An Efficient Shrinkage Estimator for the Parameters of Simple Linear Regression Model

Alaa M. Hamad\textsuperscript{1}, Muna D. Salman\textsuperscript{2}, Aseel H. Ali\textsuperscript{3}, Abbas N. Salman\textsuperscript{4}
\textsuperscript{1,2,4} (Department of Mathematics, College of Education for Pure Science/Ibn-Al-Haitham, University of Baghdad, Iraq)
\textsuperscript{3} (Computer Center - University of Baghdad, Iraq)

Abstract: This paper provided the preliminary test single stage shrinkage estimator for estimating the parameters of simple linear regression model, when a prior estimate of these parameters are available. This prior estimate has been referred in statistical literatures as guess point about the parameters. The expressions for Bias, Mean Squared Error (MSE) and Relative Efficiency of the proposed estimators are obtained. Numerical results are provided when the proposed estimators are estimators of level of significance. Comparisons with the usual estimator (O.L.S.) and existing estimators were made to show the usefulness of the proposed estimators in the sense of Relative Efficiency and Mean Squared Error.

Keywords: Simple Linear Regression Model, Least Square Estimator, Single Stage Shrinkage Estimator, Bias Ratio, Mean Squared Error, Relative Efficiency.

I. INTRODUCTION

In statistics, simple linear regression is the least squares estimator of a linear regression model with a single explanatory variable. In other words, simple linear regression fits a straight line through the set of n points in such a way that makes the sum of squared residuals of the model (that is, vertical distances between the points of the data set and the fitted line) as small as possible.

The adjective simple refers to the fact that this regression is one of the simplest in statistics. The slope of the fitted line is equal to the correlation between y and x corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that it passes through the center of mass (x, y) of the data points.

A regression models the past relationship between variables to predict their future behavior. As an example, imagine that your company wants to understand how past advertising expenditures have related to sales, stock prices, currency exchange rates, and productivity gains resulting from a training program. Some time we may have a prior estimate value (point guess) of the parameter to be estimated. If this value is in the vicinity of the true value, the shrinkage technique is useful to get an improved estimator.

Consider the following simple linear regression model:
\[ y_i = \gamma + \beta(x_i - \bar{X}) + \epsilon_i, \text{ where } \epsilon_i \sim N(0,\sigma^2) \] (1)

The preliminary test single stage shrinkage estimator \[ 1,2,3,4,5,10,12 \] is considered for estimating the parameter \( \theta \) (\( \theta \) may be refer to \( \gamma \) or \( \beta \) of previous model) when a guess point \( \theta_0 \) is available about \( \theta \) due the past experience or similar cases.

From the empirical studies it has been established that the shrinkage estimators perform better than the usual estimator when our guess point be very close to the true value of the parameter. Therefore to make sure whether \( \theta \) is closed to \( \theta_0 \) or not, we may test \( H_0: \theta = \theta_0 \) against \( H_1: \theta \neq \theta_0 \), so we denote by R to the critical region for above test.

Thompson [12] suggested shrinking the usual estimator \( \hat{\theta} \) of \( \theta \) towards the prior guess point \( \theta_0 \) and proposed the estimator \( \hat{\theta} = \psi(\hat{\theta})\hat{\theta} + (1 - \psi(\hat{\theta}))\theta_0 \), where \( (1 - \psi(\hat{\theta})) \) represents the experimenters belief in the guess point \( \theta_0 \). He was found the estimator \( \hat{\theta} \) is more efficient than \( \hat{\theta} \) if the true value \( \theta \) is close to \( \theta_0 \) (\( H_0 \) accepted) but may be less efficient otherwise, therefore to resolve the uncertainty that a guess point value is approximately the true value or not, a preliminary test of significance may be employed. So he take the usual estimator \( \hat{\theta} \) when \( \theta \) is far a way from \( \theta_0 \) (\( H_0 \) rejected) after he made the preliminary test.
Thus, preliminary test single stage shrinkage estimator has the following form
\[
\tilde{\theta} = \begin{cases} 
\psi(\hat{\theta})\hat{\theta} + (1 - \psi(\hat{\theta}))\theta_0 & , \text{if } \hat{\theta} \in R \\
\hat{\theta} & , \text{if } \hat{\theta} \not\in R 
\end{cases}
\] (2)

