International Journal of Contemporary Mathematical Sciences
Vol. 14, 2019, no. 3, 131 – 140
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ijcms.2019.9713

A Nonlinear Goal Programming Model for Determining the Bias Parameters of Liu-Type Estimator with Application to Construction Sector in Egypt

Rasha A. Farghali

Department of Mathematics, Insurance and Applied Statistics
Faculty of Commerce and Business Administration
Helwan University, Cairo, Egypt

This article is distributed under the Creative Commons by-nc-nd Attribution License.
Copyright © 2019 Hikari Ltd.

Abstract

Biased estimators such as Liu-type estimator, represents flexible way to deal with the multicollinearity problem while preserving all the explanatory variables. In this article a nonlinear goal programming model is introduced to obtain the optimal values of the bias parameters of Liu-type estimator for linear regression. The data of Construction Sector in Egypt is used to evaluate the performance of the proposed model using the estimated mean squared error and condition indices criterions, the results show that the proposed model outperforms other existing methods.

Keywords: Biased Estimator, Bias Parameters, Goal Programming, Liu-type Estimator, Mathematical programming, Multicollinearity,

Mathematics Subject Classification 62J07

1. Introduction

Suppose we have a random sample of size n, that contains the measurements of p explanatory variables , then we consider the following linear regression model:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \qquad i = 1,2,\dots,n \qquad (1)$$

Here: x_{ij} : (predetermined) the measurement of the j^{th} explanatory variable on the i^{th} observation, $j = 1,2,\dots,p$, $i = 1,2,\dots,n$.β_0, β_j , $j = 1,2,\dots,p$: (unknown) are the true parameters of model (1). ε_i: the i^{th} random error , $i = 1,2,\dots,n$ and $\varepsilon_i \sim N(0, \sigma^2 I)$. The ordinary least squares (OLS) aims to find $\hat{\beta}_{OLS}$ that minimize the sum of squared residuals. $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$, $Var(\hat{\beta}_{OLS}) = \hat{\sigma}^2 (X'X)^{-1}$, and $\hat{\sigma}^2$, is the mean of squared residuals, $\hat{\sigma}^2 = \frac{(Y-X\hat{\beta}_{OLS})'(Y-X\hat{\beta}_{OLS})}{n-p}$. $\hat{\beta}_{OLS}$, are the best estimators of true parameters, but its performance are deteriorating with the appearance of multicollinearity. Multicollinearity refers to a near-dependency between the explanatory variables $x_1, x_2, \dots, x_p$. If the design matrix $(X'X)$ suffers from ill-conditioning, then $\hat{\beta}_{OLS}$, are still unbiased estimators but with larger standard errors, larger mean squared errors, distended confidence intervals and important variables become insignificant in testing hypotheses. The most popular method used for diagnosing multicollinearity are: examining the pairwise correlations between the explanatory variables. The Variance Inflation Factors (VIFs) and Condition Index (CI) (Belsley et al, 1980).Condition index (CI) is defined as:

$$CI = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \qquad (2)$$

where, λ_{max} and λ_{min} are the maximum and the minimum eigenvalues of the matrix $(X'X)$, It is well known that, if $CI \leq 10$, it means that there is no multicollinearity problem, if $10 < CI < 30$, it means that a moderate multicollinearity problem exists and it might be corrected, finally, if $CI \geq 30$, then it refers to a severe multicollinearity problem and it must be corrected (Belsley,1991), (Gujarati, 2000). Correction methods are classified according the ability to abandon some variables to :I) subset selection, and II) biased estimators such as ridge regression, it adds a bias parameter, k , to the main diagonal of the design matrix $(X'X)$, that led to smaller regression estimators with smaller mean squared errors but CI of the matrix $(X'X + k\, I)$ is still relatively large (Liu, 1991)

Liu (2003, 2004) discussed this problem and suggested another bias parameter d in a two parameters estimator known as, Liu-type. The bias parameters k, d have two different functions; $k, k > 0$, aims to minimize the condition index of the matrix $(X'X + kI)$ to reach the desired level, and d , $-\infty < d < \infty$, aims to minimize the mean squared error $MSE(\hat{\beta}_{LT})$. Liu-type estimator for linear regression defined as:

$$\hat{\beta}_{LT} = (X'X + k\, I)^{-1}(X'X - d\, I)\hat{\beta}_{OLS} \qquad (3)$$

Several methods were introduced for calculating the bias parameters k, d separately but with limited attention to the ill-conditioning of the design matrix.

