抗差估计论文集

周江文　杨元喜　欧吉坤　王跃进　著

国家自然科学基金资助项目

测绘出版社

一九九二年二月

抗差估计论文集

周江文　杨元喜　欧吉坤　王跃进　著

国家自然科学基金资助项目

测绘出版社

一九九二年二月

(京)新登字 065 号

内 容 提 要

本论文集汇集了著者有关抗差估计的部分研究成果。主要内容有经典误差理论与抗差估计；等价权原理；相关观测抗差估计；抗差贝叶斯估计；抗差拟合推估；杠杆观测与余差；有界影响抗差估计等。文集中提出了较实用的 IGG Ⅰ、IGG Ⅱ、IGG Ⅲ方案，这些方案兼取抗差估计与最小二乘法的理论和算法，希望推进抗差估计在测量中的应用与发展。本文集所提供的研究成果可供测量理论研究人员与教学工作者，以及测量生产单位的科技人员参考与应用。其中部分成果也希望能对抗差估计理论有所补充。

抗差估计论文集

周江文　杨元喜　欧吉坤　王跃进　著

*

测绘出版社出版·发行

大兴星海印刷厂印刷

新华书店总店科技发行所经销

*

开本 787×1092　1/16·印张 10.25·字数 225 千字

1992 年 5 月第一版·1992 年 5 月第一次印刷

印数：0 001—1 000 册·定价：12.00 元

ISBN　7-5030-0525-4/P·198

前　言

抗差估计理论是一门年轻的分支学科。自本世纪60年代正式提出以来，经过统计学界二十多年的耕耘，现已逐渐形成体系，并继续向前发展。对测量学科来说，首先它在更广泛、更严密的基础上重新研究了最小二乘法和非最小二乘法，加上其它方面的丰硕成果，有利于加深经典测量理论，应当受到测量学界的重视。本文集尝试推进抗差估计理论在测量中的应用与发展，并希望能对抗差估计理论有所补充。

本文集着眼测量实际，重点研究观测参数的估计，要求估值在减免粗差影响的同时，又具有小方差。其基本方案是"抗差最小二乘法"，即在构造参数解式时，兼取抗差估计和最小二乘法的理论和算法。基于误差有界分布理论，将数据分成三个基本段：小误差取原观测权；大粗差不承认其信息，其观测予以淘汰；中间段观测采用"等价权"，使抗差估值方程形式上和最小二乘法相同（IGG Ⅰ方案）。

对于相关观测，则顾及其先验相关权，参照IGG Ⅰ方案，构造了相关等价权，也将数据分为三段(IGG Ⅲ方案)。为解决相关观测抗差问题，进行了探索。

根据参数的验前分布和观测值分布的不同情形，构造了抗差贝叶斯估计和抗差拟合推估的三种极值条件，研究了它们的解算和验后精度。

杠杆观测在估计理论中是一个棘手的问题，为此文集中较多地研究了余差，并利用余差分布构造等价权和决定数据分段(IGG Ⅱ方案)。方案中同时提出以"强淘汰"求参数初值，辅以"弱淘汰"提高其效率。

此外，结合测量实际，分析了新近出现的几种高崩溃污染率抗差估计法，研究认为，它们一般不适于测量平差计算。

作为一个实例，还讨论了高崩溃污染率坐标转换问题。假设检验作为抗差估计的一个方面也作了少量讨论。

文集共收论文13篇，部分文章兼有中英文。

著　者

1991年12月

Preface

The theory of robust estimation, first put forward in the sixties by mathematicians, has now grown into a statistics branch, and is still going ahead. In the surveying field,it resumes,in particular, the least square and nonleast square methods more extensively. Together with researches in many other topics, it seems to be bound to reinforce the classical surveying theory. This collection of paprs tries to promote its application and development in surveying. We hope also it could make some contributions to the robust estimation.

The papers facus their attention on parametric estimation of observations. The estimators are required to weaken the effects of outliers and at the same time have small variances. The basic scheme is the"robust method of least squares", that is, it gives consideration to both robust estimation and the least squares method in theory as well as algorithm. Based on the theory of bounded error distribution, data are divided into three sections: observations with small errors take the original weights; the gross errors are not accepted as information and the corresponding observations are deleted; those in the middle are given"equivalent weights", such that the observation equations lead to robust estimates (IGG I scheme).

