THEORETIC RESEARCH ON ROBUSTIFIED LEAST SQUARES ESTIMATOR BASED ON EQUIVALENT VARIANCE-COVARIANCE

LIU Jingnan
YAO Yibin
SHI Chuang

KEY WORDS robust estimation; equivalent weights; equivalent variance-covariance

ABSTRACT Depending on analyzing the abuse of equivalent weights, a set of self-contained theory system on robust estimation based on equivalent variance-covariance is established, which includes $\rho$ function, $\psi$ function, equivalent variance-covariance function, influence function and breakdown point. And an example is given to verify that the robust models proposed in this paper are reliable and correct.

1 Introduction

On the basis of large quantities of data analyzed, statisticians point out that the probability of outliers in practice and scientific experiment are approximately 1%–10%\(^1\). Outliers always affect the correctness of result. The method of Least Square is very sensitive to outliers. Some outliers will seriously affect the estimation of parameter. In order to overcome such disadvantages of LS method, the robust estimation has been brought forward. The intention is to concoct an estimating method for counteracting the effect of modeling error, especially improve the ability for counteracting the effect of outliers.

On the basis of the theory of $M$-estimation that was brought forward by Huber in 1964, Krarup and Kubik of Danish introduced the theory of robust estimation to data processing of geodesy, and also presented the famous method, called the method of Danish. Caspary and Borutta from Germany also carried out the research and applications, such as resolving simultaneously distortion model, positioning parameter and gene of SD. The scholars of China, such as Professor Li Deren, Zhou Jiangwen, Huang Youcai, Yang Yuanxi etc. have lucubrated on the robustified estimator\(^2\). Among them, Professor Zhou Jiangwen has brought forward the concept of equivalent weights. He uses the equivalent weights to translate the $M$-estimation into LS estimation, so that robust estimation can be applied to the data processing in simple form; Professor Li Deren has put forward the method on the basis of the iterative choice weights which was named the method of Li Deren; Professor Yang Yuanxi first researched the robust estimation of dependent observations and presented the concept for dependent equivalent weights. He also constructed the function of dependent equivalent weights. These are important progress of robust estimation in the data processing. It must be noted that despite robust estimation is mature for independent observations in theory and practice, the research of robust estimation for dependent observations is just rising. The technical method of application must be further discussed. In this paper we mostly describe the robust estimation for dependent observations.
2 Theory of robust estimation based on equivalent weights and its localization for dependent observations

M-estimation, known as a basic type of robust estimation, has been investigated in the field of geodesy. On the basis of the principle of M-estimation, the theory of robust estimation based on equivalent weights is discussed in this chapter.

$L_1, L_2, \cdots, L_n$ are observation samples, and $X$ are unknown parameters. Its empirical probability density may be expressed as $f(X, L_i)$. In M-estimation, we are to find the estimation of parameters under the condition

$$G = \sum_{i=1}^{n} \ln f(X, L_i) = \max$$

If we replace $-\ln f(\cdot)$ with $\rho(\cdot)$, then M-estimation requires that

$$\Omega = \sum_{i=1}^{n} \rho(X, L_i) = \min$$

Eq. (2) is the basic definition formula of M-estimation. Thus we can express $\rho(X, L_i)$ as follows:

$$\rho(X, L_i) = \rho(v_i)$$

To dependent observations, their weights are given by

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

Considering the weights of observations and assuming that the dependence of observations can be reflected by the dependent weight matrix for M-estimation, the extremum function can be constructed as

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \rho(v_i, v_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \rho(a_i X - L_i, a_j X - L_j) = \min$$

Differentiating this expression with respect to unknown $X$ yields

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \cdot \varphi_i(v_i, v_j) \cdot a = 0$$

where $\varphi_i(v_i, v_j) = \frac{\partial}{\partial v_i} \rho(v_i, v_j)$, Eq. (5) can be written in matrix as

$$A^T P V = 0$$

where $\bar{p}_{ii} = p_{ii} \cdot w_i, \bar{w}_i = \varphi_i(v_i, v_i)/v_i$

$\bar{p}_{ij} = p_{ij} \cdot w_i, \bar{w}_j = \varphi_i(v_i, v_j)/v_j$

$\bar{p}$ is called dependent equivalent weight matrix.

$$\bar{p}_{ij} = \begin{cases} p_{ij}, & v_i/\sigma < K_0 \\ \frac{p_{ij} K_0}{v_i/\sigma} \left( \frac{K_1 - v_i/\sigma}{K_1 - K_0} \right)^2, & K_0 \leqslant v_i/\sigma < K_1 \\ 0, & K_1 \leqslant v_i/\sigma \end{cases}$$

where $K_0 = 1.0-1.5, K_1 = 2.5-3.0$.

