A Tutorial Introduction to Robust Estimators with Mathematical Programming Solutions

Malcolm R. Irving

malcom.irving@brunel.ac.uk
This paper can be downloaded from http://bura.brunel.ac.uk

Brunel University, Uxbridge, UB8 3PH, UK

1. Introduction

Estimators are used when unknown states (or parameters) in a given mathematical model must be determined from available measurements. Usually, there are more measurements than are strictly needed to define the unknowns and the problem is called over-determined. This type of problem is variously referred to as state estimation, parameter estimation, multivariate regression, and curve fitting. All these terms essentially describe the same computational process. This paper illustrates and explains some robust state estimation methods and introduces original formulations that allow solutions to be obtained using general-purpose mathematical programming algorithms. Small-scale examples are used to illustrate the properties of these methods. All the examples are linear.

If the measurements have normally-distributed errors, the method of Least Squares (LS), or more generally Weighted Least Squares (WLS), provides an optimal solution. However, if some of the measurements are statistical outliers (i.e. have unexpected very large errors) then the LS estimate becomes unreliable. Various robust estimators have been proposed, which try to combat this problem by processing the measurements so that the outliers have little or no effect on the estimated states. By looking at small-scale examples, it is possible to see how these methods work, and perhaps more importantly understand conditions where some algorithms may produce unexpected results. The detection of outliers and the elimination of their effects on the estimates can provide measurement fault detection and measurement fault tolerance.

Specialised algorithms have generally been developed to solve robust estimation problems. However, it has been found that all the robust estimators considered here can be solved using standard mathematical programming algorithms. Some of these formulations are original and can provide accurate solutions with good computational efficiency.

2. Least Squares Estimator

Assume we a have a multivariate linear model:

$$A\underline{x} = \underline{b} \qquad (1)$$

where A is a known (m x n) matrix
$\underline{x}$ is an unknown state vector (n x 1), and
$\underline{b}$ is an (m x 1) vector of values which can be measured

Normally we have more measurements than states, so m > n, and we expect that each measurement will include some unknown error:

$$A\underline{x} = \underline{b} + \underline{e} \qquad (2)$$

where $\underline{b}$ is a vector of known measurements
$\underline{e}$ is a vector of unknown measurement errors

The well known least squares estimate can be found by

$$\min_{\underline{x},\, \underline{e}} \quad \underline{e}^T \underline{e} \qquad (3)$$

Subject to equation (2)

In other words, we choose values for the unknown elements of $\underline{x}$ and $\underline{e}$ that give a minimum sum of squared errors.

Although there is an analytical solution to the LS problem, equations (2) and (3) can readily be solved as a linearly constrained quadratic program. To illustrate LS solutions and to provide a benchmark for comparison with the robust methods, three simple two-dimensional linear regression problems (case 1, case 2 and case 3) have been solved and are shown in figures (1) to (3), with corresponding numerical results given in tables (1) to (3). The solutions have been obtained using the general-purpose non-linear programming package MINOS by Murtagh and Saunders, available via the NEOS public-domain web-service [1]. The three problem cases are derived from an example presented by Ryan [2]. Case 1 is intended to be a straightforward problem, where all the measurements can be fitted reasonably well by a straight line. In case 2, an additional measurement is introduced which is expected to be an ‘outlier’ having an unexpectedly high measured value. This additional measurement is further altered in case 3, to have both a high measurement value and a high coefficient value in the matrix ‘A’. The relationship between the simple two-dimensional linear regression problem and the more general multivariate case of equation (2) is explained in appendix 1. In tables 1(a), 2(a) and 3(a), each row represents a measurement, with h_i being the horizontal axis co-ordinate value, b_i being the vertical axis co-ordinate value, and e_i (LS) being the residual error from the (least squares) regression line defined by the state estimates x_1 (LS) and x_2 (LS) given in tables 1(b), 2(b) and 3(b). The vertical axis intercept of the least squares regression line is x_1 (LS) and the slope of the regression line is x_2 (LS).

h_i	b_i	e_i (LS)	e_i (LAV)	e_i (LMS)
2.5	5	-0.91046	-0.0285714	-3.83
7.5	2.6	1.39866	1.94286	-0.56333
14	2.6	1.28051	1.38571	0.56333
15	3.9	-0.0376639	0.0	-0.56333
16	5.1	-1.25584	-1.28571	-1.59
23	5.1	-1.38307	-1.88571	-0.37667
29	2.7	0.907867	0.0	3.06332