For correlated observations, the corresponding"equivalent correlated weights" are suggested (IGG II scheme). There we try to find a way for the robust estimation of such observations.

According to different prior distributions of both parameters and observations, three extreme forms for robust Bayesian estimation and robust collocation are constructed. The algorithms and posterior accuracy are discussed.

Leverage observations is a troublesome question in estimate theory. Keeping this in mind, we discuss the residuals in some detail. Equivalent weights and data division are fixed according to the distribution of residuals (IGG III scheme). In this scheme, we propose a "strong rejection" process for the preliminary determination of parameters and a supplementary "weak rejection" to improve the efficiency of the estimates.

We have also analyzed several estimators with high breakdown point develped in recent years. It is shown by investigation that they are usually not suitable for geodetic adjustment.

As a practical example, the problem of coordinate transformation with high breakdown point is studied. In addition, two papers concerning hypothesis test are included.

Of all 13 papers, a few are presented in both Chineseand English.

The authors
December, 1991

目　　录

经典误差理论与抗差估计*

周 江 文

提　　要

60 年代数学家重新从广泛的意义上提出非最小二乘法的估计问题，并已在抗差估计中取得重要的进展。本文试图从等价权出发，提出一种有效的估计方案，包括方差估计，全文立论多处借用了经典最小二乘理论。通过一个小算例，此方案和另外几种方案作了试比较，结果较好。

一、粗差与抗差估计

抗差估计（Robust estimation，有人译“稳健估计”等）从 60 年代正式提出，至今已有二十余年，数学家在这方面做了许多开拓性的工作，它的应用正在逐步扩大，其中测量将是一个重要的方面。数学家提供一定假设下的严密理论，适当应用于测量实际，可望得到有效的估计方法。

抗差估计的提出是与粗差（Outlier，有时用 Gross error）相联系的。粗差指离群的误差，由失误、观测（函数）模式差、分布模式差而来，它实际不可避免。观测模式差这里是指局部对全局性的系统差，没有有效的估计方法。就结果而言，观测模式比估计方法更重要。

测量上可取 Gauss—Markov 模式为正常分布模式。通常观测数 n 与未知量数 m 相差不很大。

小误差不可分辨，因此粗差实际指较大的局部偏离正常模式的误差。

所谓抗差估计，实际是在粗差不可免的情形下，选择估计方法使未知量估值尽可能减免粗差的影响，得出正常模式下的最佳估值。抗差估计也包括方差估计和假设检验。

最小二乘估计为粗差所吸引，使未知量估值偏离；又由于粗差的存在，方差估值往往偏大。但在正常分布模式下，此法具有优越的数学和统计性能。因此一个有效的估计方法，必须尽量保留最小二乘法的优越性，同时增强其抗差性。

子样应有一定的代表性，这是问题（包括模式）的全部根据所在。

二、等价权

设有观测子样$\{x_i\}$，其相互独立，观测权为$\{p_i\}$，i 由 1 至 n。化为单位权误差，其

* 本文发表在测绘学报第 18 卷第 2 期，1989 年。

经验概率密度可写成

$$g(x)=\frac{[p_i\delta(x-x_i)]}{[p]} \tag{1}$$

其中[ ]为求和符号，求和指数 i 常略去，如分母[p]，分子因 x_i 区别于 x，故保留指数。$\delta(x-x_i)$为集中于 x_i 的 Dirac 函数。

M 估计是要由观测$\{x_i\}$求参量$\{\theta_j\}$的估值，j 由 1 至 m，余差为$\{v_i\}$。求 θ_j 的条件是

$$\int\rho(v)g(x)dx \text{ 或}[p\rho]\text{就 } \theta_j \text{ 极小} \tag{2}$$

或

$$\left[p\frac{\partial\rho}{\partial\theta_j}\right]=\left[p\left(\frac{\partial\rho}{\partial v}\cdot\frac{1}{v}\right)\frac{\partial v}{\partial\theta_j}\cdot v\right]=\left[pw\frac{\partial v}{\partial\theta_j}\cdot v\right]=0 \tag{3}$$