where R is the preliminary test region for acceptance the null hypothesis H_0 as we mentioned above, \( \hat{\theta} \) is the usual estimator of \( \theta \), \( \psi(\hat{\theta}) \) is a shrinkage weight factor such that \( 0 \leq \psi(\hat{\theta}) \leq 1 \) which may be a function of \( \hat{\theta} \) or may be a constant (ad hoc basis).

Several authors had been studied a preliminary test shrunken estimator which is defined in (2) for special population by choosing different weight factors \( \psi(\hat{\theta}) \). See for example [1], [2], [3], [4], [5], [7], [8], [9] and [11].

The goal of this paper is to modify the preliminary test single stage shrinkage estimator which is defined in (2) for estimate the parameters \( (\theta) \) of the model (1).

Therefore, the proposed preliminary test single stage shrinkage estimator is as below:
\[
\tilde{\theta}_{PT} = \begin{cases} 
\psi_i(\hat{\theta})\hat{\theta} + (1 - \psi_i(\hat{\theta}))\theta_0 & , \text{if } \hat{\theta} \in R \\
\psi_i(\hat{\theta})\hat{\theta} + (1 - \psi_i(\hat{\theta}))\theta_0 & , \text{if } \hat{\theta} \not\in R 
\end{cases}
\] (3)

Where \( \psi_i(\hat{\theta}) \), \( i = 1, 2 \) are shrinkage weight factors such that \( 0 \leq \psi_i(\hat{\theta}) \leq 1 \).

The expressions for Bias, Mean Squared Error and Relative Efficiency of the estimator \( \tilde{\theta}_{PT} \) above are derived. Numerical results of these expressions were made to show the validity and the usefulness of the proposed estimator when it compare with the usual and existing estimators.

II. PRELIMINARY TEST SINGLE STAGE SHRUNKEN ESTIMATOR \( \tilde{\beta} \) AND \( \tilde{\gamma} \)

In this section recall the estimator which is defined in equation (3) for estimate the parameter \( \beta \) and \( \gamma \) of model (1) as below
\[
\tilde{\beta}_{PT} = \begin{cases} 
\psi_i(\hat{\beta})\hat{\beta} + (1 - \psi_i(\hat{\beta}))\beta_0 & , \text{if } \hat{\beta} \in R_i \\
\psi_i(\hat{\beta})\hat{\beta} + (1 - \psi_i(\hat{\beta}))\beta_0 & , \text{if } \hat{\beta} \not\in R_i 
\end{cases}
\] (4)

Where \( \beta_0 \) is a prior guess value of \( \beta \), \( \hat{\beta} \) is a usual estimator (O.L.S.) of \( \beta \) and R_i is a pre test region of acceptance of size \( \alpha \) for testing the hypothesis H_0: \( \beta = \beta_0 \) against the hypothesis H_1: \( \beta \neq \beta_0 \).

\[
R_i = \left[ \beta_0 - t_{n-2,\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{SS_x}}, \beta_0 + t_{n-2,\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{SS_x}} \right] 
\] (5)

Where \( \hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \), \( \hat{\beta} \sim N(\beta, \frac{\sigma^2}{SS_x}) \) and \( SS_x = \sum_{i=1}^{n} (x_i - \bar{x})^2 \) (6)

While \( t_{n-2,\frac{\alpha}{2}} \) is the 100(\( \alpha/2 \)) percentile of t-distribution with \( (n - 2) \) degree of freedom.

Now, put forward \( \psi_i(\hat{\theta}) = 0 \) and \( \psi_i(\hat{\theta}) = k = k_1/k_2 \)

Where, \( k_i = \left( \frac{2}{n-2} \right)^i \cdot \Gamma\left( \frac{n+8i-2}{2} \right)/\Gamma\left( \frac{n-2}{2} \right) \), for \( i = 1, 2 \) and \( n > 2 \).