Mathematical programming models were used to determine the optimal values of bias parameters in ridge regression, generalized ridge regression, and Liu-type estimators by (El-Dash et al, 2011), (Hamed et al, 2012), (El-Hefnawy & Farag, 2014), (Ebiad et al, 2017) and (Farghali, 2019) they showed that the estimating regression parameters using mathematical programming exceed the estimates using other classical methods such as Hoerl and Kennard (1970 a,b).

In this article, a nonlinear goal programming model is introduced for calculating the bias parameters k, d simultaneously of Liu-type estimator in linear regression. The proposed model aims to determine the optimal values of k, d , that achieve two goals: 1) minimize the sum of the residuals to be zero , 2) minimize the $MSE\left(\hat{\beta}_{LT}\right)$ to be less than $MSE\left(\hat{\beta}_{OLS}\right)$. This article is organized as follows: In section 2, the methodology is introduced. The proposed model is introduced in section 3. The application of Construction Sector of Egypt is demonstrated in section 4. Finally, conclusions are presented in section 5.

2. Methodology

Consider the following singular-variable-decomposition, that there exists an orthogonal matrix D of the degree $(p \times p)$ whose columns are the corresponding eigenvectors of $\lambda_1, \lambda_2, \dots, \lambda_p$ such that:

$$D(X'X)D' = \Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_p) \quad (4)$$

Where, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, are the ordered eigenvalues of the design matrix $(X'X)$ Rewrite model (1) in canonical form.

$$Y = Z\alpha + \varepsilon \quad (5)$$

Where, $Z = XD$ and $\alpha = D'\beta$.The Liu-type estimator is:

$$\hat{\alpha}_{LT} = (\Lambda + k\,I)^{-1}\,(\Lambda - d\,I)\hat{\alpha}_{OLS} \quad (6)$$

And the corresponding estimates: $\hat{\beta}_{OLS} = D\,\hat{\alpha}_{OLS}$, $\hat{\beta}_{LT} = D\hat{\alpha}_{LT}$.
Liu (2003, 2004) suggested to choose k,that the CI of the matrix $(X'X + k\,I)$ is reduced to 10. After k is choosen, d , that minimizes $MSE(\hat{\alpha}_{LT})$, is determined as follows:

$$d = \frac{\sum_{j=1}^{p}((\sigma^2 - k\alpha_j^2)/\left(\lambda_j + k\right)^2)}{\sum_{j=1}^{p}((\lambda_j\,\alpha_j^2 + \sigma^2)/\lambda_j\left(\lambda_j + k\right)^2)} \quad (7)$$

Several methods were introduced for the estimation of the bias parameter k see (Kibria and Banik, 2016), (El-Dereny and Rashwan,2011), (Al-Hassan, 2010).
In this article some of these methods will be reviewed to compare them with suggested method using mathematical programming techniques.

1. Hoerl et al (1975) suggested using k as: $k_{HK} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$ (8)

2. Kibria (2003) suggested the following estimators of k using the arithmetic mean, (AM), the geometric mean, (GM), and the median of the values $\hat{\sigma}^2/\hat{\alpha}_j^2$ as follows:

$$k_{AM} = \frac{1}{p}\sum_{j=1}^{p}\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2} \quad (9)$$

$$k_{GM} = \frac{\hat{\sigma}^2}{(\prod_{j=1}^{p}\hat{\alpha}_j^2)^{\frac{1}{p}}} \quad (10)$$

$$k_{MD} = \text{median}\left\{\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}\right\}, \quad j = 1,2,\dots,p \quad (11)$$

