Generalized modified least squares in the univariate linear model[*]

Xu-Qing Liu[†]

Faculty of Mathematics and Physics, Huaiyin Institute of Technology, Huai'an 223003, P.R. China

Abstract

The method of ordinary least squares is a classical procedure extensively used to build regression models in the literature. It is fulfilled by minimizing the ordinary residual sum of squares, the sum of squares of the differences between the true values and the fitted values of the response variable. In 1984, Li (The American Mathematical Monthly 91(2): 135–137) put forward the method of modified least squares, which is fulfilled by minimizing the modified residual sum of squares, the sum of squares of the perpendicular distances from the points to the fitted line.

In this short paper, we mainly aim to appeal to readers to pay close attention to the criterion of modified least squares and develop it to generalized modified least squares. The closed-form of the generalized modified least squares estimators for the intercept and the slope are derived. The results are illustrated by a numerical example. The illustration shows that generalized modified least squares is an adjusting criterion such that the resulting fitted line can be sensitive or insensitive to those outlying data values.

Keywords: Ordinary least squares; Modified least squares; Generalized modified least squares; Univariate linear model

CLC Number: O212.1 / *MSC (2000) Number:* 93E24; 62J12

1 Introduction

It is well known that the method of ***ordinary least squares*** has been widely used to make inference in the literature. Consider the univariate linear model $\mathscr{E}(Y|X) = \beta_0 + \beta_1 X$, where β_0 and β_1 are the intercept and the slope, respectively. Let $(x_1, y_1), \cdots, (x_n, y_n)$ be the observations of (X, Y) on the basis of an experiment. The method of ordinary least squares is utilized to build estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, for β_0 and β_1 such that the ordinary residual sum of squares

$$Q_{\text{ols}} \triangleq \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \tag{1.1}$$

is minimized. Denote

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2, \quad s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

By virtue of the analytical approach, the ordinary least squares estimators (OLSEs) can be expressed as:

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}}, \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{1.2}$$

The first subfigure of Figure 1, by means of a simulated data, summarizes ordinary least squares and indicates how to minimize Q_{ols}, which is not the sum of squares of the perpendicular distances from the points to the fitted line but the sum of squares of the differences between the true values and the fitted values (i.e., the longitudinal coordinates of the points of intersection of the fitted line and the plumb lines crossing the data points) of the response variable.

In 1984, Li [1] considered a new criterion, which is fulfilled by minimizing the modified residual sum of squares:

$$Q_{\text{mls}} \triangleq \frac{Q_{\text{ols}}}{1 + b_1^2} = \sum_{i=1}^{n} \frac{(y_i - b_0 - b_1 x_i)^2}{1 + b_1^2}, \tag{1.3}$$

[*] This research was supported by Grant HGC0923 and the "Green & Blue Project" Program for 2008 to Cultivate Young Core Instructors.

[†] Corresponding author. *Email address:* liuxuqing688@gmail.com (X.-Q. Liu).

which is the sum of squares of the perpendicular distances from the points to the fitted line. We call the criterion to be ***modified least squares***. The second subfigure of Figure 1 summarizes the principle of modified least squares. Li [1] offered the modified least squares estimators (MLSEs) for β_0 and β_1 as below:

$$\beta_1^* = \frac{s_{yy} - s_{xx} + \sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}}, \quad \text{and} \quad \beta_0^* = \overline{y} - \beta_1^* \overline{x}. \tag{1.4}$$

It is clear that both the empirical regression lines on the basis of OLSE and MLSE pass through the point $(\overline{x}, \overline{y})$. In other words, the fitted line based on MLSE can be derived by rotating the fitted line based on OLSE a particular angle with the point $(\overline{x}, \overline{y})$ as the center.