Now we can know that the existing functions of dependent equivalent weights are based on such a presupposition that $p_{ij}$ reflect the dependence of observations and we do not take into account the invariance of the relevance of the original dependent observations. So the existing functions of dependent equivalent weights have some shortcomings.

1) The dependent equivalent weights we designed cannot reflect the dependence of observations directly. It is only $p_{ij}$ that can reflect the dependence of observations, and $p_{ii}$ is derived from variance-covariance.

2) $p_{ij}$ can not reflect the dependence of observations directly. It is only $p_{ij}$ that can reflect the dependence of observations, and $p_{ij}$ is derived from variance-covariance.

3) If we neglect the dependent coefficient $p_{ij}$, the change of $p_{ij}$ will directly change the dependence of the observations. But the dependence of observa-
tions are determined by the observations' geometrical and physical structure. It can not be changed optionally.

3 Robustified least squares estimator based on equivalent variance-covariance

In fact, if we classify outliers to stochastic model, then the outliers will manifest the evident difference between the transcendental variance \( \sigma^2_{ii} \) and the actual variance \( \tilde{\sigma}^2_{ii} \). In this case, the outliers' distribution can be expressed as the model of variance inflation (Fig. 1), so we can inflate the variance of abnormal observations to control the influence of outliers. On the basis of such an idea, we present the robustified least squares estimator based on equivalent variance-covariance. In detail we inflate the observations' variance-covariance according to the iterative result of adjustment to ensure that the transcendental variance \( \sigma^2_{ii} \) is suitable to the actual variance \( \tilde{\sigma}^2_{ii} \), and in this way the influence of outliers is controlled.

In this case, the outliers' distribution can be expressed as the model of variance inflation (Fig. 1), so we can inflate the variance of abnormal observations to control the influence of outliers. On the basis of such an idea, we present the robustified least squares estimator based on equivalent variance-covariance. In detail we inflate the observations' variance-covariance according to the iterative result of adjustment to ensure that the transcendental variance \( \sigma^2_{ii} \) is suitable to the actual variance \( \tilde{\sigma}^2_{ii} \), and in this way the influence of outliers is controlled.

![WD (Probability density)](image)

Fig. 1. The model of variance inflation (classify outliers to stochastic model)

3.1 The dependent coefficient \( \rho_{ij} \)

The dependent coefficient of observation \( \xi_i \) and \( \xi_j \) is defined as:

\[
\rho_{ij} = \frac{D_{ij}}{\sqrt{D_{ii}D_{jj}}} = \frac{\sigma^2 Q_{ij}}{\sqrt{\sigma^2 Q_{ii} \sigma^2 Q_{jj}}} = \frac{Q_{ij}}{Q_{ii} Q_{jj}}
\]

\( D \) is the corresponding variance matrix and \( Q \) is the inverse of weight matrix of observations. \( \rho_{ij} \) reflects the dependence of observations completely.

\( Q_L \) is only related to observations' geometrical and physical structure. Whether the observations contain outliers or not, \( Q_L \) is unchanged, so is the dependent coefficient \( \rho_{ij} \). Usually, if we say the stochastic model \( P_L = \sigma_0^2 D_{L}^{-1} \) contains errors, we mean that the transcendental coefficient of variance \( \sigma_0^2 \) contain errors, or a certain observation's variance \( \sigma^2_i \) contain errors, we do not mean \( Q_L \) contain errors.

3.2 Numerical model for equivalent variance-covariance

3.2.1 Modelling

As regards the robust estimation based on equivalent weights, the core is the design of equivalent weights function. Similarly, for the robust estimation based on equivalent variance-covariance, the core is the design of equivalent variance-covariance function.

For convenience, in the deducing process, if there is no special statement, we will replace \( D_{ij}^{-1} \) with \( (D^{-1})_{ij} \), \( D_{ji}^{-1} \) with \( (D^{-1})_{ji} \), \( D_{ii}^{-1} \) with \( (D^{-1})_{ii} \), \( D_{jj}^{-1} \) with \( (D^{-1})_{jj} \), \( D_{ii}^{-1} \) with \( (D^{-1})_{ii} \), and \( D_{jj}^{-1} \) with \( (D^{-1})_{jj} \).

