An Improved Method of Prediction the Future: A Result on Improving the Stein-Rule Shrinkage Estimators of Regression Coefficients

Ching-Hui Chang*
Ming Chuan College, Taiwan
Nabendu Pal
University of Southwestern Louisiana, USA
Jyh-Jiuan Lin
China Junior College of Industrial and Commercial Management, Taiwan
Jyh-Horng Lin
Tamkang University, Taiwan

A major use of linear regression models is to predict the future. An improved shrinkage estimator of the regression coefficients leads to a better predictor of the dependent (response) variable in terms of lower prediction mean squared error. Therefore, in this paper we consider the problem of improved estimation of the unknown coefficients of a linear regression model under usual normality assumption. It is well known that the ordinary least squares (OLS) estimator of the regression coefficients can be dominated by the Stein-rule estimators which are again dominated by their "positive-part" versions. In this paper we show that the Stein-rule estimators can be dominated by a new type of estimators which are quite different from the "positive-part" estimators.

Keywords: Quadratic loss, risk function, shrinkage estimation.

AMS 1991 Subject Classification: 62J07
*Direct correspondence to Ching-Hui Chang, Department of Statistics, Ming Chuan College, Taipei, Taiwan.

1.Introduction

The inquisitive human mind has been seeking ways to predict the future since the dawn of the human civilization. This led to the development of Geometry, Astronomy and Mathematics in the early days through which our ancestors learned to study the stars and the solar system. This knowledge again helped the early mankind to forecast weather pattern which in turn helped crop harvest.

In modern science regression is a tool widely used for future prediction. Since prediction is an "educated guess", one should look into the precision associated with this prediction. In a typical multiple linear regression model we collect the data on a dependent (response) variable and several associated independent (predictor) variables. We assume that the predictor variables occur randomly over some population of values and the success of a predictor can be judged by its average performance over such a population. A predictor is evaluated by its mean squared error, called prediction mean squared error (PMSE). It has been observed (see Copas(1983)) that the PMSE of a predictor can be lowered by using a better estimator of the regression coefficients. Hence, in this paper we focus our attention to efficient estimation of regression coefficients under a quadratic loss function.

It is well known that the ordinary least squares (OLS) estimator of the regression coefficients can be dominated by she Stein-rule estimators which are again dominated by their "positive-part" versions. To predict the future with further precision our present article shows that the Stein-rule estimators can be dominated by a new type of estimators which are quite different from the "positive-part" estimators. We have also argued why our estimators are more reasonable than the positive part Stein rule estimators.

It is heartening to see that many researchers have applied shrinkage estimation in real life future prediction problems. For example, Fay and Herriot (1973) used shrinkge estimators for small area income estimation in the United States. Hebel, Faivre, Goffinet and Wallach (1993) used shrinkage estimation techniques for predicting winter wheat yield in France. Landsman and Damodaran (1989) used such a method for predicting stock market income. Also, for another applications see Copas (1983).

2.Brief review

Let us consider the linear regression model

$$\underset{\sim}{Y}=X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon} \tag{1.1}$$

where $\underset{\sim}{Y}$ is a $n\times1$ vector of observations on the dependent variable, X is a fixed matrix of order $n\times p$ (observations on p explanatory variables), $\underset{\sim}{\beta}$ is a $p\times1$ parameter vector and $\underset{\sim}{\varepsilon}$ is the error (or disturbance) vector of order $n\times1$. The error vector $\underset{\sim}{\varepsilon}$ is assumed to follow a multivariate normal distribution with mean $\underset{\sim}{0}$ and variance -covariance matrix $\sigma^2 I_n$. We further assume that $n>p$ and the matrix $(X'X)$ is of rank p. Our goal here is to estimate the regression coefficient vector $\underset{\sim}{\beta}$ efficiently.