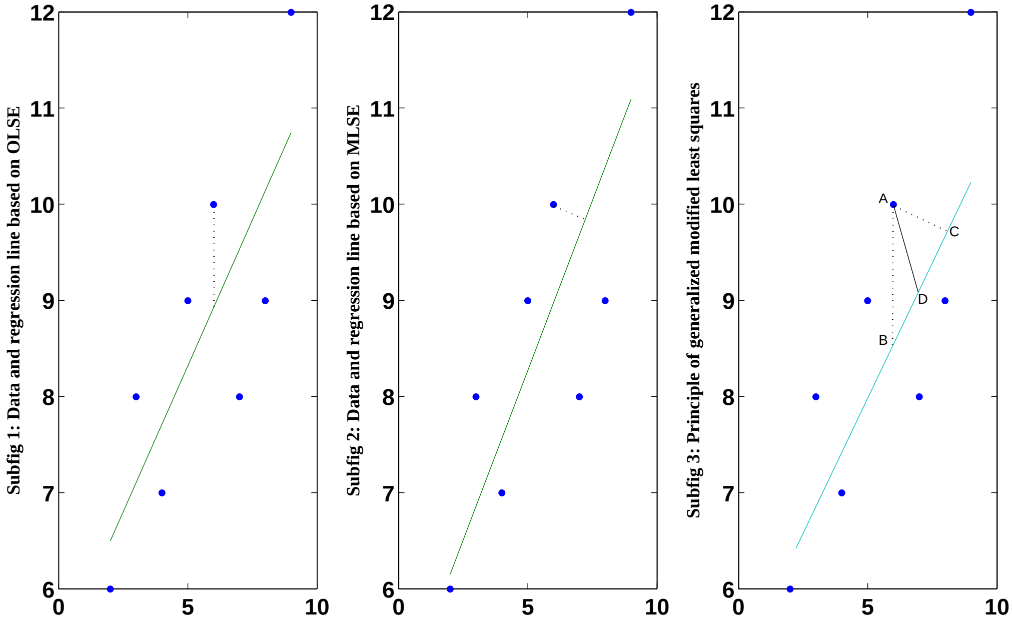

Figure 1: Plots of data points and regression lines based on OLSE and MLSE

Modified least squares was then studied by Shuchat [2] for multivariate case through linear algebra techniques. The method possesses some good aspects, which will be stated as remarks in conjunction with a numerical example in the next section. However, it is regretful that modified least squares has not been given due attention so far. One of the reasons we think is that the status of ordinary least squares in the field of statistics is deep-rooted, and the other is that the calculation of modified least squares are more complicated than that of ordinary least squares. In the paper, we appeal to readers to pay close attention to such criterion.

The rest is as follows: The advantages of the modified residual sum of squares are firstly discussed in Section 2 with the aid of a numerical example. Then we extend modified least squares to generalized modified least squares in Section 3. The representation of the new estimators is given. Finally, we apply the main result to the numerical example to show that the new criterion can adjust the fitted line such it is sensitive or insensitive to those outlying data values.

2 Modified least squares

Although MLSE is nonlinear and therefore it needs complicated calculations comparing with OLSE, there are some good aspects for it:

1. As we all know, the criterion of ordinary least squares unduly emphasizes on those large departures for a practical problem. To overcome such defect, we need a new criterion which takes the large departures into consideration

but also "weakens" the determinant degree of them to some extent. As a robust procedure, least absolute deviation regression has been widely used in recent years, due to its insensitivity to those large departures. However, linear programming techniques are required and therefore some inconveniences may occur in the process. Considering $Q_{\text{mls}} < Q_{\text{ols}}$, we believe that modified least squares can improve ordinary least squares in this sense, though the "shrinkage" from Q_{ols} to Q_{mls} is well-proportioned. The reason for this is that β_1^*, as an estimator of b_1, is not a constant before the data values are obtained, and thereafter the shrinkage factor is actually stochastic. The numerical example given below illustrates such superiority of modified least squares.

2. Modified least squares can also be applied to nonlinear models, which are denoted by $\mathscr{E}(Y|X) = f(X)$. In this case, we define the objective function as the sum of squares of the distances from each y_i to the tangent of the fitted curve, focused on x_i. That is,

$$Q_{\text{mls}} \triangleq \sum_{i=1}^{n} \frac{[y_i - f(x_i)]^2}{1 + \left([\mathrm{d}f(x)/\mathrm{d}x]\big|_{x=x_i}\right)^2}.$$

For example, we need to estimate the corresponding parameters for

$$f_1(X) = aX^2 + bX + c, \quad f_2(X) = a\exp\{bX + c\} + d, \quad f_3(X) = a\ln(bX + c) + d,$$

and so on. The modeling function can be chosen by drawing the scatter plot.