3. Khalaf & Shukur (2005) suggested a modification of Hoerl et al (1975) as follows:

$$k_{KS} = \frac{\lambda_{Max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_{Max}\,\hat{\alpha}'\hat{\alpha}} \quad (12)$$

3. The proposed nonlinear goal programming model

In this article, a nonlinear goal programming model is proposed to determine optimal values of the bias parameters (the decision variables) k, d simultaneously that achieve the following: **the first objective** is to obtain the optimal values of k, d which minimize sum residual of squares. This objective can be formulated as follows:

$$Min \quad \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\hat{\beta}_{j(LT)}x_{ij}\right) = 0 \quad (13)$$

The first goal is:

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\hat{\beta}_{j(LT)}x_{ij}\right) + s_1^- - s_1^+ = 0 \quad (14)$$

Thus, the first priority is:

$$Min\ (s_1^- + s_1^+) \quad (15)$$

where, s_1^- & $s_1^+$are the underachievement/overachievement of the first goal.

The second objective is to obtain the optimal values of k, d which minimize $MSE(\hat{\beta}_{LT})$ to be less $MSE(\hat{\beta}_{OLS})$.This objective can be formulated as follows:

$$Min \quad \sigma^2\sum_{j=1}^{p}\frac{(d-\lambda_j)^2}{\lambda_j\,(\lambda_j+k)^2} + \sum_{j=1}^{p}\frac{(d+k)^2\,\alpha_j^2}{(\lambda_j+k)^2} \le \sigma^2\sum_{j=1}^{p}\frac{1}{\lambda_j} \quad (16)$$

The second goal is:

$$\sigma^2 \sum_{j=1}^{p} \frac{(d-\lambda_j)^2}{\lambda_j\,(\lambda_j+k)^2} + \sum_{j=1}^{p} \frac{(d+k)^2\,\alpha_j^2}{(\lambda_j+k)^2} + s_2^- - s_2^+ = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \quad (17)$$

Thus, the second priority is: $Min\ \ s_2^+$ (18)

Where, s_2^- & s_2^+ are the underachievement/overachievement of the second goal. Besides these goal constraints, a rigid constraint exist which is: multicollinearity problem must be corrected i.e. CI of the matrix $(X'X + kI)$ must be less than 10. Thus, the rigid constraint will be:

$$\sqrt{\frac{\lambda_{max}+k}{\lambda_{min}+k}} \quad \leq \quad 10 \qquad (19)$$

The mathematical formulation of the proposed model is:

$$Find \qquad k, d \qquad such\ that \qquad (20)$$

$$Lexicographically\ minimize\ H = \{(s_1^- + s_1^+), s_2^+\} \qquad (21)$$

$s.t.$

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_{0(LT)} - \sum_{j=1}^{p} \hat{\beta}_{j(LT)} x_{ij}\right) + \ s_1^- - s_1^+ \ = 0 \qquad (22)$$

$$\sigma^2 \sum_{j=1}^{p} \frac{(d-\lambda_j)^2}{\lambda_j\,(\lambda_j+k)^2} + \sum_{j=1}^{p} \frac{(d+k)^2\,\alpha_j^2}{(\lambda_j+k)^2} + s_2^- - s_2^+ = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \qquad (23)$$

$$\sqrt{\frac{\lambda_{max}+k}{\lambda_{min}+k}} \quad \leq \quad 10 \qquad (24)$$

$$k > 0\ ,\ d\ \ unrestricted\ ,\ \hat{\beta}_{0(LT)},\ \hat{\beta}_{j(LT)}\ ,\ j = 1,2,\dots,p,\ unrestricted \quad (25)$$

$$s_1^-, s_1^+, s_2^-, s_2^+ \ \geq 0 \qquad (26)$$

Where: $\hat{\sigma}^2, \lambda_j, \alpha_j^2, j = 1,2,\dots,p$: parameters, as defined before. k, d: decision variables, bias parameters. Model (20)-(26) is a nonlinear goal programming model in spite of the linearity of the achievement function but the constraints are nonlinear (Ignizio, 1980). It can be solved by the Iterative method (sequential linear algorithm) with available software like the general algebraic modeling system (GAMS).

