SINET: Ethiop. J. Sci., 22(2):191-198, 1999
 ISSN: 0379-2897

SHRINKAGE ESTIMATION UNDER PITMAN NEARNESS

Eshetu Wencheko

Department of Statistics, Faculty of Science,
Addis Ababa University, PO Box 1176, Addis Ababa, Ethiopia

ABSTRACT: The Pitman Measure of Nearness criterion is applied to compare the performance of the ordinary least square estimator and the shrunken estimator. This is done by considering the weighted and the unweighted norms of both estimators. The corresponding Pitman Nearness probabilities are provided. An adaptive procedure for obtaining operational shrinkage factors is suggested.

Key words/phrases: Pitman Measure of Nearness, Pitman Nearness probability, shrinkage estimation

INTRODUCTION

Since Rao's work (1981) a large number of research publications on Pitman Measure of Nearness (PMN) has appeared. PMN was introduced by Pitman (1937), and until the beginning of the seventies his ideas did not attract the attention of researchers, partly due to the fact that many felt comfortable with the criterion of Mean Squared Error (MSE). Dyer and Keating (1983), Keating and Gupta (1984), Keating (1985), Keating and Mason (1985a; 1985b) and Rao *et al.* (1986) are but among those prominent researchers who contributed much to the popularisation of PMN. The monograph by Keating *et al.* (1993) provides an illuminating account of PMN and a long list of publications on comparisons of estimators of univariate parameters and scalar functions of the same.

This paper provides theoretical results of comparison of the unbiased estimator of the regression coefficients of the multiple linear regression model and the shrunken estimator (Mayer and Willke, 1973). At this juncture we would like to refer the readership to a related work by Conerly and Hardin (1991).

Before proceeding with the technical discussion let us introduce the following definitions.

Suppose T_1 and T_2 are two estimators of the parametric function $\tau(\theta)$, $\theta \in \Theta \subset R^k$. The Pitman Measure of Nearness of T_1 relative to T_2 is the probability

$$PN_\theta(T_1,T_2) = P[L(T_1 - \tau(\theta)) < L(T_2 - \tau(\theta))]$$

where $L(,)$ is a convex loss function of the form $(T_j - \tau(\theta))'W(T_j - \tau(\theta))$ with symmetric weight matrix $\mathbf{W}$.

Accordingly, we define the Pitman Nearness of T_1 to $\tau(\theta)$ relative to T_2 as follows:

Suppose T_1 and T_2 are two estimators of the parametric function $\tau(\theta)$. Then T_1 is said to Pitman-Nearer to $\tau(\theta)$ relative to T_2 if

$$PN_\theta(T_1,T_2; W) = P[(T_1-\tau(\theta))'W(T_1-\tau(\theta)) < (T_2-\tau(\theta))'W(T_2-\tau(\theta))] \geq 0.5$$

for all $\theta \in \Theta$ with strict inequality for at least one θ.

COMPARISON OF THE SHRUNKEN AND UNBIASED ESTIMATORS

Consider the multiple linear regression model $M\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{I}_n\}$ where $\mathbf{y} = (y_1, \ldots, y_n)'$ is a vector of observations, $\beta = (\beta_1, \ldots, \beta_p)'$ is an unknown but fixed p-vector of regression coefficients, $\mathbf{X}$ is a nxp non-stochastic regressor matrix of full column rank and $\sigma^2 > 0$ is the unknown constant variance of the error terms. Furthermore, we assume that the vector of dependent variables $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$.

The ordinary least squares estimator (OLSE) of β is given by $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. It is also the maximum likelihood estimator under normality. In both situations it is best linear unbiased in the class of linear homogenous estimators of β. In the presence of two or more near-collinear columns of $\mathbf{X}$, however, $\mathbf{b}$ does not perform well in terms of mean error squared risk. Under such circumstances the use of biased linear homogenous estimators is appealing. One such estimator, the shrunken estimator (Mayer and Willke, 1973) is simply a scalar multiple, in fact a shrinkage, of b with a shrinkage factor $c \in (0, 1)$.

In this section we compare performance of **b** and c**b** through PMN. The weight matrices we consider are the identity matrix of dimension p and the inverse of the dispersion matrix of **b**. Apparently the former introduces no weights, and therefore it compares the Euclidean norm of c**b** and **b**, while the latter utilises the Mahalanobis metric.

