

AD-A123 651

DRSAR-LEP-1
Tech Lib

AD A123 651

AD

TECHNICAL REPORT ARBRL-TR-02459

ROBUST REGRESSION: A DIAGNOSTIC TOOL

Jock Grynovicki
Jerry Thomas

TECHNICAL
LIBRARY

January 1983



US ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

Approved for public release; distribution unlimited.

Destroy this report when it is no longer needed.
Do not return it to the originator.

Secondary distribution of this report is prohibited.

Additional copies of this report may be obtained
from the National Technical Information Service,
U. S. Department of Commerce, Springfield, Virginia
22161.

The findings in this report are not to be construed as
an official Department of the Army position, unless
so designated by other authorized documents.

*The use of trade names or manufacturers' names in this report
does not constitute endorsement of any commercial product.*

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TECHNICAL REPORT ARBRL-TR-02459	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ROBUST REGRESSION: A DIAGNOSTIC TOOL		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Jock Grynovicki Jerry Thomas		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Army Ballistic Research Laboratory ATTN: DRDAR-BLB Aberdeen Proving Ground, MD 21005		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS RDT&E 1L162618AH80
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Armament Research & Development Command US Army Ballistic Research Laboratory (DRDAR-BL) Aberdeen Proving Ground, MD 21005		12. REPORT DATE January 1983
		13. NUMBER OF PAGES 34
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Robust Regression Iteratively Reweighted Least Squares Least Squares Outliers Residuals Regression		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (eal) This report discusses the usefulness of robust regression when distorted observations have undue influence on the normal least squares regression fit. Instead of throwing out these distorted observations, Iteratively Reweighted Least Squares (IRLS) techniques are used to reduce the influence these outliers have on the fitted line. These robust techniques were utilized through a Robust Statistical Estimation Computer Package (ROSEPACK) that was developed at Massachusetts Institute of Technology. A detailed description of how to run this statistical package on the BRL CYBER system is also presented.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

	Page
LIST OF FIGURES	5
I. INTRODUCTION.	7
II. ROBUST REGRESSION	11
A. Iteratively Reweighted Least Squares.	11
B. Weighting Function.	12
III. ROBUST STATISTICAL ESTIMATION PACKAGE (ROSEPACK).	16
IV. SUMMARY	17
APPENDIX A. ITERATIVELY REWEIGHTED LEAST SQUARES	19
APPENDIX B. ROSEPACK	21
DISTRIBUTION LIST	33

LIST OF FIGURES

	Page
1. Striking Velocity vs. Propellant Weight	8
2. Least Squares Regression of Striking Velocity vs. Propellant Weight	9
3. Least Squares Regression of Propellant Weight vs. Striking Velocity with Outliers Removed	10
4. Weight Function vs. Scaled Residuals	14
5. Robust Regression of Striking Velocity vs. Propellant Weight . . .	15
B1. Sample Program	22
B2. Data	24
B3. ROS Format Update.	25

I. INTRODUCTION

In experimentation, test data are collected on one or more variables of interest. The objective of the experiment usually is to determine how at least one other variable or factor is associated with or affects another variable. For example, striking velocity may be observed for a number of projectiles. Associated with this velocity may be other factors such as propellant weight, projectile weight, or propellant temperature. For this example, let's assume that we have measured and recorded the different propellant weights associated with the striking velocities. A graph of the data is given in Figure 1.

In looking at the data, there seems to be a linear relationship between striking velocity and propellant weight. A common question is, "Can this data be represented by some functional relationship?" In this era of sophisticated computers and calculators, everyone has easy access to some type of standard regression program that determines the linear relationship between two or more variables. A regression of velocity (V) on propellant weight (WT) was run on the data plotted in Figure 1. The following equation was obtained:

$$\hat{V}_i = 4648.56 - 10.58 (WT_i) \quad (1)$$

This equation was obtained by the method of least squares. This means that the sum of squares of vertical deviations of observations from the fitted line defined by (1) is smaller than the corresponding sum of squares of deviations from any other line. These deviations from the fitted regression line are commonly referred to as residuals. For this example, they are represented by the mathematical expression (residual = $(V_i - \hat{V}_i)$), where \hat{V}_i is obtained by Equation 1.

However, in analyzing the data, we notice from Figure 2 that three observations identified as a, b, and c stand out from the majority of the data and think that these observations may have unduly influenced the least squares fit. One method of determining the strength of this linear fit is to look at the Coefficient of Determination (R^2). This statistic indicates the proportion of variation in the observations (V_i 's) that are explained by the linear regression line.¹ For this example, only 16 percent of the total variability is explained by Equation 1. We feel that this is not a good fit and that most of the data can be better represented by another line. So we decided to exclude the points a, b, and c and compute another regression line. The following equation was obtained:

$$\hat{V}_i^* = 2685.22 + 152.19 (WT_i^*) \quad (2)$$

On examining this line (See Figure 3), it was found that it represents the data much better, except for the three excluded points. After excluding the three outliers, 87 percent of the variation was explained ($R^2 = .87$). However, most analysts are hesitant to throw out observations, especially without good reason. Also, if you start throwing points out, how do you decide when to stop?

¹W. Dixon, F. J. Massey, Introduction to Statistical Analysis, McGraw-Hill, Inc., 1969, P. 328.