The ordianry least squares (OLS) estimator of $\underset{\sim}{\beta}$ is given as

$$\hat{\underset{\sim}{\beta}}{}^0=(X'X)^{-1}X'\underset{\sim}{Y} \tag{1.2}$$

The fitted value of $\underset{\sim}{Y}$ and the observed residual vector $\hat{\underset{\sim}{\varepsilon}}$ are

$$\hat{\underset{\sim}{Y}}=H\underset{\sim}{Y} \text{ and } \hat{\underset{\sim}{\varepsilon}}=(I-H)\underset{\sim}{Y} \text{ where } H=X(X'X)^{-1}X' \tag{1.3}$$

Note that $\hat{\underset{\sim}{\varepsilon}}$ is independent of $\hat{\underset{\sim}{Y}}$. Also

$$\hat{\underset{\sim}{\beta}}{}^0\sim N_p(\underset{\sim}{\beta},\ \sigma^2(X'X)^{-1});\ \hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}\sim\sigma^2\chi^2_{n-p} \tag{1.4}$$

An estimator $\hat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$ is evaluated by its expected loss, called the risk function, given as

$$R(\hat{\underset{\sim}{\beta}},\ \underset{\sim}{\beta})=E\,[(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})] \tag{1.5}$$

One can use a more general loss function $(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})'Q(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})$ where Q is a known $p\times p$ positive definite matrix. But through suitable scale transformation we can reduce the problem to estimation of the regression coefficients under (1.5).

Stein (1956), and James and Stein (1961) suggested a biased estimator of $\underset{\sim}{\beta}$ of the form

$$\underset{\sim}{\hat{\beta}}^{c}=(1-C\frac{\underset{\sim}{\hat{\varepsilon}}'\underset{\sim}{\hat{\varepsilon}}}{\underset{\sim}{\hat{\beta}}^{0\prime}(X'X)\underset{\sim}{\hat{\beta}}^{0}})\underset{\sim}{\hat{\beta}}^{0} \tag{1.6}$$

and showed that $\underset{\sim}{\hat{\beta}}^{c}$ is uniformly better that $\underset{\sim}{\hat{\beta}}^{0}$(i.e., $\underset{\sim}{\hat{\beta}}^{c}$ beats/dominates $\underset{\sim}{\hat{\beta}}^{0}$ uniformly) under (1.5) provided $p\geq 3$ and $0<c\leq 2c_0$, i.e.,

$$R(\underset{\sim}{\hat{\beta}}^{c}, \underset{\sim}{\beta})\leq R(\underset{\sim}{\hat{\beta}}^{0}, \underset{\sim}{\beta})\ \forall c\epsilon(0, 2c_0), \tag{1.7}$$

where $c_0=(d-2)/(n-p+2)$, $d=\sum_{i=1}^{p}\lambda_i^*/\lambda_{\max}^*$ with $\lambda_1^*,\ldots,\lambda_p^*$ being the eigen values of $(X'X)^{-1}$ and $\lambda_{\max}^*=\max(\lambda_1^*,\ldots,\lambda_p^*)$. The estimators in (1.6) are called the Stein-rule estimators. In particular, if we take $c=c_0$ then we get the famous James-Stein estimator $\underset{\sim}{\hat{\beta}}^{JS}=\underset{\sim}{\hat{\beta}}^{c0}$. Also, $c=c_0$ is the optimal value in the range $0<c\leq 2C_0$. The exact risk expression of $\underset{\sim}{\hat{\beta}}^{JS}$ $(\equiv\underset{\sim}{\hat{\beta}}^{c0})$ is given in Judge and Bock (1976). Unfortunately, the Stein-rule estimators are again inadmissible and can be dominated by their "positive-part" versions given as (see Baranchik (1964))

$$\underset{\sim}{\hat{\beta}}^{c+}=(1-c\frac{\underset{\sim}{\hat{\varepsilon}}'\underset{\sim}{\hat{\varepsilon}}}{\underset{\sim}{\hat{\beta}}^{0\prime}(X'X)\underset{\sim}{\hat{\beta}}^{0}})^{+}\underset{\sim}{\hat{\beta}}^{0}, \tag{1.8}$$

where for any real value a, $a^{+}=\max(0, a)$.

Note that the Stein-rule estimators dominate $\underset{\sim}{\hat{\beta}}^{0}$ provided the dimension p is greater than 2 (which is needed to make sure that the risk of $\underset{\sim}{\hat{\beta}}^{c}$ exists) and $tr(X'X)^{-1}$ divided by the largest eigen valus of $(X'X)^{-1}$is larger than 2. In fact, if tr $(X'X)^{-1}/\lambda_{\max}^*\leq 2$, then for no value of $c>0$, does the Stein-rule estimators dominated the OLS estimator. Since the degree of collinearity of the columns of the design matrix X is related to the magnitude of the eigen values of $(X'X)^{-1}$, an ill conditioned $(X'X)$ matrix may affect whether or not $\underset{\sim}{\hat{\beta}}^{c}$ dominates $\underset{\sim}{\hat{\beta}}^{0}$. In the rest of the paper we will assume that $p\geq 3$ and $d=tr(X'X)^{-1}/\lambda_{\max}^*>2$.

