韓國數學敎育學會誌<數學敎育>
1983. 12. Vol. XXII, No. 1

Journal of the Korea Society of Mathematical Education
Dec. 1983. Vol. XXII, No. 1

# A Note on the Robustness of the Hodges-Lehmann Estimators Against Dependence in Regression Model

By Ung Sup Yun

*National Institute for Educational Research & Training, Seoul, Korea*

## 1. Introduction

We consider the simple linear regression model

$$Y_i=\alpha+\beta x_i+e_i,\quad i=1,\dots,n \tag{1.1}$$

where $\alpha$ and $\beta$ are unknown regression parameters, $x_i$'s are known constants, and $e_i$'s are random errors.

In this note, we are interested in the robustness of the Hodges-Lehmann type estimators (HLE) of regression parameters as compared with the ordinary least squares estimators (LSE), when the error terms are autocorrelated.

Most of the robust procedures have been developed under the assumption of independent and identically distributed observations. The HLE's are also constructed to be robust against nonnormality under the assumption of independence.

For location problems, the asymptotic relative efficiency (ARE) of the HLE is the same as the Pitman efficiency of the corresponding rank tests (see Hodges and Lehmann, 1963). Høyland(1968) and Gastwirth and Rubin (1975) studied the asymptotic behavior of the robust estimators on dependent data. The results show that, for any first order autoregressive normal process for which all the serial correlations $\rho_n$ are nonnegative, the ARE of the HLE is greater than the corresponding ARE in the case of independent observations. On the other hand, for the first order autoregressive double exponential process, the ARE of the HLE gets worse as $\rho\to 1$.

Song and Oh (1981) investigated the small sample behavior of the HLE in the simple linear regression model. The Monte Carlo study shows that the HLE is quite robust against nonnormality under the assumption of independence.

Relatively little work is done on the robustness of HLE against dependence. This is partly due to the complexity of the distributions of the estimators when the observations are correlated. In this note a preliminary study is performed on the robustness of the HLE of regression parameters in (1.1) when the errors follow first order stationary autoregressive processes, through a Monte Carlo study.

The results show that the HLE is robust with small values of $\rho$. But, for long tailed distributions the rasults indicate that the relative efficiency of the HLE is getting worse as $\rho\to 1$.

## 2. Hodges-Lehmann Type Estimators in Regression Model

We consider the simple linear regression model (1.1). Assuming that the error terms $e_i$'s are independent and identically distributed random variables with mean 0, the classical LSE of $\beta$ and $\alpha$ are given by

$$\tilde{\beta}=\frac{\sum(x_i-\bar{x})(Y_i-\bar{Y})}{\sum(x_i+\bar{x})^2} \text{ and } \tilde{\alpha}=\bar{Y}-\tilde{\beta}\bar{x},$$

respectively.

The HLE of $\beta$ based on rank tests are proposed by Scholz (1978) and Sievers (1978), among others. Song and Oh (1981) considered the following procedure to estimate $\alpha$ and $\beta$ simultaneously, which is basically equivalent to that of Scholz or Sievers.

Without loss of generality we assume that $x_1\leq\ldots\leq x_n$. Let $T_\alpha$ and $T_\beta$ be the weighted rank statistics defined by

$$T_\alpha=\sum_{i\leq j}\phi(Y_i+Y_j-\hat{\beta}(x_i+x_j)-2\alpha)$$

and

$$T_\beta=\sum_{i<j}W_{ij}\phi(Y_j-Y_i-\beta(x_j-x_i)),$$

respectively, where $\phi(\cdot)$ is an indicator function such that $\phi(t)=0$ or 1 according as $t\leq 0$ or $>0$. The weights $W_{ij}$ are nonnegative, and $W_{ij}=0$ whenever $x_i=x_j$. Let $W..=\sum_{i<j}W_{ij}$. Then HLE's of $\alpha$ and $\beta$ based on $T_\alpha$ and $T_\beta$ are defined by

$$\hat{\beta}=(\beta^*+\beta^{**})/2,\quad \hat{\alpha}=(\alpha^*+\alpha^{**})/2$$

where

$$\beta^*=\sup\{\beta : T_\beta\geq W../2\},\quad \beta^{**}=\inf\{\beta : T_\beta\leq W../2\}$$

and

$$\alpha^*=\sup\{\alpha : T_\alpha\geq n(n+1)/4\},\quad \alpha^{**}=\inf\{\alpha : T_\alpha\leq n(n+1)/4\}.$$

If we choose the weights $W_{ij}\equiv 1$, then the explicit form of the $\hat{\beta}$ is the median of

$$S_{ij}=(Y_j-Y_i)/(x_j-x_i),\quad i<j.$$

For the weights $W_{ij}=x_j-x_i$, $\hat{\beta}$ is the weighted median of $S_{ij}$ with weights $W_{ij}$, which has optimal properties (see Scholz, 1978). According to the results of a Monte Carlo study by Song and Oh (1981), the weighted median with weights $W_{ij}=x_j-x_i$ is quite robust and efficient in small sample cases.

