FIT POWER

PURPOSE

The FIT POWER command can be used in conjunction with the PRE-FIT command to assess the sensitivity of the parameter estimates to the fitting criterion (for example, least squares and L_1 fits) or to provide good starting values for L_p fits.

DESCRIPTION

Least squares estimates are optimal if the errors follow the normal distribution. However, if the errors are non-normal or if there are outliers, then the least squares estimates can be unduly affected. In this case, criterion other than least squares may be desirable.

One approach to this problem is called L_p regression. This method is based on minimizing:

$$\left(\sum_{i=1}^{n}\left|Y_i - \hat{Y}\right|^p\right)^{1/p} \qquad \textbf{(EQ 5-58)}$$

where p determines the fitting criterion (this is the p in L_p). When p is 2, this reduces to standard least squares. When p is 1, it reduces to least absolute deviations (LAD) fitting. LAD fitting provides good protection against outliers, but it is not as efficient as least squares estimation when the errors are well behaved. The value of p is generally chosen to be between 1 and 2 inclusive with least squares and LAD fitting being by far the most common.

The PRE-FIT command evaluates:

$$\left(\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}\right|^p}{n - np}\right)^{1/p} \qquad \textbf{(EQ 5-59)}$$

(where np is the number of parameters in the fit) over a lattice of user-specified parameter values. That point in the lattice for which the summation is smallest will be the “least power” fit over the lattice.

A practical method for determining the sensitivity of the estimated parameters to the goodness of fit criterion is:

1. Use the PRE-FIT command with power 2 (this is the default or it can be explicitly specified via FIT POWER 2) to determine the least squares estimates over the lattice. Several iterations can be used to determine a finer grid.
2. Use the parameter values from the PRE-FIT to generate the least squares fit via the FIT command. Perform the usual residual analysis to determine the adequacy of the model, validity of assumptions, and the existence of outliers.
3. Check to see how much of a difference there is between the obtained least squares solution and the solution that would result from using a different power. Enter FIT POWER <value> and rerun the PRE-FIT command in the region surrounding the least squares solution. Even though the PRE-FIT is restricted to the lattice, it nevertheless is strongly suggestive of the true magnitude of such differences.

The FIT command is currently limited to a power of 2 (i.e., least squares estimates only). However, if the analyst determines that an L_p fit would be appropriate, it can be estimated using a technique called iteratively weighted least squares (or IRLS). See the documentation for the WEIGHTS command (in the Support chapter) for a description of IRLS. This technique is demonstrated in the program example below. The FIT POWER and PRE-FIT commands can be used to obtain good starting values for the IRLS algorithm. More efficient methods for L_1 (i.e., least absolute deviations) regression are typically based on linear programming methods. Although DATAPLOT does have a SIMPLEX SOLUTION command for solving linear programming problems, it can be used for L_1 estimation in DATAPLOT only for small n, so the IRLS algorithm is recommended.

SYNTAX

FIT POWER <value>

where <value> is a number or parameter (typically between 1 and 2) that specifies the desired fit power.

EXAMPLES

FIT POWER 1
FIT POWER 2
FIT POWER 1.5

NOTE 1

The FIT POWER and PRE-FIT commands can be applied to both linear and non-linear fits. The IRLS algorithm relies only on the ability to do weighted fits, which the FIT command provides for both linear and non-linear fits.

NOTE 2

Iteratively re-weighted least squares is used for other approaches to robust fitting as well. See the documentation for the WEIGHTS command for details on how to do this in DATAPLOT.

NOTE 3

The PRE-FIT command is primarily used to obtain good starting values for non-linear fits.

NOTE 4

The macro in the program example below uses the Tukey weighting scheme for estimating L_p fits with IRLS. The IRLS weights go to infinity as the residuals go to zero. Tukey's method assigns a weight of 1 to sufficiently small residuals. In effect, it performs least squares fitting to small residuals and least absolute deviation fitting to moderate and large residuals.

DEFAULT

The fit power is 2 (i.e., least squares).