For M-estimation, the \( \rho \) function should satisfy:

\[
\Omega = \sum_{i=1}^{n} \rho(X_i, L_i) = \min
\]

When the variance-covariance of observations is considered, the above formula will be:

\[
\Omega = \sum_{i=1}^{n} \rho(D^{-1}; X_i, L_i) = \min
\]

For multidimensional M-estimation, the extremum function can be constructed as:

\[
\sum_{i=1}^{n} \sum_{j=1}^{m} D_{ij}^{-1} \rho(v_i, v_j) = \min
\]
Here we use the inverse of variance matrix, for the convenience of using the LS method.

Differentiating Eq. (7) with respect to \( v_i \) yields

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij}^{-1} \varphi(v_i, v_j) a_i = 0 \tag{8}
\]

In Eq. (8), we neglect differentiating Eq. (7) with respect to \( v_j \) and the result can be expressed as \( \varphi_j(v_i, v_j) \). Because we find the format of differentiating \( \frac{\partial}{\partial \sigma_l} \frac{\partial}{\partial \sigma_j} \frac{\partial}{\partial \sigma_l} \) for the convenience of using the LS method. Differentiating Eq. (7) with respect to \( \sigma_i \) yields

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij}^{-1} \varphi(v_i, v_j) = 0 \tag{9}
\]

We now directly define \( \varphi \) function and suppose that \( \tilde{D}_{ii}^{-1} = \tilde{D}_{ii}^{-1} \varphi(v_i, v_i) / \sigma_i \), \( \tilde{D}_{ij}^{-1} = \tilde{D}_{ij}^{-1} \varphi(v_i, v_j) / \sigma_j \). Then Eq. (8) can be written as

\[
A^T \tilde{D}^{-1} V = 0 \tag{10}
\]

For the convenience of computing the both sides of Eq. (10) are multiplied by \( \sigma_i^2 \), then we have

\[
A^T (\sigma_i^2 \tilde{D}^{-1}) V = 0 \tag{11}
\]

Eq. (11) can be computed by the classic LS method.

### 3.2.2 Determining adjustive coefficient of variance-covariance

We have defined the standardization error as 

\[
\omega_i = \frac{v_i}{\sqrt{(Q_{xx}P_{ii})^{1/2}}} = \frac{\sigma_i}{\sigma_0} \frac{v_i}{\sigma_0} 
\]

we take \( \omega_i \) as the adjustive coefficient of variance-covariance. Thus if observation \( l_i \) contain outlier, the adjusted variance-covariance are:

\[
\tilde{D}_{ii} = \omega_i^2 D_{ii} = \frac{v_i^2}{\sigma_i^2} D_{ii} 
\]

where \( \tilde{D}_{ii} \) are adjusted variance-covariance, \( D_{ii} \) are transcendental variance-covariance, and \( \omega_i \) are the adjustive coefficient of variance-covariance.

### 3.2.3 Numerical model for equivalent variance-covariance and its characteristic

Numerical model for equivalent variance-covariance are:

\[
\begin{align*}
\tilde{D}_{ii} &= \begin{cases}
D_{ii} & \text{if } \omega_i < k_0 \\
\omega_i^2 D_{ii} & \text{if } \omega_i \geq k_0
\end{cases} \\
\tilde{D}_{ij} &= \rho_{ij} \sqrt{D_{ii} D_{jj}} , \\
i &= 1, 2, \ldots, n; j &= 1, 2, \ldots, n; i \neq j
\end{align*}
\]

where \( k_0 \) is chosen to be 1.5-3.0, and

\[
D_{ij}^{-1} = D_{ii}^{-1} \varphi(v_i, v_i) / v_i
\]

The characteristics of numerical model for equivalent variance-covariance are as follows.

1) This model is direct embodiment of variance inflation model which classifies outliers into stochastic model.

2) For the independent observations, this model is equal to the model of equivalent weights. In other words, the equivalent weights model is a special case of equivalent variance-covariance model.

3) For the dependent observations, this model take full advantage of transcendental information (\( \rho \)) of dependent observations, thus we can keep the relevance of dependent observations constant. Compared with the former dependent equivalent weights model, this model accords more with the fact.

4) For the dependent observations, the equivalent variance-covariance matrix designed by using this model is stringently symmetrical, but the equivalent weights matrix designed by using former dependent equivalent weights model is unsymmetrical.

5) This model is simple and intuitionistic. It is easy to affiliate to the existent LS program. So it is prone to achievement by programming.