I. Weight matrix (the inverse variance-covariance matrix of b): $\sigma^{-2}X'X$

The following equivalence holds for the above choice:

$$(c\mathbf{b}-\beta)'\,\mathbf{X}'\mathbf{X}\,(c\mathbf{b}-\beta) < (\mathbf{b}-\beta)'\,\mathbf{X}'\mathbf{X}\,(\mathbf{b}-\beta)$$
$$\Leftrightarrow \mathbf{b}'\,\mathbf{X}'\mathbf{X}\,\mathbf{b} - 2(1+c)^{-1}\,\beta'\,\mathbf{X}'\mathbf{X}\beta > 0$$
$$\Leftrightarrow (\mathbf{b}-(1+c)^{-1}\,\beta)'\,\mathbf{X}'\mathbf{X}\,(\mathbf{b}-(1+c)^{-1}\,\beta) > (1+c)^{-2}\,\beta'\mathbf{X}'\mathbf{X}\beta.$$

Therefore, denoting $\sigma^{-2}\mathbf{X}'\mathbf{X}$ by $\mathbf{W}_1$ we obtain

$$PN_\beta(c\mathbf{b},\mathbf{b};\mathbf{W}_1) = P[(\mathbf{b}-(1+c)^{-1}\beta)'\,\mathbf{X}'\mathbf{X}(\mathbf{b}-(1+c)^{-1}\beta) > (1+c)^{-2}\,\beta'\mathbf{X}'\mathbf{X}\beta].$$

Since

$$(\mathbf{X}'\mathbf{X})^{1/2}\,(\mathbf{b}-(1+c)^{-1}\,\beta) \sim N(c(1+c)^{-1}\,(\mathbf{X}'\mathbf{X})^{1/2}\,\beta,\ \sigma^2\mathbf{I}_p)$$

it follows that $Q^*_W = \sigma^{-2}\,Q_W \sim \chi^2(p, \eta)$, where $Q_W = (\mathbf{b}-(1+c)^{-1}\,\beta)'\mathbf{X}'\mathbf{X}\,(\mathbf{b}-(1+c)^{-1}\,\beta)$ and the non-centrality parameter $\eta = c^2\,\sigma^{-2}\,(1+c)^{-2}\,\beta'\mathbf{X}'\mathbf{X}\beta$.

Thus, the PMN probability of c**b** relative to **b** is

$$PN_\beta\,(c\mathbf{b},\,\mathbf{b};\,\mathbf{W}_1) = P[Q^*_W > \sigma^{-2}\,(1+c)^{-2}\,\beta'\mathbf{X}'\mathbf{X}\beta].$$

II. The Euclidean Norm

Comparison of the squared unweighted norms of $c\mathbf{b}$ and the OLSE $\mathbf{b}$ yields

$$(c\mathbf{b}-\beta)'(c\mathbf{b}-\beta) < (\mathbf{b}-\beta)'(\mathbf{b}-\beta)$$
$$\Leftrightarrow c^2\mathbf{b}'\mathbf{b}-2c\mathbf{b}'\beta < \mathbf{b}'\mathbf{b}-2\mathbf{b}'\beta$$
$$\Leftrightarrow \mathbf{b}'\mathbf{b}-2(1+c)^{-1}\,\mathbf{b}'\beta > 0$$
$$\Leftrightarrow (\mathbf{b}-(1+c)^{-1}\,\beta)'(\mathbf{b}-(1+c)^{-1}\,\beta) > (1+c)^{-2}\,\beta'\beta.$$

Note that the vector $\mathbf{b}-(1+c)^{-1}\,\beta \sim N(c(1+c)^{-1}\beta,\ \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, and that its components are not independent unless $\mathbf{X}'\mathbf{X}$ is diagonal. In general, therefore, the distribution of the inner product of this vector, namely $Q_I = (\mathbf{b}-(1+c)^{-1}\,\beta)'(\mathbf{b}-(1+c)^{-1}\,\beta)$ is not a non-central χ^2 distribution. In fact, it can be shown that Q_I is a weighted sum of p independent, single degree of freedom, non-central χ^2 random variables. If we decompose the matrix $(\mathbf{X}'\mathbf{X})^{-1}$ into $\mathbf{V}\mathbf{L}^{-1}\mathbf{V}'$ where $\mathbf{V}$ contains its orthonormal eigenvectors and $\mathbf{L}$ is the diagonal matrix of the corresponding eigenvalues l_j, $j = 1, \ldots, p$, then $\mathbf{V}'(\mathbf{b}-(1+c)^{-1}\,\beta) \sim N(\mu,\ \sigma^2\,\mathbf{L}^{-1})$, with $\mu = c(1+c)^{-1}\mathbf{V}'\,\beta$. Consequently,

$$Q_I = (\mathbf{b}-(1+c)^{-1}\,\beta)'(\mathbf{b}-(1+c)^{-1}\,\beta) = \Sigma \mathbf{Z}_j^2$$

where $\mathbf{Z}_j \sim N(\mu,\ \sigma^2/\,l_j)$ so that $\mathbf{Z}_j^2 \sim \sigma^2/\,l_j\ \chi^2\ (1,\ l_j\ \mu_j^2/\sigma^2)$. Therefore,

$$\mathbf{Q}_I = \sigma^2\ \Sigma l_j^{-1}\ \mathbf{R}_j,$$

where $\mathbf{R}_j \sim \chi^2(1,\ l_j\ \mu_j^2/\sigma^2)$. Thus, we get

$$PN_\beta\ (c\mathbf{b},\ \mathbf{b};\ \mathbf{W}_2 = \mathbf{I}_p) = P[Q^*_I > \beta'\beta/\sigma^2\ (1+c)^2],$$

where $Q^*_I = Q_I/\sigma^2$.