SYNONYMS

None

RELATED COMMANDS

PRE-FIT	=	Determine a fit solution over a lattice of parameter values.
FIT	=	Carries out a least squares fit.
FIT ITERATIONS	=	Set the maximum iterations for a fit command.
FIT STANDARD DEVIATIONS	=	Set the minimum standard deviation for a fit command.
BIWEIGHT	=	Compute a bi-weight estimate for robust fitting.
WEIGHTS	=	Set the weights for a fit command.
PRED	=	A variable where predicted values are stored.
RES	=	A variable where residuals are stored.
RESSD	=	A parameter where the residual standard deviation is stored.
RESDF	=	A parameter where the residual degrees of freedom is stored.
REPSD	=	A parameter where the replication standard deviation is stored.
REPDF	=	A parameter where the replication degrees of freedom is stored.
LOFCDF	=	A parameter where the lack of fit cdf is stored.

REFERENCES

"Statistical Computing," Kennedy and Gentle, Marcel Dekker, 1980 (chapter 11).

"Elements of Statistical Computing," Thisted, Chapman and Hall, 1988 (section 3.12).

"Data Analysis and Regression," Mosteller and Tukey, Addison-Wesley, 1977.

APPLICATIONS

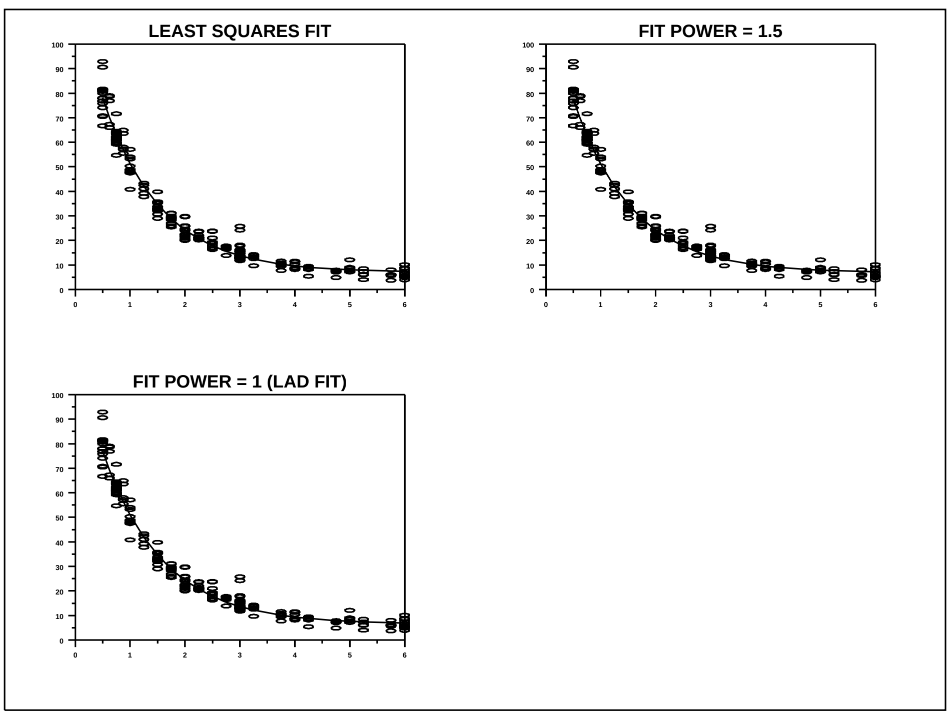
Robust Fitting

IMPLEMENTATION DATE

Pre-1987

PROGRAM 1

```
. DAN CHWIRUT ULTRASONIC REFERENCE BLOCK ANALYSIS
. PERFORM A PRE-FIT AND THEN A NON-LINEAR REGRESSION
SKIP 25
READ CHWIRUT1.DAT Y X
.
MULTIPLOT 2 2; MULTIPLOT CORNER COORDINATES 0 0 100 100
LINE BLANK SOLID; CHARACTER CIRCLE
CHARACTER HW 1.5 1.1; CHARACTER FILL OFF
.
LET FUNCTION G = A + B*EXP(-C*X)
PRE-FIT Y=G FOR A = 0 10 50 FOR B = 100 20 200 FOR C = 0.5 0.5 2
LET STRING F = FIT Y = G
^F
TITLE LEAST SQUARES FIT
PLOT Y PRED VS X
LET P = 1.5
FIT POWER P
PRE-FIT Y=G FOR A = 0 10 50 FOR B = 100 20 200 FOR C = 0.5 0.5 2
CALL LAD.DP
TITLE FIT POWER = 1.5
PLOT Y PRED VS X
LET P = 1
FIT POWER P
PRE-FIT Y=G FOR A = 0 10 50 FOR B = 100 20 200 FOR C = 0.5 0.5 2
CALL LAD.DP
TITLE FIT POWER = 1 (LAD FIT)
PLOT Y PRED VS X
END OF MULTIPLOT
```



