Chapter 8

Non-linear least squares

8.1 Introduction

In chapters 6 and 7 we considered fitting equations such as $y = a + bx$ and $y = a + b \ln x + cx^2$ to data by using linear least squares. The equations we fitted were linear in the parameters, a, b, etc.

Section 6.9 showed that, where equations are not linear in the parameters, a transformation (such as taking the logarithm of the dependent variable), may allow fitting by linear least squares to proceed.

Though transforming equations can assist in many situations, there are some equations that cannot be transformed into a form suitable for fitting by linear least squares. Examples of such equations are:

$$y = a + b \sin cx \tag{8.1}$$

$$y = a + b \ln cx \tag{8.2}$$

$$y = a + b[1 - \exp cx] \tag{8.3}$$

$$y = a \exp bx + c \exp dx. \tag{8.4}$$

For equations 8.1 to 8.4 it is not possible to obtain a set of linear equations that may be solved for best estimates of the parameters. We therefore resort to another method of finding best estimates, called *non-linear* least squares. While non-linear least squares is computationally more demanding than linear least squares, the key point, as with linear least squares, is that parameter estimates are sought that minimise the sum of squares of residuals, *SSR*. Even where transformation of data allows for fitting by linear least squares to proceed, this may not be the best approach to adopt, as the transformation affects

the error distribution in the dependent variable. As a consequence of the transformation the assumption that the errors in the dependent variable are normally distributed is likely not to be valid. This leads to bias in the estimates of the parameters.[1] The adverse effect of transformation can be avoided by fitting equations using non-linear least squares, as no transformation of raw data is required.

SSR may be considered to be a continuous function of the parameter estimates. A surface may be constructed, sometimes referred to as a hypersurface[2] in M dimensional space, where M is the number of parameters appearing in the equation to be fitted to data. As with linear least squares, parameter estimates which result in a minimisation of *SSR* are regarded as the best estimates of the parameters in the equation. Figure 8.1 shows a hypersurface which depends on estimates a and b. There is a minimum in *SSR* for $a \approx 7$. Examination of tabulated values shown in table 8.1 indicates that the minimum in *SSR* occurs for $b \approx 4$.

Fitting by non-linear least squares begins with reasonable approximations for the best estimates of the parameters. These approximations are often referred to as 'starting values'. The objective is to modify the starting values in an iterative fashion until a minimum is found in *SSR*. The computational intensiveness of the iteration process means that non-linear least squares can only realistically be carried out using a computer.

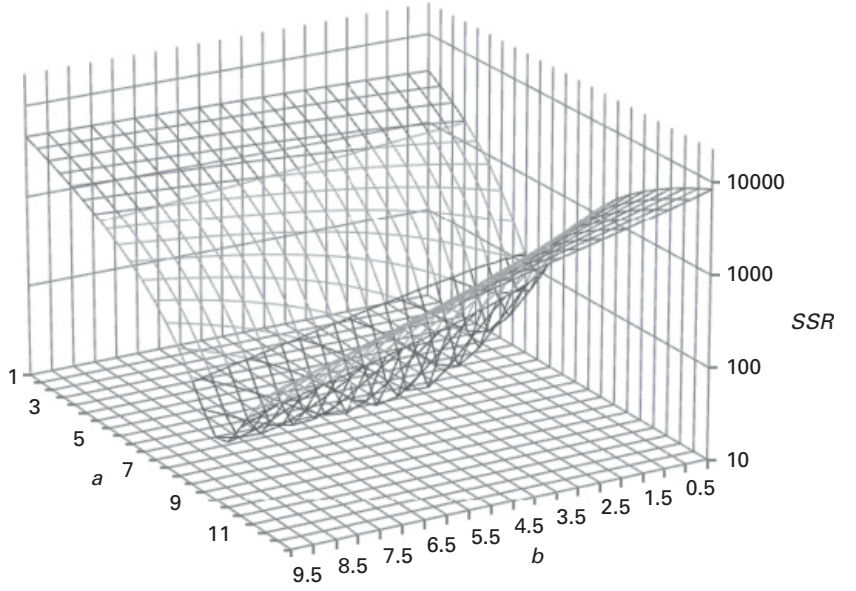


Figure 8.1. Variation of *SSR* with parameter estimates, a and b.

[1] See Motulski and Christopoulos (2005).

[2] See Bevington and Robinson (2002).

Table 8.1. *Values of SSR for a range of values of a and b. The highlighted cell contains the smallest value of SSR in the table; this indicates that best estimates of a and b are close to 7.0 and 4.0 respectively.*

		a				
		6.0	6.5	7.0	7.5	8.0
b	3.0	57.64611	21.8929	53.5454	152.6036	319.0675
	3.5	84.70766	25.42352	29.57799	97.17108	228.2028
	4.0	118.969	39.19182	19.26127	59.17733	158.94
	4.5	158.766	61.06023	19.93265	35.38321	107.4119
	5.0	202.8235	89.3755	29.5189	23.25373	70.57996
	5.5	250.1483	122.8372	46.378	20.77073	46.01537
	6.0	299.9563	160.4085	69.19152	26.30531	31.74987

There are many documented ways in which the values of *a*, *b*, *c*, etc., can be found which minimise *SSR*, including Grid Search (Bevington and Robinson, 2002), Gauss–Newton method (Nielsen-Kudsk, 1983) and the Levenberg-Marquardt algorithm (Bates and Watts, 1988).

Non-linear least squares is unnecessary when the derivatives of *SSR* with respect to the parameters are linear in those parameters. In such a situation, linear least squares offers a more efficient route to determining best estimates of the parameters (and the standard uncertainties in the best estimates). Nevertheless, a linear equation can be fitted to data using non-linear least squares. The answer obtained for best estimates of parameters and the standard uncertainties in the best estimates should agree, irrespective of whether a linear equation is fitted using linear or non-linear least squares.

Many computer-based statistical packages such as Origin[3] offer the facility to fit equations to data using non-linear least squares. A shortcoming of most packages is that the process by which the fitting takes place is hidden from the user, making it less easy to explore what is happening as fitting takes place. An alternative approach is to use a tool in Excel called Solver as this gives more control of the fitting process allowing for insights into what is happening as fitting by non-linear least squares proceeds.

[3] OriginLab Corporation, Northampton, Massachusetts, USA.

8.2 Excel's Solver add-in

Solver, first introduced in 1991, is one of many add-ins available in Excel.[4] Originally designed for business users, Solver is a powerful and flexible optimisation tool. Given a model, for example, representing the profitability of a commercial product, Excel's Solver is able to iteratively modify the parameters in that model and return the parameter estimates that would maximise the profitability. This is an example of optimisation, where parameters within a model or equation are adjusted until a desired outcome is achieved (which often involves maximising or minimising a particular quantity).

In the context of fitting equations to data, Solver is capable of finding the best estimates of parameters using least squares. It does this by iteratively altering the numerical values of variables contained in the cells of a spreadsheet until *SSR* is minimised. To solve non-linear problems, Solver uses Generalized Reduced Gradient (GRG2) code developed at the University of Texas and Cleveland State University.[5] Features of Solver are best described by reference to a particular example.

8.2.1 Example of use of Solver

Consider an experiment in which the air temperature in an enclosure (such as a living room) is measured as a function of time as heat passes into the enclosure, for example through a window. Table 8.2 contains the raw data. Figure 8.2 displays the same data in graphical form.

Through a consideration of the flow of heat into and out of an enclosure, a relationship may be derived for the air temperature, T, inside the enclosure as a function of time, t. The relationship can be expressed as

$$T = T_s + k[1 - \exp(\alpha t)], \tag{8.5}$$

where T_s, k and α are constants estimated using non-linear least squares. Using x and y as the independent and dependent variables respectively and a, b and c as best estimates of T_s, k and α respectively, equation 8.5 becomes

$$y = a + b[1 - \exp(cx)]. \tag{8.6}$$

To find best estimates, a, b and c, we proceed as follows.

(1) Enter the raw data from table 8.2 into columns A and B of an Excel worksheet as shown in sheet 8.1.

[4] See Fylstra, Lasdon, Watson and Waren (1998)

[5] See Excel's online Help. See also Smith and Lasdon (1992).

Table 8.2. *Variation of air temperature in an enclosure with time.*

Time (minutes)	Temperature (°C)
2	26.1
4	26.8
6	27.9
8	28.6
10	28.5
12	29.3
14	29.8
16	29.9
18	30.1
20	30.4
22	30.6
24	30.7

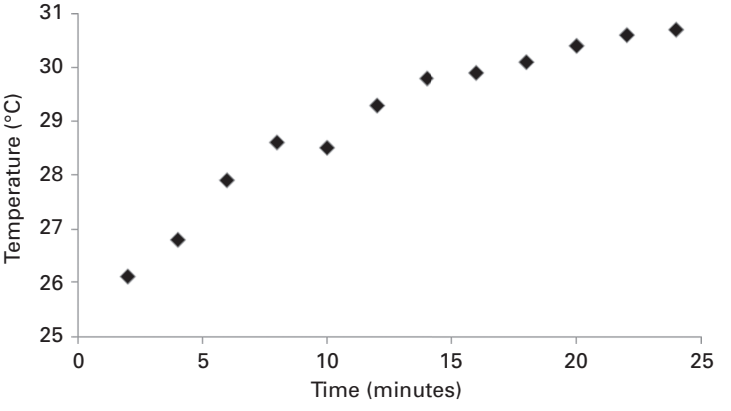


Figure 8.2. Temperature variation with time inside an enclosure.