The expressions for Bias and Mean Square Error (MSE) of \( \tilde{\beta} \) are respectively given by

\[
\text{Bias}(\tilde{\beta}_{PT} | \beta, R_i) = E(\tilde{\beta}_{PT} - \beta) = -\frac{\sigma}{\sqrt{SS_x}} \left\{ -\lambda_i [1 - k + kJ(\alpha, b^*)] - kJ(\alpha, b^*) \right\} 
\] (8)
An Efficient Shrinkage Estimator for the Parameters of Simple Linear Regression Model

Where $J_\ell(a^*,b^*) = \frac{1}{\sqrt{2\pi}\alpha}\int_0^\infty t^{\ell-1/2}e^{-t/2}dt$, \(\ell = 0,1,2\) \hfill (9)

And $t = \frac{\sqrt{SS_{\alpha}}(\beta - \beta)}{\sigma}, \lambda_i = \frac{\sqrt{SS_{\alpha}}(\beta - \beta_0)}{\sigma}, a^* = -\lambda_i - t_{\alpha/2,n-2}, b^* = -\lambda_i + t_{\alpha/2,n-2}$ \hfill (10)

We denote to the Bias ratio of $\hat{\beta}_{pr}$ as $B(\hat{\beta}_{pr})$ and defined as below

$$B(\hat{\beta}_{pr}) = \frac{\text{Bias}(\hat{\beta}_{pr}|\beta, R_1)}{\sigma/\sqrt{SS_{\alpha}}}$$ \hfill (11)

And,

$$\text{MSE}(\hat{\beta}_{pr}|\beta, R_1) = \text{E}(\hat{\beta}_{pr} - \beta)^2$$

$$= \frac{\sigma^2}{SS_{\alpha}} \left\{ k^2(1 + \lambda_i^2) - \lambda_i^2(2k - 1) - k^2[J_1(a^*,b^*)+2J_1(a^*,b^*)J_0(a^*,b^*)] \right\}$$

$$2k\lambda_i[J_1(a^*,b^*)+\lambda_iJ_0(a^*,b^*)]$$ \hfill ...(12)

The Efficiency of the proposed estimator $\tilde{\beta}_{pr}$ relative to $\hat{\beta}$ is defined as

$$\text{R.Eff}(\tilde{\beta}_{pr}|\beta) = \frac{\text{MSE}(\hat{\beta})}{\text{MSE}(\tilde{\beta}_{pr}|\beta, R_1)}$$ \hfill (13)

See [3], [4] and [5].

Also, preliminary test single stage shrinkage estimator of $\gamma$ using equation (3) will be:

$$\tilde{\gamma}_{pr} = \begin{cases} \psi_1(\hat{\gamma})\hat{\gamma} + (1 - \psi_1(\hat{\gamma}))\gamma_0 & , \text{if } \hat{\gamma} \in R_2 \\ \psi_2(\hat{\gamma})\hat{\gamma} + (1 - \psi_2(\hat{\gamma}))\gamma_0 & , \text{if } \hat{\gamma} \notin R_2 \end{cases}$$ \hfill (14)

Where $\psi_i(\hat{\gamma})$, $i = 1,2$ are shrinkage weight factors such that $0 < \psi_i(\hat{\gamma}) < 1$ and $\hat{\gamma}$ is an unbiased estimator (O.L.S.) of $\gamma$ as well as $R_2$ is the pretest region for acceptance of testing the hypothesis $H_{0}: \gamma = \gamma_0$ vs. the hypothesis $H_1$: $\gamma \neq \gamma_0$ with level of significance $\alpha$.

\[ R_2 = [-\gamma_0 - Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}, -\gamma_0 + Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}] \] \hfill (15)

Where $(Z_{\alpha/2})$ is the 100$(\alpha/2)$ percentile point of the standard normal distribution.