4. Construction Sector in Egypt

Construction sector is the foundation of the overall development and the leading sector for the vision of Egypt 2030. It is contributing more and more to the GDP as it averaged 23047.06 EGP Million from 2007 until 2018, in the fourth quarter of 2018 reached 59846.30 EGP Million. To determine the factors affecting the cost of constructing projects in Egypt a multiple linear regression model is set where the dependent variable y (cost of constructing projects) depends on many explanatory variables such as: reinforcing steel x_1 , cement x_2 , bricks x_3 , sand & gravel x_4 , wood x_5 , walls & ready mix concrete x_6 , tiles x_7 , pipes x_8 , bathroom tools x_9 , metal tools x_{10} and electrical tools x_{11} and miscellaneous materials x_{12} . The dataset taken from the Egyptian Annual Bulletin of Construction by public sector of the year 2015/2016. It is clear that these variables are connected as they used together to have a finished constructed project. OLS estimates and its standard errors are as follows:

$$\hat{y}_i^* = \frac{928.252}{(767.659)} - \frac{12.613}{(4.487)} x_{1i}^* + \frac{16.001}{(10.671)} x_{2i}^* + \frac{51.448}{(23.757)} x_{3i}^* + \frac{6.784}{(3.527)} x_{4i}^* - \frac{30.964}{(58.936)} x_{5i}^* + \frac{44.400}{(10.560)} x_{6i}^* - \frac{16.916}{(28.169)} x_{7i}^* + \frac{7.913}{(1.321)} x_{8i}^* + \frac{11.866}{(4.843)} x_{9i}^* + \frac{327.056}{(59.158)} x_{10i}^* - \frac{29.654}{(17.161)} x_{11i}^* + \frac{1.838}{(1.014)} x_{12i}^* \quad ,i = 1,2,..,27 \quad , MSE(\hat{\beta}_{OLS}) = 1773.778$$

The fitted model has three inverse relationship that contradict the theoretical signs between these variables, this refers to the existence of multicollinearity problem, so bivariate correlations was calculated. Since that, the explanatory variables are highly correlated, then we may conclude that a multicollinearity problem might be present, to diagnose it the CI was calculated by equation (2) $C.I. = 43.822$, which means that there is a severe multicollinearity and must be corrected. The biased estimator, Liu-type estimator for linear regression was used. For determining bias parameters methods in equations (8)-(12) were used in comparison with proposed model.

1	0.940	0.844	0.747	0.830	0.894	0.977	0.707	0.830	0.975	0.892	0.636
0.940	1	0.622	0.689	0.728	0.899	0.689	0.933	0.730	0.889	0.616	0.849
0.844	0.622	1	0.854	0.801	0.651	0.958	0.706	0.880	0.878	0.615	0.847
0.747	0.689	0.854	1	0.804	0.893	0.723	0.956	0.837	0.690	0.808	0.746
0.830	0.728	0.801	0.804	1	0.638	0.667	0.652	0.808	0.767	0.695	0.863
0.894	0.899	0.651	0.893	0.638	1	0.833	0.922	0.738	0.889	0.970	0.758
0.977	0.689	0.958	0.723	0.667	0.833	1	0.934	0.633	0.661	0.839	0.765
0.707	0.933	0.706	0.956	0.652	0.922	0.934	1	0.762	0.954	0.886	0.693
0.830	0.730	0.880	0.837	0.808	0.738	0.633	0.762	1	0.357	0.330	0.876
0.975	0.889	0.878	0.690	0.767	0.889	0.661	0.954	0.357	1	0.976	0.726
0.892	0.616	0.615	0.808	0.695	0.970	0.939	0.886	0.330	0.976	1	0.923
0.636	0.849	0.847	0.764	0.863	0.758	0.765	0.693	0.876	0.726	0.923	1

The parameter estimates and its standard errors between brackets are shown in table (4-1).

