).

Another interesting possibility is when we choose $K_1>0$ and $K_2=0$ in (2.3), this provides the class of estimators as

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

We shall concentrate our attention to the general class $\hat{\beta}$ and study its properties under the concentration probability criterion. There from we shall explore 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.

3. The small disturbance 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

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

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

$$f(r) = \left[1 + \sigma e_1 + \sigma^2 e_2 + \sigma^3 e_3 + O(\sigma^{4+j})\right]\xi(r) \quad ; \quad j \geq 0 \qquad (3.2)$$

Where

$$e_1 = -\frac{k_1(\alpha' r)}{(k_3 + \theta_o)}$$
$$e_2 = \frac{K_1}{(K_3 + \theta_o)T}\left[(p - r'r) - \frac{(K_1\frac{n+2}{n} + 2)}{2(K_3 + \theta_o)}(\alpha'\alpha - r'\alpha'\alpha r)\right] \qquad (3.3)$$
$$e_3 = \frac{K_1(\alpha' r)}{(K_3 + \theta_o)^2}\Bigg[3r'r - 4\alpha'\alpha - 2(p+2) + \frac{n+2}{n}\{K_1(r'r - p - 2) + K_2\} - (\alpha' r)^2 - 3\alpha'\alpha\frac{K_1(n+2)}{6n(K_3 + \theta_o)} + 4\left\{\frac{K_1(n+4)}{n} + 12\right\}\Bigg]$$

The small disturbance 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 the appendix

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

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

Where $\overline{m} = col(m_1, m_2, ...m_p)$ 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 $\left\{\left|\tilde{\beta}_j-\beta_j\right| \le m_j \ ; \ j=1,2,.......p\right\}$ in the p-dimensional Euclidean space.

Let

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

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

$$CP(b) = \phi(m) \qquad (4.3)$$

<u>Theorem 4.1</u>: The small disturbance approximation for the concentration probability of estimator $\hat{\beta}$ around β in the region $(\left|r_j\right| \le m_j \, ; j=1,2,.....p)$ of the p dimensional Euclidean space, to the order $O(\sigma^3)$, is given by

$$CP(\hat{\beta}) = \left\{ 1 + \frac{K_1\sigma^2}{(K_3+\theta_o)} trE - \frac{K_1\left\{\frac{K_1(n+2)}{n}+4\right\}\sigma^2}{(K_3+\theta_o)^2}\alpha' E\alpha \right\} \phi(m) \qquad (4.4)$$

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

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

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 $O(\sigma)$. However if we go up to the order $O(\sigma^3)$ the difference in their concentration is given by

$$CP(\hat{\beta}) - CP(b) = \frac{K_1\sigma^2}{(K_3+\theta_o)T}\left\{ trE - \frac{\frac{K_1(n+2)}{n}+4}{2(K_3+\theta_o)}\alpha' E\alpha \right\}\phi(m) \qquad (5.1)$$

Thus the class of estimator $\hat{\beta}$ will have greater concentration around β to the order $O(\sigma^3)$, if and only if, we have

$$0 < K_1 < 2\left(\frac{n}{n+2}\right)\left[\frac{tr\ E}{\left(\frac{\alpha' E\alpha}{\theta_o}\right)}\left(\frac{K_3+\theta_o}{\theta_o}\right) - 2\right]$$

The condition will definitely hold true as long as

$$0 < K_1 < \frac{2\left(\frac{trE}{\lambda_{\max}(E)} - 2\right)}{\frac{n+2}{n}} \qquad (5.2)$$

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

$$0 < K_1 < \frac{2\left(\frac{\sum_{j=1}^{p} e_j}{e_p} - 2\right)}{n+2 \big/ n} \qquad (5.3)$$

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

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

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 $O(\sigma^3)$ does not involve the characterizing scalar K_2. The pair of estimators $\tilde{\beta} \equiv (\hat{\beta}, \hat{\beta}_3)$ as well as $\tilde{\tilde{\beta}} \equiv (\hat{\beta}_0, \hat{\beta}_2)$ has the same concentration probability approx. up to the order $O(\sigma^3)$ .The difference in their concentration probability approximations is given by

$$CP(\tilde{\tilde{\beta}}) - CP(\tilde{\beta}) = \frac{K_1 K_3}{(K_3+\theta_o)\theta_o} \left\{ trE - \frac{\left(\frac{K_1(n+2)}{n} + 4\right)\left(\frac{K_3 + 2\theta_o}{K_3+\theta_o}\right)}{2\theta_o} \alpha' E \alpha \right\} \phi(m) \quad 5.5)$$

The difference will be positive if and only if

$$0 < K_1 < 2\left(\frac{n}{n+2}\right)\left(\frac{trE}{\alpha' E\alpha / \alpha'\alpha} \frac{K_3+\theta_o}{K_3+2\theta_o} - 2\right) \tag{5.6}$$

