Generalized shrunken type-GM estimator and its application

C Z Ma\(^1\) and Y L Du\(^1\)

\(^{1}\)Institute of Science, Information Engineering University, Zhengzhou 450001, China

E-mail: marcz@163.com

Abstract. The parameter estimation problem in linear model is considered when multicollinearity and outliers exist simultaneously. A class of new robust biased estimator, Generalized Shrunken Type-GM Estimation, with their calculated methods are established by combination of GM estimator and biased estimator include Ridge estimate, Principal components estimate and Liu estimate and so on. A numerical example shows that the most attractive advantage of these new estimators is that they can not only overcome the multicollinearity of coefficient matrix and outliers but also have the ability to control the influence of leverage points.

1. Introduction

It is well known that the classical least squares (LS) estimation have many desirable statistical properties, particularly when error are normally distributed the variance of LS estimations are the least in all sort of unbiased estimation class. But LS estimator frequently are influenced by two class of problems that is outliers and multicollinearity, which in turn lead to highly unstable LS estimates. To circumvent influence of multicollinearity, many biased estimators such as stain estimator, order ridge estimator, principal components estimator root root estimator and Liu estimator etc. have been developed [1-5], which have successfully applied to many fields. However, these biased estimators are not robust, this is, they are sensitive to observations contaminated by outlier. So robust estimation has been investigated [6-9] such as M estimation, which can overcome the impact of observation outliers. When multicollinearity and outliers coexisting parameter estimation is more difficult, some biased robust estimation are proposed [10-18], which is robust against outliers an well, multicollinearity. However, parameter estimation when high leverage point [19, 20], multicollinearity and outliers are simultaneous, has been nearly ignored. Few works have been reported [21, 22].

The purpose of this paper is to propose a class of new estimators which are robust against multicollinearity and , as well, outliers and high leverage point. After a brief review over the generalized M (GM) estimation in Section 2, we define a class of estimators, generalized shrunken type GM estimators, by rafting the GM estimation technique into biased estimation in Section 3. A number numerical example is presented in section4 to illustrate that these new estimators are better than the biased estimators and the robust estimators when multicollinearity, outliers and high leverage point exist simultaneously.

2. Generalized M estimation

Consider the linear regression model of the form

\[
\begin{align*}
y &= X\beta + e \\
E(e) &= 0, \text{Cov}(e) &= \sigma^2 I
\end{align*}
\]
where \( y=(y_1, y_2, \cdots, y_n)^T \) is an \( n \times 1 \) vector of response observations, \( X=(x_1, x_2, \cdots, x_n)^T \) is an \( n \times p \) matrix of the levels of the regression variables, \( x_i \) is the \( i \)-th row of \( X \). \( \beta \) is \( p \times 1 \) vector of unknown parameters, \( e=(e_1, e_2, \cdots, e_n)^T \) is an \( n \times 1 \) vector of random errors, and \( I \) is \( n \times n \) identity matrix. The errors are assumed to have \( \text{E}(e_i)=0 \) and \( \text{D}(e_i)=\sigma^2 \), and assumed to be uncorrelated. The classic least squares (LS) estimates for the model coefficient is given by

\[
\hat{\beta}_{LS} = (X^T X)^{-1} X^T y
\]

Note that GM estimators are the solution of the equations

\[
\sum_{i=1}^{n} \psi(y_i - x_i^T \beta)x_i = 0
\]  
(3)

where \( \psi(\bullet) \) is a bounded real-valued function defined on \( \mathbb{R}^p \times \mathbb{R}^p \). When the weighted functions are \( w_i = \psi(r_i/s) \), \( w_i = \psi(r_i) \), \( \sigma \) is estimated value of \( \sigma \), therefore equations (3) is described by

\[
X^T W r = 0
\]

(4)

where \( r = y - X \hat{\beta} \) is residual vector and \( W = \text{diag}(w_1, w_2, \cdots, w_n) \) is weight matrix. So equations (4) can be written in weighted LS form as

\[
\hat{\beta}_{GM} = (X^T W X)^{-1} X^T W y
\]

(5)

or

\[
X^T W y - X^T W X \beta = 0
\]  
(6)