(2) Type **=$B$15+$B$16*(1-EXP($B$17*A2))** into cell C2 as shown in sheet 8.1. Cells B15 to B17 contain the starting values for a, b and c respectively.
(3) Use the cursor to highlight cells C2 to C13.
(4) Click on Home Ribbon, then choose Editing>Fill>Down.[6]

Sheet 8.2 shows the calculated values in the C column. As the squares of the residuals, $(y - \hat{y})^2$, are required, these are calculated in column D.

[6] As in chapter 2 we abbreviate these steps to Home>Editing>Fill>Down.

Sheet 8.1. *Temperature (y) and time (x) data from table 8.1 entered into a spreadsheet.*[7]

	A	B	C
1	*x* (mins)	*y* (℃)	$\hat{y}$ (℃)
2	2	26.1	=$B$15+$B$16*(1-EXP($B$17*A2))
3	4	26.8	
4	6	27.9	
5	8	28.6	
6	10	28.5	
7	12	29.3	
8	14	29.8	
9	16	29.9	
10	18	30.1	
11	20	30.4	
12	22	30.6	
13	24	30.7	
14			
15	a	1	
16	b	1	
17	c	1	

Sheet 8.2. *Calculation of sum of squares of residuals.*

	A	B	C	D
1	*x* (mins)	*y* (℃)	$\hat{y}$ (℃)	$(y-\hat{y})^2$ (℃^{2})
2	2	26.1	−5.38906	991.560654
3	4	26.8	−52.5982	6304.066229
4	6	27.9	−401.429	184323.2129
5	8	28.6	−2978.96	9045405.045
6	10	28.5	−22024.5	486333300.3
7	12	29.3	−162753	26498009287
8	14	29.8	−1202602	1.44632E+12
9	16	29.9	−8886109	7.89635E+13
10	18	30.1	−6.6E+07	4.31124E+15
11	20	30.4	−4.9E+08	2.35385E+17
12	22	30.6	−3.6E+09	1.28516E+19
13	24	30.7	−2.6E+10	7.01674E+20
14			SSR =	7.14765E+20
15	a	1		
16	b	1		
17	c	1		

The sum of the squares of residuals, *SSR*, is calculated in cell D14 by summing the contents of cells D2 through to D13. Starting values for *a*, *b* and *c*

[7] The estimated values of the dependent variable must be distinguished from values obtained through experiment. Estimated values are represented by the symbol, $\hat{y}$, and experimental values by the symbol, *y*.

are extremely poor, as the predicted values, $\hat{y}$, in column C of sheet 8.2 bear no resemblance to the experimental values in column B. As a consequence, *SSR* is very large. Choosing good starting values for parameter estimates is often crucial to the success of fitting equations using non-linear least squares and we will return to this issue in section 8.3.2.

SSR in cell D14 is reduced by altering the contents of cells B15 through to B17. Solver is able to adjust the parameter estimates in cells B15 to B17 until the number in cell D14 is minimised. To accomplish this, choose the Data Ribbon on Excel, go to the Analysis group and click on Solver. If Solver does not appear, then it can be added in as follows.

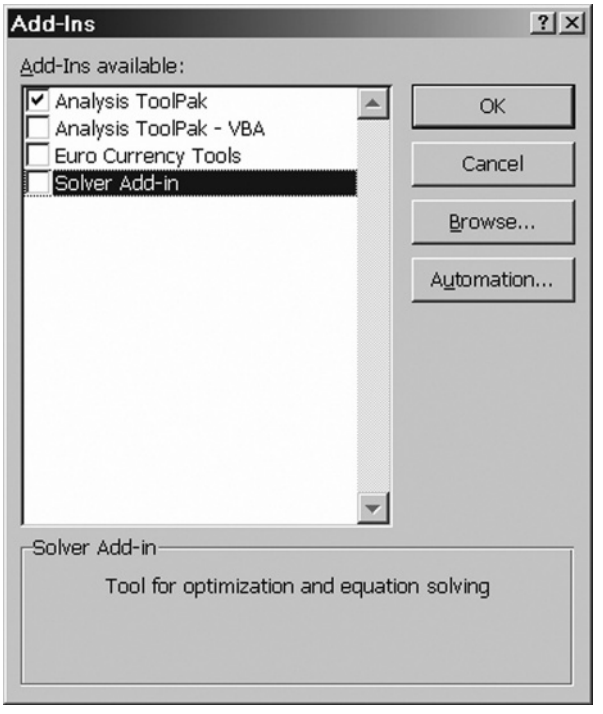
(1) Click on the File tab, File, at the top left of the screen.

(2) Click on Options.

(3) Click on Add-Ins.

(4) Select Manage: Excel Add-ins Go... then click on Go.

(5) The following the dialog box should appear.



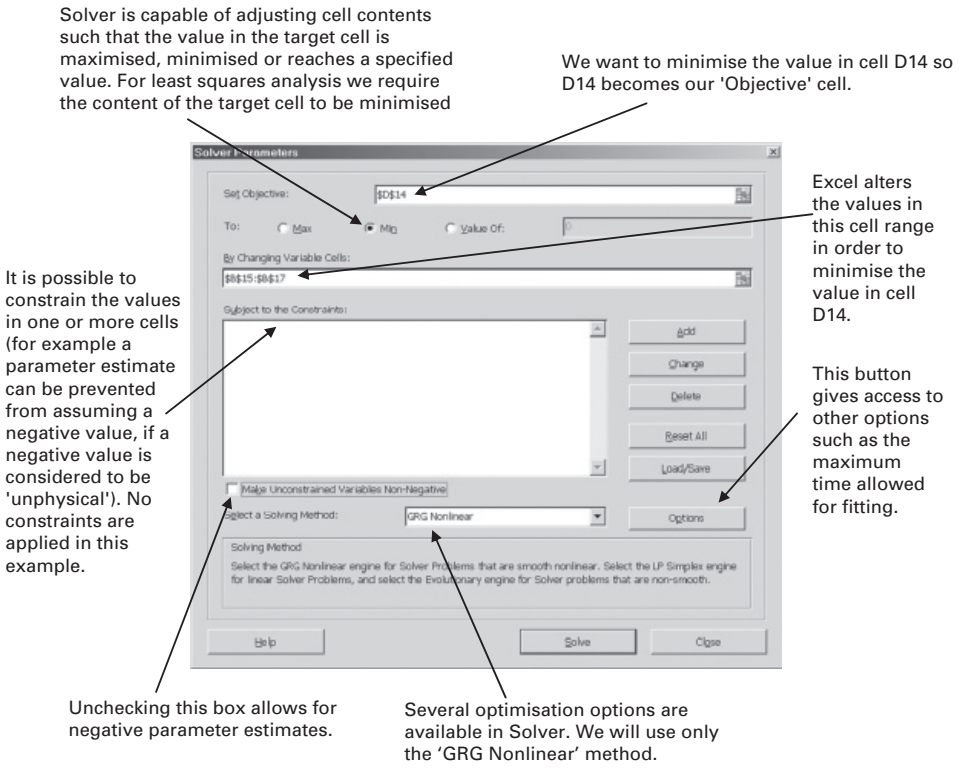


Figure 8.3. Solver dialog box with cell references inserted.

Tick the Solver Add-in box then press OK.

Solver should be visible on the right hand side of the Data Ribbon.

Click on Solver. The dialog box shown in figure 8.3 should appear. The annotations in figure 8.3 describe some of the features of Solver accessible through the dialog box.

After entering the information into the dialog box, click on the Solve button. After a few seconds Solver returns with the dialog box shown in figure 8.4. Click on OK.

Inspection of cells B15 to B17 in the spreadsheet indicates that Solver has adjusted the parameters. Sheet 8.3 shows the new parameters, predicted y values and *SSR*.

SSR in cell D14 in sheet 8.3 is almost 20 orders of magnitude smaller than that in cell D14 in sheet 8.2. However, all is not as satisfactory as it might seem. Consider the best line through the points shown in figure 8.5 which utilises the parameter estimates in cells B15 through to B17 of sheet 8.3.

Sheet 8.3. *Best values for a, b and c returned by Solver when starting values are poorly chosen.*[8]

	A	B	C	D
1	x (mins)	y (°C)	$\hat{y}$ (°C)	$(y - \hat{y})^2$ (°C^2)
2	2	26.1	24.750128	1.822154
3	4	26.8	27.996494	1.431597
4	6	27.9	29.10535	1.45287
5	8	28.6	29.484101	0.781635
6	10	28.5	29.613471	1.239817
7	12	29.3	29.657659	0.12792
8	14	29.8	29.672753	0.016192
9	16	29.9	29.677908	0.049325
10	18	30.1	29.679669	0.176678
11	20	30.4	29.68027	0.518011
12	22	30.6	29.680476	0.845525
13	24	30.7	29.680546	1.039286
14			SSR	9.501009
15	a	15.2458		
16	b	14.4347		
17	c	−0.5371		
18				

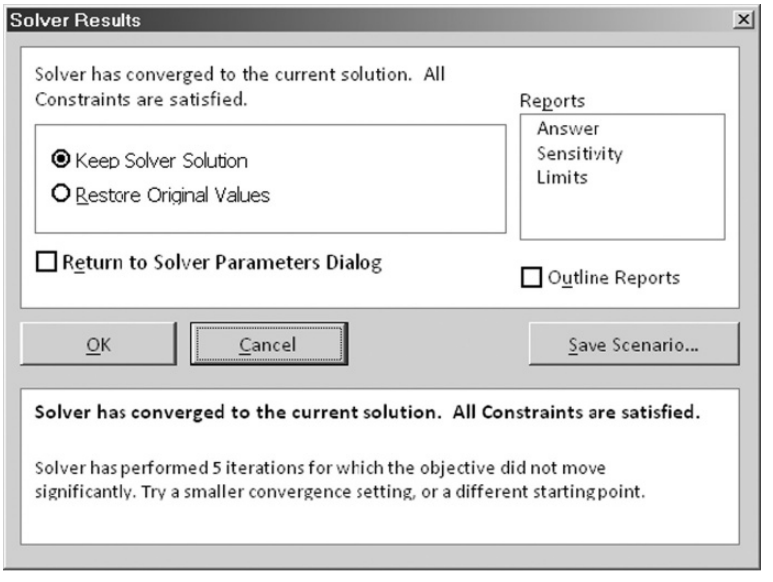


Figure 8.4. Solver dialog box indicating that fitting has been completed.