The following output is generated (only the last iteration in the fit cycle is printed):

```
      ***************************************************************************
      **  PRE-FIT Y=G   FOR A = 0 10 50 FOR B = 100 20 200 FOR C = 0.5 0.5 2  **
      ***************************************************************************


LEAST SQUARES NON-LINEAR PRE-FIT
      SAMPLE SIZE N =      214
      MODEL--Y=(A + B*EXP(-C*X))
      REPLICATION CASE
      REPLICATION STANDARD DEVIATION =     0.3281762600D+01
      REPLICATION DEGREES OF FREEDOM =         192
      NUMBER OF DISTINCT SUBSETS     =          22

      LATTICE VALUES FOR A        =   0.0000000E+00  0.1000000E+02  0.5000000E+02
      LATTICE VALUES FOR B        =   0.1000000E+03  0.2000000E+02  0.2000000E+03
      LATTICE VALUES FOR C        =   0.5000000E+00  0.5000000E+00  0.2000000E+01

      NUMBER OF LATTICE POINTS    =         144

  STEP                  RESIDUAL  *  PARAMETER
 NUMBER                 STANDARD  *  ESTIMATES
                        DEVIATION *
----------------------------------*-----------
    1--               0.89193E+01 * 0.00000E+00 0.10000E+03 0.50000E+00
    6--               0.81236E+01 * 0.00000E+00 0.12000E+03 0.10000E+01
   10--               0.67238E+01 * 0.00000E+00 0.14000E+03 0.10000E+01
   26--               0.49619E+01 * 0.10000E+02 0.10000E+03 0.10000E+01
   30--               0.44007E+01 * 0.10000E+02 0.12000E+03 0.10000E+01

       FINAL PARAMETER ESTIMATES
       1  A                    10.0000
       2  B                    120.000
       3  C                    1.00000

      RESIDUAL    STANDARD DEVIATION =         4.4007382393
      RESIDUAL    DEGREES OF FREEDOM =         211
      REPLICATION STANDARD DEVIATION =         3.2817625999
      REPLICATION DEGREES OF FREEDOM =         192
      LACK OF FIT F RATIO =  1327.1779 = THE 100.0000% POINT OF THE
      F DISTRIBUTION WITH     19 AND    192 DEGREES OF FREEDOM


      *********************************
      **  FIT Y = (A + B*EXP(-C*X))  **
      *********************************


LEAST SQUARES NON-LINEAR FIT
      SAMPLE SIZE N =      214
      MODEL--Y =(A + B*EXP(-C*X))
      REPLICATION CASE
      REPLICATION STANDARD DEVIATION =     0.3281762600D+01
      REPLICATION DEGREES OF FREEDOM =         192
      NUMBER OF DISTINCT SUBSETS     =          22
```

```
ITERATION  CONVERGENCE  RESIDUAL  *  PARAMETER
 NUMBER      MEASURE    STANDARD  *  ESTIMATES
                        DEVIATION *
----------------------------------*-----------
    1--  0.10000E-01  0.44007E+01 * 0.10000E+02 0.12000E+03 0.10000E+01
    2--  0.50000E-02  0.33695E+01 * 0.70832E+01 0.11387E+03 0.94537E+00
    3--  0.25000E-02  0.33691E+01 * 0.69854E+01 0.11389E+03 0.94192E+00

        FINAL PARAMETER ESTIMATES           (APPROX. ST. DEV.)    T VALUE
       1  A                    6.98114       (0.5187    )          13.
       2  B                    113.877       ( 1.753    )          65.
       3  C                   0.941724       (0.2553E-01)          37.

      RESIDUAL    STANDARD DEVIATION =         3.3691074848
      RESIDUAL    DEGREES OF FREEDOM =         211
      REPLICATION STANDARD DEVIATION =         3.2817625999
      REPLICATION DEGREES OF FREEDOM =         192
      LACK OF FIT F RATIO =       1.5990 = THE  94.0437% POINT OF THE
      F DISTRIBUTION WITH     19 AND    192 DEGREES OF FREEDOM

      *******************
      **  FIT POWER P  **
      *******************


THE FIT POWER HAS JUST BEEN SET TO   0.1500000E+01

      ***************************************************************************