3. If we assume additionally that the dependent variable, Y, has the variance σ^2 and the observations $y_1, \cdots, y_n$ are from an independent and identically distributed sample, a forthcoming naive problem is how to estimate σ^2 on the basis of modified least squares other than ordinary least squares? Considering that MLSE has no concise algebraic properties, we structure directly an estimator for σ^2, following the form of $\hat{\sigma}^2$, which is based on ordinary least squares, where

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \quad \text{with} \quad \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

Let

$$\sigma^{*2} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - y_i^*)^2, \quad \text{with} \quad y_i^* = \beta_0^* + \beta_1^* x_i.$$

By direct operations, σ^{*2} is also expressible as

$$\sigma^{*2} = \frac{s_{yy} - 2s_{xy}\beta_1^* + s_{xx}\beta_1^{*2}}{n-2}. \tag{2.1}$$

In the following, we will call σ^{*2} to be the MLSE of σ^2. Although we can not explain from the angle of theoretical analysis that the MLSE is better than the OLSE, the simulated numerical example can illustrate the expected result of us, however.

Now, we apply the result of MLSE to a dataset, which is concerned with the relationship between accounting rates on stocks and market returns; cf. [3, Example 2.1, p. 16]. Fifth-four companies were chosen as a sample. Let X be the mean yearly accounting rate for the period 1959 to 1974, and Y be the corresponding mean market rate. The data are given in [3, Table 2.1, p. 16]. See also Table 1, for convenience. We assume in addition that suggesting a linear relationship, $Y \approx \beta_0 + \beta_1 X$, between the two variables is reasonable. With the aid of Matlab 7.0, the values of the OLSEs and MLSEs for β_0 and β_1 are

$$\hat{\beta}_0 = 0.8480, \quad \hat{\beta}_1 = 0.6103, \quad \text{and} \quad \beta_0^* = -9.6415, \quad \beta_1^* = 1.4214,$$

respectively. The empirical regression equations on the basis of ordinary and modified least squares follow immediately as below: $\hat{y}_{\text{ols}} = 0.8480 + 0.6103x$, and $y_{\text{mls}}^* = -9.6415 + 1.4214x$. Figure 2 illustrates the plot of the data and the regression lines based on ordinary and modified least squares. By the figure, we can find that the first line (based on ordinary least squares) is sensitive to the point (32.58, 14.73), which seems to be an outlying data value. Accordingly,

Table 1: **Accounting rates and market rates from 1959 to 1974**

Company	Accounting Rate	Market Rate
McDonnell Douglas	17.96	17.73
NCR	8.11	4.54
Honeywell	12.46	3.96
TRW	14.70	8.12
Raytheon	11.90	6.78
W.R. Grace	9.67	9.69
Ford Motors	13.35	12.37
Textron	16.11	15.88
Lockheed Aircraft	6.78	-1.34
Getty Oil	9.41	18.09
Atlantic Richfield	8.96	17.17
Radio Corporation of America	14.17	6.78
Westinghouse Electric	9.12	4.74
Johnson and Johnson	14.23	23.02
Champion International	10.43	7.68
R.J. Reynolds	19.74	14.32
General Dynamics	6.42	-1.63
Colgate-Palmolive	12.16	16.51
Coca-Cola	23.19	17.53
International Business Machines	19.20	12.69
Allied Chemical	10.76	4.66
Uniroyal	8.49	3.67
Greyhound	17.70	10.49
Cities Service	9.10	10.00
Philip Morris	17.47	21.90
General Motors	18.45	5.86
Philips Petroleum	10.06	10.81
FMC	13.3	5.71
Caterpillar Tractor	17.66	13.38
Georgia Pacific	14.59	13.43
Minnesota Mining & Manufacturing	20.94	10.00
Standard Oil (Ohio)	9.62	16.66
American Brands	16.32	9.40
Aluminum Company of America	8.19	0.24
General Electric	15.74	4.37
General Tire	12.02	3.11
Broaden	11.44	6.63
American Home Products	32.58	14.73
Standard Oil (California)	11.89	6.15
International Paper	10.06	5.96
National Steel	9.60	6.30
Republic Steel	7.41	0.68
Warner Lambert	19.88	12.22
U.S. Steel	6.97	0.90
Bethlehem Steel	7.90	2.35
Armco Steel	9.340	5.03
Texaco	15.40	6.13
Shell Oil	11.95	6.58
Standard Oil (Indiana)	9.560	14.26
Owens Illinois	10.05	2.60
Gulf Oil	12.11	4.97
Tenneco	11.53	6.65
Inland Steel	9.920	4.25
Kraft	12.27	7.30

the second line (based on modified least squares) is not very sensitive to that point. In this sense, we think that the method of ordinary least squares has been improved by that of modified least squares.