[8] If some of the Solver options have been modified then Solver may return slightly different parameter estimates to those shown in sheet 8.3. Solver options can be accessed by clicking on Options in the Solver Parameters dialogue box. For more details on Solver options, go to section 8.5.

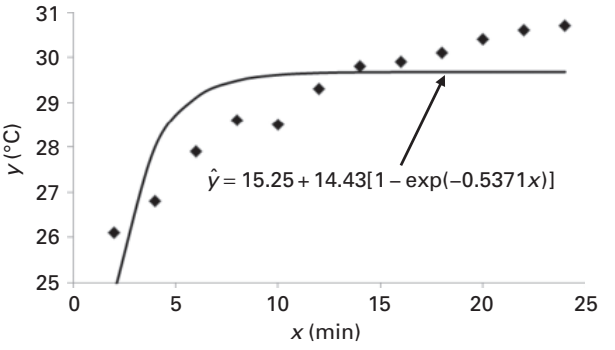


Figure 8.5. Graph of y versus x showing the line based on equation 8.6 where a, b and c have the values given in sheet 8.3.

As discussed in section 6.7, a plot of residuals (i.e. a plot of $(y_i - \hat{y}_i)$ versus x_i) is often used as an indicator of the 'goodness of fit' of an equation to data, with trends in the residuals indicating a poor fit.[9] However, no plot of residuals is required in this case to reach the conclusion that the line on the graph in figure 8.5 is **not** a good fit to the experimental data. Solver has found a minimum in *SSR*, but this is a *local* minimum[10] and the parameter estimates are of little worth. The source of the problem can be traced to the arbitrarily chosen starting values for a, b and c appearing in sheet 8.2 (i.e. $a = b = c = 1$). Working from these starting values, Solver has discovered a minimum in *SSR*. However, there is another combination of parameter estimates that will give a lower value for *SSR*.

Methods by which good starting values for parameter estimates may be obtained are considered in section 8.3.2. In the example under consideration here, we note (by reference to equation 8.6) that when $x = 0$, $y = a$. Drawing a line 'by eye' through the data in figure 8.2 indicates that, when $x = 0$, $y \approx 25.5$ °C. Starting values for b and c may be established by further preliminary analysis of data which we consider in section 8.3.2. Denoting starting values by a_0, b_0 and c_0, we find[11]

$$a_0 = 25.5,\ b_0 = 5.5 \text{ and } c_0 = -0.12.$$

Inserting these values into sheet 8.2 and running Solver again gives the output shown in sheet 8.4 and in graphical form in figure 8.6.

[9] See Cleveland (1994) for a discussion of residuals.

[10] Local minima are discussed in section 8.3.1.

[11] All parameter estimates and standard uncertainties in parameters estimates in this example have units (for example the unit of c is min^{-1}, assuming time is measured in minutes). For convenience, units are omitted until the analysis is complete.

Sheet 8.4. *Best values for a, b and c returned by Solver when starting values for parameter estimates are good.*

	A	B	C	D
1	x (mins)	y (℃)	$\hat{y}$ (℃)	$(y-\hat{y})^2$ (℃²)
2	2	26.1	26.072473	0.000758
3	4	26.8	26.977367	0.031459
4	6	27.9	27.727663	0.0297
5	8	28.6	28.349774	0.062613
6	10	28.5	28.8656	0.133663
7	12	29.3	29.293299	4.49E-05
8	14	29.8	29.647927	0.023126
9	16	29.9	29.941968	0.001761
10	18	30.1	30.185773	0.007357
11	20	30.4	30.387925	0.000146
12	22	30.6	30.55554	0.001977
13	24	30.7	30.694519	3E-05
14			SSR	0.292635
15	a	24.98113		
16	b	6.387886		
17	c	−0.09367		
18				

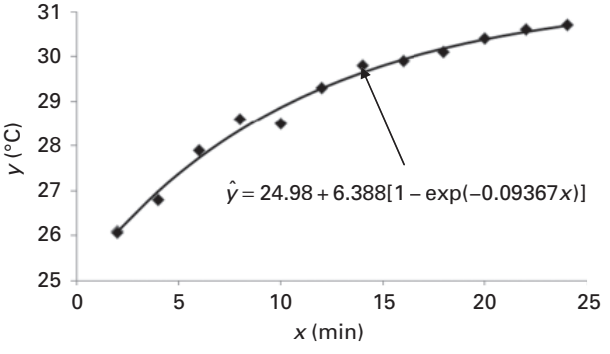


Figure 8.6. Graph of y versus x showing line and equation of line based on estimates a, b and c appearing in sheet 8.4.

The sum of squares of residuals in cell D14 of sheet 8.4 is less than that in cell D14 of sheet 8.3. This points to the parameter estimates obtained using Solver when good starting values are used being rather better than those obtained when the starting values are poorly chosen. More

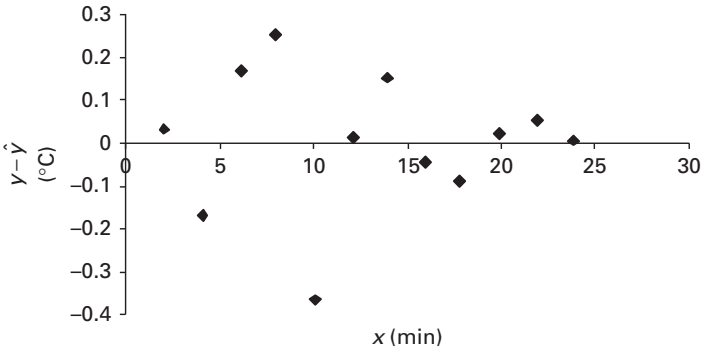


Figure 8.7. Plot of residuals based on the data and equation in figure 8.6.

Table 8.3. *Data for exercise A.*

x (rad)	y
0	1.77
0.1	2.28
0.2	3.83
0.3	5.47
0.4	7.50
0.5	8.60
0.6	8.59
0.7	10.56
0.8	10.11
0.9	10.36
1.0	10.15

importantly, the line fitted to the data in figure 8.6 (where the line is based upon the new best estimates of the parameters) is far superior to the line fitted to the same data as shown in figure 8.5. This is further reinforced by the plot of residuals shown in figure 8.7 which exhibits a random scatter about the x axis.

Exercise A

Consider the data in table 8.3.

Assuming the best line through the data in table 8.3 is of the form $y = a + b \sin cx$, use Solver to determine *a, b* and *c*. Try a range of starting values for *a, b* and *c*. Does Solver consistently converge to the same final values for *a, b* and *c*?

8.2.2 Limitations of Solver

Solver is able to efficiently solve for the best estimates of parameters in an equation, such as those appearing in equation 8.6. However, Solver does not provide standard uncertainties in the parameter estimates. Standard uncertainties in estimates are extremely important, as without them it is not possible to quote a coverage interval for the estimates and so we cannot decide if the estimates are good enough for any particular purpose to which they may be put.

The standard uncertainties in the parameter estimates *a*, *b* and *c* can be determined with the aid of the matrix of partial derivatives given by[12]

$$\mathbf{E} = \begin{bmatrix} \sum \left(\frac{\partial y_i}{\partial a}\right)^2 & \sum \left(\frac{\partial y_i}{\partial a}\frac{\partial y_i}{\partial b}\right) & \sum \left(\frac{\partial y_i}{\partial a}\frac{\partial y_i}{\partial c}\right) \\ \sum \left(\frac{\partial y_i}{\partial a}\frac{\partial y_i}{\partial b}\right) & \sum \left(\frac{\partial y_i}{\partial b}\right)^2 & \sum \left(\frac{\partial y_i}{\partial b}\frac{\partial y_i}{\partial c}\right) \\ \sum \left(\frac{\partial y_i}{\partial a}\frac{\partial y_i}{\partial c}\right) & \sum \left(\frac{\partial y_i}{\partial b}\frac{\partial y_i}{\partial c}\right) & \sum \left(\frac{\partial y_i}{\partial c}\right)^2 \end{bmatrix}. \tag{8.7}$$

Standard uncertainties in *a*, *b* and *c* are obtained using the diagonal elements of the $\mathbf{E^{-1}}$ matrix. Explicitly,

$$s_a = s\left(E_{11}^{-1}\right)^{1/2} \tag{8.8}$$

$$s_b = s\left(E_{22}^{-1}\right)^{1/2} \tag{8.9}$$

$$s_c = s\left(E_{33}^{-1}\right)^{1/2}, \tag{8.10}$$

[12] Note that this approach can be extended to any number of parameters. Kutner, Nachtsheim and Neter (2003) discuss the use of matrices to determine the variance in the parameter estimates.