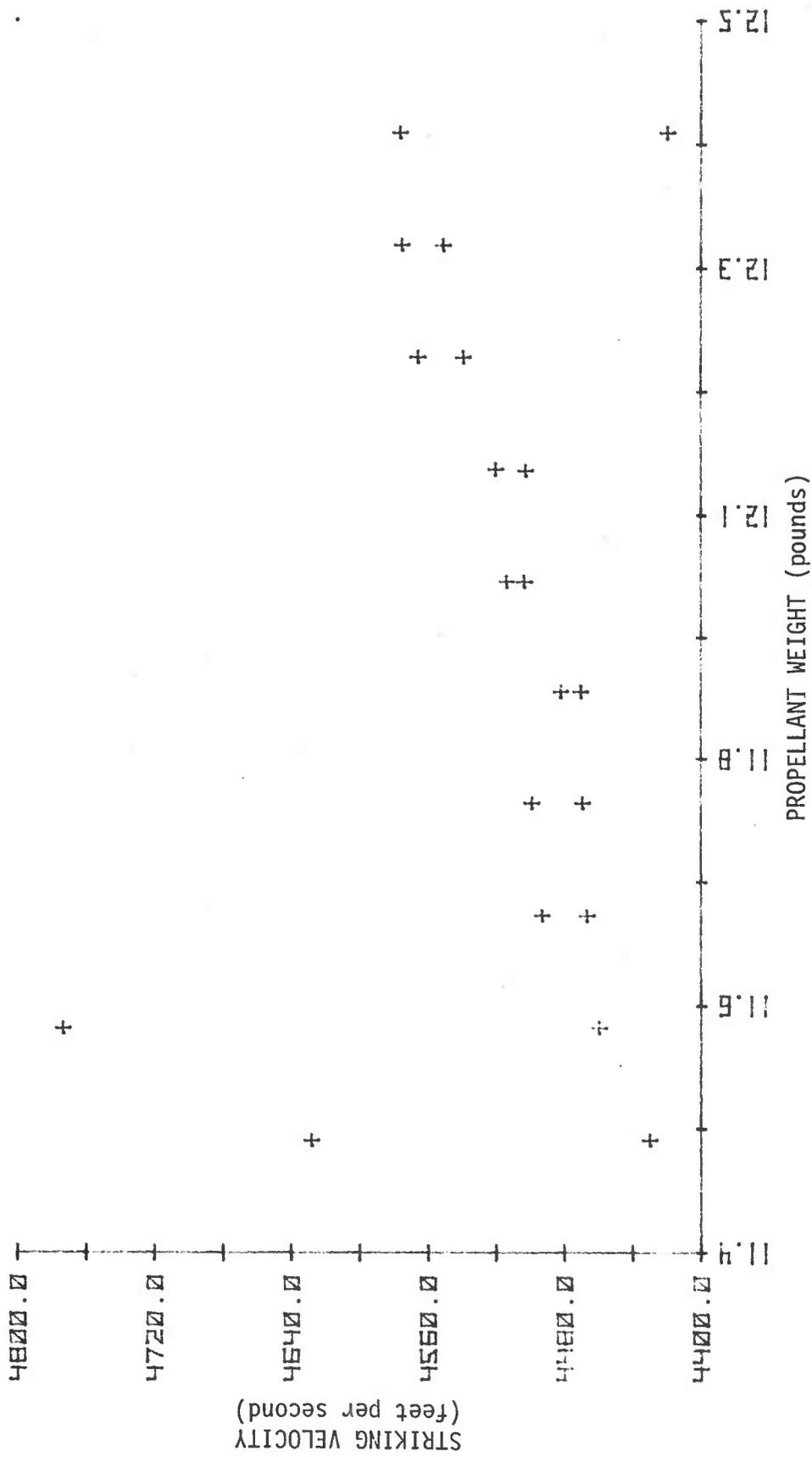


FIGURE 1. Striking Velocity vs. Propellant Weight

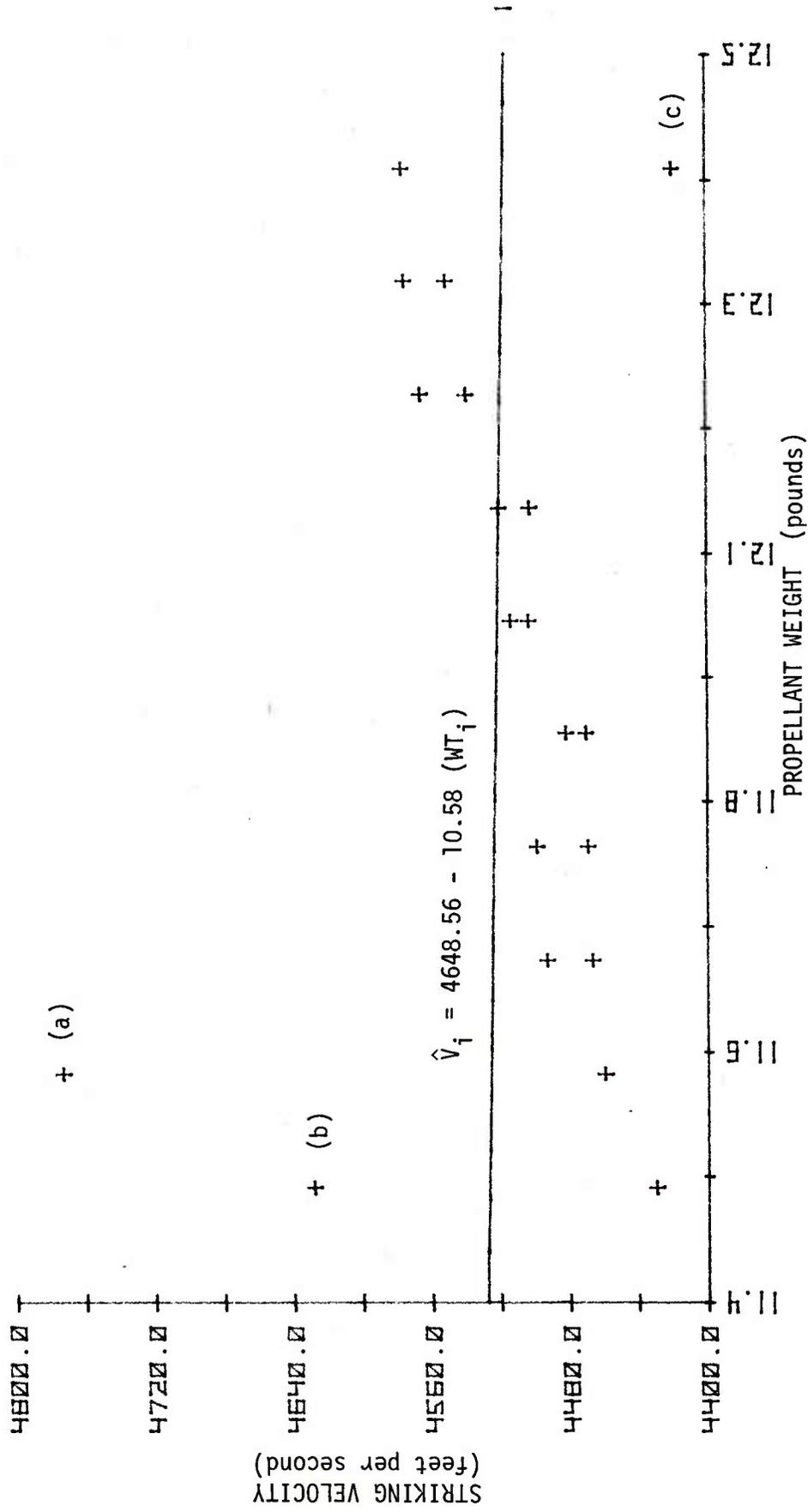


Figure 2. Least Squares Regression of Striking Velocity vs. Propellant Weight

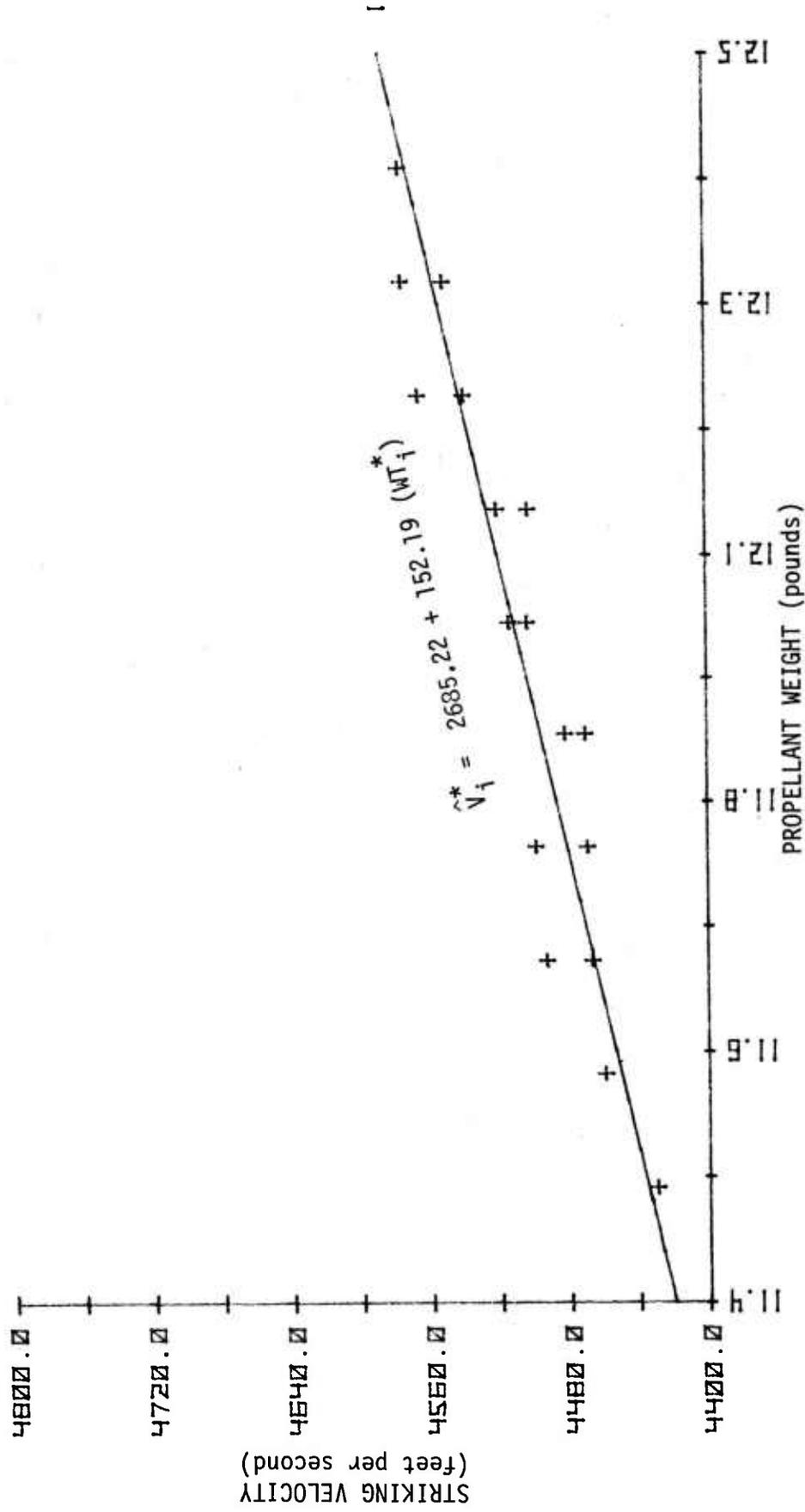


Figure 3. Least Squares Regression of Propellant Weight vs. Striking Velocity with Outliers Removed

One solution to this problem is to amend the usual least squares procedure by transforming the original data so that distorted observations will not have undue influence on the fitted least squares regression line. This leads us to the topic of the paper, robust regression using Iteratively Reweighted Least Squares (IRLS). A computer package, Robust Statistical Estimation Package (ROSEPACK), which computes solutions to the iteratively reweighted least squares problem and is available on the Ballistic Research Laboratory's CDC computer, is discussed.

11. ROBUST REGRESSION

A well known² statistician by the name of George E. P. Box coined the term "robustness" where robust techniques are defined to be techniques that are insensitive to moderate departures from the underlying assumptions. The assumptions for least squares are that the dependent variable, (for this example, striking velocity) is normally distributed and has a constant variance ($\sigma^2_{y \cdot x}$) for any given independent variable (i.e., propellant weight).