### 3.3 Dependent \( \varphi \) function

In 3.2.1, we have used the following formula to establish the numerical model for equivalent variance-covariance:

\[
\tilde{D}_{ii}^{-1} = D_{ii}^{-1} \varphi(v_i, v_i) / v_i
\]
Theoretic Research on Robustified Least Squares Estimator Based on Equivalent Variance-Covariance

So the dependent ϕ function of equivalent variance-covariance model is:

\[ \phi_i(v_i, v_j) = D_{ii}^{-1}v_i/D_{ii}^{-1} \]

where \( D_{ii} \) and \( D_{ij} \) are defined by Eq. (12).

For the model established in this paper, the dependent ϕ function is as follows:

\[ \phi_i(v_i, v_j) = \begin{cases} v_i, & |wi| < k_0 \\ v_i/\sigma^2_i, & |wi| \geq k_0 \end{cases} \]

where \( k_0 \) is chosen to be 1.5-3.0, and \( \sigma_i = \frac{\sigma_i}{\sigma_i} \). In fact, \( \phi_i(v_i, v_j) \) is a special case of \( \phi_i(v_i, v_j) \), in other words, \( \phi_i(v_i, v_j) \) contain \( \phi_i(v_i, v_i) \) and \( \phi_i(v_i, v_j) \) can be derived from \( \phi_i(v_i, v_i) \). If we consider the symmetry of the dependent coefficient \( \rho_{ij} \), non-diagonal element can be expressed by diagonal element. So in this paper, we emphasize on adjusting the diagonal element of variance-covariance matrix when establishing the dependent ρ function and the coefficient of dependent equivalent variance-covariance.

3.4 Coefficient of dependent equivalent variance-covariance

For convenience, we specially define the coefficient of dependent equivalent variance-covariance \( w \) as \( w(r_i) = \frac{r_i}{\phi(r_i)} \), so according to the model established in this paper, the coefficient of dependent equivalent variance-covariance \( w \) are:

\[ w_i(v_i, v_j) = \begin{cases} 1, & |wi| < k_0 \\ \omega_i^2/(2 \times \omega_i^2), & |wi| \geq k_0 \end{cases} \]

where \( k_0 \) is chosen to be 1.5-3.0, and \( \omega_i = \frac{\sigma_i}{\sigma_i} \). The curve of \( w_i \) is shown in Fig. 2.

![Fig. 2 Curve of coefficient of dependent equivalent variance-covariance](image)

3.5 Dependent ρ function

In practice, ϕ function is the differential function of ρ function, and ρ is a symmetric, continuous, stringent convexity function with a unique minimum at zero. So the dependent ρ function of equivalent variance-covariance model is:

\[ \rho_i(v_i, v_j) = \int_0^\infty \phi_i(v_i, v_j)dv_i = D_{ii}^{-1}v_i^2/(2D_{ii}^{-1}) \]

where \( D_{ii} \) is defined by Eq. (12).

For the model established in this paper, the dependent ρ function is as follows:

\[ \rho_i(v_i, v_j) = \begin{cases} v_i^2/2, & |wi| < k_0 \\ v_i^2/(2 \times \omega_i^2), & |wi| \geq k_0 \end{cases} \]

where \( k_0 \) is chosen to be 1.5-3.0, and \( \omega_i = \frac{\sigma_i}{\sigma_i} \).

3.6 Influence function

According to the definition of influence function, for the model established in this paper, the influence function is as follows:
\[ IF_j(L_j; \hat{X}, F^{(j)}) = \begin{cases} - N^{-1} \sum_{i=1}^{n} a_i^T D_i^{-1} v_i, & |\omega_i| < k_0 \\ - (A^T A)^{-1} \sum_{i=1}^{n} a_i^T D_i^{-1} v_i / \omega_i, & |\omega_i| \geq k_0 \end{cases} \]

where \( k_0 \) is chosen to be 1.5-3.0, and \( \omega_i = \frac{v_i}{\sigma_v} \).

3.7 Break-down point

Break-down point, the same as influence function, is a significant index of quantificational robustness. Generally speaking, break-down point is the least ratio to make the estimation out of the control. In fact the reason of estimation break-down may be more than one kind. In some case, the limitless of \( \phi \) function is not the only reason of estimation break-down. According to the definition of break-down point, the break-down point of the model based on equivalent variance-covariance is 8%-10% by using the data from "Xi'an crustal deformation and surveying control GPS net".

4 Example and analysis

Nowadays, GPS is vastly used in surveying, and the adjustment of GPS net is of practical significance in geodesy adjustment. GPS net adjustment takes the baseline of synchronous net as adjustment objects, and in the same synchronous net all the baselines are dependent. So the robustified least squares estimator based on equivalent variance-covariance is significant in GPS net adjustment.