where[13]

$$s \approx \left[\frac{1}{n-3}\sum (y_i - \hat{y}_i)^2\right]^{1/2}. \tag{8.11}$$

A convenient way to calculate the elements of the **E** matrix is to write

$$\mathbf{E} = \mathbf{D^T D}. \tag{8.12}$$

$\mathbf{D^T}$ is the transpose of the matrix, **D**, where **D** is given by

$$\mathbf{D} = \begin{bmatrix} \frac{\partial y_1}{\partial a} & \frac{\partial y_1}{\partial b} & \frac{\partial y_1}{\partial c} \\ \frac{\partial y_2}{\partial a} & \frac{\partial y_2}{\partial b} & \frac{\partial y_2}{\partial c} \\ \frac{\partial y_i}{\partial a} & \frac{\partial y_i}{\partial b} & \frac{\partial y_i}{\partial c} \\ \frac{\partial y_n}{\partial a} & \frac{\partial y_n}{\partial b} & \frac{\partial y_n}{\partial c} \end{bmatrix}. \tag{8.13}$$

The partial derivatives in equation 8.13 are evaluated on completion of fitting an equation using Solver, i.e. at values of a, b and c that minimise *SSR*. It is possible in some situations to determine the partial derivatives analytically. A flexible approach, and one that is generally more convenient, is to use the method of 'finite differences' to find $\frac{\partial y_1}{\partial a}$, $\frac{\partial y_2}{\partial a}$, etc. In general,

$$\left(\frac{\partial y_i}{\partial a}\right)_{b,c,x_i} \approx \frac{y[a(1+\delta), b, c, x_i] - y[a, b, c, x_i]}{a(1+\delta) - a}. \tag{8.14}$$

As double precision arithmetic is used by Excel, the perturbation, δ, in equation 8.14 can be as small as $\delta = 10^{-6}$ or 10^{-7}.

Similarly, the partial derivatives, $\frac{\partial y_i}{\partial b}$ and $\frac{\partial y_i}{\partial c}$, are approximated by using

$$\left(\frac{\partial y_i}{\partial b}\right)_{a,c,x_i} \approx \frac{y[a, b(1+\delta), c, x_i] - y[a, b, c, x_i]}{b(1+\delta) - b} \tag{8.15}$$

and

$$\left(\frac{\partial y_i}{\partial c}\right)_{a,b,x_i} \approx \frac{y[a, b, c(1+\delta), x_i] - y[a, b, c, x_i]}{c(1+\delta) - c}. \tag{8.16}$$

[13] The expression $n-3$ in the denominator of the term in the square brackets of equation 8.11 appears because the estimate of the population standard deviation in the y values about the fitted line requires that the sum of squares of residuals be divided by the number of degrees of freedom. The number of degrees of freedom is the number of data points, n, minus the number of parameters, M, in the equation. In this example, $M = 3$.

8.2.3 Spreadsheet for determining standard uncertainties in parameter estimates

To illustrate the process of estimating standard uncertainties, we describe a step-by-step approach using Excel.

To find good approximations of the derivatives $\frac{\partial y_1}{\partial a}$ and $\frac{\partial y_2}{\partial a}$, it is necessary to perturb a slightly (say to $1.000001 \times a$) while leaving the parameter estimates b and c at their optimum values. Sheet 8.5 shows the optimum values, as obtained by Solver for a, b and c in cells G20 to G22. Cell H20 contains the value $1.000001 \times a$. Cell I21 contains the value $1.000001 \times b$ and cell J22 contains the value $1.000001 \times c$.

We use the modified parameter estimates to calculate the numerator in equation 8.14. The denominator in equation 8.14 may be determined by entering the formula = $H$20-$G$20 into a cell on the spreadsheet.

Sheet 8.5. *Modification of best estimates of parameters.*

	F	G	H	I	J
19		from solver	*b,c* constant	*a,c* constant	*a,b* constant
20	*a*	24.981126	24.9811512	24.98112622	24.98112622
21	*b*	6.3878859	6.387885919	6.387892307	6.387885919
22	*c*	−0.0936749	−0.093674867	−0.09367487	−0.093674961
23					

The partial derivative $\left(\frac{\partial y_1}{\partial a}\right)_{b,c,x_1}$ is calculated by entering the formula = (H2-C2)/($H$20-$G$20) into cell L2 of sheet 8.6.[14] By using Home>Editing> Fill>Down, the formula may be copied into cells in the L column so that the partial derivative is calculated for every x_i. To obtain $\frac{\partial y_i}{\partial b}$, $\frac{\partial y_i}{\partial c}$, etc., this process is repeated for columns M and N, respectively, of sheet 8.6. The contents of cells L2 to N13 become the elements of the **D** matrix given by equation 8.13.

Excel's TRANSPOSE() function is used to transpose the **D** matrix. We proceed as follows.

- Highlight cells B24 to N26.
- In cell B24 type =**TRANSPOSE(L2:N13)**.
- While holding down the Ctrl and Shift keys press Enter to transpose the contents of cells L2 to N13 into cells B24 to M26.
- Multiply $\mathbf{D}^{\mathbf{T}}$ with **D** (using the MMULT matrix function in Excel) to give **E**, i.e.

[14] The values in the C column of the spreadsheet are shown in sheet 8.4.

$$\mathbf{E} = \mathbf{D}^{\mathbf{T}}\mathbf{D} = \begin{bmatrix} 12 & 7.65923 & 246.036 \\ 7.65923 & 5.49390 & 160.861 \\ 246.036 & 160.861 & 5251.55 \end{bmatrix}. \tag{8.17}$$

The MINVERSE() function in Excel is used to find the inverse of **E**, i.e.

$$\mathbf{E}^{-1} = \begin{bmatrix} 2.239638 & -0.48571 & -0.09005 \\ -0.48571 & 1.870409 & -0.034537 \\ -0.09005 & -0.034537 & 0.005467 \end{bmatrix}. \tag{8.18}$$

Two more steps are required to calculate the standard uncertainties in the parameter estimates. The first is to calculate the square root of each diagonal element of the matrix $\mathbf{E}^{-1}$. The second is to calculate s using equation 8.11. Using the sum of squares of residuals appearing in cell D14 of sheet 8.4, we obtain

$$s = \left[\frac{1}{12-3} \times 0.2926\right]^{1/2} = 0.1803.$$

It follows that

$$s_a = s(E_{11}^{-1})^{1/2} = 0.1803 \times (2.240)^{1/2} \quad = 0.270 \tag{8.19}$$

$$s_b = s(E_{22}^{-1})^{1/2} = 0.1803 \times (1.870)^{1/2} \quad = 0.247 \tag{8.20}$$

$$s_c = s(E_{33}^{-1})^{1/2} = 0.1803 \times (0.005467)^{1/2} \quad = 0.0133. \tag{8.21}$$

Sheet 8.6. *Calculation of partial derivatives.*

	H	I	J	K	L	M	N
1	$\hat{y}$ with b,c, constant	$\hat{y}$ with a,c constant	$\hat{y}$ with a,b, constant		$\partial y/\partial a$	$\partial y/\partial b$	$\partial y/\partial c$
2	26.07249786	26.07247397	26.07247387		1	0.170846	−10.5931
3	26.97739198	26.977369	26.97736864		1	0.312504	−17.5666
4	27.72768829	27.72766606	27.72766536		1	0.42996	−21.8481
5	28.34979926	28.34977764	28.34977654		1	0.527349	−24.1539
6	28.86562487	28.86560377	28.86560223		1	0.6081	−25.0341
7	29.29332358	29.29330291	29.29330093		1	0.675055	−24.9085
8	29.64795155	29.64793124	29.64792883		1	0.73057	−24.0952
9	29.94199265	29.94197263	29.94196981		1	0.776601	−22.8327
10	30.18579791	30.18577814	30.18577493		1	0.814768	−21.2983
11	30.38794995	30.38793038	30.38792681		1	0.846414	−19.6217
12	30.55556506	30.55554565	30.55554176		1	0.872654	−17.8964
13	30.69454375	30.69452448	30.69452029		1	0.894411	−16.1878

8.2.4 Coverage intervals for parameter estimates

We use parameter estimates and their respective standard uncertainties to quote a coverage interval for each parameter.[15] For the parameters appearing in equation 8.5,

$$T_s = a \pm t_{X\%,v} s_a \tag{8.22}$$

$$k = b \pm t_{X\%,v} s_b \tag{8.23}$$

$$\alpha = c \pm t_{X\%,v} s_c, \tag{8.24}$$

where $t_{X\%,v}$ is the critical value of the t distribution for $X\%$ level of confidence with v degrees of freedom. Values of $t_{X\%,v}$ can be found in appendix 1. In this example $v = n - 3$ where n is the number of data points. In table 8.2 there are 12 points, so that $v = 9$. If we choose a level of confidence of 95% (the commonly chosen level),

$$t_{95\%,9} = 2.262.$$

Restoring the units of measurement and quoting a 95% coverage interval gives,

$$T_s = (24.98 \pm 2.262 \times 0.270)^\circ\text{C} = (24.98 \pm 0.61)^\circ\text{C}$$
$$k = (6.388 \pm 2.262 \times 0.247)^\circ\text{C} = (6.39 \pm 0.56)^\circ\text{C}$$
$$\alpha = (-0.09367 \pm 2.262 \times 0.0133)\ \text{min}^{-1} = (-0.094 \pm 0.030)\ \text{min}^{-1}.$$

Exercise B

The amount of heat entering an enclosure through a window may be reduced by applying a reflective coating to the window. An experiment is performed to establish the effect of a reflective coating on the rise in air temperature within the enclosure. The temperature within the enclosure as a function of time is shown in table 8.4.