Statistical functional of GM estimation is given by the solution of the equations

\[
\int \psi(y - X^T \beta) X dF(X, y) = 0
\]

(7)

where \( F \) is joint distribution function of \( (X, y) \). And influence function of GM estimation is

\[
\text{IF}(\beta_{GM}, F) = \psi(y - X^T \beta_{GM}) B_{GM}^{-1} X
\]

(8)

where \( B_{GM} = \int \psi'(X, y - X^T \beta_{GM}) XX^T dF(X, y) \cdot B_{GM} \) is generally a positive definite matrix. By letting

\[
\psi(X, y, y - x_i^T \beta) = \gamma(x_i, \eta(y_i, y - x_i^T \beta / \gamma(x_i)))
\]

(9)

where \( \gamma(x_i, \eta(x_i)) = (1 - h_i) / \sqrt{h_i} \). \( h_i \) is the diagonal element of the matrix \( H = X(X^T X)^{-1} X^T \) and \( \eta(\bullet) \) is a bounded function, then the influence function of GM estimation is function of bounded, when high leverage points are present.

Formally, the expression of the robust estimator (5) is identical to the M estimator, the only difference is that the weight matrix in the M estimation is replaced by the equivalent weight matrix in the GM estimator. It is this substitution that enables the robust estimation capable of resisting the influence of high leverage point, because we can choose a suitable weight matrix that has exactly this ability.

But Gui (2001) derived the asymptotic estimator of variance-covariance matrix for the robust estimator \( \hat{\beta}_{GM} \), which is given by

\[
\sum_{h_{na}} = \sigma^2 \cdot (X^T WX)^{-1}
\]

(10)
\[
\hat{\sigma}^2 = \frac{1}{n-p} \left( \frac{1}{n} \sum_{i=1}^{n} w_i \psi^2 (r_i) \right) .
\] (11)

It is easily seen from (10) that the asymptotic variance-covariance matrix of robust estimator is of the same form as the variance–covariance matrix of the LS estimator. In particular, when multicollinearity among the columns of the matrix \(X\) is present, the matrix \(X^T W X\) maybe ill-conditioned, implying relatively large asymptotic variances. Hence, we expect the GM estimator to perform asymptotically, similar to the LS estimator in the presence of multicollinearity. So it becomes necessary to modify the robust estimator in the line of biased estimator.

3. Generalized shrunken type GM estimation

In response to the above perceived deficiency with the robust estimator, we propose a new estimator by grafting the biased estimation technique (Massy 1965; Hoerl and Kennard 1970a,1970b; Liu 1993; Gruber 1998) philosophy into the robust estimator, which may be designated the generalized shrunken type GM (GSGM) estimator and be denoted by \(\hat{\beta}_{GSGM}\).

\[
\hat{\beta}_{GSGM} = \hat{\beta}_{GSGM} (S) = \widetilde{Q} S \tilde{\beta}_{GSGM} = \widetilde{Q} S \tilde{\lambda}^T X^T \tilde{\lambda} \tilde{y}
\] (12)

where \(\widetilde{Q}\) is an orthogonal matrix satisfying

\[
\widetilde{Q}^T X^T \tilde{W} X \widetilde{Q} = \tilde{A} = \text{diag} (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p) : \tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_p > 0.
\]

That is the spectral decomposition of matrix \(X^T W X\), and

\[
S = \text{diag} (s_1, \ldots, s_p) \quad 0 \leq s_i \leq 1, i = 1,2, \cdots, p
\]
is called a shrinking parameter matrix.