Table 1(a): Measurements and Residuals using LS, LAV and LMS methods on Case 1

	(LS)	(LAV)	(LMS)
x_1	4.13498	5.18571	0.736667
x_2	-0.0181763	-0.0857143	0.173333

Table 1(b): State Estimates using LS, LAV and LMS methods on Case 1

h_i	b_i	e_i (LS)	e_i (LAV)	e_i (LMS)
2.5	5	-1.19696	0.0	-17.4
7.5	2.6	1.64229	1.96604	-8.5
14	2.6	2.21332	1.40189	-0.05
15	3.9	1.00117	0.0150943	-0.05
16	5.1	-0.110981	-1.2717	0.05
22	13	-7.48388	-9.69245	-0.05
23	5.1	0.503969	-1.87925	9.15
29	2.7	3.43107	0.0	19.35

Table 2(a): Measurements and Residuals using LS, LAV and LMS methods on Case 2

	(LS)	(LAV)	(LMS)
x_1	3.58342	5.21698	-15.65
x_2	0.08785	-0.0867925	1.3

Table 2(b): State Estimates using LS, LAV and LMS methods on Case 2

h_i	b_i	e_i (LS)	e_i (LAV)	e_i (LMS)
2.5	5	-3.03261	-3.26667	-4.32353
7.5	2.6	0.286362	0.0	-0.45294
14	2.6	1.48103	1.12667	1.45883
15	3.9	0.364822	0.0	0.45294
16	5.1	-0.651384	-1.02667	-0.45294
45	13	-3.22134	-3.9	0.17649
23	5.1	0.635178	0.186667	1.60589
29	2.7	4.13795	3.62667	5.7706

Table 3(a): Measurements and Residuals using LS, LAV and LMS methods on Case 3

	(LS)	(LAV)	(LMS)
x_1	1.5079	1.3	-0.058824
x_2	0.183795	0.173333	0.294118

Table 3(b): State Estimates using LS, LAV and LMS methods on Case 3

The least squares fit for case 1, as expected, gives a good compromise among the available measurements. The result for case 2 is still reasonable, but it is apparent that the outlier has distorted the fit slightly. In case 3, the outlier has significantly distorted the fit. In fact, this outlier has been deliberately chosen to be a 'leverage point'. These are measurements that can have an undue influence on estimators, due to their relatively greater distance from other measurements in the factor space [3]. Leverage point outliers are particularly problematic as they are hard to detect.

Further statistical processing of least squares estimates can be undertaken by examining the residuals $\underline{e}$ to test whether they follow expected error statistics. For each measurement we can test the hypothesis that the residual error is consistent with its calculated variance or otherwise. We will not pursue this approach further here, but will note in passing that in case 3 the outlier measurement does not have the largest residual.

3. Least Absolute Values Estimator

The existence of unexpectedly large residuals is associated with the presence of outliers in the measurement set. The least squares principle could be said to give outliers excessive weight by squaring the value of the residual. An alternative approach is to minimise the sum of absolute values of residuals. By taking the absolute value (or modulus) of the residual, the effect of outliers on the estimate is reduced. A property of Least Absolute Value (LAV) estimates is that at least 'n' of the measurements will be fitted exactly (with zero residuals).

An efficient algorithm for LAV estimation is via the solution of the following linear program:

$$\min_{\underline{x}, \underline{e}, \underline{f}} \quad \Sigma (e_i + f_i) \tag{4}$$

$$\text{Subject to:} \quad A\underline{x} - \underline{e} + \underline{f} = \underline{b} \tag{5}$$

$$\underline{e} \geq \underline{0} , \ \underline{f} \geq \underline{0} \tag{6}$$

where $\underline{e}$ and $\underline{f}$ are non-negative vectors of unknown measurement errors

The LAV regressions are shown together with the LS estimates in figures (1) to (3) and tables (1) to (3). In case (1) the LAV regression fits the first and last measurements exactly but has less well-balanced residuals when compared with the LS regression. For case (2), LAV is not influenced by the 'bad' measurement and it could be argued provides a better regression that the LS estimate. However, in case (3) the leverage point does influence the LAV estimate, and in this case the LS and LAV regressions are quite similar. The linear programs were solved using MINOS.