Fit equation 8.6 to the data in table 8.4. Find a, b and c and their respective standard uncertainties.

[15] We could replace $t_{X\%,v}$ by the coverage factor, k, for v degrees of freedom with $X\%$ level of confidence (see section 5.9.1).

Table 8.4. *Data for exercise B.*

Time (minutes)	Temperature (°C)
2	24.9
4	25.3
6	25.4
8	25.8
10	26.0
12	26.3
14	26.4
16	26.6
18	26.5
20	26.8
22	27.0
24	26.9

8.3 More on fitting using non-linear least squares

There are several challenges to fitting equations using non-linear least squares, including:

(1) choosing an appropriate model to describe the relationship between x and y;
(2) avoiding local minima in *SSR*;
(3) establishing good starting values prior to fitting by non-linear least squares.

We consider (2) and (3) next.

8.3.1 Local minima in *SSR*

When data are noisy, or starting values are far from the best estimates, a non-linear least squares fitting routine can become 'trapped' in a local minimum.

To illustrate this situation, we draw on the analysis of data appearing in section 8.2.1. Equation 8.6 is fitted to the data in table 8.2 using the starting values given in sheet 8.2 and the best estimates, a, b and c are obtained for the parameters. For clarity, the relationship between only one parameter estimate (c) and *SSR* is considered. Solver finds a minimum in *SSR* when c is about –0.53 and terminates the fitting procedure. The variation of *SSR* with c is shown in figure 8.8.[16]

[16] For clarity, units of measurement are omitted from the graph.

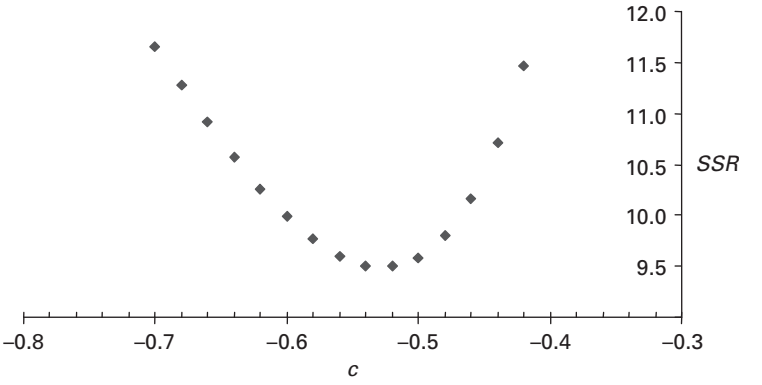


Figure 8.8. Variation of *SSR* with *c* near a local minimum for equation 8.6 is fitted to the data in table 8.1.

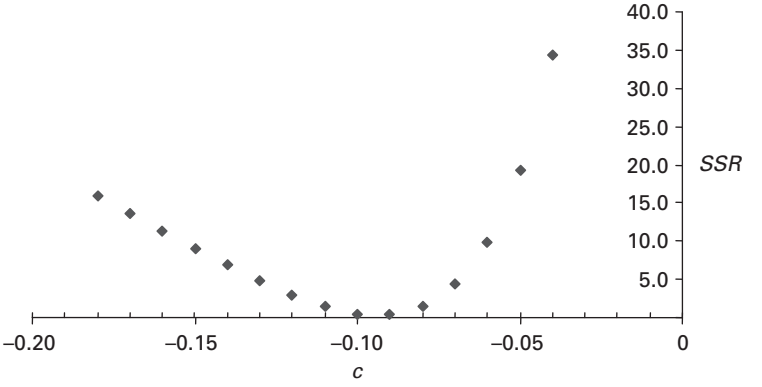


Figure 8.9. Variation of *SSR* with *c* near to the global minimum for equation 8.6 is fitted to the data in table 8.1.

The minimum in *SSR* in figure 8.8 is referred to as a *local* minimum as there is another combination of parameter estimates that will give a lower value for *SSR*. The lowest value of *SSR* obtainable corresponds to a *global* minimum. It is the global minimum that we seek in all fitting using least squares.

When starting values are used that are closer to the final values,[17] the non-linear fitting routine finds parameter estimates that produce a lower final value for *SSR*. Figure 8.9 shows the variation of *SSR* with *c* in the interval $(-0.18 < c < -0.04)$.

[17] See section 8.2.1.

A number of indicators can assist in identifying a local minimum, though there is no infallible way of deciding whether a local or global minimum has been discovered. A good starting point is to plot the raw data along with the fitted line (as illustrated in figure 8.5). A poor fit of the line to the data could indicate:

- a local minimum has been found;
- an inappropriate model has been fitted to the data.

When a local minimum in *SSR* is found, the standard uncertainties in the parameter estimates tend to be large. As an example, best estimates appearing in sheet 8.3 (resulting from being trapped in a local minimum), their respective standard uncertainties and the magnitude of the ratio of these quantities (expressed as a percentage) are:

$$a = 15.25\,^{\circ}\text{C}, s_a = 9.04\,^{\circ}\text{C}, \text{so that } |s_a/a| \times 100\% = 59\%;$$
$$b = 14.43\,^{\circ}\text{C}, s_b = 8.93\,^{\circ}\text{C}, \text{so that } |s_b/b| \times 100\% = 62\%;$$
$$c = -0.5371\ \text{min}^{-1}, s_c = 0.276\ \text{min}^{-1}, \text{so that } |s_c/c| \times 100\% = 51\%.$$

When the global minimum in *SSR* is found (see sheet 8.4), the best estimates of the parameters and standard uncertainties are:

$$a = 24.98\,^{\circ}\text{C}, s_a = 0.270\,^{\circ}\text{C}, \text{so that } |s_a/a| \times 100\% = 1.1\%;$$
$$b = 6.388\,^{\circ}\text{C}, s_b = 0.247\,^{\circ}\text{C}, \text{so that } |s_b/b| \times 100\% = 3.9\%;$$
$$c = -0.09367\ \text{min}^{-1},\ s_c = 0.0133\ \text{min}^{-1}, \text{so that } |s_c/c| \times 100\% = 14\%.$$

There is merit in fitting the same equation to data several times, each time using different starting values for the parameter estimates. If, after fitting, there is consistency between the final values obtained for the best estimates, then it is likely that the global minimum has been identified.

8.3.2 Starting values

There are no general rules that may be applied in order to determine good starting values[18] for parameter estimates prior to fitting by non-linear least squares. Familiarity with the relationship being studied can assist in deciding what might be reasonable starting values for some of the parameters in an equation.

A useful approach to determining starting values is to begin by plotting the experimental data. Consider the data in figure 8.10, which has a smooth line drawn through the points 'by eye'.

If the relationship between x and y is given by equation 8.25, then we are able to estimate a and b by considering the data displayed in figure 8.10 and a

[18] Starting values are sometimes referred to as *initial estimates*.

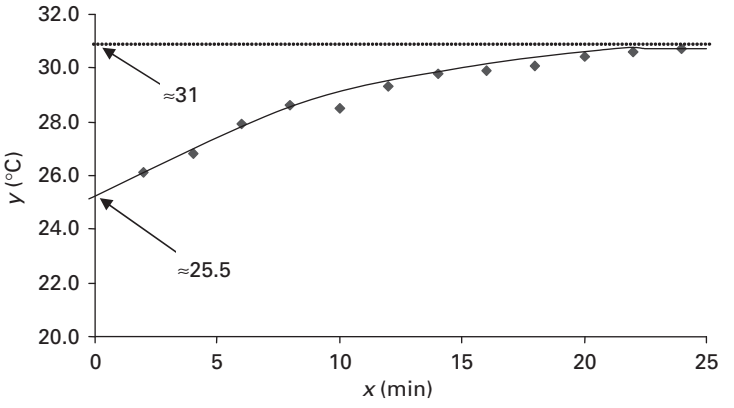


Figure 8.10. Line drawn 'by eye' through the data given in table 8.1.

'rough' line drawn through the data. Equation 8.25 predicts that $y = a$ when x is equal to zero:

$$y = a + b[1 - \exp(cx)]. \tag{8.25}$$

From figure 8.10 we see that when $x = 0$, $y \approx 25.5$ °C, so that $a \approx 25.5$ °C. When x is large (and assuming c is negative), then $\exp(cx)$ tends to zero so that, $y = a + b$. Inspection of the graph in figure 8.10 indicates that when x is large, $y \approx 31.0$ °C, i.e. $a + b \approx 31.0$ °C. It follows that $b \approx 5.5$ °C. If we write the starting values for a and b as a_0 and b_0 respectively, then $a_0 = 25.5$ °C and $b_0 = 5.5$ °C.

In order to determine a starting value for c, c_0, equation 8.25 is rearranged into the form[19]

$$\ln\left[1 - \left(\frac{y - a_0}{b_0}\right)\right] = c_0 x. \tag{8.26}$$

Equation 8.26 has the form of the equation of a straight line passing through the origin. It follows that plotting $\ln\left[1 - \left(\frac{y-a_0}{b_0}\right)\right]$ versus x should give a straight line with slope, c_0.