The most common solution approach is iteratively reweighted least squares (IRLS), we assume that the estimator \(\hat{\beta}_{GSGM}^{(m)}\), the residual \(r_{(m)}^{(m)} = y - X \hat{\beta}_{GSGM}^{(m)}\) and spectral decomposition of matrix \(\widetilde{Q}^{(m)} X^T \tilde{W}^{(m)} X \widetilde{Q}^{(m)} = \tilde{A}^{(m)}\) to the m-th step estimation are known, then we can give the (m+1)-th step estimator

\[
\hat{\beta}_{GSGM}^{(m+1)} = \left( \widetilde{Q}^{(m)} S^{(m)} \tilde{A}^{(m)}^{-1} \right)^T \tilde{Q}^{(m)} X^T \tilde{W}^{(m+1)} y
\] (13)

where

\[
\tilde{y}_{i}^{(m+1)} = \frac{\psi (r_i)}{\tilde{\sigma}_{MAD}^{(m)} (x_{(i)})} \tilde{r}_{i}^{(m)}
\] (14)

\[
\tilde{r}_{i}^{(m)} = y_i - X_{(i)}^T \tilde{\beta}_{GSGM}^{(m)}
\] (15)

\[
\tilde{\gamma}^{(m)} (x_{(i)}) = \frac{1 - \tilde{\gamma}_{i}^{(m)}}{\sqrt{\tilde{\gamma}_{i}^{(m)}}}
\] (16)

\[
\tilde{H}^{(m)} = (\tilde{h}_{(m)})_{n \times n} = X (X^T \tilde{W}^{(m)} X)^{-1} X^T \tilde{W}^{(m)}
\] (17)
Obviously, the GM estimator $\hat{\beta}_{GM}$ refer to the case where $S=I$. Further, according to concrete problems in practical application, we can generate many useful estimators by appropriate of the shrinking parameter matrix $S$. Several important estimators are given next.

### 3.1 Ridge type GM estimator

When $S = \tilde{\Lambda}(\tilde{\Lambda} + kl)^{-1}$, we can get

$$
\hat{\beta}_{ORG} (k) = (X^T \tilde{W}X + kl)^{-1} X^T \tilde{W}y
= (X^T \tilde{W}X + kl)^{-1} X^T \tilde{W}x \hat{\beta}_{GM}
$$

which is called a ridge type GM (ORGM) estimator, where $k \geq 0$ is called a ridge parameter.

### 3.2 Principal components type GM estimator

When $S = \tilde{\Lambda}^{-1} \tilde{\Lambda}^{-1}$, we get

$$
\hat{\beta}_{PCGM} (r) = \tilde{Q}_r \tilde{\Lambda}^{-1} \tilde{Q}_r^T X^T \tilde{W}y = \tilde{Q}_r \tilde{\Lambda}^{-1} \tilde{Q}_r^T \hat{\beta}_{GM}
$$

which is called a principal-component-type GM (PCGM) estimator, where

$$
\tilde{Q}_r = \tilde{Q} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, 1 \leq r \leq p
$$

### 3.3 Liu type GM estimator

When $S = (\tilde{\Lambda} + dI)^{-1} (\tilde{\Lambda} + dI)^{-1}$, we get

$$
\hat{\beta}_{LGM}(d) = \tilde{Q}_r \tilde{\Lambda}^{-1} \tilde{Q}_r^T X^T \tilde{W}y = (X^T \tilde{W}X + dI)^{-1} (X^T \tilde{W}X + dI) \hat{\beta}_{GM}
$$

which is called a Liu type GM (LGM) estimator, where $0 < d < 1$.

It is also an important problem in application that selection ways of partial parameters above estimators, according to ways of selected partial parameters in biased estimators, we propose two ways to each estimator.

Ridge parameter $k$ of ORGM estimator is obtained by

$$
\hat{k} = \frac{p \hat{\sigma}_{MAD}^2}{\hat{\beta}_{ORG}^T X^T X \hat{\beta}_{ORG}} = \frac{p \hat{\sigma}_{MAD}^2}{\hat{\beta}_{ORG}^T \hat{\beta}_{ORG}}
$$

Principal component number $r$ of PCGM estimator is obtained by

$$
\tilde{\lambda}_r >> \tilde{\lambda}_{r+1} > 0.001 \quad \text{or} \quad \sum_{i=1}^r \tilde{\lambda}_i / \sum_{i=1}^p \tilde{\lambda}_i \geq 75\% \text{ or } 80\%
$$