Departures from the assumption of normality and/or homogeneity of variance may distort the least square fit in representing the true linear relationship that is demonstrated by the bulk of the data. The most common of these departures is attributed to what is generally referred to as an "outlier" or "high leverage point." There are many definitions of outliers and many techniques for identifying them.¹ For the purpose of this paper, let's define an outlier as an observation that has undue influence on the fitted regression line. One way of limiting the effects of these outliers is to use a robust regression technique called Iteratively Reweighted Least Squares.³

A. Iteratively Reweighted Least Squares

In simple terms, Iteratively Reweighted Least Squares (IRLS) is a reweighted least squares problem where the weights are functions of the scaled residuals. The basic idea is to transform the observations by multiplying them by values of a weighting function so that they appear to satisfy the usual assumptions. Then, one uses normal least squares techniques on these weighted observations. The weighting values are not fixed but are found as part of an iterative process. There are many different weighting functions, but all of them exhibit the following property: small weighting values are assigned to observations that have large residuals associated with them. The solution to the problem is obtained by:

- 1) determining an initial fit of the data,
- 2) selecting a weight function,
- 3) calculating the scaled residuals,

²R. V. Hogg, "Statistical Robustness: One View of its Use for Application Today," The American Statistician, August 1979, V33, P. 108-115.