4. Least Median of Squares

Rousseeuw [4] introduced a new robust estimation principle referred to as Least Median of Squares (LMS). This is a generalisation of the idea that the median of a set of real values is a more robust estimate than the mean. For example, if we measure temperature using five different thermometers and obtain the readings 12.7, 12.5, 19.8, 12.6, 12.8, the median (12.7) is a more robust estimate than the mean (14.08). This idea is generalised to the multivariate estimator problem by finding an estimate that minimises the median of the squared residuals. Roughly speaking, the median is unaffected even if up to half of the residuals are very high. (When larger problems with more than two state variables are considered, it is customary to minimise the $(n+m+1)/2$ ordered squared residual, since non-zero residuals only exist for $m > n$.)[5] A characterisation of an LMS estimate is that it seeks a regression that minimises the value of a tolerance 't' whereby the majority of the measurements fall within tolerance. This interpretation motivates an original implementation of an LMS estimator via a mixed integer program, formulated as follows:

$$\min_{\underline{x}, \underline{k}, t} \quad t \tag{7}$$

$$\text{Subject to:} \quad \underline{b} - t - M\,\underline{k} \leq A\underline{x} \leq \underline{b} + t + M\,\underline{k} \tag{8}$$

$$k_1 + k_2 + \ldots + k_m \leq K \tag{9}$$

where $\underline{k}$ is an unknown binary integer vector (each element is either 0 or 1)
t is an unknown scalar tolerance
M is a specified arbitrary large positive scalar
K is specified as m/2 (if m is even)
(m-1)/2 (if m is odd)

In the above mixed integer linear program, the binary integer variables $\underline{k}$ allow some of the measurements to be 'switched off' or 'rejected' (in other words to be outside the tolerance 't'). The value of 'M' is chosen to be large enough so that when a measurement is switched off (by k_i being 1) the expanded tolerance '$t + M\underline{k}$' is large enough to avoid that measurement having any effect on the estimate $\underline{x}$. The specification of K, together with constraint (9), is such that a majority of the measurements cannot be switched off.

The three test case problems were solved using MINTO by Savelsbergh, Nemhauser and Linderoth [1]. Figures (1) to (3) and tables (1) to (3) show the results obtained. The regression obtained by LMS in case 1 is the result of 'rejecting' measurements 1, 5 and 7; while achieving a close fit to the remaining measurements. Case 2, based on an example from Ryan [2], includes 4 measurements that are approximately co-linear. The co-linearity here represents an 'accidental' pattern of errors, but has a dominating effect on the LMS estimate. In case 3, LMS is rejecting too many measurements and consequently fits the outlier leverage point.

Some of the problems associated with LMS are related to its property of rejecting up to half of the measurements. Generally, these rejected measurements may include some outliers but will usually include many good measurements. An obvious generalisation of LMS is to reduce the value of K in the formulation above. This would force fewer measurements to be rejected. This approach has been tried on the test cases and some good results are show for case 2 with K=3 (see table 5) and for case 3 with K=1 (table 6). We refer to these results as GLMS-K (generalised LMS with up to K rejected measurements). Table (6) shows that other choices of K do not have such a useful effect. The main drawback of this generalisation is that it is not easy to specify in advance what an appropriate value for K should be.

e_i (LTS-3)	e_i (LMR)
-3.5408	-1.5
-0.29603	1.03207
0.80216	1.20377
-0.32888	-0.06981
-1.35993	-1.2434
-0.17726	-1.05849
3.23646	1.5

Table 4(a): Residuals using LTS-3 and LMR methods on Case 1

	LTS-3	LMR
x_1	1.03682	3.43396
x_2	0.168953	0.0264151

Table 4(b): State Estimates using LTS-3 and LMR methods on Case 1

e_i (GLMS-3)	e_i (LTS-1)	e_i (LTS-2)	e_i (LTS-3)	e_i (LTS-4)	e_i (LMR)
0.91428	-0.91046	-0.23360	-3.39501	-17.4026	-1.5
2.8857	1.39866	1.91797	-0.10016	-8.49285	0.93704
2.3286	1.28051	1.59502	1.06314	-0.03024	0.98518
0.94286	-0.03766	0.24534	-0.05789	-0.0283	-0.30741
-0.34286	-1.25584	-1.00435	-1.07893	0.07364	-1.5
-8.75714	-9.2649	-9.20246	-7.90511	-0.01472	-9.35556
-0.94286	-1.38307	-1.35214	0.17386	9.18722	-1.44815
0.94286	0.90787	0.74975	3.64767	19.3989	0.996295