Here we take "Xi'an crustal deformation and surveying control GPS net" as an example to analyze the validity and the feasibility of the theory and the model proposed in this paper. This net has been past multiplicate outliers analysis and inspection, so we consider there is no outlier in the net. For the convenience of discussion, we add four outliers into the net, shown in Table 1.

For the outliers we added are small, we ignore the may be more than one kind. In some case, the limitless of \( \phi \) function is not the only reason of estimation break-down. According to the definition of break-down point, the break-down point of the model based on equivalent variance-covariance is 8%-10% by using the data from "Xi'an crustal deformation and surveying control GPS net".

Three schemes are performed:
1) \( S_1 \): LS adjustment without any outliers
2) \( S_2 \): LS adjustment with four outliers
3) \( S_3 \): Robustified Least Squares adjustment with four outliers (using the dependent equivalent variance-covariance model suggested in this paper).

Taking A001 as fixed point, we use the above three schemes to process 3D unlimited adjustment, and compare the results. The accuracy of coordinate are shown in Fig. 3-Fig. 5. The accuracy of baseline portion are shown in Fig. 6-Fig. 8.

| No. | Synchronous net     | Baseline  | Portion | Outliers/m |
|-----|---------------------|-----------|---------|------------|
| 1   | B9834101.REG        | A001-A003 | \( \Delta x \) | 0.05       |
| 2   | B9834201.REG        | A001-A006 | \( \Delta y \) | 0.05       |
| 3   | B9834301.REG        | A001-A009 | \( \Delta x \) | 0.05       |
| 4   | B9834302.REG        | A001-A013 | \( \Delta y \) | 0.05       |

![Fig. 3 Comparing of RMS of X coordinate](image)
Theoretic Research on Robustified Least Squares Estimator Based on Equivalent Variance-Covariance

Fig. 4 Comparing of RMS of Y coordinate

Fig. 5 Comparing of RMS of Z coordinate

Fig. 6 Comparing of RMS of ΔX portion of baseline

Fig. 7 Comparing of RMS of ΔY portion of baseline
From the above figures, following knowledges can be drawn:

1) The classical least squares adjustment is not robust. If observations contain outliers, the accuracy will drop distinctly because of the influence of outliers.

2) The model established in this paper can effectively deal with the outliers of dependent observations.

5 Conclusion

By analyzing the defect of equivalent weights, a set of self-contained theory system on robust estimation based on equivalent variance-covariance are established, which include $\rho$ function, $\varphi$ function, equivalent variance-covariance function, influence function and breakdown point. We use a small example to proof-test the robust model proposed in this paper. Through repetitious computation, we can draw such conclusions:

1) Using equivalent variance-covariance function to design dependent robustified least squares estimator is an effective method to process outliers.

2) If observations do not contain outlier, the parameter estimation gained by this model is the same as that by the LS method. Consequently, it is agonie and optimal.

3) If observations contain outliers, the parameter estimation gained by this model will less be influenced by the outliers.

4) This model can automatically process the outlier of dependent observation and the known data. It also can apply to process the outliers of independent observation, and the processing result is corresponded with that by the model of equivalent weights.

5) This model can apply to all kinds of surveying data.

The authors would like to give the notice that in this paper, we use standardization error to locate outliers, it isn’t the optimal method. If we locate outliers by reliability theory and the distinguishability theory of outliers, and then use the model of equivalent variance-covariance to process outliers, we can get better result. For how to locate outliers by reliability theory and the distinguishability theory of outliers, we will discuss in other papers.

References

1 Huber P J. Robust Statistics. New York: Wiley, 1981
2 Li D R. Error disposal and the theory of dependability. Beijing: Publishing House of Surveying and Mapping, 1988 (in Chinese)
3 Zhou J W. Classical theory of errors and robust estimation. Acta Geodaetica et Cartographica Sinica, 1989, 18 (2): 115-120 (in Chinese)
4 Huang Y C. Data snooping and robust estimation. Beijing: Publishing House of Surveying and Mapping, 1990 (in Chinese)
5 Shi C, Liu J N. Correspondence based on outlier analysis. Wuhan; Journal of Wuhan Technical University of Surveying and Mapping, 1998(1) (in Chinese)
6 Yang Y X. Robust estimation and its application. Beijing: Bayi Publishing House, 1993 (in Chinese)
7 Zhou J W, Huang Y C, Yang Y X, Ou J K. Robust least squares method. Wuhan: Publishing House of Huazhong University of Science and Technology, 1995 (in Chinese)
8 Tukey J W. Study of robustness by simulation, particularly in improvement by adjustment and combination. In: Launer R L, Wilkinson, G N, eds Robustness in Statistics, 1975