Another improvement is with respect to the MLSE of error variance. By Matlab 7.0, the values of the OLSE and the MLSE are given as $\hat{\sigma}^2 = 25.8644$ and $\sigma^{*2} = 41.8321$, respectively. In the following, we make two simulation studies based on $x_1, \cdots, x_{54}$, the data values of the mean yearly accounting rates of the fifth-four companies, and the normal distribution. The first one is to generate stochastically fifty-four "observations" of Y (each observation is derived by

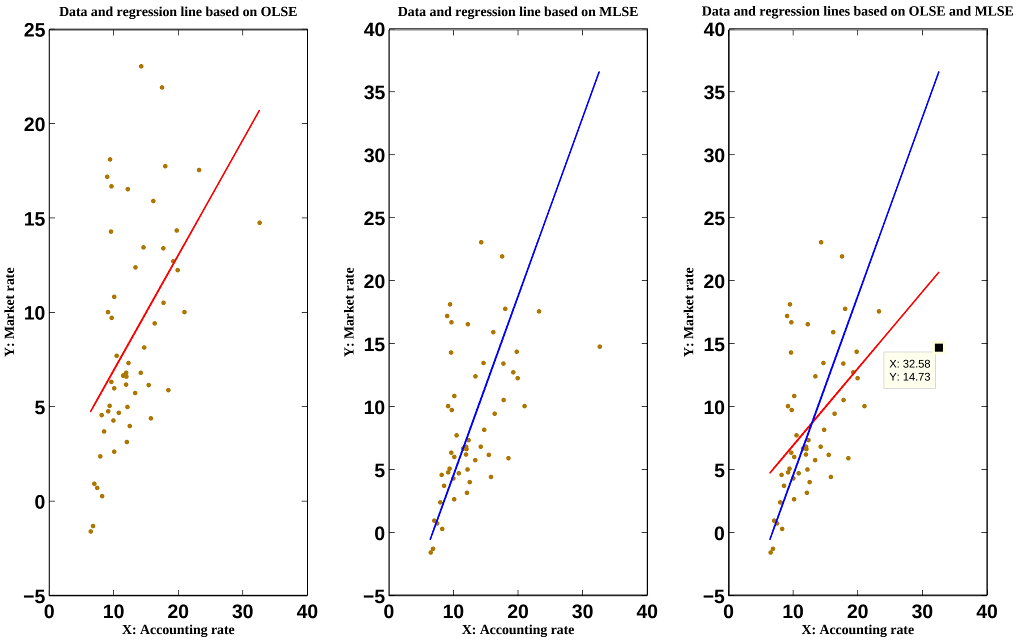

Figure 2: Plots of data points and regression lines based on ordinary and modified least squares

generating three points and averaging them) by means of $x_1, \cdots, x_{54}$, $\hat{\beta}_0, \hat{\beta}_1$. Then calculate the simulated OLSE and the simulated MLSE. Finally, we get the absolute errors of them. Repeat the above procedures fifty times. Replacing $\hat{\beta}_0, \hat{\beta}_1$ with β_0^*, β_1^* in the first simulation study gives the second one. Figure 3 shows the plots of the two simulations. It is seen that the absolute error of MLSE is not larger than that of OLSE.

3 Generalized modified least squares

By the previous section, the fitted line based on MLSE can be derived by rotating the fitted line based on OLSE a particular angle with the point $(\bar{x}, \bar{y})$ as the center. A natural problem is that: if each line located between the two lines (as shown in the third subfigure of Figure 2) can be used as the fitted line or not. The answer is YES, since one can choose a line which is “closer” to the fitted line based on OLSE if inclining to ordinary least squares and choose a line which is “closer” to the fitted line based on MLSE if inclining to modified least squares. The direct consequence is that the estimators of β_1 and β_0 can be written $\lambda\hat{\beta}_1 + (1-\lambda)\beta_1^*$, the convex combination of OLSE and MLSE, for some $\lambda \in [0,1]$ and $\bar{y} - [\lambda\hat{\beta}_1 + (1-\lambda)\beta_1^*]\bar{x} = \lambda\hat{\beta}_0 + (1-\lambda)\beta_0^*$. Adjusting the value of λ gives different results and corresponding fitted lines. Further, the fitted line passes though the point $(\bar{x}, \bar{y})$ inherently. In that way, what is the criterion the resulting estimators follow?