      **  PRE-FIT Y=G   FOR A = 0 10 50 FOR B = 100 20 200 FOR C = 0.5 0.5 2  **
      ***************************************************************************


NON-LINEAR PRE-FIT (FIT POWER =    1.50000)
      SAMPLE SIZE N =      214
      MODEL--Y=(A + B*EXP(-C*X))
      REPLICATION CASE
      REPLICATION STANDARD DEVIATION =     0.3281762600D+01
      REPLICATION DEGREES OF FREEDOM =         192
      NUMBER OF DISTINCT SUBSETS     =          22

      LATTICE VALUES FOR A        =   0.0000000E+00  0.1000000E+02  0.5000000E+02
      LATTICE VALUES FOR B        =   0.1000000E+03  0.2000000E+02  0.2000000E+03
      LATTICE VALUES FOR C        =   0.5000000E+00  0.5000000E+00  0.2000000E+01

      NUMBER OF LATTICE POINTS    =         144

  STEP                  RESIDUAL  *  PARAMETER
 NUMBER                   NORM    *  ESTIMATES
                                  *
----------------------------------*-----------
    1--               0.49150E+01 * 0.00000E+00 0.10000E+03 0.50000E+00
    6--               0.46797E+01 * 0.00000E+00 0.12000E+03 0.10000E+01
   10--               0.40237E+01 * 0.00000E+00 0.14000E+03 0.10000E+01
   26--               0.29583E+01 * 0.10000E+02 0.10000E+03 0.10000E+01
   30--               0.28311E+01 * 0.10000E+02 0.12000E+03 0.10000E+01
```

```
 FINAL PARAMETER ESTIMATES
 1  A                    10.0000
 2  B                    120.000
 3  C                    1.00000

RESIDUAL    NORM               =         2.8310678005
RESIDUAL    DEGREES OF FREEDOM =         211
REPLICATION STANDARD DEVIATION =         3.2817625999
REPLICATION DEGREES OF FREEDOM =         192
LACK OF FIT F RATIO =   210.2156 = THE 100.0000% POINT OF THE
F DISTRIBUTION WITH     19 AND    192 DEGREES OF FREEDOM

*************************************
**      FIT Y = (A + B*EXP(-C*X))  **
*************************************


LEAST SQUARES NON-LINEAR FIT
      SAMPLE SIZE N =      214
      MODEL--Y =(A + B*EXP(-C*X))
      REPLICATION CASE
      REPLICATION STANDARD DEVIATION =     0.3281762600D+01
      REPLICATION DEGREES OF FREEDOM =         192
      NUMBER OF DISTINCT SUBSETS     =          22

ITERATION  CONVERGENCE  RESIDUAL  *  PARAMETER
 NUMBER      MEASURE    STANDARD  *  ESTIMATES
                        DEVIATION *
----------------------------------*-----------
    1--  0.10000E-01  0.12610E+02 * 0.67890E+01 0.11297E+03 0.15537E+01
    2--  0.50000E-02  0.10702E+02 * 0.90247E+01 0.92084E+02 0.56341E+00
    3--  0.25000E-02  0.40178E+01 * 0.10002E+02 0.10229E+03 0.87426E+00
    4--  0.12500E-02  0.26625E+01 * 0.67667E+01 0.11257E+03 0.93293E+00
    5--  0.62500E-03  0.26605E+01 * 0.67873E+01 0.11293E+03 0.93332E+00

       FINAL PARAMETER ESTIMATES          (APPROX. ST. DEV.)    T VALUE
       1  A                    6.78763       (0.4212    )          16.
       2  B                    112.927       ( 1.481    )          76.
       3  C                   0.933333       (0.2096E-01)          45.

      RESIDUAL    STANDARD DEVIATION =         2.6605234146
      RESIDUAL    DEGREES OF FREEDOM =         211
      REPLICATION STANDARD DEVIATION =         3.2817625999
      REPLICATION DEGREES OF FREEDOM =         192
***** NON-FATAL DIAGNOSTIC--THE 1ST INPUT ARGUMENT TO THE FCDF   SUBROUTINE IS NEGATIVE
*****
***** THE VALUE OF THE ARGUMENT IS -0.28065093E+01 *****
      LACK OF FIT F RATIO =      -2.8065 = THE   0.0000% POINT OF THE
      F DISTRIBUTION WITH     19 AND    192 DEGREES OF FREEDOM


      *******************
      **  FIT POWER P  **
      *******************
```

```