Partial parameter $d$ of LGM estimator is obtained by
\[
\hat{d} = \frac{\sum_{i=1}^{p} \hat{\sigma}_{i}^2 - \hat{\sigma}_{MAD}^2}{\sum_{i=1}^{p} \hat{\lambda}_{i} (\hat{\lambda}_{i} + 1)^2} \quad \text{or} \quad \hat{d} = 1 - \frac{\sum_{i=1}^{p} \hat{\sigma}_{MAD}^2}{\sum_{i=1}^{p} \hat{\lambda}_{i} (\hat{\lambda}_{i} + 1)^2}
\]

where \( \hat{\alpha}_{\text{out}} = (\hat{\alpha}_{1}, \ldots, \hat{\alpha}_{p})^{T} = \tilde{Q}^T \hat{\mu}_{\text{out}} \) and \( \tilde{Q}^{T} \hat{x}^{T} \tilde{W} \tilde{Q} = \tilde{\Lambda} = \text{diag}(\tilde{\lambda}_{1}, \ldots, \tilde{\lambda}_{p}) \).

### 4. Numerical example

In order to illustrate the performance of the proposed GSGM estimators, we consider a set of data that reflect the information take from Myers (1990) as shown in table 1. They are 17 hospitals human resource data of U.S. Navy at various sites around the world. There are five regressors and one response variable. The aim here is to produce an empirical equation that will estimate manpower needs for Naval hospitals. Where \( y \) is average monthly workload; \( x_{1} \) is the average number of patients is received a daily; \( x_{2} \) is average number of the x-ray is used; \( x_{3} \) is bed occupancy days per month; \( x_{4} \) is the eligible population in local district /1000; \( x_{5} \) is the average length of the patients stay in hospital.

The correlation matrix and condition number (77593) indicate that there is a serious linear dependency among the regressors, Therefore we should use ridge estimators to reduce the effect of the multicollinearity on the parameter estimates.

Further analysis of the data indicates that the cases 6,10,14,16 and 17 are identified as outlier, and the cases 10,15,16 and 17 have very high leverage. These analyses give that the dataset has simultaneously multicollinearity and outliers in both direction.

| Site | X1  | X2  | X3  | X4  | X5  | y     |
|------|-----|-----|-----|-----|-----|-------|
| 1    | 15.57 | 2463 | 472.92 | 18.0 | 4.45 | 566.52 |
| 2    | 44.02 | 2048 | 1339.75 | 9.5  | 6.92 | 696.82 |
| 3    | 0.42  | 3940 | 620.25  | 12.8 | 4.28 | 1033.15|
| 4    | 18.74 | 6305 | 568.33  | 36.7 | 3.90 | 1603.62|
| 5    | 49.20 | 5723 | 1497.6  | 35.7 | 5.50 | 1611.37|
| 6    | 5.48  | 11520| 1365.83 | 24.0 | 4.60 | 1613.27|
| 7    | 59.28 | 5579 | 1687.00 | 43.3 | 5.62 | 1854.17|
| 8    | 94.39 | 5969 | 1639.92 | 46.7 | 5.15 | 2160.55|
| 9    | 128.02| 8461 | 2872.33 | 78.7 | 6.18 | 2305.58|
| 10   | 128.02| 20106| 3655.08 | 180.5| 6.15 | 3503.93|
| 11   | 96.00 | 133131| 2912.00 | 60.9 | 5.88 | 3571.89|
| 12   | 13.42 | 10771| 3921.00 | 103.7| 4.88 | 3741.40|
| 13   | 127.21| 15543| 3865.67 | 157.7| 7.00 | 10343.81|
| 14   | 252.90| 36194| 7684.10 | 157.7| 7.00 | 10343.81|
| 15   | 409.20| 34703| 12446.33| 169.4| 10.78| 11932.17|
| 16   | 463.70| 39204| 14098.40| 331.4| 7.05 | 15414.94|
| 17   | 510.22| 86533| 15524.00| 371.6| 6.35 | 18854.45|
Since the real value of parameters are unknown, we will measure the accuracy of estimations by forecast period mean squared error (MSEP) (Chen 1987). First we select any one as forecast objectives except for the cases 6,10,14,15,16 and 17, then to compute estimators of parameter, at last, to calculate the MSEP. According to this criterion, we consider the LS estimators , ridge estimators, M estimators, ridge type M (ORM) estimators, Principal components type M (PCM) estimators, Liu type M (LM) estimators, ORGM estimators, PCGM estimators and LGM estimators of parameter, then calculate their MSEP separately in Table 2.