Many researchers have suggested various types of shrinkage estimators of $\underset{\sim}{\beta}$ (for example, see Ullah and Ullah (1978) for double k-class estimators and other references), but so far only the "positive -part" estimators are known to dominate the Stein-rule estimators

$\hat{\underset{\sim}{\beta}}^{c}$ uniformly. But the "positive-part" estimators also have some drawbacks,-both theoretical as well as practical. First of all, the "positive-part" estimators are nonanalytic and hence inadmissible. Next note that we can write $\hat{\underset{\sim}{\beta}}^{c+}$ as

$$\hat{\underset{\sim}{\beta}}^{c+} = \begin{cases} \underset{\sim}{0} & \text{if } \underset{\sim}{Y}'H\underset{\sim}{Y} \leq c\underset{\sim}{Y}'(I-H)\underset{\sim}{Y} \\ \hat{\underset{\sim}{\beta}}^{c} & \text{if otherwise,} \end{cases} \tag{1.9}$$

i.e., on the set $\{\underset{\sim}{Y}'H\underset{\sim}{Y} \leq c\underset{\sim}{Y}'(I-H)\underset{\sim}{Y}\}$ we estimate $\underset{\sim}{\beta}$ by $\underset{\sim}{0}$. So, if there is a small measurement error in our observations, then $\hat{\underset{\sim}{\beta}}^{c+}$ can lead us to a null estimate even if our $\underset{\sim}{Y}$ is non-null. The probability $\pi_c = P\,\{\underset{\sim}{Y}'H\underset{\sim}{Y} \leq c\underset{\sim}{Y}'(I-H)\underset{\sim}{Y}\}$ that $\hat{\underset{\sim}{\beta}}^{c+}$ can lead us to a null value can be substantial depending on the values of n, p and c.

In the next section we derive a new type of estimators which are uniformly better than the Stein-rule estimators (including the famous James-Stein estimator).

3.Improved Shrinkage Estimators of $\underset{\sim}{\beta}$.

To derive improved estimators of $\underset{\sim}{\beta}$, we start with the structure

$$\hat{\underset{\sim}{\beta}}^{r} = \left(1-(d-2)\frac{(\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}})\gamma(T)}{\hat{\underset{\sim}{\beta}}^{0\prime}(X'X)\hat{\underset{\sim}{\beta}}^{0}}\right)\hat{\underset{\sim}{\beta}}^{0}, \tag{2.1}$$

where $\gamma(\cdot)$ is a suitable nonnegative function of $T=(\hat{\underset{\sim}{\beta}}^{0\prime}X'X\hat{\underset{\sim}{\beta}}^{0})/(\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}})$. Note that $\gamma(T) \equiv (n-p+2)^{-1}$ gives the James-Stein estimator $\hat{\underset{\sim}{\beta}}^{JS}$. For simplicity we will use the following notations for the rest of the paper.

<u>Notations:</u> Let $M=(X'X)$. Recall that $\lambda_1^*,\ldots,\lambda_p^*$ are the eigen values of M^{-1}. Also, $\lambda_{\max}^*$ and $\lambda_{\min}^*$ are the largest and the smallest eigen values of M^{-1} respectively. Let $\lambda_1,\ldots,\lambda_p$ be the eigen values of M, and $\lambda_{\max} = \max\,(\lambda_1,\ldots,\lambda_p)$; $\lambda_{\min} = \min\,(\lambda_1,\ldots,\lambda_p)$.

There exists a suitable orthogonal matrix P such that $M=P'\Lambda P$ and $M^{-1}=P'\Lambda^{-1}P$ where $\Lambda = \mathrm{diag}\,\{\lambda_1,\ldots,\lambda_p\}$ and $\Lambda^{-1}=\mathrm{diag}\,\{\lambda_1^*,\ldots,\lambda_p^*\}$, i.e., $\lambda_i^*=1/\lambda_i$, $i=1,2,\ldots,p$. Define $\underset{\sim}{W}=(W_1,\ldots,W_p)'=P\hat{\underset{\sim}{\beta}}^{0}$. Then $\underset{\sim}{W} \sim N_p(\underset{\sim}{\eta}, \sigma^2\Lambda^{-1})$ where $\underset{\sim}{\eta}=P\underset{\sim}{\beta}$. Also, let $S=\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}$. Then $T=(\underset{\sim}{W}'\Lambda$

$\underset{\sim}{W}/S)$.