Table 5(a): Residuals using GLMS, LTS and LMR methods on Case 2

	GLMS-3	LTS-1	LTS-2	LTS-3	LTS-4	LMR
x_1	6.12857	4.13498	4.89061	1.15757	-15.6574	3.48148
x_2	-0.085714	-0.018176	-0.049685	0.178969	1.30194	0.0074074

Table 5(b): State Estimates using GLMS, LTS and LMR methods on Case 2

e_i (GLMS-1)	e_i (GLMS-2)	e_i (GLMS-3)	e_i (LTS-1)	e_i (LTS-2)	e_i (LTS-3)	e_i (LTS-4)	e_i (LMR)
-1.23333	-4.62222	-4.688	-0.91046	-4.67767	-3.39501	-4.03319	-1.5
1.2	-0.83333	-0.90133	1.39866	-0.83852	-0.10016	-0.22280	0.93704
1.24333	0.97222	0.90133	1.28051	1.03237	1.06314	1.61071	0.98518
-0.05	-0.05	-0.121338	-0.03766	0.020203	-0.05790	0.59279	-0.30741
-1.24333	-0.97222	-1.044	-1.25584	-0.89197	-1.07893	-0.32514	-1.5
-8.95	-0.81666	-0.90135	-9.68295	-0.44490	-3.78883	-0.04487	-9.18519
-1.19667	0.97223	0.89733	-1.38307	1.12284	0.173857	1.64941	-1.44815
1.24333	5.0389	4.96132	0.90787	5.24982	3.64767	5.74188	0.996295

Table 6(a): Residuals using GLMS, LTS and LMR methods on Case 3

	GLMS-1	GLMS-2	GLMS-3	LTS-1	LTS-2	LTS-3	LTS-4	(LMR)
x_1	3.75	-0.31666	-0.381333	4.13498	-0.39725	1.15757	0.261616	3.48148
x_2	0.006666	0.277778	0.277333	-0.018176	0.28783	0.178969	0.282078	0.0074074

Table 6(b): State Estimates using GLMS, LTS and LMR methods on Case 3

5. Least Trimmed Squares

The principle of Least Trimmed Squares (LTS), also proposed by Rousseeuw [4], is to consider the sum of squared errors for the (m-K) smallest residuals only. Equivalently, the K largest residuals are rejected and the remaining residuals are considered in a least squares objective. An original mathematical programming formulation for LTS is as follows:

$$\min_{\underline{x},\, \underline{e},\, \underline{k}} \quad \underline{e}^T \underline{e} \tag{10}$$

Subject to:

$$\underline{b} - M\,\underline{k} \;\leq\; A\underline{x} - \underline{e} \;\leq\; \underline{b} + M\,\underline{k} \tag{11}$$

$$k_1 + k_2 + \ldots + k_m \;\leq\; K \tag{12}$$

where $\underline{k}$ is an unknown binary integer vector (each element is either 0 or 1)
M is a specified arbitrary large positive scalar
K is a specified number of measurements that may be rejected
$\underline{e}$ is a vector of unknown measurement errors

In this formulation, the binary vector $\underline{k}$ allows up to K of the measurements to be switched off. At the solution, any measurement that is switched off (k_i being 1) will have its associated e_i at zero (and not contributing to the objective function). This formulation is a Mixed Integer Nonlinear Program, which can be efficiently solved via the NEOS server, using the MINLP algorithm of Fletcher and Leyffer [1].

Results for cases (1) to (3) are given in tables (4), (5) and (6). As with the GLMS-K method, good results can sometimes be obtained, depending rather crucially on the choice of K.