³P. Holland, R. Welsch, "Robust Regression Using Iteratively Reweighted Least Squares," Communications in Statistics: Theory and Methods, 1977.

- 4) calculating the weighting values based on the scaled residuals,
- 5) determining a least squares regression fit based on the weighted observations,
- 6) testing to see if this fit differs from the previous fit.

Steps 4 through 6 are continued until the linear expression of the data converges. Further theoretical discussion on robust regression estimates using Iteratively Reweighted Least Squares is outlined in Appendix A.

B. Weighting Function

Generally speaking, one of eight weight functions³ is commonly used as part of the iterative process (see Table 1). The default tuning constant associated with each function provides 95 percent asymptotic efficiency of the estimated regression coefficients with respect to ordinary least squares when the errors are normally distributed. A graph of the values of each weight function vs. the scaled residuals is given in Figure 4. Weight values of one indicate ordinary least squares. The Fair weight function deviates the fastest from normal least squares, and in fact, was derived to approximate the sum of the absolute residual regression. On the other hand, the Welsch weight function provides estimates that resemble the normal least square's process. The Talwar weight function behaves like normal linear regression except that observations which have large standardized residuals are excluded from the analysis. All the other weight functions provide estimates that lie between these extremes.

Going back to the previous data, an Iteratively Reweighted Least Squares Regression is performed using the Cauchy Weighting Function. The following equation was obtained:

$$\hat{V}_i = 2887.1 + 135.33 (WT_i) \quad (3)$$

This regression line represents the linear trend of the data (Figure 5) without having to disregard any of the observations. The equation was obtained by giving the observations the following weights.

TABLE 1. WEIGHT FUNCTIONS

<u>NAME</u>	<u>$\rho(r)$</u>	<u>$\rho'(r)$</u>	<u>$w(r)$</u>	<u>RANGE</u>	<u>TUNING CONSTANT</u>
Andrews	$A^2 [1 - \cos(r/A)]$	$A \sin(r/A)$	$(r/A)^{-1} \sin(r/A)$	$ r \leq \pi A$	$A = 1.339$
	$2 A^2$	0	0	$ r > \pi A$	
Bi Square	$(B^2/6) [1 - [1 - (r/B)^2]^3]$	$r [1 - (r/B)^2]^2$	$(1 - (r/B)^2)^2$	$ r \leq B$	$B = 4.685$
	$B^2/2$	0	0	$ r > B$	
Talwar	$r^2/2$	r	1	$ r \leq T$	$T = 2.795$
	$T^2/2$	0	0	$ r > T$	
Cauchy	$(C^2/2) \log[1 + (r/C)^2]$	$r [1 + (r/C)^2]^{-1}$	$[1 + (r/C)^2]^{-1}$		$C = 2.385$
Welsch	$(W^2/2) [1 - \exp[-(r/W)^2]]$	$r \exp[-(r/W)^2]$	$\exp[-(r/W)^2]$		$R = 2.985$
Huber	$r^2/2$	r	1	$ r \leq H$	$H = 1.345$
	$H r - H^2/2$	$\text{sgn} \cdot (r) H$	$H r ^{-1}$	$ r > H$	
Logistic	$L^2 \log[\cosh(r/L)]$	$L \tanh(r/L)$	$(r/L)^{-1} \tanh(r/L)$		$L = 1.205$
Fair	$F^2 [r /F - \log(1 + r /F)]$	$r(1 + r /F)^{-1}$	$(1 + r /F)^{-1}$		$F = 1.400$