Using the above notations, the risk of $\hat{\underset{\sim}{\beta}}^{r}$(in(2.1)) is

$$R(\hat{\underset{\sim}{\beta}}^{r}, \underset{\sim}{\beta})=E\,[\,\|\,\hat{\underset{\sim}{\beta}}^{0}-(d-2)\frac{S_r(T)}{\underset{\sim}{W}'\Lambda\underset{\sim}{W}}\hat{\underset{\sim}{\beta}}^{0}-\underset{\sim}{\beta}\,\|^{2}]$$

$$=E\,[\,\|\,\underset{\sim}{W}-(d-2)\frac{S_r(T)}{\underset{\sim}{W}'\Lambda\underset{\sim}{W}}\underset{\sim}{W}-\underset{\sim}{\eta}\,\|^{2}]$$

$$=E\,[\,\|\,\underset{\sim}{W}-\underset{\sim}{\eta}\,\|^{2}+(d-2)^{2}\frac{S^{2}\gamma^{2}(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^{2}}\,\|\,\underset{\sim}{W}\,\|^{2}$$

$$-2(d-2)\frac{S\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}\underset{\sim}{W}'(\underset{\sim}{W}-\underset{\sim}{\eta})] \quad (2.2)$$

We now simplify the last term of the expression (2.2) by using Stein's

(1981) normal identity. Note that for a random variable $Z\sim N(\mu,\sigma^{2})$, $E(g(Z)(Z-\mu))=\sigma^{2}E(g'(Z))$ for any real valued function $g(\cdot)$ provided (a) the expectations exist; and (b) $g(Z)\phi(Z)\rightarrow 0$ as $Z\rightarrow\pm\infty$ ($\phi(\cdot)$ is the N(0,1) *pdf*). So,

$$E\,[\frac{S\gamma(T)}{\underset{\sim}{W}'\Lambda\underset{\sim}{W}}\underset{\sim}{W}'(\underset{\sim}{W}-\underset{\sim}{\eta})]=\sum_{i=1}^{p}E\,[\frac{S\gamma(T)}{\underset{\sim}{W}'\Lambda\underset{\sim}{W}}W_i(W_i-\eta_i)]$$

$$=\sum_{i=1}^{p}\sigma^{2}\lambda_i^{-1}E\,[\frac{\partial}{\partial W_i}\{\frac{S\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}W_i\}\,]$$

$$=\sum_{i=1}^{p}\sigma^{2}\lambda_i^{-1}E\,[\frac{S\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}+W_i\frac{\partial}{\partial W_i}\{\frac{S\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}\}\,]$$

$$=\sum_{i=1}^{p}\sigma^{2}\lambda_i^{-1}E\,[S\frac{\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}+SW_i\frac{\partial}{\partial W_i}\{\frac{\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}\}\,]$$

$$=\sum_{i=1}^{p}\sigma^{2}\lambda_i^{-1}E\,[S\frac{\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}+SW_i\{\frac{1}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}\frac{\partial}{\partial W_i}\gamma(T)+\gamma(T)\frac{\partial}{\partial W_i}(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^{-1}\}\,]$$

$$=\sum_{i=1}^{p}\sigma^{2}\lambda_i^{-1}E\,[S\frac{\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}+SW_i\{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^{-1}\gamma'(T)\cdot\frac{\partial T}{\partial W_i}-\gamma(T)\frac{2\lambda_i W_i}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^{2}}\}\,]$$

$$=\sigma^{2}E\,[\frac{S\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}(\mathrm{tr}\Lambda^{-1})+2\gamma'(T)\frac{\|\,\underset{\sim}{W}\,\|^{2}}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}-2\gamma(T)\frac{1}{T}\frac{\|\,\underset{\sim}{W}\,\|^{2}}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}] \quad (2.3)$$