Sufficient condition to hold (5.6) good is

$$0 < K_1 < \frac{\frac{trE}{\lambda_{\max}(E)} - 4}{\frac{n+2}{n}} \tag{5.7}$$

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

$$0 < K_1 < \frac{(p-4)}{\frac{n+2}{n}} \quad ; \quad p > 4 \tag{5.8}$$

For selecting the best estimator in the pair $\tilde{\tilde{\beta}} \equiv (\hat{\beta}_0, \hat{\beta}_2)$, the difference in the concentration probabilities of estimators $\hat{\beta}_0$ and $\hat{\beta}_2$ to the order $O(\sigma^4)$ comes out to be

$$CP(\hat{\beta}_0) - CP(\hat{\beta}_2) = \frac{\sigma^4 K_1 K_2 (n+2)}{\theta_o^{\ 2}} \left\{ trE - \frac{\left(\frac{K_1(n+4)}{n} + 4\right)}{\theta_o} \alpha' E \alpha \right\} \phi(m) \qquad (5.9)$$

This is positive if and only if we have

$$0 < K_1 < \left(\frac{n}{n+4}\right)\left(\frac{trE}{\alpha' E \alpha / \alpha'\alpha} - 4\right) \quad , \quad K_2 > 0 \qquad (5.10)$$

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

$$0 < K_1 < \frac{\left(\frac{trE}{\lambda_{\max}(E)} - 4\right)}{\left(\frac{n+4}{n}\right)} \quad , \quad K_2 > 0 \qquad (5.11)$$

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

$$0 < K_1 < \frac{(p-4)}{\frac{n+4}{n}} \quad ; \quad p > 4 \qquad (5.12)$$

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.

6. Appendix

Observing that

$$s - \sigma^2 = \lambda_{-\frac{1}{2}} + \lambda_{-1} \qquad (A.1)$$

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

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

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

$$E(\varepsilon) = 0$$
$$E(\varepsilon^2) = \frac{2\sigma^4}{T} \qquad \text{(A.2)}$$

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

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

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

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

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

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

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.

References

1. Adkins, L.C. and J.B.Eells(1995): "*Improved estimators of energy models*", Energy Econ. 17, pp. 15-25.
2. Bao, H. and A.T.K.Wan(2007): *Improved estimators of hedonic housing price models*, J. Real Estate Res. 29, pp 267-304.
3. Carter, R.A.L. (1989): "*Improved Stein-rule Estimator for Regression Problems*", Journal of econometrics, 17, pp. 113-123.
4. Grauer,R.R. and N.H. Hakansson(2001*): Applying portfolio change and conditional performance measure: the case of industry rotation via dynamic investment model*, Rev. Quant. Finance Account 17, pp. 237-265.
5. 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.
6. Knight, J.R., R.C. Hill and C.F. Sirmans(1992):*Biased prediction of housing values*,", J. Amer. Real Estate Urban Econ. Assoc. 20,pp. 427-456.
7. Knight, J.R., R.C. Hill and C.F. Sirmans(1993):*Estimation of hedonic price models using non-sample information: A Monto Carlo study*", J. Urban Econ.34,pp. 319-346.
8. Knight, J.R., R.C. Hill and C.F. Sirmans(1993a):*Stein rule estimation in real state appraisal*", The Appraisal Journal 15,pp. 539-544.
9. McClatchey,C.A. and S.P.Vandenful(2005):*The efficacy of optimization modeling as a retirement strategy in the presence of estimation error*, Financ. Serv. Rev. 14, pp. 173-189

10. Menjoge, S.S. (1984):"*On double K-Class Estimators of Coefficients in Linear Regression*", Economic Letters, 15, pp. 295-300.

11. Rao, C.R. (1981): *Some comments on the minimum Mean Square as a Criterion of Estimation*, Statistics and Related Topics edited by M.Csorgo, D.A.Dawson, J.N.K.Rao, and A.K.M.E.Saleh, Amesterdam, North Holland, 123-147.

12. Robert, C.P., J.T.Hwang and W.E. Strawderman (1993): *Is Pitman Closeness a Reasonable Criterion?*, Journal of American Statistical Association, 88, 57–76.

13. Shalabh, H.Toutenberg and C.Heumann, (2009): *Stein-rule estimation under an extended balanced loss function*, Journal of Statistical Computation and Simulation,10, pp. 1259-1273.

14. Srivastava, V.K. and S. Upadhyaya, (1986): "*A Necessary and Sufficient Condition for the Dominance of An Improved Family of Estimators in Linear Regression Models*", Economic Letters, 20, pp. 345-349.

15. Srivastava, V.K. and S. Upadhyaya, (1977): "*Properties of Stein-Like Estimators in the Regression Model when Disturbances are Small"*, Journal of Statistical Research, 11, pp. 5-12.

16. Stevenson,S : *Bayes-Stein estimators and international real estate asset allocation*, J. Real Estate Res.,2001, 21, pp. 89-103.

17. Ullah, A. and S. Ullah(1978): "*Double K-Class Estimators of Coefficients in Linear Regression*", Econometrica, 44,pp. 705-722.

18. Vinod, H.D. (1980):"*Improved Stein-Rule Estimator for Regression Problem*", Journal of Econometrics, 12, pp. 143-150.

19. Wan, A.T.K., A.Chaturvedi and G.H.Zou(2003): *Unbiased estimation of the MSE matrices of improved estimators in linear regression*, J. Appl. Stat. 30, pp. 173-189.