其中 ρ 是挑选的极值函数。

（3）是估值方程，直接解算往往很困难。但它可以变化的形式改写为

$$\left(\frac{\partial V}{\partial\theta}\right)'PV=A'PV=O \tag{4}$$

其中

$$\frac{\partial V}{\partial\theta}=\begin{Bmatrix}\frac{\partial v_1}{\partial\theta_1} & \cdots & \frac{\partial v_1}{\partial\theta_m}\\ \vdots & & \vdots\\ \frac{\partial v_n}{\partial\theta_1} & \cdots & \frac{\partial v_n}{\partial\theta_m}\end{Bmatrix}\text{，记为}A$$

$$P=(p_{ii}),\quad p_{ii}=p_i\left(\frac{\partial\rho}{\partial v}\cdot\frac{1}{v}\right)_i=p_iw_i \tag{5}$$

w_i 称为权因子。

（4）可看作最小二乘解的法方程，相应观测方程

$$\underset{n\,m\,m1}{A\,\theta}=\underset{n1}{X}+\underset{n1}{V}\qquad\text{等价权}P \tag{6}$$

计算 P 要知道 v，它可取用适当的近似值，权的精度要求不高。我们称 P 为等价权，因为取它作为观测方程（6）的权所得出的法方程，正是估值方程（3）。

这样利用等价权 P 可将 M 估计化为最小二乘估计，这无论在计算、估算方案制定上都带来很大的便利。我们将充分利用它。

通过权因子，可以对不同的极值函数 ρ 进行对比。反之，若规定了权因子，也可以找出相应的极值函数。

下面列举几种通常有效的估计方案。这里作了适当的改化。在$|v|=k\sigma$时，权因子均为 1，σ 为观测权中误差，k 为倍数。

	极值函数 ρ	权因子 w	
最小二乘法	$v^2/2$	1	（7）
绝对和极小	$k\sigma\|v\|$	$\frac{k\sigma}{\|v\|}$	（8）

一种 Huber 方案[1]

$$\begin{cases} |v|<k\sigma; & v^2/2 \\ |v|\geqslant k\sigma: & k\sigma|v|-\dfrac{(k\sigma)^2}{2} \end{cases} \qquad \begin{matrix} 1 \\ \dfrac{k\sigma}{|v|} \end{matrix} \tag{9}$$

Fair 函数

$$2(k\sigma)^2\left\{\frac{|v|}{k\sigma}-\ln\left(1+\frac{|v|}{k\sigma}\right)\right\} \qquad \frac{2}{1+\dfrac{|v|}{k\sigma}} \tag{10}$$

丹麦法

$$\begin{cases} v^2/2, & |v|<k\sigma \\ -(k^2\sigma^2+k\sigma|v|)\exp\left(-\dfrac{|v|}{k\sigma}+1\right), & \text{其它}|v| \end{cases} \qquad \begin{matrix} 1 \\ 1/\exp\left(\dfrac{|v|}{k\sigma}-1\right) \end{matrix} \tag{11}$$

$\left(v\text{ 作迭代计算，权因子累乘因子 }\exp\left(-\dfrac{|v|}{k\sigma}\right)\right)$

取等价权作为已知值，经典公式可近似用于估算单位权方差 σ_0 及平差函数的积差。

按[4]pp44—48，注意 P 是近似权，$P=(p_i w_i)$，而 $Q=Q_s=\left(\dfrac{1}{p_i}\right)$ 是观测权逆，则可得

$$\sigma_0^2 \doteq \frac{V'PV}{n-m} \qquad \text{(不计淘汰观测)} \tag{12}$$

$$Q_\theta \doteq C^{-1}A'PQPAC^{-1}=C^{-1}A'\overline{P}AC^{-1}, \quad \overline{P}=(p_i w_i^2) \tag{13}$$

$$Q_{B\theta} \doteq BQ_\theta B' \tag{14}$$

例如观测平差值 $Y=A\cdot\theta$，则

$$Q_Y=AC^{-1}A'\overline{P}AC^{-1}A' \tag{15}$$

式中 $C=A'PA$。

三、抗差方案的选择——IGG I 方案

从上节列举的几种估计方案看，一个有效的抗差方案应作如下考虑。

有一界限 $k\sigma$，$|v|$ 在限内采用最小二乘法，权因子为 1；限外权因子随 $|v|$ 的增大由 1 逐渐减小。

绝对和极小解的最简单情形联系于中位数，正负余差权之和相等。观测变动只须保持所有余差符号不变，解不受影响，因此具有优越的抗差性。抗差理论证明，它的影响函数 (Influence function) 绝对值不变(不因粗差而异)；其崩溃污染率 (Breakdown point) 为极大值 1/2 (污染率在此限内，估值在界内)。这和最小二乘解 (平均值) 相比，具有显明的优越性。但由界限 $k\sigma$ 向内，权因子由 1 无限增大，这与观测权大大不符。