Combining (2.2) and (2.3) we get

$$R(\underset{\sim}{\hat{\beta}}^{r}, \underset{\sim}{\beta})=A_1+A_2(\gamma), \tag{2.4}$$

where

$A_1=E\,[\,\|\underset{\sim}{W}-\underset{\sim}{\eta}\|^2]\;=\sigma^2 tr\Lambda^{-1}$; and

$$\begin{aligned}
A_2(\gamma)=&E\,[(d-2)^2\frac{S^2\gamma^2(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^2}\|\underset{\sim}{W}\|^2-2\sigma^2(d-2)\frac{S\gamma(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}(tr\Lambda^{-1})\\
&+4\sigma^2(d-2)\frac{\gamma(T)}{T}\frac{\|\underset{\sim}{W}\|^2}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}-4\sigma^2(d-2)\gamma'(T)\frac{\|\underset{\sim}{W}\|^2}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}]\\
=&E\,[(d-2)^2\frac{S^2\gamma^2(T)}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^2}\|\underset{\sim}{W}\|^2-2\sigma^2(d-2)\frac{\gamma(T)}{T}(tr\Lambda^{-1})]\\
&+E\,[4\sigma^2(d-2)\frac{\gamma(T)}{T}\frac{\|\underset{\sim}{W}\|^2}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}]\;-E\,[4\sigma^2(d-2)\gamma'(T)\frac{\|\underset{\sim}{W}\|^2}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}]\\
=&A_{21}(\gamma)+A_{22}(\gamma)+A_{23}(\gamma),\ (\text{say});
\end{aligned} \tag{2.5}$$

where

$$A_{23}(\gamma)=-4\sigma^2(d-2)E\,[\gamma'(T)\,a\frac{\|\underset{\sim}{W}\|^2}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}]\;;$$

$$A_{22}(\gamma)=4\sigma^2(d-2)E\,[\frac{\gamma(T)}{T}\;\{\frac{\|\underset{\sim}{W}\|^2}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}\}\;]\;;\text{and}$$

$$A_{21}(\gamma)=\sigma^2 E\,[d-2)^2\frac{\gamma^2(T)}{\sigma^2T^2}\|\underset{\sim}{W}\|^2-2(d-2)\frac{\gamma(T)}{T}tr\Lambda^{-1}]\;.$$

We now simplify $A_{21}(\gamma)$ further as

$$\begin{aligned}
A_{21}(\gamma)&=\sigma^2 E\,[\frac{\sigma^2}{\|\underset{\sim}{W}\|^2}\{\frac{(d-2)\gamma(T)\|\underset{\sim}{W}\|^2}{\sigma^2T}-tr\Lambda^{-1}\}^2]\;-\sigma^4E(\frac{tr\Lambda^{-1}}{\|\underset{\sim}{W}\|})^2\\
&=A_{211}(\gamma)+A_{212},\ (\text{say});
\end{aligned} \tag{2.6}$$

where

$A_{212}=-\sigma^4E(\frac{tr\Lambda^{-1}}{\|\underset{\sim}{W}\|})^2$; and

$$A_{211}(\gamma)=\sigma^2E\,[\frac{\sigma^2}{\|\underset{\sim}{W}\|^2}\{\frac{(d-2)\gamma(T)\|\underset{\sim}{W}\|^2}{\sigma^2T}-tr\Lambda^{-1}\}^2] \tag{2.7}$$

From (2.4)$-$(2.7) we have

$$\begin{aligned}R(\hat{\underset{\sim}{\beta}}^{r}, \underset{\sim}{\beta}) &= A_1 + A_2(\gamma)\\ &= A_1 + (A_{21}(\gamma) + A_{22}(\gamma) + A_{23}(\gamma))\\ &= A_1 + \{(A_{211}(\gamma) + A_{212}) + A_{22}(\gamma) + A_{23}(\gamma)\} \quad (2.8)\end{aligned}$$

Obviously the terms A_1 and A_{212} do not depend on $\gamma(T)$. We now deal with the term $A_{211}(\gamma)$ only. We can write

$$\sigma^{-2}A_{211}(\gamma)$$

$$= E\left[E\left\{\frac{\sigma^2}{\|\underset{\sim}{W}\|^2}\left((d-2)\frac{\gamma(T)\|\underset{\sim}{W}\|^2}{\sigma^2 T} - \mathrm{tr}\Lambda^{-1}\right)^2 \mid T\right\}\right]$$