THE FIT POWER HAS JUST BEEN SET TO   0.1000000E+01

      *************************************************************************
      **  PRE-FIT Y=G   FOR A = 0 10 50 FOR B = 100 20 200 FOR C = 0.5 0.5 2  **
      *************************************************************************


NON-LINEAR PRE-FIT (FIT POWER =    1.00000)
      SAMPLE SIZE N =      214
      MODEL--Y=(A + B*EXP(-C*X))
      REPLICATION CASE
      REPLICATION STANDARD DEVIATION =     0.3281762600D+01
      REPLICATION DEGREES OF FREEDOM =         192
      NUMBER OF DISTINCT SUBSETS     =          22

      LATTICE VALUES FOR A        =   0.0000000E+00  0.1000000E+02  0.5000000E+02
      LATTICE VALUES FOR B        =   0.1000000E+03  0.2000000E+02  0.2000000E+03
      LATTICE VALUES FOR C        =   0.5000000E+00  0.5000000E+00  0.2000000E+01

      NUMBER OF LATTICE POINTS    =         144

  STEP                  RESIDUAL  *  PARAMETER
 NUMBER                   NORM    *  ESTIMATES
                                  *
----------------------------------*-----------
    1--               0.27678E+01 * 0.00000E+00 0.10000E+03 0.50000E+00
    6--               0.27479E+01 * 0.00000E+00 0.12000E+03 0.10000E+01
   10--               0.24647E+01 * 0.00000E+00 0.14000E+03 0.10000E+01
   26--               0.18884E+01 * 0.10000E+02 0.10000E+03 0.10000E+01

       FINAL PARAMETER ESTIMATES
       1  A                    10.0000
       2  B                    100.000
       3  C                    1.00000

      RESIDUAL    NORM               =         1.8884271383
      RESIDUAL    DEGREES OF FREEDOM =         211
      REPLICATION STANDARD DEVIATION =         3.2817625999
      REPLICATION DEGREES OF FREEDOM =         192
      LACK OF FIT F RATIO =    26.5459 = THE 100.0000% POINT OF THE
      F DISTRIBUTION WITH     19 AND    192 DEGREES OF FREEDOM


      *************************************
      **      FIT Y = (A + B*EXP(-C*X))  **
      *************************************


LEAST SQUARES NON-LINEAR FIT
      SAMPLE SIZE N =      214
      MODEL--Y =(A + B*EXP(-C*X))
      REPLICATION CASE
      REPLICATION STANDARD DEVIATION =     0.3281762600D+01
      REPLICATION DEGREES OF FREEDOM =         192
      NUMBER OF DISTINCT SUBSETS     =          22
```

```
 ITERATION  CONVERGENCE  RESIDUAL  *  PARAMETER
  NUMBER      MEASURE    STANDARD  *  ESTIMATES
                         DEVIATION *
 ----------------------------------*-----------
     1--  0.10000E-01  0.10492E+02 * 0.65145E+01 0.11151E+03 0.14935E+01
     2--  0.50000E-02  0.89452E+01 * 0.84801E+01 0.91384E+02 0.56841E+00
     3--  0.25000E-02  0.30828E+01 * 0.94650E+01 0.10186E+03 0.86728E+00
     4--  0.12500E-02  0.17941E+01 * 0.65044E+01 0.11121E+03 0.91717E+00
     5--  0.62500E-03  0.17927E+01 * 0.65120E+01 0.11145E+03 0.91666E+00

         FINAL PARAMETER ESTIMATES           (APPROX. ST. DEV.)    T VALUE
        1  A                  6.51158       (0.3069    )          21.
        2  B                  111.448       ( 1.137    )          98.
        3  C                 0.916640       (0.1551E-01)          59.

       RESIDUAL    STANDARD DEVIATION =         1.7926622629
       RESIDUAL    DEGREES OF FREEDOM =         211
       REPLICATION STANDARD DEVIATION =         3.2817625999
       REPLICATION DEGREES OF FREEDOM =         192
 ***** NON-FATAL DIAGNOSTIC--THE 1ST INPUT ARGUMENT TO THE FCDF   SUBROUTINE IS NEGATIVE
*****
 ***** THE VALUE OF THE ARGUMENT IS -0.67915759E+01 *****
       LACK OF FIT F RATIO =      -6.7916 = THE   0.0000% POINT OF THE
       F DISTRIBUTION WITH     19 AND    192 DEGREES OF FREEDOM
```