By the third subfigure of Figure 1, we consider minimizing the sum of squares of such AD, which is not larger than AB and also not smaller then AC. We call the criterion to be ***generalized modified least squares***. Denote by

$$Q_{\text{gmls}} \triangleq \frac{Q_{\text{ols}}}{1+b_1^2\tau} = \sum_{i=1}^{n} \frac{(y_i - b_0 - b_1 x_i)^2}{1+b_1^2\tau}$$

the residual sum of squares based on generalized modified least squares, where $\tau \in [0,1]$ is any fixed arbitrary real scalar. Clearly, Q_{gmls} reduces to Q_{ols} if $\tau = 0$ and Q_{mls} if $\tau = 1$. We call $\tilde{\beta}_0$ and $\tilde{\beta}_1$ the ***generalized modified least squares estimators*** (GMLSEs) for β_0 and β_1, if $\tilde{\beta}_0$ and $\tilde{\beta}_1$ minimize Q_{gmls} with respect to b_0 and b_1, for given $\tau \in (0,1)$. We mention here that the GMLSEs yielded by generalized modified least squares may have different version from that of $\lambda\hat{\beta}_1 + (1-\lambda)\beta_1^*$ and $\lambda\hat{\beta}_0 + (1-\lambda)\beta_0^*$. Without loss of generality, we assume that $s_{xy} \neq 0$ and $s_{xx} \neq s_{yy}$. By direct

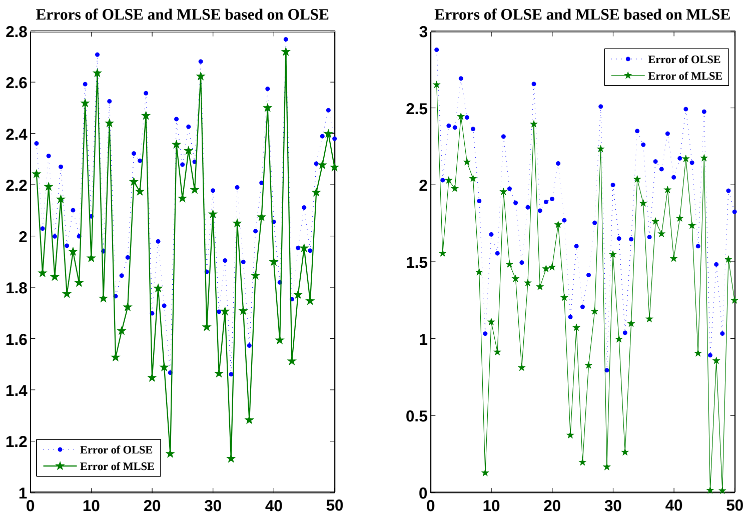

Figure 3: Plots of the absolute errors of OLSE and MLSE

operations (by manual or with the aid of Matlab 7.0), we have

$$\frac{\partial Q_{\text{gmls}}}{\partial b_0}=0 \quad\Leftrightarrow\quad b_0+b_1\bar{x}=\bar{y}, \quad \text{and}$$
$$\frac{\partial Q_{\text{gmls}}}{\partial b_1}=0 \quad\Leftrightarrow\quad \left(s_{xx}+n\bar{x}^2-s_{yy}\tau-n\bar{y}^2\tau-nb_0^2\tau+2nb_0\bar{y}\tau\right)b_1+\left(s_{xy}+n\overline{xy}-nb_0\bar{x}\right)\left(b_1^2\tau-1\right)=0.$$

Further, we obtain

$$s_{xy}\tau b_1^2+\left(s_{xx}-s_{yy}\tau\right)b_1-s_{xy}=0. \tag{3.1}$$

It follows that $b_0=\bar{y}-b_1\bar{x}$, with $b_1=b_1^{(+)}(\tau)$ or $b_1=b_1^{(-)}(\tau)$, where

$$b_1^{(+)}(\tau)\triangleq\frac{s_{yy}\tau-s_{xx}+\sqrt{(s_{yy}\tau-s_{xx})^2+4s_{xy}^2\tau}}{2s_{xy}\tau}, \quad b_1^{(-)}(\tau)\triangleq\frac{s_{yy}\tau-s_{xx}-\sqrt{(s_{yy}\tau-s_{xx})^2+4s_{xy}^2\tau}}{2s_{xy}\tau}.$$