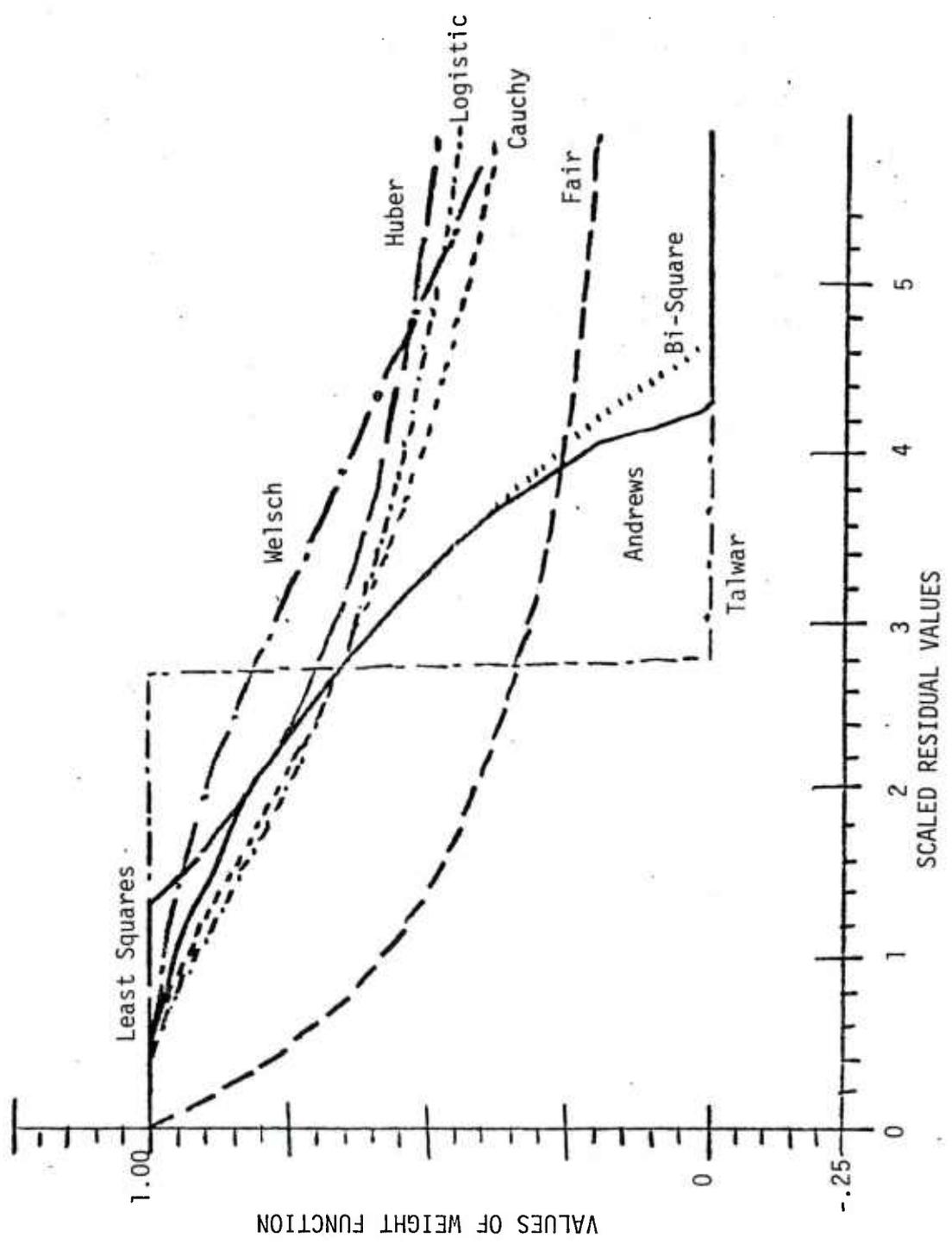


Figure 4. Weight function vs. Scaled Residuals

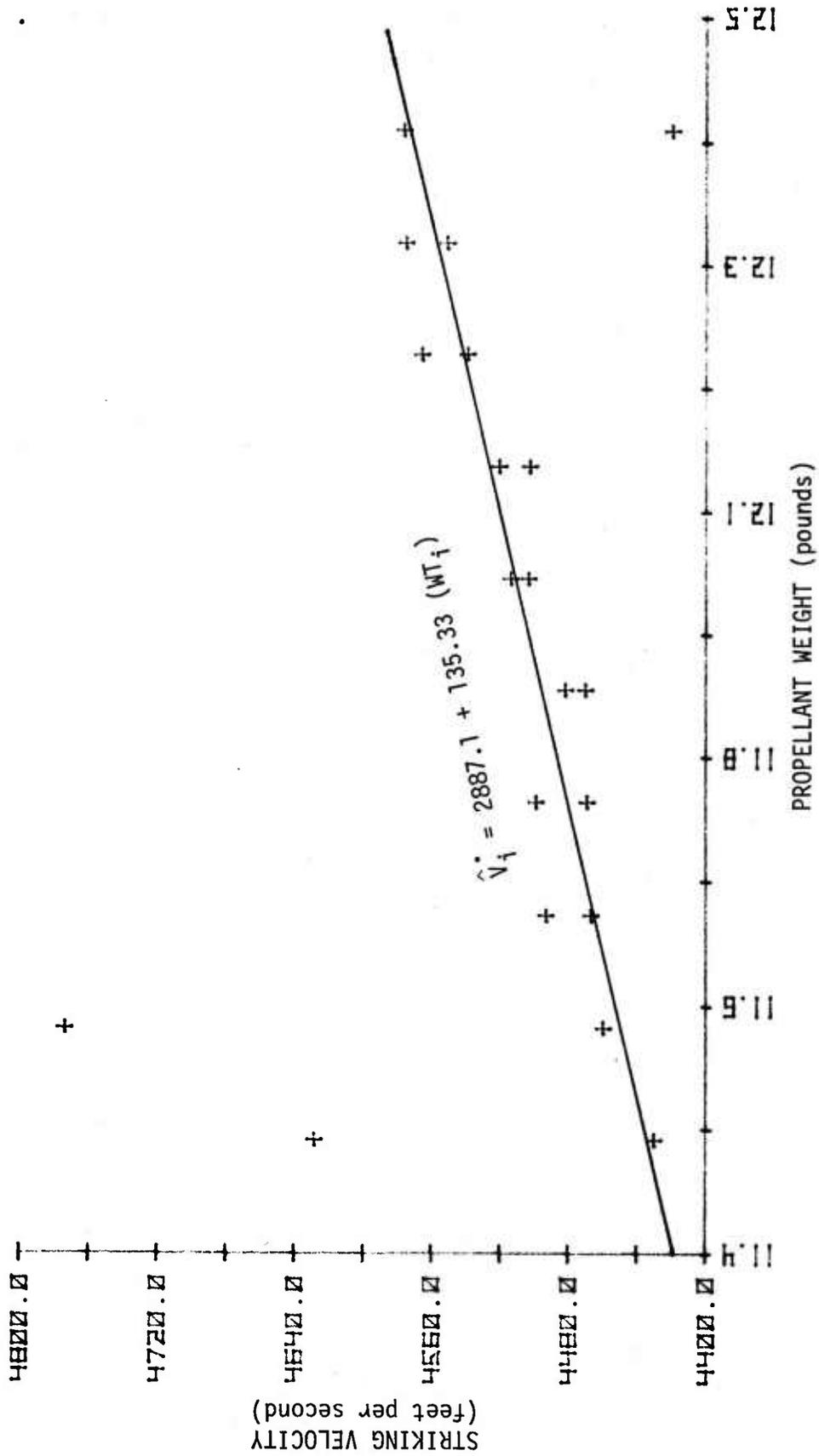


Figure 5. Robust Regression of Striking Velocity vs. Propellant Weight

<u>Observations</u>	<u>Weight Values</u>
1	.2689253E+00
2	.9679182E+00
3	.1609051E+00
4	.9986705E+00
5	.9165155E+00
6	.9973460E+00
7	.9607954E+00
8	.9607727E+00
9	.8837554E+00
10	.9570672E+00
11	.9985002E+00
12	.9891447E+00
13	.9205748E+00
14	.9970075E+00
15	.8814305E+00
16	.9997461E+00
17	.9092763E+00
18	.9999391E+00
19	.9772802E+00
20	.3343834E+00

All of the observations have weighted values close to one except for the three outliers. These high leverage points have large absolute residuals associated with them. Consequently, the Cauchy weight values for these three points are less than 0.4. In essence, the influence of these observations on the regression line has been reduced without having to exclude them from the sample.