Figure 8.11 shows a plot of $\ln\left[1 - \left(\frac{y-a_0}{b_0}\right)\right]$ versus x. The best straight line and the equation of the line has been added using the Trendline option in Excel.[20] The slope of the line is approximately –0.12. The starting values may now be stated for this example, i.e.:

[19] Starting values, a_0 and b_0, are substituted into equation 8.26.

[20] For details on using Trendline, see section 2.7.1. Note that for a straight line forced through the origin, the slope would be –0.11 min^{-1}. Taking this to be c_0 rather than –0.12 min^{-1} would have no influence on the final fitted values.

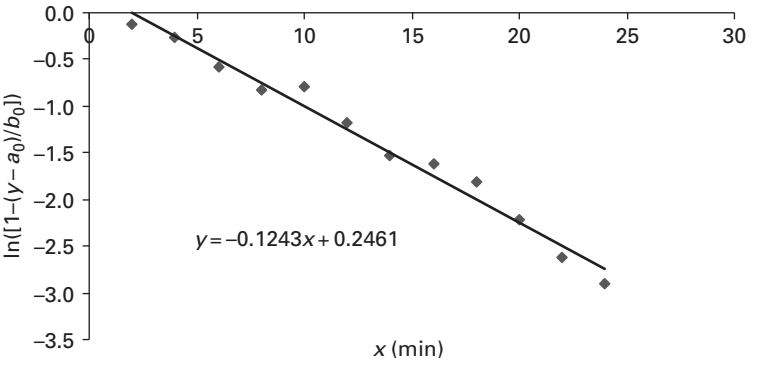


Figure 8.11. Line of best fit used to determine a starting value for *c*.

$$a_0 = 25.5^{\circ}\text{C};\ b_0 = 5.5^{\circ}\text{C};\ c_0 = -0.12\,\text{min}^{-1}.$$

These starting values were used in the successful fit of equation 8.6 to the data given in table 8.1 (the output of the fit is shown in sheet 8.4).

8.3.3 Starting values by curve stripping

Establishing starting values in some situations is quite difficult and may require a significant amount of pre-processing of data before employing fitting by non-linear least squares. For example, the fitting to data of an equation consisting of a sum of exponential terms, such as

$$y = a \exp bt + c \exp dt \tag{8.27}$$

or

$$y = a \exp bt + c \exp dt + e \exp ft, \tag{8.28}$$

where b, d and f are negative with $|b| > |d| > |f|$ and $t \geq 0$, is particularly challenging especially when data are noisy and/or the ratio of the parameters within the exponentials is less than approximately 3 (for example, when the ratio b/d in equation 8.27 is less than 3).[21] Fitting of equations such as equation 8.27 and equation 8.28 is quite common, for example the kinetics of drug transport through the human body is routinely modelled using 'compartmental analysis'.[22]

[21] See Kirkup and Sutherland (1988).

[22] See Jacquez (1996).

Compartmental analysis attempts to predict concentrations of drugs as a function of time (e.g. in blood or urine). The relationship between concentration and time is often well represented by a sum of exponential terms. In analytical chemistry, excited state lifetime measurements offer a means of identifying components in a mixture. The decay of phosphorescence with time that occurs after illumination of the mixture may be captured. The decay can be represented by a sum of exponential terms. Fitting a sum of exponentials by non-linear least squares allows for each component in the mixture to be discriminated.[23]

If an equation to be fitted to data consists of a sum of exponential terms, good starting values for parameter estimates are extremely important if local minima in *SSR* are to be avoided. It is also possible that, if starting values for the parameter estimates are too far from the optimum values, the *SSR* will increase during the iterative process to such an extent that the *SSR* exceeds the maximum floating point number that a spreadsheet (or other program) can handle. In this situation, fitting is terminated and an error message is returned by the spreadsheet.

Data in figure 8.12 have been gathered in an experiment in which the decay of photo-generated current in the wide band gap semiconductor cadmium sulphide (CdS) is measured as a function of time after photo-excitation of the semiconductor has ceased. There appears to be an exponential decay of the photocurrent with time. Theory indicates[24] that there may be more than one decay mechanism for photoconductivity. That, in turn, suggests that an equation of the form given by equation 8.27 or equation 8.28 may be appropriate.

If equation 8.27 is to be fitted to data, how are starting values for parameter estimates established? If b is large (and negative) then the contribution of the first term in equation 8.27 to y is small when t exceeds some value, which we will designate as t'. Equation 8.27 can now be written, for $t > t'$,

$$y \approx c \exp dt. \tag{8.29}$$

Equation 8.29 can be linearised by taking natural logarithms of both sides of the equation. The next step is to fit a straight line to the transformed data to find (approximate values) for c and d which we will designate as c_0 and d_0 respectively.

Now we revisit equation 8.27 and write, for $t < t'$,

$$y - c_0 \exp d_0 t = a \exp bt. \tag{8.30}$$

[23] See Demas (1983).

[24] See Bube (1960), chapter 6.

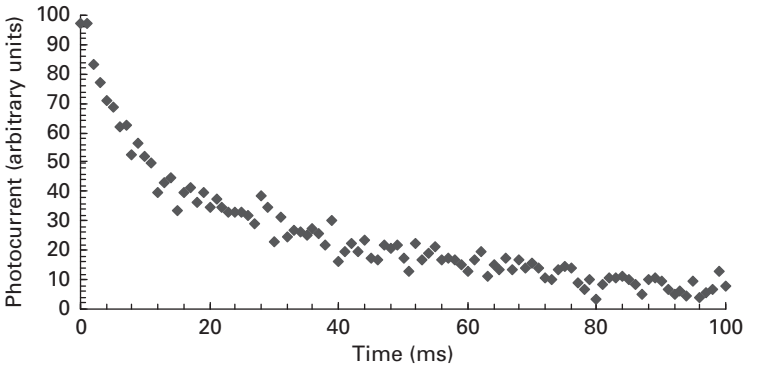


Figure 8.12. Photocurrent versus time data for cadmium sulphide.

Transforming equation 8.30 by taking natural logarithms of both sides of the equation then fitting a straight line to the transformed data will yield approximate values for a and b which can serve as starting values in a non-linear fit.

8.4 Weighted non-linear least squares

There are some situations in which the standard deviation of the residuals, $(y - \hat{y})$, is not constant (see section 6.7). These residuals are said to be heteroscedastic. Such a situation may be revealed by plotting[25] $(y - \hat{y})$ versus x. If residuals *are* heteroscedastic, then *weighted* fitting is required. The purpose of weighted fitting is to obtain best estimates of the parameters by forcing the line close to the data with least uncertainty, while giving much less weight to those data with larger uncertainty.

As described in section 6.10, the starting point for weighted fitting using least squares is to define a sum of squares of residuals that takes into account the standard deviation in the y values. We write

$$\chi^2 = \sum \left(\frac{y_i - \hat{y}_i}{s_i} \right)^2, \tag{8.31}$$

where χ^2 is the weighted sum of squares of residuals,[26] and s_i is the standard deviation in the ith y value. The purpose of the weighted fitting is to find best estimates of parameters that minimise χ^2 in equation 8.31.

[25] See section 6.7 for a discussion of residuals.

[26] $\sum \left(\frac{y_i - \hat{y}_i}{s_i} \right)^2$ follows a chi-squared distribution, hence the use of the symbol χ^2.

8.4.1 Weighted fitting using Solver

In order to establish best estimates of parameters using Solver when weighted fitting is performed, we use an approach similar to that described in section 8.2.3. For weighted fitting, an extra column in the spreadsheet containing the standard deviations s_i is required. It is possible that the absolute values of s_i are unknown and that only relative standard deviations are known. For example, equations 8.32 and 8.33 are sometimes used when weighted fitting is required,

$$s_i \propto \sqrt{y_i} \tag{8.32}$$

$$s_i \propto y_i. \tag{8.33}$$

Weighted fitting can be carried out so long as:

- the absolute standard deviations in values are known, or
- the relative standard deviations are known.

In order to accomplish weighted non-linear least squares, we proceed as follows.

(1) Fit the desired equation to data by calculating χ^2 as given by equation 8.31. Use Solver to modify parameter estimates so that χ^2 is minimised.
(2) Determine the elements in the **D** matrix, as described in section 8.2.2.
(3) Construct the weight matrix, **W**, in which the diagonal elements of the matrix contain the weights to be applied to the y values.
(4) Calculate the weighted standard deviation s_w, where s_w is given by

$$s_w = \left(\frac{\chi^2}{n - M}\right)^{1/2}, \tag{8.34}$$

and where χ^2 is given by equation 8.31, n is the number of data points and M is the number of parameters in the equation to be fitted to the data.
(5) Calculate the standard uncertainties in parameter estimates, given by[27]

$$\mathbf{s}(\mathbf{B}) = s_w \left[\left(\mathbf{D}^{\mathrm{T}}\mathbf{W}\mathbf{D}\right)^{-1}\right]^{1/2}. \tag{8.35}$$

B is the matrix containing elements equal to the best estimates of the parameters, and s_w is the weighted standard deviation, given by equation 8.34.
(6) Calculate the coverage interval for each parameter appearing in the equation at a specified level of confidence (usually 95%).

To illustrate steps (1) to (6), we consider an example of weighted fitting using Solver.

[27] See Kutner, Nachtsheim and Neter (2003).