### Table 2. The MSEP of various estimators.

| estimation | LS | R | M | ORM | PCM | LM | ORGM | PCGM | LGM |
|------------|----|---|---|-----|-----|----|------|------|-----|
| MSEP       | 356.2 | 277.8 | 271.3 | 173.7 | 265.0 | 80.2 | 99.6 | 74.4 | 131.3 |

It is obvious to see that MSEP of GSGM estimation is most little in all kinds of estimators in Table 2 it illuminates that GSGM estimators can not only overcome the multicollinearity of coefficient matrix and outliers but also have the ability to control the influence of leverage points.

5. Acknowledgments

Natural Science Foundation of China, NO.41174005, 40974009; Foundation and Advanced Technology Research Program of Henan Province, NO.082300410240.

### References

[1] Gruber M H J 1998 Improving Efficiency by Shrinkage: The James Stein and Ridge Regression Estimators (NewYork: Marcel Dekker)

[2] Hoerl A E and Kennard R W 1970a Ridge Regression: Biased Estimation for Non-orthogonal Problems Technometrics 12 55

[3] Hoerl A E and Kennard R W 1970b Ridge Regression: Applications to Non-orthogonal Problems Technometrics 12 69

[4] Massy W F 1965 Principal components regression in exploratory statistical research JASA 60 234

[5] Liu K A 1993 New class of biased estimate in linear regression CSTM 22(2) 393

[6] Huber P J 1977 Robust Covariances Statistical Decision Theory and Related Topics II ed Gupta and Moore (New York: Academic Press) pp 165

[7] Huber P J 1981 Robust Statistics (New York: Wiley)

[8] Hampel F R 1973 Ridge Estimation: A Condensed Partial Survey ZWYG 27 87

[9] Maronna R A 1976 Robust M-estimators of Multivariate Location and Scatter Annals of Statistics 4 51

[10] Holland P W 1973 Weighted Ridge Regression: Combining Ridge and Robust Regression Methods NBER Working Paper Series Working Paper #11

[11] Pariente S and Welsch R E 1977 Ridge and Robust Regression Using Parametric Linear Programming Working Paper MIT Alfred P Sloan School of Management

[12] Askin R G and Montgomery D C 1980 Augmented Robust Estimators Technometrics 22 333

[13] Pfaffenberger R C and Dielman T E 1985 A Comparison of Robust Ridge Estimators BESPA 631

[14] Silivapulle M J 1991 Robust Ridge Regression Based on an M-Estimator ARJS 33(3) 319

[15] Gui Q M, Duan Q T, Zhou Q Y and Guo J F 2001 A New Class of Biased Linear Estimators in Deficient-rank Linear Models CJM 16(1) 71

[16] Xia J L 1988 A Kind of New Biased Estimator of Regression Parameters: Root Estimator CJMSAP 3 21

[17] Nyquist H 1985 Ridge Type M-estimators LSITCWK (Berlin: Springer) p 246

[18] Olcay A and Nedret B 2000 Robust Liu estimator for regression based on an M-estimator JAS 27 1
[19] Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A 1986 *Robust Statistics: The Approach Based on Influence Function* (New York: John Wiley)

[20] Rousseeuw P J and Zomeren B C 1990 Unmasking Multivariate Outliers and Leverage Points *JASA* **85** 633

[21] Olcay A and Nedret B 1996 Roust Ridge Regression Estimation Biased on the GM-estimators *JM&CS* **22**(3) 1

[22] Simpson J R and Montgomery D C 1996 A biased-robust regression technique for the combined outlier-multicollinearity problem *JSCS* **56**(1) 1

[23] Myers R H 1990 *Classical and Modern Regression with Applications* second edition (Boston: Duxbury Press)

[24] Chen X R and Wang S G 1987 *Advanced Regression Analysis* (Hefei: Anhui Education Publishing House)