6. Least Measurements Rejected

Whereas the LMS method pre-determines the number of measurements to reject and then seeks a regression that minimises the tolerance on the retained measurements, a new approach has been proposed by the present author, which follows the converse principle. This has been implemented using a genetic algorithm in reference [6] and as a mathematical program in [7]. This approach requires the user to pre-specify a tolerance for each measurement and then seeks a regression that minimises the number of measurements unable to satisfy their tolerance. The tolerance value of each measurement should be chosen according to the range of error within which the measurement can still be regarded as ‘good’. For example, a temperature measurement of 12.5 ^oC might have a tolerance of $\pm$ 1.0 ^oC. This is compatible with the usual engineering approach for specifying transducer accuracy. (If required, the tolerance can be asymmetrical, e.g. + 1.0 to –1.5.) The new approach is referred to as Least Measurements Rejected (LMR).

The mathematical programming formulation for LMR is as follows:

$$\min_{\underline{x}, \underline{k}} \quad \Sigma k_i \tag{13}$$

$$\text{Subject to:} \quad \underline{b} - M\,\underline{k} - \underline{t} \le A\underline{x} \le \underline{b} + M\,\underline{k} + \underline{t} \tag{14}$$

where $\underline{k}$ is an unknown binary integer vector (each element is either 0 or 1)
M is a specified arbitrary large positive scalar
$\underline{t}$ is a specified vector of tolerances on measurement errors

As before, the binary vector $\underline{k}$ allows some measurements to be 'switched off', but in this case the solution will be a regression which maximises the number of measurements that are within tolerance (i.e. minimises the number of measurements which need to be switched off with $k_i = 1$). This formulation is a Mixed Integer Linear Program, which can be efficiently solved via the NEOS server, using the MINTO algorithm.

Results obtained using LMR for cases (1), (2) and (3) are shown in tables (4), (5) and (6) and figures (1), (2) and (3). A tolerance of $\pm$ 1.5 was used throughout. The LMR approach gives a consistent regression estimate for all three cases.

In common with the LMS, GLMS and LTS formulations, LMR is formulated here as a mixed integer program, which has traditionally been regarded as computationally intensive. However, it may be noted that mixed integer programs with thousands of variables can now be solved routinely in less than one minute [1]. Although the test cases considered here are too small to draw definitive conclusions about computational performance, it was noticed that (using MINTO) the LMR method only required the solution of 1 linear program (LP) sub-problem in each case, whereas LMS required the solution of 47, 73 and 96 LP sub-problems in cases (1), (2) and (3) respectively. Complete enumeration of the branch-and-bound tree would require the solution of 128, 256 and 256 LP sub-problems respectively.

7. Conclusions

Some robust estimation methods have been discussed. Formulation of each method as the solution of a mathematical programming problem has provided a unifying framework and in some instances offers an original solution algorithm. The small test cases considered indicate that some of the robust estimation methods can be disturbed by the presence of leverage point outliers, or spurious co-linearity of some measurements. The generalised LMS approach and the LTS approach can provide good results but are critically dependent on the user-selected criteria of how many measurements to reject. A new approach termed LMR, which in some ways is the converse of the LMS principle, automatically decides how many measurements to reject, but does require the user to specify a suitable tolerance for each measurement. By considering the characteristics of a measurement process, it is possible to selecting a robust estimation algorithm that is capable of reliably detecting measurement faults and providing valid estimates in their presence. This can be vital where measurements are processed automatically within an on-line system.

References

[1] NEOS, “NEOS sever for optimisation”, http://neos.mcs.anl.gov/neos, Feb. 2008

[2] T. P. Ryan, *Modern Regression Methods,* Wiley, 1997, ISBN 0-471-52912-5

[3] A. Abur and A. G. Exposito, *Power System State Estimation Theory and Implementation*, Marcel Dekker, 2004, ISBN 0-8247-5570-7

[4] P.J. Rousseeuw, “Least median of squares”, *Jnl. of American Statistical Association*, 79, pp 871-880, 1984

[5] A. Monticelli, *State Estimation in Electric Power Systems A Generalized Approach*, Kluwer, 1999, ISBN 0-7923-8519-5

[6] A. K. Al-Othman and M. R. Irving, “Robust state estimator based on maximum constraints satisfaction of uncertain measurements”, *Measurement*, Vol. 40, pp. 347 - 359, 2007

[7] M. R. Irving, “Robust state estimation using mixed integer programming”, *IEEE trans. Power Systems*, 23, 3, August 2008, pp 1519-1520