III. ROBUST STATISTICAL ESTIMATION PACKAGE (ROSEPACK)

The Massachusetts Institute of Technology (M.I.T.) developed a Robust Statistical Estimation Package (ROSEPACK),⁴ which the Ballistic Research Laboratory (BRL) obtained and installed on the CDC computer. The package contains internal documentation, a primer which explains the control statements,

⁴D. Coleman, P. Holland, N. Kaden, V. Klema, S.C. Peters, "A System of Subroutines for Iteratively Reweighted Least Squares Computation," Massachusetts Institute of Technology (MIT), December 1977.

an interactive driver (IRLSDR), and computational tools for computing solutions to the Iteratively Reweighted Least Squares problem. These include an option for choosing an initial estimate, a default convergence level, a robust estimate for the scale of the residuals, and the eight weight functions that were previously stated. In addition, a host of optional statistical output is available. The package can be used interactively or in batch mode. Further details, including the job stream used to run our example, are provided in Appendix B.

IV. SUMMARY

We have illustrated the effects that outliers can have on normal least square estimates and have presented robust regression using ROSEPACK as a method of reducing their influence on the regression fit. By no means do we suggest that robust regression should replace normal least squares, but we believe it should be incorporated as a part of the analysis in most least squares problems. Each analysis should include both least squares regression and robust regression using ROSEPACK. If the estimates obtained from these two procedures are similar, then the inferences drawn from least squares should be correct. If, however, the estimates do not agree, one should reexamine the data, paying particular attention to those observations with low weights from the robust fit. This approach will force the analyst to examine his data and assure that the functional relationship being fitted represents the bulk of the data.

Another problem that occurs in regression analyses is high correlations between the various predictors or independent variables in the model. This problem could cause unstable parameter estimates. Although it is not addressed in this paper, it will be addressed in a future work on biased regression estimation.

APPENDIX A. ITERATIVELY REWEIGHTED LEAST SQUARES

Iteratively Reweighted Least Squares^{3,A1,A2} (IRLS) is a computational procedure to produce robust linear regression estimates. It is simply a reweighted least squares problem where the weights are functions of the scaled residuals. Consider the standard regression model

$$Y = X\beta + \epsilon \quad (\text{A-1})$$

where Y is the n vector of observations, X is the $n \times q$ matrix of independent variables, β is the q vector of coefficients, and ϵ is the n vector of random errors.

A robust estimate of β , $\hat{\beta}$ minimizes

$$\sum_{i=1}^n \rho((Y_i - X_i\beta)/S) \quad (\text{A-2})$$

with respect to β , where ρ is a robust loss function. S is a robust estimate of scale for the residuals $Y_i - X_i\beta$. Since ρ is a robust loss function, the solution to the fit is a minimization problem. Equating the first partial derivatives, with respect to the elements of β (β_j) equal to zero, is equivalent to finding the maximizing solution associated with the q equations

$$\sum_{i=1}^n X_{ij} \rho' \left(\frac{Y_i - X_i\hat{\beta}}{S} \right) = 0 \text{ for } j = 1, 2, \dots, q$$

Since

$$\begin{aligned} \frac{d(\rho((Y_i - X_i\beta)/S))}{d(\beta_j)} &= \\ \frac{d(\rho((Y_i - X_i\beta)/S))}{d((Y_i - X_i\beta)/S)} \times \frac{d((Y_i - X_i\beta)/S)}{d(\beta_j)} &= \\ = \rho' \left(\frac{Y_i - X_i\beta}{S} \right) \left(-\frac{X_{ij}}{S} \right) \end{aligned}$$

The solution to this problem is obtained by taking Gauss-Newton iterations:

$$\hat{\beta}_{k+1} = \hat{\beta}_k + S(X^T P''_k X)^{-1} X^T P'_k \quad (\text{A-3})$$

where $P'_k = [\rho'_k(\frac{Y_1 - X_1\hat{\beta}_k}{S}), \dots, \rho'_k(\frac{Y_n - X_n\hat{\beta}_k}{S})]^T$; X^T is the transpose of X and P''_k is an $n \times n$ diagonal matrix.

^{A1} A.E. Benton, J.W. Tukey, "The Fitting of Power Series Meaning Polynomials, Illustrated on Band Spectroscopic Data," *Technometrics*, V16, P. 147-185.

^{A2} L. Penly, W. A. Larson, "Robust Regression: Communications in Statistics, Theory and Methodology," A6(4), 1977, P. 335-362.

To avoid the computational expense of evaluating P''_k , approximations are often made. ROSEPACK approximates ρ'' as

$$\rho''(x) = \frac{\rho'(x)}{x}$$

The weighting function is defined as

$$w_i \left(\frac{Y_i - X_i \hat{\beta}_k}{S} \right) = \frac{\rho' \left(\frac{Y_i - X_i \hat{\beta}_k}{S} \right)}{\left(\frac{Y_i - X_i \hat{\beta}_k}{S} \right)} \quad (\text{A-4})$$

Substituting back into equation (A-3) we obtain

$$\hat{\beta}_{k+1} = \hat{\beta}_k + S (X^T W_K X)^{-1} X^T W_K \left(\frac{Y_i - X_i \hat{\beta}_k}{S} \right)$$

where

$$W_K = \left[w_1 \left(\frac{Y_1 - X_1 \hat{\beta}_k}{S} \right), \dots, w_n \left(\frac{Y_n - X_n \hat{\beta}_k}{S} \right) \right]^T.$$

This simplifies to

$$\hat{\beta}_{k+1} = (X^T W_K X)^{-1} X^T W_K Y$$

Since the usual least squares estimate is

$$\hat{\beta}_{LS} = (X^T X)^{-1} X^T Y$$

at each iteration we are solving the weighted least squares problem:

$$W^{1/2(k)} Y = W^{1/2(k)} X \hat{\beta}_{k+1}.$$

The iteration process requires a start value for $\hat{\beta}$, $\hat{\beta}_0$, which can be obtained from ordinary least squares, or from least absolute residual regression.

Holland and Welsh³ suggest using the least absolute residual estimator as a "good starting value" and median absolute deviation as a robust estimate of scale. The iteration process is continued until convergence is obtained.

APPENDIX B. ROSEPACK

ROSEPACK (Robust Statistical Estimation Package)¹ is a system of portable FORTRAN programs and an interactive driver to solve iteratively reweighted least squares problems. The interactive driver (IRLSDR), through a set of control statements that can be set up as a batch job, governs the options available to solve the weighted least squares problem.