$$= E\left[E\left\{\frac{\sigma^2}{\|\underset{\sim}{W}\|^2}(d-2)^2\frac{S^2\gamma^2(T)\|\underset{\sim}{W}\|^4}{\sigma^4(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^2} \mid T\right\} - 2(d-2)\mathrm{tr}\Lambda^{-1}E\left\{\frac{\sigma^2}{\|\underset{\sim}{W}\|^2\cdot(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}\frac{S\gamma(T)\|\underset{\sim}{W}\|^2}{\sigma^2} \mid T\right\} + E\left\{\frac{\sigma^2}{\|\underset{\sim}{W}\|^2}(\mathrm{tr}\Lambda^{-1})^2 \mid T\right\}\right]$$

$$= E\left[\triangle(\gamma(T))\right], \text{ (say)} \tag{2.9}$$

The expression $\triangle(\gamma(T))$ in (2.9) can be rewritten as

$$\triangle(\gamma(T)) = \gamma^2(T)(d-2)^2E\left\{\frac{S^2\|\underset{\sim}{W}\|^2}{\sigma^2(\underset{\sim}{W}'\Lambda\underset{\sim}{W})^2} \mid T\right\} - 2(d-2)\mathrm{tr}\Lambda^{-1}\frac{\gamma(T)}{T}$$

$$+ (\mathrm{tr}\Lambda^{-1})^2E\left[\frac{\sigma^2}{\|\underset{\sim}{W}\|^2} \mid T\right] \tag{2.10}$$

which is quadratic in $\gamma(T)$. For fixed T, the optimal value of $\gamma(T)$ which minimizes $\triangle(\gamma(T))$ is

$$\begin{aligned}\gamma_{\mathrm{opt}}(T) &= \left(\frac{\mathrm{tr}\Lambda^{-1}}{d-2}\right)T^{-1}\left\{E\left(\frac{\|\underset{\sim}{W}\|^2}{\sigma^2} \mid T\right)\right\}^{-1}\\ &= \left(\frac{\mathrm{tr}\Lambda^{-1}}{d-2}\right)\left\{E\left(\frac{S\|\underset{\sim}{W}\|^2}{\sigma^2\underset{\sim}{W}'\Lambda\underset{\sim}{W}} \mid T\right)\right\}^{-1}\\ &\leq \frac{\mathrm{tr}\Lambda^{-1}}{(d-2)\lambda^*_{\min}}\left\{E\frac{S}{\sigma^2} \mid T)\right\}^{-1} \quad (2.11)\end{aligned}$$

The following lemma gives us an upper bound of (2.11).

Lemma 2.1 Let S and T are definde as above (See the notations at the beginning of this section). Then $\{E(S/\sigma^2 \mid T)\}^{-1} \leq (1+T)/n$.

Proof. Note that $V = S/\sigma^2 \sim X^2_{n-p}$ and V is independent of $\underset{\sim}{W} \sim N_p(\underset{\sim}{\eta}, \sigma^2\Lambda^{-1})$. Define $U = (\underset{\sim}{W}'\Lambda\underset{\sim}{W}/\sigma^2)$. Then $U \sim x^2_p(\lambda)$ where $\lambda = (\underset{\sim}{\eta}'$

$\Lambda \underset{\sim}{\eta}/\sigma^2) = \underset{\sim}{\beta}'(X'X)\underset{\sim}{\beta}/\sigma^2$. Let $h_V(\cdot)$ and $h_V(\cdot \mid \lambda)$ denote the *pdfs* of V and U respectively. If we make the transformation (V, U) → (V, T) where T=U/V, then the joint *pdf* $g(v, t)$ of (V, T) is $g(v, t) = v h_U(v) h_U(vt \mid \lambda)$. It is easy to see that

$$E(V \mid T=t) = \frac{\int_0^\infty v^2 h_V(v) h_U(vt \mid \lambda) dv}{\int_0^\infty v h_V(v) h_U(vt \mid \lambda) dv} = E\lambda(B), \text{ (say)}$$

where $B \sim f(b \mid \lambda) \propto b\, h_V(b) h_U(bt \mid \lambda)$. Notice that $f(b \mid \lambda > 0)/f(b \mid \lambda = 0) = h_U(bt \mid \lambda > 0)/h_U(bt \mid \lambda = 0)$ is an increasing function in b, i.e., $f(b \mid \lambda > 0)$ is stochastically larger than $f(b \mid \lambda = 0)$. Therefore, $E_{\lambda \geq 0}(B) \geq E_{\lambda=0}(B) = \eta(1+t)^{-1}$. Hence $(E(V \mid T))^{-1} \leq (1+T)/n$.