Table 8.5. *Current–voltage data for a germanium tunnel diode.*

V (mV)	I (mA)
10	4.94
20	6.67
30	10.57
40	10.11
50	10.44
60	12.90
70	10.87
80	9.73
90	7.03
100	5.61
110	3.80
120	2.36

8.4.2 Example of weighted fitting using Solver

The relationship between the current, I, through a tunnel diode[28] and the voltage, V, across the diode may be written[29]

$$I = AV(B - V)^2. \tag{8.36}$$

A and B are constants to be estimated using least squares. Table 8.5 shows current–voltage data for a germanium tunnel diode.

Equation 8.36 could be fitted to data using unweighted non-linear least squares (in the first instance it is prudent to carry out an unweighted fit, as the residuals may show marginal evidence of heteroscedasticity and so there is little point in performing a more complex analysis).

In this example we are going to assume that the error in the y quantity is proportional to the size of the y quantity, i.e. equation 8.33 is valid for these data.

The data in table 8.5 are entered into a spreadsheet as shown in sheet 8.7 and plotted in figure 8.13.

[28] A tunnel diode is a semiconductor device with characteristics that make it useful in microwave circuits.

[29] See Karlovsky (1962).

Sheet 8.7. *Data from table 8.5 entered into a spreadsheet.*

	A	B
1	*x*(mV)	*y*(mA)
2	10	4.94
3	20	6.67
4	30	10.57
5	40	10.11
6	50	10.44
7	60	12.90
8	70	10.87
9	80	9.73
10	90	7.03
11	100	5.61
12	110	3.80
13	120	2.36
14		

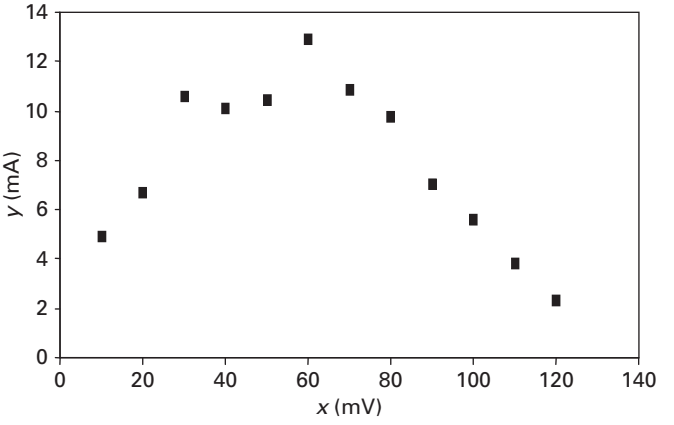


Figure 8.13. Current–voltage data for a germanium tunnel diode.

8.4.3 Best estimates of parameters using Solver

Consistent with symbols used in other least squares analyses, we rewrite equation 8.36 as

$$y = ax(b - x)^2. \tag{8.37}$$

We can obtain a reasonable value for b, which we will use as a starting value, b_0, by noting that equation 8.37 predicts that $y = 0$ when $x = b$. By inspection of

Sheet 8.8. *Fitted values and weighted sum of squares of residuals* before *optimisation occurs.*

	B	C	D
1	y (mA)	$\hat{y}$ (mA)	$\left(\frac{y-\hat{y}}{y}\right)^2$
2	4.94	4.752	0.001448311
3	6.67	7.986	0.038927822
4	10.57	9.9	0.004017905
5	10.11	10.692	0.003313932
6	10.44	10.56	0.000132118
7	12.90	9.702	0.061457869
8	10.87	8.316	0.055205544
9	9.73	6.6	0.103481567
10	7.03	4.752	0.105001811
11	5.61	2.97	0.221453287
12	3.80	1.452	0.381793906
13	2.36	0.396	0.692562482
14		sum	1.668796555
15			
16			From solver
17		a	3.30E-05
18		b	130

figure 8.13 we find that when $y = 0$, $x \approx 130$ mV, so that $b_0 = 130$ mV. Equation 8.37 is rearranged to give

$$a = \frac{y}{x(b-x)^2}. \tag{8.38}$$

An approximate value for a (which we take to be the starting value, a_0) can be obtained by choosing any data pair from sheet 8.7 (say, $x = 50$ mV and $y = 10.44$ mA) and substituting these into equation 8.38 along with $b_0 = 130$ mV. This gives (to two significant figures) $a_0 = 3.3 \times 10^{-5}$ mA/(mV)^3.

Sheet 8.8 shows the cells containing the calculated values of current ($\hat{y}$) in column C based on equation 8.37. The parameter estimates are the starting values (3.3×10^{-5} and 130) in cells D17 and D18. Column D of sheet 8.8 contains the weighted sum of squares of residuals. The sum of these residuals appears in cell D14.

Running Solver using the default settings (see section 8.5.1) gives the output shown in sheet 8.9.

The weighted standard deviation is calculated using equation 8.34, i.e.

$$s_w = \left(\frac{\chi^2}{n-p}\right)^{1/2} = \left(\frac{0.0895702}{12-2}\right)^{1/2} = 0.09464. \tag{8.39}$$

Sheet 8.9. *Fitted values and weighted sum of squares of residuals* after *using Solver to minimise the value in cell D14 by changing the parameter estimates in cells D17 and D18.*[30]

	B	C	D
1	y (mA)	$\hat{y}$ (mA)	$\left(\frac{y-\hat{y}}{y}\right)^2$
2	4.94	4.4579265	0.009523
3	6.67	7.68209988	0.0230248
4	10.57	9.81027104	0.0051662
5	10.11	10.9801908	0.0074084
6	10.44	11.3296102	0.007261
7	12.90	10.99628	0.0217784
8	10.87	10.117951	0.0047867
9	9.73	8.83237427	0.0085107
10	7.03	7.2773006	0.0012375
11	5.61	5.59048087	1.211E−05
12	3.80	3.90966598	0.0008329
13	2.36	2.3726068	2.854E−05
14		sum	0.0895702
15			
16			From solver
17		a	2.2958E−05
18		b	149.34612

χ^2

8.4.4 Determining the D matrix

In order to determine the matrix of partial derivatives, we calculate

$$\left(\frac{\partial y_i}{\partial a}\right)_{b,x_i} \approx \frac{y[a(1+\delta), b, x_i] - y[a, b, x_i]}{a(1+\delta) - a} \tag{8.40}$$

and

$$\left(\frac{\partial y_i}{\partial b}\right)_{a,x_i} \approx \frac{y[a, b(1+\delta), x_i] - y[a, b, x_i]}{b(1+\delta) - b}, \tag{8.41}$$

where δ is chosen to be 10^{-6} (see section 8.2.2). Sheet 8.10 shows the values of the partial derivatives in the **D** matrix.

8.4.5 The weight matrix, W

The weight matrix is a square matrix with diagonal elements proportional to $\frac{1}{s_i^2}$ and other elements equal to zero.[31] In this example, s_i is taken to be equal to y_i, so that the diagonal matrix is as given in sheet 8.11.

[30] Note that prior to fitting, the Use Automatic Scaling box was checked in the Options dialogue box in Solver. It is recommended that this box be checked ahead of any fitting using Solver (see section 8.5.1 for details).

[31] For details on the weight matrix, see Kutner, Nachtsheim and Neter (2003).

Sheet 8.10. *Calculation of partial derivatives used in the* **D** *matrix.*

	E	F	G	H	I
1	$\hat{y}$ (*b* constant)	$\hat{y}$ (*a* constant)		$\partial y/\partial a$	$\partial l/\partial b$
2	4.457930955	4.457936052		194173.4121	0.063984
3	7.682107563	7.682117621		334608.3762	0.118784
4	9.810280846	9.810295588		427304.8921	0.1644
5	10.98020182	10.98022084		478262.9600	0.200834
6	11.32962152	11.32964425		493482.5800	0.228084
7	10.99629095	10.99631671		478963.7516	0.24615
8	10.11796114	10.11798911		440706.4753	0.255034
9	8.832383106	8.832412317		384710.7509	0.254733
10	7.277307872	7.277337222		316976.5784	0.24525
11	5.590486459	5.590514708		243503.9579	0226583
12	3.909669886	3.909695656		170292.8892	0.198732
13	2.372609174	2.372630951		103343.3725	0.161699
14					
15					
16	*b* constant	*a* constant			
17	2.295850E−05	2.295848E−05			
18	149.3461202	149.3462695			

D

Sheet 8.11. *Weight matrix for tunnel diode analysis (while the weights are shown to only three decimal places, Excel retains all figures for the calculations).*

	C	D	E	F	G	H	I	J	K	L	M	N
24	0.041	0	0	0	0	0	0	0	0	0	0	0
25	0	0.022	0	0	0	0	0	0	0	0	0	0
26	0	0	0.009	0	0	0	0	0	0	0	0	0
27	0	0	0	0.010	0	0	0	0	0	0	0	0
28	0	0	0	0	0.009	0	0	0	0	0	0	0
29	0	0	0	0	0	0.006	0	0	0	0	0	0
30	0	0	0	0	0	0	0.008	0	0	0	0	0
31	0	0	0	0	0	0	0	0.011	0	0	0	0
32	0	0	0	0	0	0	0	0	0.020	0	0	0
33	0	0	0	0	0	0	0	0	0	0.032	0	0
34	0	0	0	0	0	0	0	0	0	0	0.069	0
35	0	0	0	0	0	0	0	0	0	0	0	0.180

W

8.4.6 Calculation of $(\mathbf{D^T W D})^{-1}$

To obtain standard uncertainties in estimates a and b, we must determine $(\mathbf{D^T W D})^{-1}$. Sheet 8.12 shows the several steps required to determine $(\mathbf{D^T W D})^{-1}$. The steps consist of the following.