To facilitate the understanding of ROSEPACK, the example given in this report will be explained in detail. This job is listed in Figure B-1 and was set up on the BRL/CYBER 170 computer using SYSTEM DATA of the SENATOR editor. The first nine lines are the job control cards needed to attach ROSEPACK and access the data file. These statements are explained below and must be submitted in the following order:

100 NAME.

The job name of this run. Can be any alpha-numeric name consisting of up to fourteen characters.

110 USER (XXXX, XXXX)

CYBER user identification. Your user name and password must be supplied in the parentheses.

120 CHARGE (XXXXXX, XX)

CYBER charge card. Your account number must be supplied in the parentheses.

130 GET (ROSE = ROESG0)

Attaches executable source ROESG0 through ROSE to run ROSEPACK.

140 GET (RL = ROSELIB)

Attaches ROSEPACK library of subroutines through RL.

150 LIBRARY, RL.

Loads ROSEPACK library.

160 GET (TAPE10 = DATA)

This statement copies the data file named DATA to TAPE 10 for access to ROSEPACK. Any data file can be used by simply replacing DATA with the name of your permanent data file. Of course, this is assuming your file is stored on MFA of the CYBER system. The data file should be in fixed column format with the values of the independent variable or response being in the last column. The first line of the data file must contain information representing the number of observations, the number of independent variables, and the acceptable rank tolerance level of the data matrix. A matrix whose determinant is less than the rank tolerance level is considered to be singular.

100	NAME.	360	03
110	USER(XXXX,XXXX)	370	MODE
120	CHARGE(XXXXX,XX)	380	1
130	GET(ROSE=ROESGO)	390	1HMA
140	GET(RL=ROSELIB)	400	1
150	LIBRARY,RL.	410	PRIN
160	GET(TAPE10=DATA)	420	STAR
170	ROSE(PL=20000)	430	2
180	*EOR	440	1WGT
190	PRCO	450	02
200	0	460	TUNI
210	PRCO	470	2.38500000
220	1	480	STAT
230	PRCO	490	3
240	2	500	STEM
250	PRCO	510	1
260	3	520	MAXI
270	PRCO	530	20
280	5	540	CONV
290	PRCO	550	1
300	6	560	ALGO
310	PRCO	570	1
320	7	580	OPTI
330	PRCO	590	ITER
340	4	600	QUIT
350	STEP		

Figure B-1. Sample Program.

A rank tolerance of $-1.0 E0$ was used for the example. The format of this line is 2I5, E16.7. The data file used for this report is listed in Figure B-2.

The ROSEPACK subroutine, MATRDR, which reads the input data file, excluding the first line, must be modified to comply with the format of the input data. A short program named ROS, which is available as a public file on the front end of the BRL/CYBER system, was written to make this change. The program is listed in Figure B-3. In using this program to make the necessary format change, one should do the following before submitting the ROSEPACK program:

- 1) Attach ROS in IAF mode
GET(ROS/UN=PUBLIC)
- 2) Enter Senator
SENATOR
- 3) Attach ROS in Senator mode
OLD,/ROS
- 4) Retype lines 110 and 120 of ROS. Supply your name, password and account number. Be sure you retype the line number.
- 5) Retype line 230 of ROSE to comply with the format of the dependent and independent variable. Use standard FORTRAN format statements. If additional lines are needed use line 231, etc.
- 6) List, and check for errors
LIST
- 7) Submit program
SUBMIT
- 8) Exit Senator mode
END

170 ROSE (PL=2000)

This statement executes ROSEPACK and allows for a maximum of 2000 pages.

180 EOR

Signifies the end of the job control cards and the beginning of the control statement that governs the options available to solve the weighted least squares problem and option output information.

190 PRCO

200 0

100	20	2	-1.0E0
110	1.011.5	4628.4	
120	1.011.5	4430.1	
130	1.011.6	4773.7	
140	1.011.6	4459.7	
150	1.011.7	4493.1	
160	1.011.7	4466.8	
170	1.011.8	4499.0	
180	1.011.8	4469.2	
190	1.011.9	4470.3	
200	1.011.9	4482.0	
210	1.012.0	4514.0	
220	1.012.0	4503.5	
230	1.012.1	4502.8	
240	1.012.1	4520.7	
250	1.012.2	4565.9	
260	1.012.2	4539.4	
270	1.012.3	4575.4	
280	1.012.3	4551.2	
290	1.012.4	4576.5	
300	1.012.4	4419.8	

Figure B-2. Data.

```
100 ROSE.
110 USER (XXX, XXXX)
120 CHARGE (XXXX, XX)
130 GET (DLDPL=ROSEPL)
140 UPDATE (Q)
150 FTN (I=COMPILE)
160 REPLACE (LGO=ROSESGO)
170 *EOR
180 *IDENT JOCK03
190 *COMPILE IRLSDR, MATOA, MATOB, MATRDR
200 *INSERT MATRDR .117
210     M=MSAVE
220 *DELETE MATRDR.125
230 9002 FORMAT(F3.1, F4.1, X, F6.1)
240 *DELETE MATOA.60
250     CALL MATRDR(MM, NN, M, N, A, SIGDIG)
260 *DELETE MATOB.60
270     CALL MATRDR(MM, NN, M, N, B, SIGDIG)
```

Figure B-3. ROS Format Update.

The PRCO command with its associated number controls the print control vector that governs the optional output. The command should be one of the first ROSEPACK control statements specified and should be repeated for each output option desired followed by its appropriate option number. The PRCO command followed by 0 acts as an on-off switch. The first time these two commands are encountered, the print vector is turned on. The second time these commands are read the print vector is turned off. The remaining print vector options are summarized in Table B-1.

350 STEP

360 03

The STEP command followed by a positive integer specifies the number of iterations between the printing of intermediate results which must be specified by MODE command.

370 MODE

380 1

Used to request printing of intermediate results. Unless specified, there will be no printing of intermediate results.

390 IHMA

400 1

Specifies the computing and printing of the diagonal of the hat matrix.

410 PRINT

Allows printing of the desired output specified by the print options.