In the estimator $\tilde{\gamma}_{pr}$ which is defined in (14), we assume that $\psi_1(\hat{\gamma}) = 0$ and $\psi_2(\hat{\gamma}) = h$ (constant).

The expressions for Bias and Mean Square Error (MSE) of $\tilde{\gamma}_{pr}$ are respectively given as below:-

$$\text{Bias}(\tilde{\gamma}_{pr}|\gamma, R_2) = \text{E}(\tilde{\gamma}_{pr} - \gamma)$$

$$= \frac{\sigma}{\sqrt{n}} \left\{-\lambda_2[1 - h + hJ_0(a_1,b_1)] - hJ_1(a_1,b_1)\right\}$$ \hfill (16)

Where $J_\ell(a_1,b_1) = \frac{1}{\sqrt{2\pi}\alpha}\int_a^b \int_0^\infty z^{\ell-1/2}e^{-z/2}dZ, \ell = 0,1,2$ \hfill (17)

And $Z = \frac{\sqrt{n}(\hat{\gamma} - \gamma)}{\sigma}, \lambda_2 = \frac{\sqrt{n}(\gamma - \gamma_0)}{\sigma}, a_1 = -\lambda_2 - Z_{\alpha/2}, b_1 = -\lambda_2 + Z_{\alpha/2}$ \hfill (18)

We denote to the Bias ratio of $\tilde{\gamma}_{pr}$ as $B(\tilde{\gamma}_{pr})$ which is defined as

$$B(\tilde{\gamma}_{pr}) = \frac{\text{Bias}(\tilde{\gamma}_{pr}|\gamma, R_2)}{\sigma/\sqrt{n}}$$ \hfill (19)
And
\[ \text{MSE}(\tilde{\gamma}_n, R_2) = E(\tilde{\gamma}_n - \gamma)^2 \]
\[ = \frac{\sigma^2}{n} \left\{ h^2(1 + \lambda_1^2) - \lambda_1^2(2h - 1) - h^2 [J_1(a, b) + 2\lambda_1 J_2(a, b) + \lambda_1^2 J_3(a, b)] - 2h\lambda_1 [J_1(a, b) + \lambda_1 J_2(a, b)] \right\} \]

...(20)

The Efficiency of the proposed estimator \( \tilde{\gamma}_{PT} \) relative to estimator \( \hat{\gamma} \) is defined as
\[ \text{R.Eff}(\tilde{\gamma}_n | \gamma, R_2) = \frac{\text{MSE}(\hat{\gamma})}{\text{MSE}(\tilde{\gamma}_n | \gamma, R_2)} \] (21)

III. NUMERICAL RESULTS

1. The computation of Relative Efficiency \([\text{R.Eff}(\cdot)]\) and Bias Ratio \([B(\cdot)]\) were used for the estimator \( \hat{\beta}_{PT} \), these computations were performed for \( \alpha = 0.01, 0.05, 0.1, \lambda_1 = 0.0(0.1)2 \) and \( n = 8, 10, 12, 20 \). Some of these computations are displayed in the attached table 1 which leads to the following results.

i. The Relative Efficiency \([\text{R.Eff}(\cdot)]\) of \( \hat{\beta}_{PT} \) are adversely proportional with small value of \( \alpha \) and those of \( n \) and \( k \).

ii. \( \text{R.Eff}(\hat{\beta}_{PT}) \) has a maximum value when \( \beta = \beta_0 \) (\( \lambda_1 = 0 \)).

iii. The Bias Ratio \([B(\cdot)]\) of \( \hat{\beta}_{PT} \) are reasonably small when \( \beta = \beta_0 \) and vice versa otherwise.

iv. The Bias Ratio \([B(\cdot)]\) of \( \hat{\beta}_{PT} \) are increasing function with increases value of sample size (n).

v. The Effective Interval [The value of \( \lambda_1 \) which make the R.Eff(\cdot) of \( \hat{\beta}_{PT} \) greater than 1] is \([0, 1]\).