Table (4-2): estimated MSE and condition indices

$Measures$	OLS	$Liu-Type\ Estimator$					
		k_{HK}	k_{AM}	k_{GM}	k_{MD}	k_{KS}	k_{GP}
$MSE(\beta)$	773.778	36.468	39.430	35.218	28.872	30.872	**19.223**
CI	43.822	78.226	28.849	30.463	22.129	15.020	**8.912**

Table (4-1): estimated regression parameters and its standard errors

$Estimates$	OLS	$Liu-Type\ Estimator$					
		k_{HK}	k_{AM}	k_{GM}	k_{MD}	k_{KS}	k_{GP}
$\hat{\beta}_0$	928.252 (767.659)	1.981 (28.235)	0.645 (26.013)	0.318 (22.004)	0.569 (19.695)	0.776 (25.171)	**0.241 (18.013)**
$\hat{\beta}_1$	-12.613 (4.487)	0.818 (3.651)	0.943 (3.019)	1.115 (2.556)	0.722 (2.438)	0.745 (3.006)	**0.661 (1.998)**
$\hat{\beta}_2$	16.001 (10.671)	0.511 (2.913)	0.863 (3.425)	0.433 (2.761)	0.523 (2.014)	0.560 (3.002)	**0.360 (2.006)**
$\hat{\beta}_3$	51.448 (23.757)	0.430 (15.062)	0.295 (13.370)	0.105 (10.206)	0.321 (12.033)	0.333 (11.803)	**0.187 (8.652)**
$\hat{\beta}_4$	6.784 (3.527)	0.825 (2.036)	0.679 (1.998)	0.142 (1.406)	0.723 (2.001)	0.310 (1.567)	**0.386 (0.908)**
$\hat{\beta}_5$	-30.964 (58.936)	0.421 (22.141)	0.519 (23.480)	0.956 (20.631)	0.491 (21.018)	0.382 (20.493)	**0.267 (18.551)**
$\hat{\beta}_6$	44.400 (10.560)	0.951 (4.052)	3.501 (3.798)	0.804 (3.004)	0.937 (4.231)	0.822 (3.992)	**0.719 (2.059)**
$\hat{\beta}_7$	-16.916 (28.169)	2.792 (13.401)	2.003 (12.887)	1.954 (13.021)	0.819 (13.063)	3.570 (11.691)	**0.769 (9.152)**
$\hat{\beta}_8$	7.913 (1.321)	0.732 (0.963)	0.889 (0.845)	0.930 (0.901)	1.012 (0.799)	0.880 (0.693)	**0.624 (0.602)**
$\hat{\beta}_9$	11.866 (4.843)	0.454 (1.754)	0.396 (2.003)	0. 627 (1.989)	3.701 (1.649)	1.706 (2.701)	**0.323 (1.019)**
$\hat{\beta}_{10}$	327.056 (59.158)	4.112 (21.004)	6.308 (19.725)	3.021 (20.369)	5.213 (18.998)	3.336 (20.912)	**2.542 (16.294)**
$\hat{\beta}_{11}$	-29.654 (17.161)	1.102 (10.429)	2.015 (9.052)	0.943 (8.001)	1.592 (10.069)	2.714 (9.411)	**0.632 (7.405)**
$\hat{\beta}_{12}$	1.838 (1.014)	0.221 (0.206)	0.101 (0.317)	0.099 (0.284)	0.119 (0.339)	1.221 (0.143)	**0.065 (0.115)**

The performance of the proposed model was evaluated by: Mean Squared Errors for estimates and CI as follows:

From tables (4-1) and (4-2), we can observe that, parameter estimates and its standard errors using the proposed model are the lowest. Also, the proposed model performance is much better than other estimators in terms of smaller $MSE\ and\ CI$. The proposed model has some advantages that: It determines the bias parameters of

Liu-type estimator simultaneously and the priorities of its objectives can be changed due to the preferences of decision maker.

5. Conclusions

In this article, a nonlinear goal programming model is introduced for calculating the bias parameters k, d simultaneously of Liu-type estimator in linear regression. The proposed model aims to determine the optimal values of k, d , that achieve two goals: 1) minimize the sum of the residuals to be zero , 2) minimize the $MSE(\hat{\beta}_{LT})$ to be less than $MSE(\hat{\beta}_{OLS})$. The data of construction sector in Egypt set is used to compare between the proposed model and some methods for determining bias. Regression parameters estimates, its standard errors, estimated mean squared error of Regression parameters and condition indices are used as performance criterions. Results show that the proposed method exceeds other methods.