从测量误差理论来看，界限 $k\sigma$ 之 k 可取 1.5(按正态分布，误差在 $\pm 1.5\sigma$ 以外的概率仅为 0.13)，限外之观测既不能完全否定，又要限制其有害作用，采用抗差权因子

$$w=\frac{1+b}{1+b\dfrac{|v|}{k\sigma}} \tag{16}$$

以降低观测权是可取的。式中 b 取正值。

当余差超出 $\pm 2.5\sigma$ 时(正常模式下，概率为0.01)，在观测模式可用的情形下，不应作为观测信息，即取 $w=0$ (从抗差估计看，粗差也不能过大)。如按绝对和方案(8)，当 $|v|=2.5\sigma$ 时，w 仅达3/5，权因子缩小嫌慢。

(16)式求导数，$\frac{\partial \omega}{\partial b}=\frac{1-|v|/k\sigma}{(1+b|v|/k\sigma)^2}$，恒负。因此加大 b 可使 w 减小，$\lim_{b\to\infty} w=\frac{1}{|v|/k\sigma}$，即(8)已是减小最甚的因子。

Fair 函数不设界限，小 v 之权与观测权不符，但限内最大权因子为2，优于绝对和法。

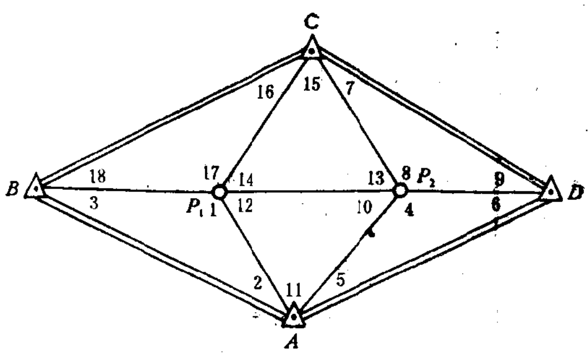
丹麦法权因子采用 $\exp\left(1-\frac{|v|}{k\sigma}\right)$，且在迭代计算中累乘因子，没有抗差上的论证，它实质上是淘汰法。

综上所述，我们的结论是：余差在 $\pm 1.5\sigma$ 以内，采用原观测权，即此段用最小二乘法；$\pm 2.5\sigma$ 以外，观测不用，即淘汰法；在中间段即 $\pm 1.5\sigma \sim \pm 2.5\sigma$ 之间(包括 $\pm 2.5\sigma$)，按绝对和极小取权因子 $w=\frac{1}{|v|/k\sigma}$，作为抗差措施。这个方案以下记为IGG Ⅰ方案。

为了防止余差的显著偏离，在提供余差的平差中应避免孤立的观测，丢掉可疑的观测；必要时计算 $J_A=A(A'PA)^{-1}A'P$ [4]，其对角元素 J_i 近于1时，对应之观测应丢掉即)所谓杠杆点 leverage point(作者将有另文论及)。也还可作少数几次迭代计算。

IGG Ⅰ方案权因子之间变化比较平缓，因此 k 的规定及余差值的小变动影响不大。

四、算　例



一测角网如图，等权测角18个(n)，A、B、C、D为已知点，P_1、P_2为未知点，未知量数 $m=4$。观测数据取自《测量平差基础》第282页。计算结果取自陆晓鸣同志的研究生论文。

观测方程如下：

δx_1	δy_1	δx_2	δy_2	l
−5.61	.18	.00	.00	−.20
2.46	1.32	.00	.00	−.60

3.15	−1.50	.00	.00	3.10
.00	.00	−3.53	4.77	−.90
.00	.00	.33	−3.47	−.50
.00	.00	3.20	−1.30	2.60
.00	.00	−2.45	−1.30	−3.10
.00	.00	5.65	.00	8.50
.00	.00	−3.20	1.30	−1.90
2.62	−.89	−2.29	−2.58	−1.20
−2.46	−1.32	−.33	3.47	2.00
−.16	2.21	2.62	−.89	−3.30
−2.62	.89	.17	−2.19	−4.00
2.33	2.60	−2.62	.89	−8.50
.29	−3.49	2.45	1.30	13.20
−.29	3.49	.00	.00	−9.60
3.44	−4.99	.00	.00	10.70
−3.15	1.50	.00	.00	−3.10