We can also very easily compare two shrunken estimators $c_1\mathbf{b}$ and $c_2\mathbf{b}$, $c_1 < c_2$, by applying either of the above norms. For the purpose of demonstration let us use the Euclidean norm. Observe that

$$(c_1\mathbf{b}-\beta)'\ (c_1\mathbf{b}-\beta) < (c_2\mathbf{b}-\beta)'(c_2\mathbf{b}-\beta)$$

$$\Leftrightarrow(\mathbf{b}-(c_1 + c_2)^{-1}\,\beta)'(\mathbf{b}-(c_1 + c_2)^{-1}\,\beta) > (c_1 + c_2)^{-2}\,\beta'\beta.$$

Since $\mathbf{b}-(c_1 + c_2)^{-1}\,\beta \sim N((c_1 + c_2 - 1)\,(c_1 + c_2)^{-1}\beta,\ \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ the vector $\mathbf{V}'(\mathbf{b} - (c_1 + c_2)^{-1}\,\beta) \sim N(\xi,\ \sigma^2\,\mathbf{L}^{-1})$. The inner product of $(\mathbf{b}-(c_1 + c_2)^{-1}\,\beta)$ can be represented as a sum

$$(\mathbf{b}-(c_1 + c_2)^{-1}\,\beta)'(\mathbf{b}-(c_1 + c_2)^{-1}\,\beta) = \Sigma \mathbf{A}_j^2,$$

where $\mathbf{A}_j \sim N(\xi_j,\ \sigma^2/l_j)$, with $\xi_j = (c_1 + c_2 - 1)\,(c_1 + c_2)^{-1}\mathbf{P}'\,\beta$. Note that $\mathbf{A}_j^2 \sim \sigma^2/l_j\ \chi^2(1,\ l_j\ \xi_j^2/\sigma^2)$. Therefore, $\Sigma\mathbf{A}_j^2$ can be expressed as

$$\mathbf{F} = \sigma^2\ \Sigma l_j^{-1}\mathbf{H}_j,$$

with $\mathbf{H}_j \sim \chi^2(1,\ l_j\xi_j^2/\sigma^2)$. Thus, the PMN of $c_1\mathbf{b}$ relative to $c_2\mathbf{b}$ is

$$\mathrm{PN}_\beta(c_1\mathbf{b},\ c_2\mathbf{b}) = \mathrm{P}[\mathbf{D} > \beta'\beta/\sigma^2\ (c_1 + c_2)^2],$$

where $\mathbf{D} = \mathbf{F}/\sigma^2$.

OPERATIONAL SHRINKAGE FACTORS

The discussion in the previous section gives theoretical conditions under which the PMN could be used to compare the shrunken regression estimator with the standard unbiased estimator. Those conditions depend on unknown parameters of the model through the non-centrality parameter and the shrinkage factor.

For application purposes we may be interested in a set of values of the shrinkage factors which guarantee PMN dominance of either of the estimators. Operational shrinkage factors would be obtained if appropriate estimators could be used to replace the non-centrality parameter. Because there are no analytical methods that lend themselves for the derivation of parametric functions like the ones that appear as non-centrality parameters, which are non-linear in β and σ^2, we opt to use plug-in strategies in order to get operational substitutes for these. The plug-in strategy is a straightforward method that relies on direct substitutions of estimators for their respective parameters. Direct substitution, however, cannot always promise good estimators.

In order to arrive at improved estimators for the non-centrality parameters we can go for a further change of the resultant adaptive estimators of the non-centrality parameters by modifying the deterministic parts of that estimator which is obtained from a direct plug-in.

Referring to case I above a direct substitution of $\mathbf{b}$ for β, and the unbiased estimator or maximum likelihood estimator for σ^2 in $\phi = \beta'\mathbf{X}'\mathbf{X}\beta/\sigma^2$ gives the estimator $\hat{\varphi} = \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}/s^2$. By the same analogy estimators for the signal $\zeta = \beta'\beta$ from a class of $\mathbf{b}'\Omega(\mathbf{X})\mathbf{b}$, where $\Omega(\mathbf{X})$ is a weight matrix depending only on the regression matrix, can be obtained (see Gnot *et al.*, 1995).