Appendix 1

For the case of linear regression in two-dimensions, we can write the equation of a straight line as:

$$x_1 + x_2\, h = b \qquad \text{(A1)}$$

where h is the horizontal axis co-ordinate
b is the vertical axis co-ordinate
x_1 is the intercept on the vertical axis
x_2 is the slope

If we have ‘m’ measurement points, each defined by horizontal and vertical components (h_i, b_i), then we can write a set of measurement equations:

$$x_1 + x_2\, h_i = b_i + e_i \qquad (i = 1 , \dots , m) \qquad \text{(A2)}$$

which can be written in matrix form:

$$A\, \underline{x} = \underline{b} + \underline{e} \qquad \text{(A3)}$$

Where $\underline{x} = [x_1, x_2]$
$\underline{b} = [b_1, \dots , b_m]$
$\underline{e} = [e_1, \dots , e_m]$

$$A = \begin{vmatrix} 1 & h_1 \\ : & : \\ 1 & h_m \end{vmatrix}$$

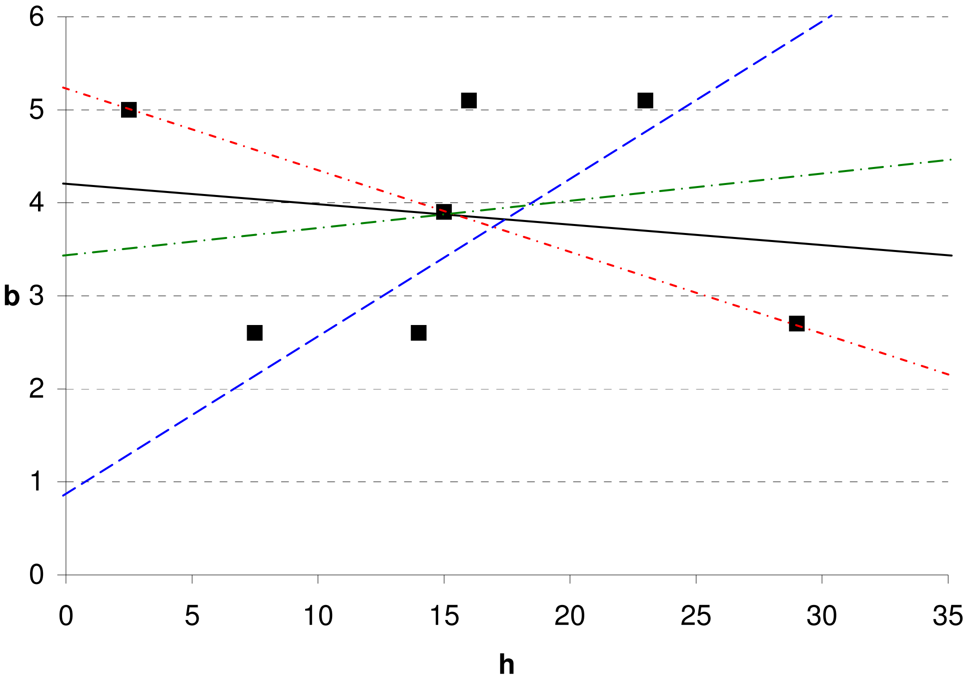
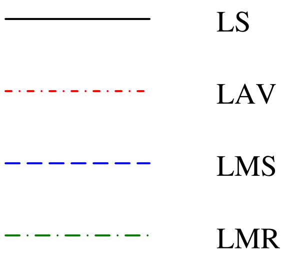