未知量的权逆阵 $Q_X = C^{-1}$如下：

.0120	.0043	.0023	.0025
.0043	.0161	.0024	.0032
.0023	.0024	.0117	.0041
.0025	.0032	.0041	.0169

单位权中误差 $\sigma_0 = 1.3''$。由上表可算得

$$\sigma_{x_1} = 0.14\text{dm} = \sigma_{x_2}$$

此外还算了 $J_A = AC^{-1}A'P$，($C = A'PA$)，得到对角元素 J_i(参看[4]pp·44—48)：

0.37， 0.13， 0.12， 0.39， 0.20， 0.11， 0.12， 0.37， 0.11，
(j_1)
0.26， 0.26， 0.16， 0.17， 0.27， 0.25， 0.19， 0.40， 0.12，
(j_{18})

其值均不大，因此没有杠杆点问题。

现在以粗差 $-10''(7.7\sigma_0)$分别加于 l_6，l_9，l_{17}，用不同方案进行估算，结果如下：

从这个小算例，可作如下试结论：

1. 不加粗差时，最小二乘估值LS，Huber 方案，IGG I 方案估值相同，可以作为较好的参考值。

2. 有粗差时，直接采用最小二乘法，结果很差(较差和较好的估值在表中加了方框)。

3. 最小绝对和（LAS）的结果也不佳。

	不加粗差 LS Huber, IGG	不加粗差 LAS	加于 l_6 LS	Huber	IGG I	LAS
δ_{x_1}	−0.10	0.04	−0.14	−0.12	−0.10	0.04
δ_{y_1}	2.32	2.38	2.29	2.31	2.32	2.22
δ_{x_2}	−1.21	−1.01	−1.53	−1.31	−1.23	−1.11
δ_{y_2}	−0.53	−0.49	−0.45	−0.51	−0.53	−0.29
σ_0*	1.3 (1.4)		2.7	1.6	1.2	1.9
v_6			8.3	9.1	9.4	9.4

	加于 l_9 LS	Huber	IGG I	LAS	加于 l_{17} LS	Huber	IGG I	LAS
δ_{x_1}	−0.06	−0.10	−0.10	0.04	−0.30	−0.23	−0.13	−0.16
δ_{y_1}	2.35	2.33	2.32	2.40	2.98	2.73	2.47	2.74
δ_{x_2}	−0.89	−1.17	−1.25	−0.99	−1.17	−1.18	−1.20	−1.08
δ_{y_2}	−0.62	−0.55	−0.52	−0.51	−0.46	−0.49	−0.52	−0.61
σ_0*	3.1	1.6	1.2	1.6	2.0	1.7	1.2	1.5
v_9 / v_{17}	10.1	11.1	11.4	10.6	4.8	6.3	7.9	6.5

4．IGG I 方案的估值和其他方案比较，一般与参考值最靠近。从加了粗差的余差看，此法也较好。(算例曾按未知量第四位小数达到稳定统计迭代次数，结果 LAS 法用了 23 次，Huber 方案用了 9 次，IGG I 方案用了 7 次。注意这样高的稳定位数在实际应用中没有必要。)

5．粗差改用 −5″，也进行了同样的计算，结论基本相同，只是情况要缓和些。

上述 IGG I 方案只是一个原型，我们还将提出能包含杠杆观测和相关观测的方案。

参 考 文 献

1. Huber, P.J.: Robust Statistics. John Wiley & Sons, 1981.

2. Hampel, F.R.et.al: Robust Statistics. John Wiley & Sons, 1986.

3. Rey, W.J.J: Introduction to Robust and Quasi-Robust Statistical Methods. Springer Verlag, 1983.

4. 周江文：越组约化理论，见《误差理论》，测绘出版社，1979.

* Huber 方案的方差计算公式是[1]p180，取 k=1.5。IGGI、LAS 的方差估计用近似权法。由于方差定义上的差异，这里列出的数字，只供参考。

Classical Theory of Errors and Robust Estimation*

Zhou Jiangwen

Abstract

In the sixties mathematicians resumed the non-least-squares problems and have made prominent progress in robust estimation. The present paper, starting from"equivalent weights",proposes a robust scheme(IGG I)for estimation of parameters and covariances of their functions, in reasoning much use is made of the classical least-squares theory. The results of trial computation on a small net show that the IGG scheme is good as compared with some other ones.