To obtain the operational shrinkage factors for the two cases considered in the preceding section we demand that:

$$PN_\beta(c\mathbf{b}, \mathbf{b}; \mathbf{W}_1) = P[c^{-2} (1+c)^2 V^*_W > \hat{\varphi}] \geq 0.5$$

as well as

$$PN_\beta(c\mathbf{b}, \mathbf{b}; \mathbf{W}_2) = P[c^{-2} (1+c)^2 V^*_I > \hat{\xi}] \geq 0.5$$

where V^*_W and V^*_I , respectively, are $\chi^2(p, \hat{\eta})$ and $\chi^2(p, \hat{\nu})$ distributed random variables. The estimated degrees of freedom $\hat{\eta}$ and $\hat{\nu}$ are $\hat{\eta} = c^2 (1+c)^{-2} \hat{\varphi}$ and $\hat{\nu} = c^2 (1+c)^{-2} \hat{\xi}$.

In either case the minimum value of the operational factor is obtained when the PMN probability is exactly one-half. All values between such a minimum and 1 favour the shrunken estimator in terms of PMN and vice versa. In opting for such an adaptive procedure we are fully aware of the fact that the lower bound of the operational steering factor depends on the choices of estimators for ϕ and ξ.

SUMMARY

In this paper PMN performance comparison of a biased linear homogenous estimator of the regression vector, namely the shrunken estimator, and the OLSE was discussed. As an anonymous referee pointed out the comparison involving weights, with the inverse of the variance-covariance matrix as weight matrix, is the natural and useful one. On the other hand the inclusion of the result of PMN comparison based on unweighted squared norm is interesting theoretically, and it makes the undertaking holistic.

The section dealing with determination of operational adaptive shrinkage factors by way of using plug-in is only suggestive of what one could do under circumstances when practical ad-hoc solutions would be necessary. We recognise that the resultant factors are stochastic, and, as a consequence statisticians may have their differences on whether such a non-deterministic factor is acceptable. Nonetheless, such adaptive steering factors obtained through plug-in could be practically useful. In connection with this issue we would like to address the work of Conerly and Hardin (1991) where deterministic shrinkage factors are suggested. The special issue of *Communications in Statistics* (Volume 20, Number 11) consists of a wide variety of latest research on Pitman Measure of Nearness, and the author would like to mention that this special issue was the motivation for the present undertaking. The present work complements those works related to PMN comparisons of regression estimators.

REFERENCES

1. Conerly, M.D. and Hardin, M.J. (1991). Shrunken estimators via a Pitman nearness criterion. *Communications in Statistics - Theory and Methods* 20(11):3591–3604.

2. Dyer, D. and Keating, J.P. (1983). On the comparison of estimators in the rectangular distribution. *Metron* 41:155–165.

3. Gnot, S., Trenkler, G. and Zmyslony, R. (1995). Non-negative minimum biased quadratic estimation in the linear regression models. *Journal of Multivariate Analysis* 54:113–125.

4. Keating, J.P. (1985). More on Rao's phenomenon. *Sankhya Series* **B** **47**:18–21.

5. Keating, J.P. and Gupta, R.C. (1984). Simultaneous comparison of scale parameters. *Sankhya Series* **B 46**:275–280.

6. Keating, J.P. and Mason, R.L. (1985a). Pitman's measure of closeness. *Sankhya Series* **B 47**:22–32.

7. Keating, J.P. and Mason, R.L. (1985b). Practical relevance of an alternative criterion in estimation. *The American Statistician* **39**:203–205.

8. Keating, J.P., Mason, R.L. and Sen, P.K. (1993). Pitman Measure of Closeness: A Comparison of Statistical Estimators. Society for Industrial and Applied Mathematics (SIAM) (monograph).

9. Mason, R.L., Keating, J.P., Sen, P.K. and Blaylock, N.W. (1990). Comparison of linear estimators using Pitman's Measure of Closeness. *Journal of American Statistical Association* **85**:579–581.

10. Mayer, L.S. and Willke, T.A. (1973). On biased estimation in linear models. *Technometrics* **15**:497–508.

11. Pitman, E.J.G. (1937). The closest estimates of statistical parameters. *Proceedings of the Cambridge Philosophical Society* **33**:212–222.

12. Rao, C.R. (1981). Some comments on the minimum mean square error as a criterion of estimation. In: *Statistics and Related Topics*, pp. 123–143, (Csorgo, M., Dawson, D.A. Rao, J.N.K. and Saleh, M.E., eds). Amsterdam, North-Holland.

13. Rao, C.R., Keating, J.P. and Mason, R.L. (1986). The Pitman Nearness criterion and its determination. *Communications in Statistics* **A15**:3173–3191.