vi. The proposed estimator \( \hat{\beta}_{PT} \) dominates the usual estimator \( \hat{\beta} \) with large sample size n.

i.e. \( \lim_{n \to \infty} [\text{MSE}(\hat{\beta}_{PT} | \beta, R_1) - \text{MSE}(\hat{\beta})] \leq 0 \)

vii. \( \hat{\beta}_{PT} \) is consistent estimator

i.e.; \( \lim_{n \to \infty} \text{MSE}(\hat{\beta}_{PT} | \beta, R_1) = 0 \).

viii. The considered estimator \( \hat{\beta}_{PT} \) is better than the usual estimator and also than the estimator introduced by [2] and [3] in the sense of Mean Squared Error.

2. The computation of Relative Efficiency \([\text{R.Eff}(\cdot)]\) and Bias Ratio \([B(\cdot)]\) of the proposed estimator \( \tilde{\gamma}_{PT} \) were made on different constants involved in it, some of these computations are given in annexed tables (2)-(6) for samples of these constant e.g. \( \alpha = 0.01, 0.05, 0.1, h = 0.1(0.1)0.5 \) and \( \lambda_2 = 0.0(0.1)2 \). The following results from the mentioned table were made

i. The Relative Efficiency \([\text{R.Eff}(\cdot)]\) of \( \tilde{\gamma}_{PT} \) has a maximum value when \( \gamma \) very close to \( \gamma_0 \) (\( \lambda_2 = 0 \)) and decreases with increases value of \( \lambda_2 \) and \( h \).

ii. \( \text{R.Eff}(\tilde{\gamma}_{PT}) \) increasing function with small value of \( \alpha \) [level of significance of acceptance region \( R_2 \)].

iii. The Bias Ratio \([B(\cdot)]\) of \( \tilde{\gamma}_{PT} \) are reasonably small when \( \gamma \) close to \( \gamma_0 \) (\( \lambda_2 = 0 \)) and increases otherwise

iv. \( B(\tilde{\gamma}_{PT}) \) are increases when \( \alpha \) increases.

v. The Effective Interval [The value of \( \lambda_2 \) which make the R.Eff(\cdot) of \( \tilde{\gamma}_{PT} \) greater than 1] is \([0, 1]\).

vi. The considered estimator \( \tilde{\gamma}_{PT} \) is consistent estimator and dominates the usual estimator \( \hat{\gamma} \).

vii. The considered estimator \( \tilde{\gamma}_{PT} \) is better than the estimator \( \hat{\gamma} \) (least square method) and some existing estimator e.g. [2] and [3] in terms of higher Efficiency especially at \( \gamma \approx \gamma \).
### Figures and Tables

#### Table 1

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

#### Table 2

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

#### Table 3

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

An Efficient Shrinkage Estimator for the Parameters of Simple Linear Regression Model
From the above discussions it is obvious that by using guess point value one can improve the usual estimator. It can be noted that if the guess point \( \theta_0 \) is very close to the true value of the parameter \( \theta \) (i.e., \( \lambda \) is approximate close to one), the proposed estimators perform better than the usual estimator. If one has no confidence in the guessed value then preliminary test shrinkage estimators can be suggested. We can safely use the proposed estimators for small sample size at usual level of significance \( \alpha \) and moderate value of shrinkage weight factor \( \hat{\phi}(\theta) \).

### ACKNOWLEDGEMENTS

The authors wanted to provide thanks to the referees and to the editor for constructive suggestion and valuable comments which resulted in the improvement of this article.

### V. CONCLUSION

From the above discussions it is obvious that by using guess point value one can improve the usual estimator. It can be noted that if the guess point \( \theta_0 \) is very close to the true value of the parameter \( \theta \) (i.e., \( \lambda \) is approximate close to one), the proposed estimators perform better than the usual estimator. If one has no confidence in the guessed value then preliminary test shrinkage estimators can be suggested. We can safely use the proposed estimators for small sample size at usual level of significance \( \alpha \) and moderate value of shrinkage weight factor \( \hat{\phi}(\theta) \).
REFERENCES