Using the above lemma in (2.11) we get

$$\gamma_{opt}(T) \leq \left(\frac{1+T}{n}\right)\left(\frac{tr\Lambda^{-1}}{(d-2)\lambda^*_{\min}}\right) \tag{2.12}$$

The upper bound of $\gamma_{opt}(T)$ in (2.12) is free from the unknown parameters. Now, given any nonnegative function $\gamma(T)$, define a new function $\gamma^*(T)$ as

$$\gamma^*(T) = \min\left\{\gamma(T), \left(\frac{1+T}{n}\right)\left(\frac{tr\Lambda^{-1}}{(d-2)\lambda^*_{\min}}\right)\right\} \tag{2.13}$$

Obviously $\gamma^*(T) \leq \gamma(T)\ \forall T \geq 0$. Also, for fixed T, the expression $\triangle(\gamma(T))$ in (2.10) is quadratic in $\gamma(T)$ with unique minimum at $\gamma(T) = \gamma_{opt}(T)$. Using the convexity of $\triangle(\gamma(T))$ it is now trivial that

$$\triangle(\gamma^*(T)) \leq \triangle(\gamma(T)) \tag{2.14}$$

We are now ready to state the main result of this paper.

Theorem 2.1. Assume p>2 and d>2. Any shrinkage estimator $\hat{\underset{\sim}{\beta}}^r$ in (2.1) is uniformly dominated by $\hat{\underset{\sim}{\beta}}^{r*}$ provided $E\{(\|\underset{\sim}{W}\|^2/\underset{\sim}{W}'\Lambda\underset{\sim}{W})(\gamma'(T) - \gamma^{*\prime}(T)\} \leq 0$ where γ^* is defined in (2.13).

Proof. It is enough to show that the risk difference (RD) $= R(\hat{\underset{\sim}{\beta}}^{r*}, \underset{\sim}{\beta} - R(\hat{\underset{\sim}{\beta}}^r, \underset{\sim}{\beta}) \leq 0\ \forall \underset{\sim}{\beta}, \sigma^2 > 0$. Using (2.8) it is clear that

$RD = \{A_{211}(\gamma^*) - A_{211}(\gamma)\} + \{A_{22}(\gamma^*) - A_{22}(\gamma)\} + \{A_{23}(\gamma^*) - A_{23}(\gamma)\}$.

From (2.9) and (2.14) it is obvious that the first term of RD (A_{211}

$(\gamma^*)-A_{211}(\gamma))\leq 0$. Also, the second term is

$$(A_{22}(\gamma^*)-A_{22}(\gamma))=4\sigma^2(d-2)E\left[\frac{(\gamma^*(T)-\gamma(T))}{T}\cdot\frac{\|\underset{\sim}{W}\|^2}{(\underset{\sim}{W}'\Lambda\underset{\sim}{W})}\right]$$

and this is ≤ 0 since $\gamma^*(T)\leq\gamma(T)$. Finally, $RD\leq 0$ provided the third term of $RD\leq 0$.

Corollary 2.1. For any constant $c\epsilon(0, 2c_0)$ (where $c_0=(d-2)/(n-p+2)$) define $\gamma_c^*(T)=\min\{c/(d-2), (\mathrm{tr}\Lambda^{-1}/(d-2)\lambda^*_{\min})(1+T)/n)\}$. Then $\hat{\underset{\sim}{\beta}}^{r^*_c}=(1-(d-2)(\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}})\gamma_c^*(T)/(\hat{\underset{\sim}{\beta}}^{0\prime}(X'X)\hat{\underset{\sim}{\beta}}^0))\hat{\underset{\sim}{\beta}}^0$ is uniformly better that the Stein-rule estimator $\hat{\underset{\sim}{\beta}}^c$.

Proof. Take $\gamma(T)\equiv\gamma_c(T)=c/(d-2)$. Define $\gamma_c^*(T)$ as given in the corollary. Observe that $\gamma_c'(T)\equiv 0$ and $\gamma_c^{*\prime}(T)\geq 0$. Therefore, the result holds easily.