PROGRAM 2 (this is the LAD.DP macro used by PROGRAM 1)

```
. This macro computes Least Absolute Deviations (LAD) and other Lp fits for powers
. between 1 and 2 using iteratively re-weighted least squares. The following assumes
. that a string F has been defined before calling this macro to define the type of fit. E.g.,
.       LET STRING F = FIT Y X
. The power of the fit is specified with the following command before calling this macro:
.       LET P = <VALUE>
. The convergence criterion and the maximum number of iterations can also be modified.
.
. The following algorithm is used:
.       a) Perform an initial unweighted least squares estimate
.       b) Scale the residuals (where s = Median Absolute Deviation/0.6745):
.               Ui = Ei/s
.       c) Apply the specified weight function
.       d) Check for convergence
.
WEIGHT
^F
IF P > 2
    LET P = 2
END OF IF
IF P < 1
    LET P = 1
END OF IF
LET MAXITER = 10
LOOP FOR K = 1 1 MAXITER
    LET RESOLD = RES
    LET MED = MEDIAN RES
    LET TEMP = ABS(RES - MED)
    LET MAD = MEDIAN TEMP
    LET S = MAD/0.6745
    LET U = RES/S
    LET TEMP = ABS(RES)
    LET C = MEDIAN TEMP
    LET TAG = ABS(Y - PRED)**(2-P)
    LET WT = C/TAG SUBSET TAG > C
    LET WT = 1 SUBSET TAG <= C
    WEIGHTS WT
    ECHO ON; FEEDBACK ON
    ^F
    .
    LET DELTA = (RESOLD - RES)**2
    LET NUM = SUM DELTA
    LET NUM = SQRT(NUM)
    LET DELTA2 = RESOLD*RESOLD
    LET DENOM = SUM DELTA2
    LET CONV = NUM/DENOM
    IF CONV <= 0.0001
        BREAK LOOP
    END OF IF
END OF LOOP
```