P2.14 HETEROSKEDASTICITY AND THE WEIGHTED LEAST SQUARE METHOD IN RADAR RAINFALL ESTIMATE

Pu Wenyao*[1,3] ，Wei Ming[2]，Li Hongbin[3]，He Yuke[3]，Yu Shixiu[3]，Huang Zhen[4]，LiNan[1]
（1. Department of Atmospheric Sciences，Nanjing University，Nanjing 210093；
2. Nanjing Sino-America Cooperative Remote Sensing Laboratory, JKLMD, NUIST, Nanjing ；
3. Weather bureau of Dalian，Liaoning province 116001）

1. INTRODUCTION

Recent studies show that the Z-I relations depend on the calculation methods[1], with which the same data may yield significant different values of A and b[2] [3] [4] [5]. Campos（2000）[6] concluded that the I-Z relation is fitted better through power-function regression for rain accumulations, whereas the linear relationship of lgI and lgZ is better for instantaneous rain rates. Brook E.Martner（2005）[7] show that for the same sample data, different fitting methods may lead to different Z-I relations. As for the power function fitting, (Z, I) values in big-value areas may have big weight, although the validity of these points are questionable.

Based on the basic hypothesis of the linear least squares model, we compare the difference among the error terms in different regression models and their influence on the results.

2. BASIC HYPOTHESIS IN THE LINEAR ORDINARY LEAST SQUARES MODEL

Corresponding Author address: Weather bureau of Dalian, Liaoning, China,
E-mail:pwysy@163.com

For the linear regression model with two variables, the regression calculation usually refers to least squares or its derived methods. It is the basic idea that the sum of squared residuals is taken as the criterion for the goodness of the fit with the regression line. the sum of squared residuals can represent, as a whole, the level of error between all of the samples and the regression line.

In order to obtain the effective regression analysis, we should bring some limitations to the two-variable linear regression model. It is crucial to make some hypothesis on the properties of the error term, which mainly includes basic items as follows [8]:

1) The relation $y_i = a + bx_i + \varepsilon_i$ between variable X and Y comes into existence for all of the samples, where ε_i is the random error term.

2) The error term ε_i corresponding to each group of samples is a random variable with zero mean values. Namely, the mathematical expectation of E(ε_i)=0 comes into existence when i=1,...n.

3) The variance of the error term ε_i is a constant. Namely, $\mathrm{Var}(\varepsilon_i) = \sigma^2$ comes into

existence when i=1,...n.

In different Z-I regression models, the description of error terms, expectations of the error terms and variance of the error terms are different, and the physical meanings are completely different. Whether or not their distribution satisfies the above basic hypothesis is the precondition for the use of least squares.

3.REQUIREMENTS OF THE BASIC HYPOTHESIS IN PRECIPITATION

1) The error term and its expectation and variance should show the relationship of the real and retrieval rainfall intensity. The aim of the regression calculation is to make the fitting effect of the rainfall intensity I optimal.

2) The basic idea of the least squares is to minimize the sum of squared residuals. The test standard[9] of the Z-I regression effect is $\sum_{\mathrm{i}}\varepsilon_{\mathrm{i}}=\sum_{\mathrm{i}}(I_i-\hat{I}_i)\rightarrow 0$ or $\varepsilon_{\mathrm{i}}=I_i-\hat{I}_i\rightarrow 0$, and the calculation result should satisfy the requirement of the test standard.

3) In the process of the Z-I fitting, because the samples with intense rainfall intensity are more essentiality to Z-I relation and their number is fewer, their contribution to the total variance is more prominent. Therefore, reasonable heteroscedasticity should be showed.

4. CONCLUSIONS

1) The comparison of the I-Z regression model is described as follows:

Table 1. comparison of the I-Z regression model

Regression mode	$\lg I_i=\lg\mathrm{c}+\mathrm{d}\cdot\lg\mathrm{Z_i}+\varepsilon_{\mathrm{i}}$	$I_i=\mathrm{c}\cdot\mathrm{Z_i^d}+\varepsilon_{\mathrm{i}}$		
		$\mathrm{w_i}=1$	$\mathrm{w_i}=\mathrm{Z_i^{-d}}$	$\mathrm{w_i}=\mathrm{Z_i^{-2d}}$
Equation's property	Linear equation	Non-linear equation		
Error term	$\varepsilon_{\mathrm{i}}=\lg I_i-\lg\hat{I}_i$	$\varepsilon_{\mathrm{i}}=I_i-\hat{I}_i$		
sum of squared residuals	$\sum_{\mathrm{i}}(\lg I_i-(\lg\mathrm{c}+\mathrm{d}\cdot\lg\mathrm{Z_i}))^2$	$\sum_{\mathrm{i}}(I_i-\mathrm{c}\cdot\mathrm{Z_i^d})^2$	$\sum_{\mathrm{i}}\mathrm{Z_i^{-d}}\cdot(I_i-\mathrm{c}\cdot\mathrm{Z_i^d})^2$	$\sum_{\mathrm{i}}\mathrm{Z_i^{-2d}}\cdot(I_i-\mathrm{c}\cdot\mathrm{Z_i^d})^2$
least-square results	$\sum_{\mathrm{i}}(\lg I_i-\lg\hat{I}_i)=0$ $\sum_{\mathrm{i}}(\lg I_i-\lg\hat{I}_i)\cdot\lg\mathrm{Z_i}=0$	$\sum_{\mathrm{i}}(I_i-\hat{I}_i)\cdot\mathrm{Z_i^d}=0$ $\sum_{\mathrm{i}}(I_i-\hat{I}_i)\cdot\mathrm{Z_i^d}\cdot\lg\mathrm{Z_i}=0$	$\sum_{\mathrm{i}}(I_i-\hat{I}_i)=0$ $\sum_{\mathrm{i}}(I_i-\hat{I}_i)\cdot\lg\mathrm{Z_i}=0$	$\sum_{\mathrm{i}}(I_i-\hat{I}_i)\cdot\mathrm{Z_i^{-d}}=0$ $\sum_{\mathrm{i}}(I_i-\hat{I}_i)\cdot\mathrm{Z_i^{-d}}\cdot\lg\mathrm{Z_i}=0$

Heteroscedasticity	none	strong	moderate	none
Effect test	$\sum_{i}(I_i - \hat{I}_i) \neq 0$	$\sum_{i}(I_i - \hat{I}_i) \neq 0$	$\sum_{i}(I_i - \hat{I}_i) = 0$	$\sum_{i}(I_i - \hat{I}_i) \neq 0$

2) When estimating the precipitation by Radar, we hope there is a comparable result between the real rainfall intensity and the retrieval rainfall intensity. Considering $\sum_{i}\varepsilon_{i} = \sum_{i}(I_i - \hat{I}_i) \to 0$, the physical meaning of error term is more definite when we take the rainfall intensity I as the dependent variable in power-function regression mode.

3) The error term in $\lg I_i = \lg c + d \cdot \lg Z_i + \varepsilon_i$ is related to lgI and lgZ, and the fitting result satisfy $\sum_{i}(\lg \frac{I_i}{\hat{I}_i}) = 0$. The error term $\varepsilon_i = (I_i - \hat{I}_i)$ in $I_i = c \cdot Z_i^{d} + \varepsilon_i$ directly represent the difference between real rainfall intensity and retrieval one. When the weight factor is made to $w_i = Z_i^{-d}$, the fitting result satisfies $\sum_{i}\varepsilon_i = \sum_{i}(I_i - \hat{I}_i) = 0$.

4) The calculation results of the two precipitation processes show that the hypothesis that the variances of the error term in Z-I relation has the same order of magnitude agrees more with the real situation.

As a result, there is no heteroscedasticity in $\lg I_i = \lg c + d \cdot \lg Z_i + \varepsilon_i$, The heteroscedasticity is relatively strong when weight factor is made 1 in $I_i = c \cdot Z_i^{d} + \varepsilon_i$, and few samples in big-value areas may greatly affect the fitting result. There is no heteroscedasticity when weight factor is made to $w_i = Z_i^{-2d}$. considering the real distribution of the Z-I samples, it is reasonable to choose $w_i = Z_i^{-d}$.

References：

[1] Zhangpeichang,Dudingyu,Daitiepi. Radar Meteorology [M]. Beijing ： Meteorology press，2001.

[2] Zawadzki, I. Factors affecting the precision of radar measurements of rain[R]. 22P Conf. on radar meteorology, Zurich, Switzerland, AMS, 1984, 251-256.

[3] Ciach, G.J., and W. F. Krawjewski. Radar-rain gauge comparisons under observational uncertainties[J]. J. Appl. Meteor., 1999, 38, 1519-1525.

[4] Campos, E. and I. Zawadski. Instrumental uncertainties in Z-I relations[J]. J. Appl. Meteor., 2000, 39, 1088-1102.

[5] Tokay,A., A.Kruger, and W.F.Krajewski. Comparison of drop size distribution measurements by impact and optical disdrometers[J]. J. Appl. Meteor., 2001 40, 2083-2097.

[6] Campos, E. and I. Zawadski. Instrumental uncertainties in Z-I relations[J]. J. Appl. Meteor., 2000, 39, 1088-1102.

[7] Brooks E. Martner, Vladimir Dubovskiy, and Sergey Y. Matrosov. Z-I relations from raindrop disdrometers: sensitivity to regression methods and DSD data

refinements[R]. 32nd Conf. on Radar Meteor., Albuquerque, NM, 2005.

[8] Russell Davidson &James G.Mackinnon. Econometric Theory and Methods [M]. Shanghai university of finance and economics press 2006.

[9] yuxiaoding, yaoxiuping, xiongtinhui. Basic principles and applications of the NEXTRAD Doppler weather radar[M]. Beijing: China Meteorological Administration training center, 2004.