Remark 2.1. For the James-Stein estimator $\hat{\underset{\sim}{\beta}}^{JS}(\equiv\hat{\underset{\sim}{\beta}}^{c0})$ we write $\hat{\underset{\sim}{\beta}}^{JS}=\hat{\underset{\sim}{\beta}}^{r0}$where $\gamma_0(T)=(n-p+2)^{-1}$. Then by defining$\gamma_0^*(T)=\min\{(n-p+2)^{-1}, \mathrm{tr}\Lambda^{-1}((d-2)\lambda^*_{\min})^{-1}(1+T)/n\}$, we have $\hat{\underset{\sim}{\beta}}^{r^*_0}$ which is uniformly better than $\hat{\underset{\sim}{\beta}}^{JS}$.

A question which arises quite naturally here is-"Is it possible to improve upon the estimator $\hat{\underset{\sim}{\beta}}^{r^*_0}$ further?" The answer is "yes". A "positive-part" version $\hat{\underset{\sim}{\beta}}^{r^*_0+}$ is even better than $\hat{\underset{\sim}{\beta}}^{r^*_0}$. Note that the usual "positive-part" James-Stein estimator $\hat{\underset{\sim}{\beta}}^{JS+}$ takes the null value with probability $\pi_{c0}=P\{\underset{\sim}{Y}'H\underset{\sim}{Y}\leq c_0\underset{\sim}{Y}'(I-H)\underset{\sim}{Y}\}$. On the other hand, the estimator $\hat{\underset{\sim}{\beta}}^{r^*_0+}$ takes the null value with probability $\pi^*_{c0}=P\{\underset{\sim}{Y}'H\underset{\sim}{Y}\leq c_0\phi^*\underset{\sim}{Y}'(I-H)\underset{\sim}{Y}\}$, where $\phi^*=\min\{1, (n-p+2)\mathrm{tr}\Lambda^{-1}((d-2)\lambda^*_{\min})^{-1}(1+T)/n\}\leq 1$. As a result $\pi^*_{c_0}\leq\pi_{c_0}$, i.e., even though both $\hat{\underset{\sim}{\beta}}^{JS+}$ and $\hat{\underset{\sim}{\beta}}^{r^*_0+}$ are uniformly better than the popular James-Stein estimator, the latter is less likely to estimate $\underset{\sim}{\beta}$ by a null value when $\underset{\sim}{Y}$ is not so (i.e., $\hat{\underset{\sim}{\beta}}^0$ is not so).

Remark 2.2. The above theorem is a generalization of Kubokawa et al's (1993) result where the simpler problem of estimating a multivariate normal mean vector was considered. In the simpler case of a multivariate normal mean estimation Kubokawa et al. (1993) reduced the mean estimation problem to a variance estimation problem. Using improved variance estimation techniques Pal, Lin and Chang (1997) extended Kubokawa et al's (1993) results which we strongly feel can be adopted in a regression setup. This is currently under investigation.

Reference

Baranchik, A. J.. 1964. "Multiple regression and the estimation of the mean of a multivariate normal distribution." *Technical Report 51*. Department of Statistics, Stanford University.

Copas, J. B.. 1983. "Regression, prediction and shrinkage." *Journal of Royal Statistical Society,* Series B 45:311-54.

Fay, R. E. and Herriot, R. A.. 1979. "Estimates of income for small places: An application of James-Stein procedures to census data." *Journal of American Statistical Association* 74:269-77.

Hebel, P. et al.. 1993, "Shrinkage estimators applied to prediction of French winter wheat yield." *Biometrics* 49:281-93.

James, W. and Stein, C.. 1961. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics & Probability.* University of California Press. Pp. 361-79.

Judge, G. G. and Bock, M. E.. 1976. "A comparison of traditional and Stein-rule estimators under weighted squared error loss." *International Economic Review* 17:234-40.

Kubokawa, T. et al.. 1993. "Estimation of the variance and its applications." *Journal of Statistical Planning and Inference* 35:319-33.

Landsman, W. R. and Damodaran, A.. 1989. "Using shrinkage estimators to improve upon time series model proxies for security market's expectation of earnings." *Journal of Accounting research* 27:97-115.

Pal, N. et al.. 1997. Applications of improved variance estimators in a multivariate normal mean vector estimation. (Submitted).

Stein, C.. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third*

Berkeley Symposium on Mathematical Statistics & Probability. University of California Press. Pp.197-206.

Stein, C.. 1981. "Estimation of the mean of a multivariate normal distribution." Annals of Statistics 9:1135-151.

Ullah, A. and Ullan, S.. 1978. "Double K-class estimators of coefficients in linear regression." *Econometrica,* Vol. 46,3:705-22.