Figure 1: Selected Regressions for Case 1



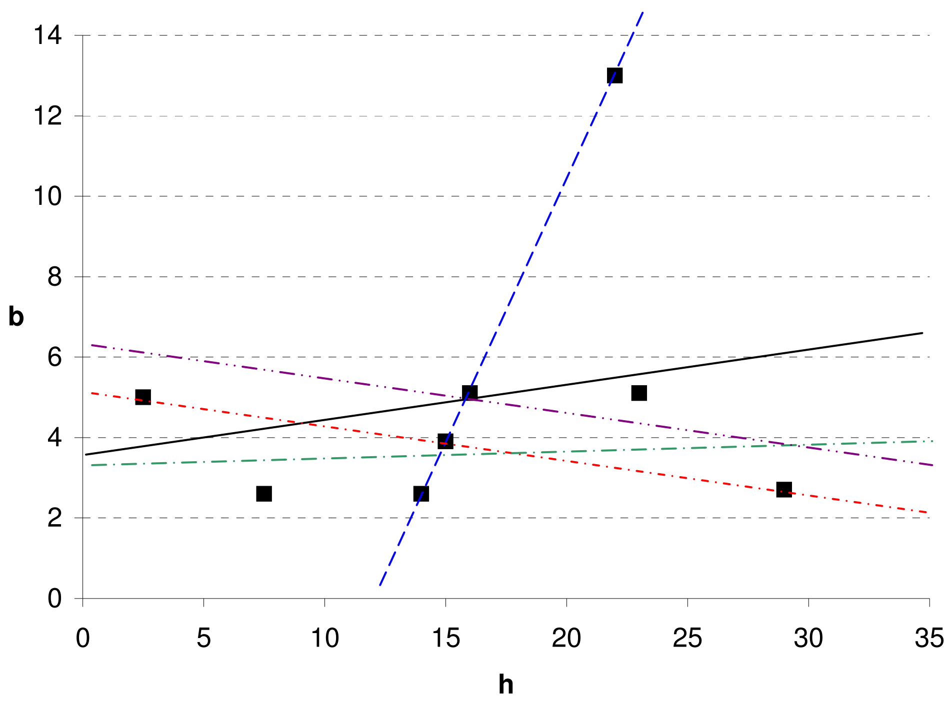
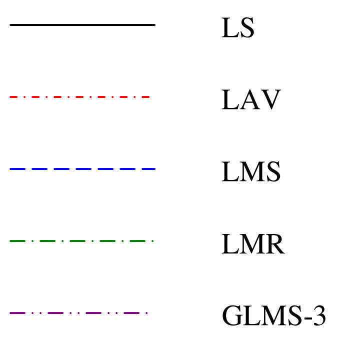


Figure 2: Selected Regressions for Case 2



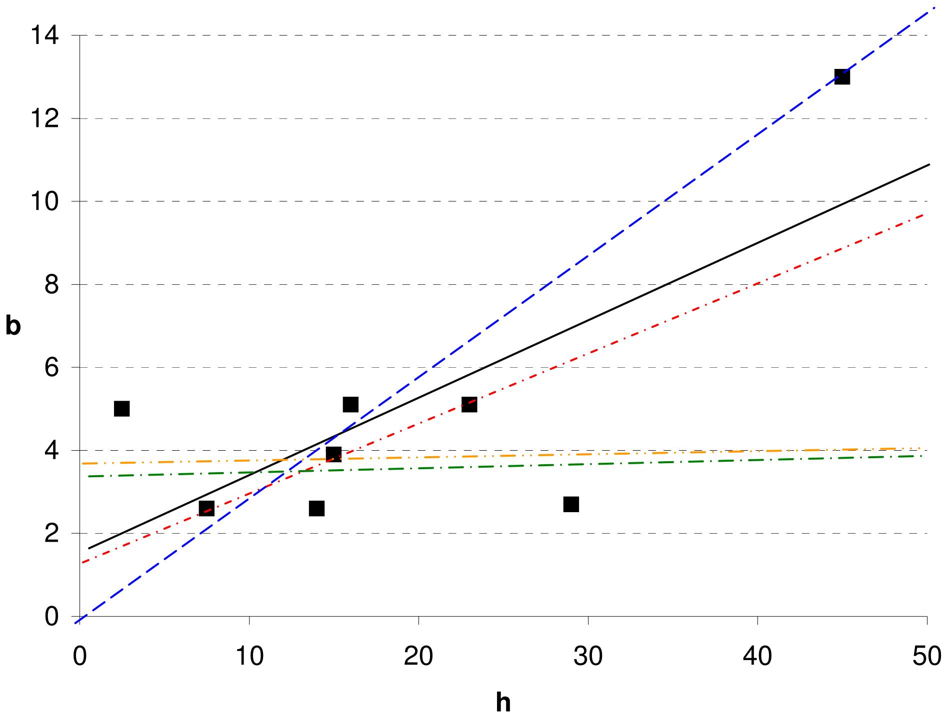


Figure 3: Selected Regressions for Case 3

LS

LAV

LMS

LMR

GLMS-1