420 STAR

430 2

Allows the user to choose the type of start used to calculate initial estimates of the regression coefficients. There are three starts available to the user that are specified by option number 1, 2 or 3. They are summarized below:

- 1) Least Absolute Residual Start
- 2) Least Squares Start using singular value decomposition
- 3) Least Squares Start using the householder algorithm (QR)

440 IWGT

450 02

TABLE B-1. PRCO PRINT OPTIONS

<u>CONTROL STATEMENT</u>	<u>COMMAND</u>
PRCO	
0	Initiates Print vector for following settings
1	Sets solution vector
2	Sets residuals, weighted diagonal elements of hat matrix
3	Sets convergence level
4	Sets original data matrix
5	Sets singular values of singular value decomposition
6	Sets the alpha associated with QR decomposition
7	Sets upper triangular matrix of QR decomposition

The IWGT command and associated option number allows the user to select one of eight weight functions or a user defined weight function. The different options for this command are summarized in Table B-2.

460 TUNI

470 2.38500000

The TUNI command allows the user to specify a tuning constant different from the default tuning constant associated with each weight function. The default tuning constants are given in Table B-2.

480 STAT

490 3

Specifies the printing of various statistics outlined in Table B-3.

500 STEM

510 1

The STEM command with option number 1 specifies a Stem and Leaf representation of the independent variable.

520 MAX I

530 20

Specifies the maximum number of iterations per 'ITER' command if convergence is not reached.

540 CONV

550 1

The CONV command allows convergence checking after each iteration.

560 ALGO

570 1

The computational algorithm used during each iteration is specified by this command. A zero specifies the householder algorithm. A one represents the Singular Value Decomposition algorithm.

580 OPTI

Prints list of options in effect.

590 ITER

TABLE B-2. WEIGHT FUNCTIONS AND ASSOCIATED TUNING CONSTANT

<u>COMMAND</u>	<u>WEIGHT</u>	<u>DEFAULT TUNING CONSTANT</u>
IWGT		
0	Andrews	1.339
1	Bi Square	4.685
2	Cauchy	2.385
3	Fair	1.345
4	Huber	1.400
5	Logistic	1.205
6	Talwar	2.795
7	Welsch	2.985
8	User Supplied	-
9	Previously Defined	-

TABLE B-3. STATISTIC OUTPUT OPTIONS GOVERNED BY STAT COMMAND

<u>Option Number</u>	<u>Statistic</u>
0	Number of Observations Number of Variables Sum of Weights Condition Number of Data Matrix Maximum Residual Minimum Weight Maximum Diagonal Entry of $X^T(X^T X)^{-1}X$
1	All of Option 0 Sum of Square Residuals Weighted Sum of Square Residuals Sum of Absolute Residuals
2	All of Options 0 and 1 R-Square Weighted R-Square Standard Error Weighted Standard Error
3	All of Options 1, 2, and 3 F Statistic Weighted F Statistic

Initiates the Iteratively Reweighted Least Squares Process until convergence or the maximum number of iterations is reached. May be repeated with the same or different options.

600 QUIT

Exit from ROSEPACK

Summarizing, we have demonstrated how to run ROSEPACK on the BRL/CYBER system. Our sample run performed Iteratively Reweighted Least Squares using the Cauchy weight function with its default tuning constant. A maximum of 20 iterations have been requested with intermediate results being printed out after every third step. All of the print options and statistics have been requested.

DISTRIBUTION LIST

<u>No. of</u> <u>Copies</u>	<u>Organization</u>	<u>No. of</u> <u>Copies</u>	<u>Organization</u>
12	Administrator Defense Technical Info Center ATTN: DTIC-DDA Cameron Station Alexandria, VA 22314	1	Commander US Army Communications Rsch and Development Command ATTN: DRDCO-PPA-SA Fort Monmouth, NJ 07703
1	Commander US Army Materiel Development and Readiness Command ATTN: DRCMDM-ST 5001 Eisenhower Avenue Alexandria, VA 22333	1	Commander US Army Electronics Research and Development Command Technical Support Activity ATTN: DELSD-L Fort Monmouth, NJ 07703
1	Commander US Army Armament Research and Development Command ATTN: DRDAR-TDC Dover, NJ 07801	1	Commander US Army Missile Command ATTN: DRSMI-R Redstone Arsenal, AL 35898
2	Commander US Army Armament Research and Development Command ATTN: DRDAR-TSS Dover, NJ 07801	1	Commander US Army Missile Command ATTN: DRSMI-YDL Redstone Arsenal, AL 35898
1	Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-LEP-L, Tech Lib Rock Island, IL 61299	1	Commander US Army Tank Automotive Rsch and Development Command ATTN: DRDTA-UL Warren, MI 48090
1	Director US Army ARRADCOM Benet Weapons Laboratory ATTN: DRDAR-LCB-TL Watervliet, NY 12189	1	Director US Army TRADOC Systems Analysis Activity ATTN: ATAA-SL, Tech Lib White Sands Missile Range NM 88002
1	Commander US Army Aviation Research and Development Command ATTN: DRDAV-E 4300 Goodfellow Blvd. St. Louis, MO 63120	1	Commander US Army Research Office ATTN: DRXRO-MA, Dr. Robert Launer P.O. Box 12211 Research Triangle Park, NC 27709
1	Director US Army Air Mobility Research and Development Laboratory Ames Research Center Moffett Field, CA 94035	1	Commander US Army Combat Development Experimentation Command ATTN: ATEC-SA, Dr. Marion R. Bryson Fort Ord, CA 93941

DISTRIBUTION LIST

Aberdeen Proving Ground

Dir, USAMSAA
ATTN: DRXSY-D
DRXSY-MP, H. Cohen

Cdr, USATECOM
ATTN: DRSTE-TO-F

Dir, USACSL, Bldg. E3516, EA
ATTN: DRDAR-CLB-PA

USER EVALUATION OF REPORT

Please take a few minutes to answer the questions below; tear out this sheet, fold as indicated, staple or tape closed, and place in the mail. Your comments will provide us with information for improving future reports.

1. BRL Report Number _____

2. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which report will be used.)

3. How, specifically, is the report being used? (Information source, design data or procedure, management procedure, source of ideas, etc.) _____

4. Has the information in this report led to any quantitative savings as far as man-hours/contract dollars saved, operating costs avoided, efficiencies achieved, etc.? If so, please elaborate.

5. General Comments (Indicate what you think should be changed to make this report and future reports of this type more responsive to your needs, more usable, improve readability, etc.) _____

6. If you would like to be contacted by the personnel who prepared this report to raise specific questions or discuss the topic, please fill in the following information.

Name: _____

Telephone Number: _____

Organization Address: _____