References

[1] Y. AL-Hassan, Performance of new ridge regression estimator, *Journal of the Association of Arab Universities for Basic and Applied Sciences*, **9** (2010), 23–26. https://doi.org/10.1016/j.jaubas.2010.12.006

[2] D. Belsley, E. Kuh, and R. Welsch, *Regression diagnostics: identifying influential data and sources of multicollinearity*, John Wiley & Sons, New York, USA, 1980.

[3] D. Belsley, A guide to using the collinearity diagnostics, *Computer Science in Economics and Management*, **4** (1991), 33-50.

[4] R. Ebiad, R. Farghali and S. Abo-El-Hadid, A mathematical programming approach for Liu-type estimator, *Advances and Applications in Statistics*, **50** (2017), 293-314. https://doi.org/10.17654/as050040293

[5] A. El-Dash, A. El-Hefnawy and R. Farghali, Treating multicollinearity problem using goal programming technique, *The Egyptian Statistical Journal*, **55** (2011), 67-92.

[6] M. El-Dereny and N. Rashwan, Solving multicollinearity problem using ridge regression models, *International Journal of Contemporary Mathematical Sciences*, **6** (2011), 585-600.

[7] A. El-Hefnawy and A. Farag, A combined nonlinear programming model and Kibria method for choosing ridge parameter regression, *Communications in*

Statistics: Simulation and Computations, **43** (2014), 1442-1470. https://doi.org/10.1080/03610918.2012.735317

[8] R. Farghali, New shrinkage parameters for the generalized Liu-type estimator using mathematical programming, *International Journal of Contemporary Mathematical Sciences,* **14** (2019), 91-102. https://doi.org/10.12988/ijcms.2019.9410

[9] D. Gujarati, *Essentials of econometrics*, McGraw-Hill, New York, USA, 2000.

[10] R. Hamed, A. El-Hefnawy and A. Farag, Selection of the ridge parameter using mathematical programming, *Communications in Statistics: Simulation and Computations,* **42** (2012), 1409-1432. https://doi.org/10.1080/03610918.2012.659821

[11] A., Hoerl and W., Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12** (1970a), 55-67. https://doi.org/10.1080/00401706.1970.10488634

[12] A. Hoerl and W. Kennard, Ridge regression: applications to nonorthogonal problems, *Technometrics,* **12** (1970b), 69-82. https://doi.org/10.1080/00401706.1970.10488635

[13] J. Ignizio, *Linear programming in single and multiple-objective systems*, Prentice-Hall, INC., Englewood Cliffs, USA, 1980.

[14] G. Khalaf and G. Shukur, Choosing ridge parameters for regression problems, *Communications in Statistics: Theory and Methods,* **34** (2005), 1177-1182. https://doi.org/10.1081/sta-200056836

[15] G. Kibria, Performance of some new ridge regression estimators, *Communications in Statistics: Simulation and Computations,* **32** (2003), 419-435. https://doi.org/10.1081/sac-120017499

[16] G. Kibria and S. Banik, Some ridge regression estimators and their performance, *Journal of Modern Applied Statistical Methods,* **15** (2016), 206-238. https://doi.org/10.22237/jmasm/1462075860

[17] K. Liu, A new class of biased estimate in linear regression, *Communications in Statistics-Theory and Methods,* **22** (1993), 393-402. https://doi.org/10.1080/03610929308831027

[18] K. Liu, Using Liu-type estimator to combat collinearity, *Communications in Statistics-Theory and Methods,* **32** (2003), 1009-2003. https://doi.org/10.1081/sta-120019959

[19] K., Liu, More on Liu-type estimator in linear regression, *Communications in Statistics-Theory and Methods,* **33** (2004), 2723-2733. https://doi.org/10.1081/sta-200037930

Received: June 25, 2019; Published: July 20, 2019