1. Outliers and Robust Estimation

Robust Estimation was formally put forward in the sixties. In over twenty years, mathematicians have perfected lots of originating works, finding applications in ever more fields, in which surveying will be an important one. Mathematicians provide rigid theories under definite conditions, effective estimation methods can hopefully be found dy proper use of them.

The motivation of rodust estimation is connected with outliers (some times "gross errors"). They are, as the name implies, outlying errors, stemming from mistakes, inadequacy in observation functions and in distribution models, all inevitable practically.Here the observation model deviations are restricted to local ones,noefficient estimation methods are available to cope with whole model deviations. So far as results are concerned, the observation model is even more important than the method of estimation.

In surveying, we may take Gauss-Markov distribution as the model one.The amount of observation n is usually not very large as compared with the number of parameters m.

Small errors are not distinguishable in origin, so the outliers practically refer to the large ones, locally deviated from models.

The robust estimation refers to those methods, by which the estimates or

* Published in Acta Geodetica et Cartographica Sinica Vol. 18, No 2, 1989

others, are to large extent free from the influence of outliers, i. e. the robust estimator obtains the best estimates of model parameters and others.

In the method of least squares estimates of parameters are attracted by outliers, variances estimates are usually to be larger owing to the existance of outliers. Nevertheless, under normal distribution, it is superior in mathematical and statistical properties. Therefore an effective estimation method will preserve the superiority of the method of least-squares, while at the same time improving its robustness.

After all, the whole rests on a well representative sample.

2. Equivalent weights

Given observation sample $\{x_i\}$, mutually independent, with observation weights $\{p_i\}$, i from 1 to n. Its empirical probability density may be written as

$$g(x) = \frac{[p_i \delta(x-x_i)]}{[p]} \qquad (1)$$

[ ] means summation, wherein index i generally omitted, like $[p]$ in the denominator; in the numerator, to distinguish x_i from x, i remains. $\delta(x-x_i)$ means Dirac function concentrated at x_i.

In M estimation, we are to find the estimation of parameters $\{\theta_j\}$, j from 1 to m, under the condition

$$\text{minimizing } \int \rho(v) g(x) dx \text{ or } [p\rho] \text{ with respect to } \theta_j \qquad (2)$$

or

$$\left[p \frac{\partial \rho}{\partial \theta_j} \right] = \left[p \left(\frac{\partial \rho}{\partial V} \frac{1}{V} \right) \frac{\partial V}{\partial \theta_j} \cdot V \right] = \left[pw \frac{\partial V}{\partial \theta_j} \cdot V \right] = 0 \qquad (3)$$

in which ρ some chosen function, $\{v_i\}$ the residuals.

To directly solve (3), the equation of estimates, is difficult, but the equation can be transformed into

$$\left(\frac{\partial V}{\partial \theta} \right)' PV = A' PV = 0 \qquad (4)$$

$$\frac{\partial V}{\partial \theta} = \begin{Bmatrix} \frac{\partial v_1}{\partial \theta_1} & \cdots & \frac{\partial v_1}{\partial \theta_m} \\ \vdots & & \\ \frac{\partial v_n}{\partial \theta_1} & \cdots & \frac{\partial v_n}{\partial \theta_m} \end{Bmatrix} = A \qquad (5)$$

$$P = (p_{ii}), \quad p_{ii} = p_i \left(\frac{\partial \rho}{\partial V} \frac{1}{V} \right)_i = p_i w_i$$

w_i will be called the weight factor.

(4) may be taken as normal equations in the least squares method, corresponding to observation equations

$$\underset{n\,m\,1}{A}\,\underset{n\,1}{\theta} = \underset{n\,1}{X} + \underset{n\,1}{V}, \qquad \text{equivalent weights } P \tag{6}$$

V is indispensable to calculate P, but we may take adequate approximate values of them, seeing that it is not very strict to assign weights. We shall call P equivalent weights, because taking them as weights of the observation equations (6), the normal equations thus found are the right estimate equations (3).

Thus by use of equivalent weights P, M estimation can be transformed into least squares estimation, which brings in convenience both in computation and in estimation scheme design. We shall make good use of it.