The explicit form of the $\hat{\alpha}$ is the median of

$$1/2\{(Y_i-\hat{\beta}x_i)+(Y_j-\hat{\beta}x_j)\},\quad i\leq j$$

which is the HLE of location parameter based on the Wilcoxon test.

## 3. An Empirical Study on the Robustness Against Dependence

In order to investigate the small sample behavior of the HLE in linear regression model (1.1) with autocorrelated errors, we consider the case when the error terms $\{e_i\}$ follow a first order autoregressive process

$$e_i=\rho e_{i-1}+\varepsilon_i \quad (3.1)$$

where $\varepsilon_i$'s are white noise.

In our simulation study we compared the HLE with the LSE on the normal, double exponential, and Cauchy processes. In each case the location and scale parameters used are 0 and 1, respectively. The observed values are simulated from the model (3.1) and (1.1) with $n=20$; $\alpha=0$, $\beta=1$; $x_i=i$, $i=1, \dots, 20$. The values of $\rho$ used in the Monte Carlo study are 0(0.1) 0.5 and 0.8. In each case 500 observations are generated.

The uniform random numbers were generated by using the intrinsic subroutine RANDU in PDP 11, and converted to normal variates by standardizing sum of twelve random numbers. The inverse integral transformation was applied to generate double exponential and Cauchy processes. In each case $e_0$ is set by a random variate from the corresponding distribution.

**Table** 1. Empirical Means and Relative Efficiencies.

| $\rho$ | $\tilde{\beta}$(LSE) | $\hat{\beta}$(HLE) | RE($\hat{\beta}, \tilde{\beta}$) | $\tilde{\alpha}$(LSE) | $\hat{\alpha}$(HLE) | RE($\hat{\alpha}, \tilde{\alpha}$) |
|---|---|---|---|---|---|---|
| | | | Normal | | | |
| .0 | 1.00 | 1.00 | .94 | −.031 | −.030 | .92 |
| .1 | .996 | .996 | .93 | .039 | .036 | .92 |
| .2 | .999 | .999 | .96 | .033 | .039 | .96 |
| .3 | 1.00 | 1.00 | .94 | −.011 | −.009 | .96 |
| .4 | 1.00 | 1.00 | .96 | −.021 | −.016 | .96 |
| .5 | .997 | .997 | .96 | .043 | .042 | .97 |
| .8 | .994 | .994 | .97 | .013 | .019 | .99 |
| | | | Double Exponential | | | |
| .0 | .998 | .999 | 1.27 | .010 | .004 | 1.25 |
| .1 | 1.00 | 1.00 | 1.14 | −.032 | −.014 | 1.15 |
| .2 | 1.00 | 1.00 | 1.16 | .015 | .008 | 1.15 |
| .3 | 1.00 | 1.00 | 1.16 | −.000 | .004 | 1.13 |
| .4 | 1.00 | 1.00 | 1.18 | −.015 | −.029 | 1.17 |
| .5 | .995 | .994 | 1.09 | .070 | .070 | 1.10 |
| .8 | .983 | .983 | .99 | .180 | .193 | .98 |
| | | | Cauchy | | | |
| .0 | .892 | .995 | 3496. | −.221 | .053 | 5711. |
| .1 | 1.09 | 1.01 | 157. | −.917 | −.037 | 128. |
| .2 | .990 | .992 | 3787. | −2.92 | .096 | 9392. |
| .3 | 1.40 | 1.02 | 680. | 5.02 | −.071 | 224. |
| .4 | .985 | .996 | 17. | −.201 | −.010 | 29. |
| .5 | 1.06 | 1.01 | 13. | −.495 | −.071 | 6. |
| .8 | 1.22 | 1.14 | 1.9 | .284 | .235 | 1.8 |

The results of the Monte Carlo study are summarized in Table 1. The values of empirical means of $\tilde{\beta}$ (LSE), $\hat{\beta}$ (HLE), $\tilde{\alpha}$ (LSE), and $\hat{\alpha}$ (HLE) are presented. The relative efficiency (RE) of two estimators are computed as the inverse ratio of empirical variances.

The em pirical results show that the HLE is quite robust against a weak dependence. But, as $\rho \to 1$, the relative efficiencies are getting worse for double exponential and Cauchy proccsses, which have rather long tails.

## References


1. Gastwirth, J.L. and Rubin, H. (1975). The behavior of robust estimators on dependent data, *Ann. Statist.*, **3**, 1070-1100.
2. Hodges, J.L. and Lehmann, E.L. (1963). Estimates of location based upon rank tests, *Ann. Math. Statist.*, **34**, 598-611.
3. H$\phi$yland, A. (1968). Robustness of the Wilcoxon estimate of location against a certain dependence, *Ann. Math. Statist.*, **38**, 1196-1201.
4. Scholz, F.W. (1978). Weighted median regression estimates, *Ann. Statist.*, **6**, 603-609.
5. Sievers, G.L. (1978). Weighted rank statistic for simple linear regression, *J. Amer. Statist. Assoc.*, **73**, 628-631.
6. Song, M.S. and Oh, C.H. (1981). On a robust subset selection procedure for the slopes of regression equations, *J. Korean Statist. Soc.*, **10**, 105-121.