Inserting $b_0=\bar{y}-b_1^{(+)}(\tau)\bar{x},\ b_1=b_1^{(+)}(\tau)$ and $b_0=\bar{y}-b_1^{(-)}(\tau)\bar{x},\ b_1=b_1^{(-)}(\tau)$ into Q_{gmls}, it follows that

$$Q_{\text{gmls}}^{(+)}-Q_{\text{gmls}}^{(-)}=-\frac{1}{\tau}\sqrt{(s_{yy}\tau-s_{xx})^2+4s_{xy}^2\tau}<0,$$

and therefore $\bar{y}-b_1^{(-)}(\tau)\bar{x}$ and $b_1^{(-)}(\tau)$ are definitely not the MLSEs of β_0 and β_1. Denote now

$$\tilde{\beta}_1(\tau)=b_1^{(+)}(\tau), \quad \text{and} \quad \tilde{\beta}_0(\tau)=\bar{y}-b_1^{(+)}(\tau)\bar{x}, \tag{3.2}$$

respectively. We have the following theorem:

Theorem 3.1 *$\tilde{\beta}_0(\tau)$ and $\tilde{\beta}_1(\tau)$ are the GMLSEs for β_0 and β_1, respectively.*

Proof. On the basis of the above analysis, it suffices to justify $a>0$ and $b^2-ac<0$, where

$$a=\left.\frac{\partial^2 Q_{\text{gmls}}}{\partial b_0^2}\right|_{\substack{b_0=\tilde{\beta}_0(\tau)\\ b_1=\tilde{\beta}_1(\tau)}}, \quad b=\left.\frac{\partial^2 Q_{\text{gmls}}}{\partial b_0\partial b_1}\right|_{\substack{b_0=\tilde{\beta}_0(\tau)\\ b_1=\tilde{\beta}_1(\tau)}}, \quad c=\left.\frac{\partial^2 Q_{\text{gmls}}}{\partial b_1^2}\right|_{\substack{b_0=\tilde{\beta}_0(\tau)\\ b_1=\tilde{\beta}_1(\tau)}}.$$

With the aid of Matlab 7.0, it can be concluded that $a = 2n\Big/\left[1+\tau\left(\tilde{\beta}_1(\tau)\right)^2\right] > 0$ and

$$\begin{aligned} b^2 - ac &= \frac{4n}{\left(1+\tau\left(\tilde{\beta}_1(\tau)\right)^2\right)^4}\left[2s_{xy}\tau^2\left(\tilde{\beta}_1(\tau)\right)^3 + 3\tau\left(s_{xx}-s_{yy}\tau\right)\left(\tilde{\beta}_1(\tau)\right)^2 - 6s_{xy}\tau\tilde{\beta}_1(\tau) - \left(s_{xx}-s_{yy}\tau\right)\right] \\ &= \frac{4n}{\left(1+\tau\left(\tilde{\beta}_1(\tau)\right)^2\right)^4}\left\{\left[s_{xy}\tau\left(\tilde{\beta}_1(\tau)\right)^2 + \left(s_{xx}-s_{yy}\tau\right)\tilde{\beta}_1(\tau) - s_{xy}\right]\left(2\tau\tilde{\beta}_1(\tau) + \frac{s_{xx}-s_{yy}\tau}{s_{xy}}\right)\right\} \\ &\quad - \frac{4n}{\left(1+\tau\left(\tilde{\beta}_1(\tau)\right)^2\right)^4}\cdot\frac{\left(s_{yy}\tau - s_{xx}\right)^2 + 4s_{xy}^2\tau}{s_{xy}}\cdot\tilde{\beta}_1(\tau), \end{aligned}$$

which combined with (3.1) yields that

$$b^2 - ac = -\frac{4n}{\left(1+\tau\left(\tilde{\beta}_1(\tau)\right)^2\right)^4}\cdot\frac{\left[\left(s_{yy}\tau - s_{xx}\right)^2 + 4s_{xy}^2\tau\right]\cdot\left(s_{yy}\tau - s_{xx} + \sqrt{(s_{yy}-s_{xx})^2 + 4s_{xy}^2\tau}\right)}{2s_{xy}^2\tau} < 0$$

holds inherently. The proof is thus completed. ▮

We mention two facts, one of which is that $\lim\limits_{\tau\to 0^+}\tilde{\beta}_k(\tau) = \hat{\beta}_k$ holds for $k = 0, 1$, and therefore ordinary and modified least squares are extended in this sense. The other is that Li [1] did not give the strict proof for β_1^* and β_0^* to be the MLSEs of β_1 and β_0. We have offered the supplement here.