Through weight factors, comparison among different extremum functions ρ can be made. Reversely, given weight factors, the corresponding ρ may be found.

The following outlines some generally effective estimation schemes with appropriate modification, so that the weight factor at $|V| = k\sigma$ is 1, σ the observation weight s. e., k the multiplier.

	Extremal function ρ	Weight factors w	
least squares	$v^2/2$	1	(7)
least absolute sum	$k\sigma\lvert v\rvert$	$k\sigma/\lvert v\rvert$	(8)
One of Huber's schemes	$\lvert v\rvert < k\sigma$: $v^2/2$	1	
	$\lvert v\rvert \geqslant k\sigma$: $k\sigma\lvert v\rvert - \frac{(k\sigma)^2}{2}$	$\frac{k\sigma}{\lvert v\rvert}$	(9)
Fair function	$2(k\sigma)^2\left\{\frac{\lvert v\rvert}{k\sigma} - \ln\left(1 + \frac{\lvert v\rvert}{k\sigma}\right)\right\}$	$\frac{2}{1 + \frac{\lvert v\rvert}{k\sigma}}$	(10)

Danish method

$$\begin{cases} v^2/2, & |v| < k\sigma \qquad\qquad 1 \\ -(k^2\sigma^2 + k\sigma|v|\exp\left(-\frac{|v|}{k\sigma} + 1\right), & |v| \text{ otherwise} \qquad 1/\exp\left(\frac{|v|}{k\sigma} - 1\right) \end{cases} \tag{11}$$

Take equivalent weights as known, classical formula may be used for approximate assessment of variance of unit weight σ_0 and covariances of estimated functions.

According to references [4] pp44-48, distinguishing $P = (p_i w_i)$, the equivalent weights from $Q = Q_x = (1/p_i)$ the weight inverse of observation we find

$$\sigma_0^2 = V'PV/(n-m) \qquad \text{rejected observations out of account} \tag{12}$$

$$Q_\theta = C^{-1}A'PQPAC^{-1} = C^{-1}A'\overline{P}AC^{-1}, \quad \overline{P} = (p_i w_i^2) \tag{13}$$

$$Q_{B\theta} = BQ_\theta B \tag{14}$$

for example for $Y=A\theta$, we have

$$Q_v = AC^{-1}A'\overline{P}AC^{-1}A' \tag{15}$$

in which $C=A'PA$.

3. Choice of Robust Schemes——IGG I Scheme

Viewing from the preceding, an effective robust scheme should be designed as follows.

There will be a bound $k\sigma$, within which least squares method remains, weight factor being 1; Outside it the factor decreases from 1 gently as $|v|$ increases.

The method of least absolute sum is connected with the median in the simplest case, where sums of weights, respectively of positive and negtive residuals are equal. If changes of observations do not alter symbols of all residuals, the solution will be kept unchanged, therefore the method is superior. In robust theory, its absolute influence functions are shown constant, the breakdown point is the maximum 1/2. This confirms its superiority, as compared with the least squares method. Unfortunately, from the bound $k\sigma$ inward, the factor increases boundlessly, thus conflicting drastically with the observation weights.

According to the theory of errors, the multiplier k may be taken 1.5 (in normal distribution, the probability of an error outside $\pm 1.5\sigma$ is 0.13), observation with an error beyond the bound can not be neglected entirely, yet its harm should be prevented. Therefore it is advisable to set the weight factor as

$$w = \frac{1+b}{1+b\dfrac{|v|}{k\sigma}} \qquad b \text{ positive} \tag{16}$$

to lower the observation weights.

Once the residuals lie beyond $\pm 2.5\sigma$ (probability 0.01 in model distribution), they should not be taken as observation information i.e. taken $w=0$, provided the observation model is appropriate. In the method of least absolute sum, w is only 3/5 when $|v|=2.5\sigma$, thus the factor decreases somewhat too slowly.

Differentiate(16), $\dfrac{\partial w}{\partial b} = \dfrac{1-|v|/k\sigma}{(1+b|v|/k\sigma)^2}$, being negtive. Therefore to enlarge b will decrease w, $\lim\limits_{b\to\infty} w = 1/\dfrac{|v|}{k\sigma}$, thus(8) is already the fast decreasing factor.

Fair function sets no bound for small v, where it gets weight still contradicting observation weight, though its maximal weight factor is only 2, better than