Table 2: **GMLSEs for** $\tau = 0, 0.1, 0.2, \cdots, 0.9, 1.0$

τ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\tilde{\beta}_0(\tau)$	0.8480	−0.0517	−1.0649	−2.1746	−3.3475	−4.5400	−5.7083	−6.8181	−7.8482	−8.7894	−9.6415
$\tilde{\beta}_1(\tau)$	0.6103	0.6799	0.7582	0.8441	0.9348	1.0270	1.1173	1.2031	1.2828	1.3556	1.4214

Let us now apply the result of 3.1 to the dataset considered in Section 2. We still assume that suggesting the seemingly linear relationship $Y \approx \beta_0 + \beta_1 X$ is reasonable. With the aid of Matlab 7.0, the values of the GMLSEs for β_0 and β_1 are given in Table 2. Figure 4 illustrates the plot of the data and the regression lines based on generalized modified least squares. By the figure, we can find that generalized modified least squares is indeed an adjusting criterion.

4 Concluding summary

In the short paper, we developed the method of modified least squares. The illustration of Figure 2 shows that modified least squares is not very sensitive to those outlying data values, while the illustration of Figure 4 reflects that generalized modified least squares can adjust the fitted line such it is sensitive or insensitive to those outlying data values. As a adjusting criterion, we think that, generalized modified least squares should be used widely in practical problems by choosing conformable value of τ in the range from 0 to 1.

As we can see, univariate regression models have relatively limited value in some practical applications. However, as a trigger, this paper may lead to more better means that could be sprang out as it should be. Finally, we mention that one potential possible direction of the paper is to generalize the results to multivariate regressions.

References

[1] H.C. Li. A generalized problem of least squares, *The American Mathematical Monthly*, 91(2) (1984), 135–137.

[2] A. Shuchat, Generalized least squares and eigenvalues, *The American Mathematical Monthly*, 92(9) (1985), 656–659.

[3] R.H. Myers. Classical and modern regression with application (2nd Edition), *Higher Education Press*, Beijing, 2005 (*Photocopy version*).

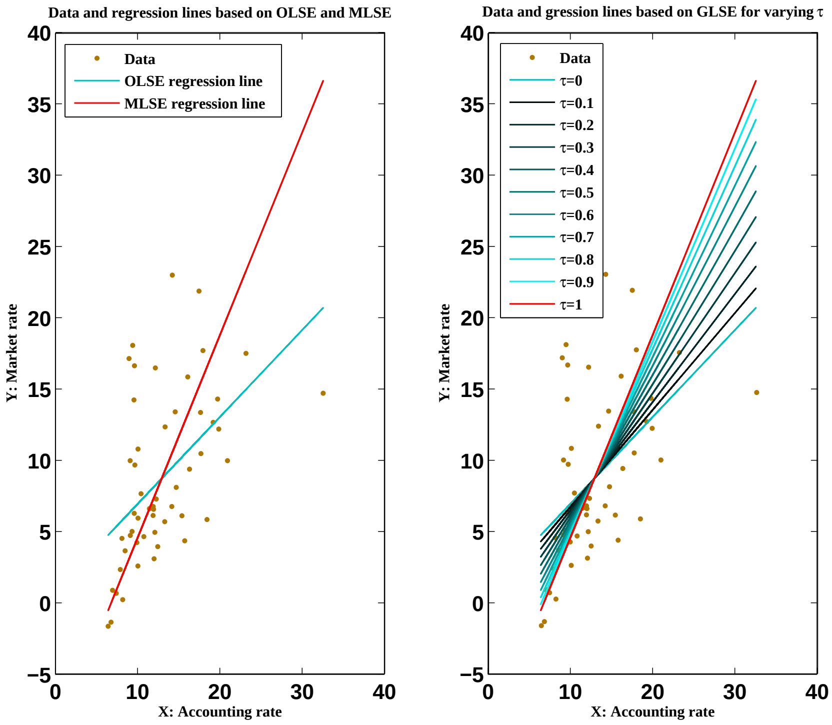

Figure 4: Plots of data points and regression lines based on OLSE, MLSE and GMLSE