(a) Calculation of the matrix **WD**. The elements of this matrix are shown in cells C37 to D48. (**W** is multiplied with **D** using the MMULT() function in Excel).

Sheet 8.12. *Calculation of* $(\mathbf{D^T W D})^{-1}$.

	B	C	D	E	F	G	H
37	**WD**	7956.75278	0.0026219		$\mathbf{D^T WD}$	2.2597E + 10	15368.5701
38		7521.16542	0.00267			15368.5701	0.01347659
39		3824.61646	0.0014715				
40		4679.12273	0.0019649				
41		4527.62896	0.0020926		$\mathbf{(D^T WD)^{-1}}$	1.9722E–10	–0.0002249
42		2878.21496	0.0014792			–0.0002249	330.689201
43		3729.84121	0.0021584				
44		4063.57839	0.0026907				
45		6413.81639	0.0049625				
46		7737.13727	0.0071995				
47		11793.1364	0.0137626				
48		18554.9003	0.0290323				

(b) Calculation of the matrix $\mathbf{D^T WD}$. The elements of this matrix are shown in cells G37 to H38.

(c) Inversion of the matrix, $\mathbf{D^T WD}$. The elements of the inverted matrix are shown in cells G41 to H42.

8.4.7 Bringing it all together

To calculate the standard uncertainties in a and b, the weighted standard deviation (given by equation 8.34) is multiplied by the square root of the diagonal elements of the $(\mathbf{D^T WD})^{-1}$ matrix, i.e.

$$s_a = s_w\left(1.9722 \times 10^{-10}\right)^{1/2}$$
$$= 0.094\,641 \times \left(1.9722 \times 10^{-10}\right)^{1/2} = 1.329 \times 10^{-6}$$

and

$$s_b = s_w(330.69)^{1/2} = 0.094\,641 \times (330.69)^{1/2} = 1.721.$$

It follows that the 95% coverage intervals for A and B are

$$A = a \pm t_{95\%,v} s_a \tag{8.42}$$

$$B = b \pm t_{95\%,v} s_b. \tag{8.43}$$

In this example, the number of degrees of freedom, $v = n - p = 12 - 2 = 10$. From table 2 in appendix 1,

$$t_{95\%,10} = 2.228.$$

It follows that (inserting units and rounding the expanded uncertainty to two significant figures)

$$A = (2.30 \pm 0.30) \times 10^{-5}\text{mA}/(\text{mV})^3$$
$$B = (149.3 \pm 3.8)\,\text{mV}.$$

Exercise C

Equation 8.36 may be transformed into a form suitable for fitting by linear least squares.

(i) Show that equation 8.36 can be rearranged into the form

$$\left(\frac{I}{V}\right)^{1/2} = A^{1/2}B - A^{1/2}V. \tag{8.44}$$

(ii) Plot a graph of $\left(\frac{I}{V}\right)^{1/2}$ versus V.

(iii) Use unweighted least squares to obtain best estimates of A and B and standard uncertainties in the best estimates.[32]

(iv) Why is it preferable to use non-linear least squares to estimate parameters rather than to linearise equation 8.36 followed by using linear least squares to find these estimates?

8.5 More on Solver

Solver uses the General Reduced Gradient (GRG) method for solving non-linear problems and this section is devoted to describing features of Solver that relate to this.[33]

Though optimisation can be carried out successfully with the default settings in the Solver Option dialog box, Solver possesses several options that can be adjusted by the user to assist in the optimisation process and we will describe those next.

The Solver dialog box, as shown in figure 8.3, offers the facility to constrain parameter estimates so that unrealistic estimates can be avoided. In any case the best estimates returned by Solver need to be compared with 'physical reality' before being accepted. Consider an example in which a parameter in an equation represents the speed of sound, v, in air. If, after fitting by least squares, the best estimate of v is –212 m/s, it is fair to question whether this value is 'reasonable'. If it is not, then one course of action is to try new starting values for the parameter estimates. Alternatively, we could use the 'Subjects to Constraints box' in the Solver dialog box to constrain the estimate

[32] Care must be taken when calculating the uncertainty in the estimate of B as this requires use of both slope and the intercept and these are correlated.

[33] See Smith and Lasdon (1992). Solver includes two other methods for solving optimisation problems: one incorporates the Simplex method for solving linear problems, and the other 'Evolutionary Solver' which is capable of being used with any Excel formula or function. We will not consider those two methods in this book.

of v so that it cannot take on negative values. This cannot guarantee that a physically meaningful value will be found for v, only that the value will be non-negative.

If we are quite sure that a parameter will lie between certain limits, then specifying the constraints can reduce the fitting time. Perhaps even more useful, is that constraints can be used with an option new to Excel 2010 referred to as 'Multi-start' (described in the next section) which improves the probability of finding a global minimum in *SSR*. To use this option, bounds for all the parameters appearing in the equation to be fitted to data must be specified using the constraints box.

8.5.1 Solver Options

To view Solver options shown in figure 8.14, it is necessary to click on the Option button in the Solver dialog box. This dialog box may be used to modify, for example, the method by which the optimisation takes place. This, in turn, may provide for a better fit or reduce the fitting time over that obtained using the default settings.

We now consider some of the options in the Options dialog box.

Constraint Precision	Solver evaluates constraints within the tolerance specified within the Constraint Precision box. For example, suppose the constraint precision is set to 0.0001. If we specified that the parameter in cell $D$18 >=0, and after fitting the content of cell $D$18 was –0.000 015, then Solver would consider the constraint to be satisfied.
Use Automatic Scaling	In certain problems there may be many orders of magnitude difference between the data, the parameter estimates and the value in the cell which is referenced in the Set Objective box. This can lead to rounding problems owing to the finite precision arithmetic performed by Excel. If the 'Use Automatic Scaling' box is ticked, then Solver will scale values before carrying out optimisation (and 'unscale' the solution values before entering them into the spreadsheet). It is advisable to tick this box for all problems.
Show Iteration Results	Ticking this box causes Solver to pause after each iteration, allowing new estimates of parameters and the value in the cell referenced in the Set Objective box to be viewed. If parameter estimates are used to draw a line of best fit

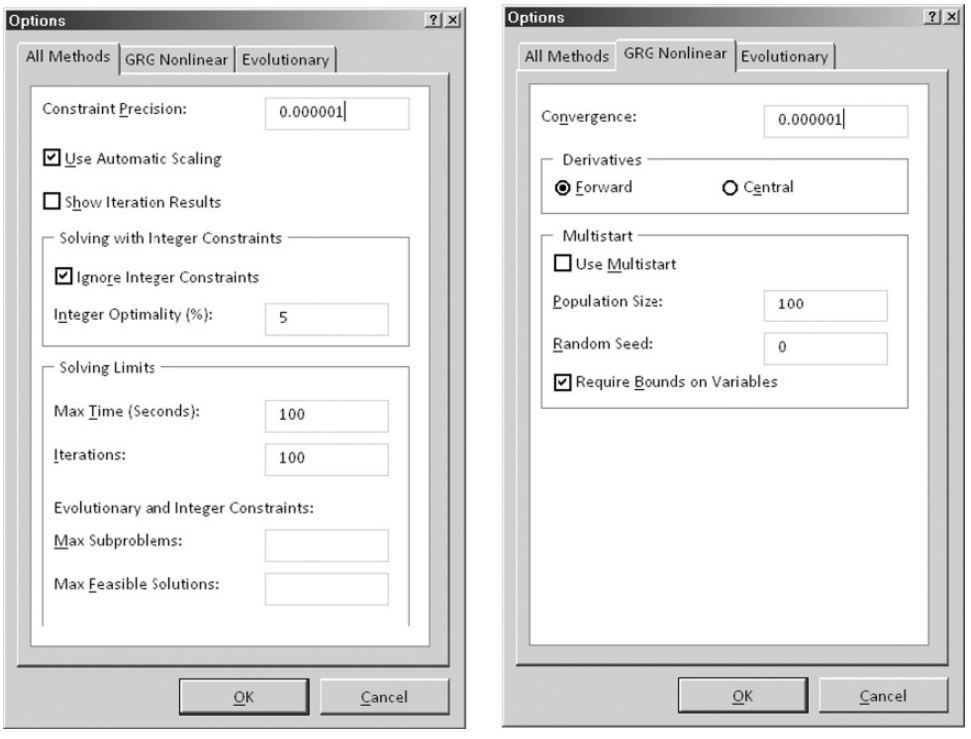


Figure 8.14. Solver Options dialog box illustrating default settings.

through the data, then the line will be updated after each iteration. Updating the fitted line on the graph after each iteration gives a valuable insight into the progress made by Solver to find best estimates of the parameters in an equation.

Max Time — This restricts the total time Solver spends searching for an optimum solution. Unless there are many data, the default value of 100 s is usually sufficient. If the maximum time is set too low, such that Solver has not completed its search, then a message is returned 'The maximum time limit was reached; continue anyway?'. Clicking on the Continue button will cause Solver to carry on searching for a solution.

Iterations — This is the maximum number of iterations that Solver will execute before terminating its search. The default value is 100, but this can be increased to a limit of 32767. Solver is likely to find an optimum solution before reaching such a

	limit or return a message that an optimum solution cannot be found. If the number of iterations is set too low, such that Solver has not completed its search, then a message will be returned 'The maximum iteration limit was reached; continue anyway?'. Clicking on the Continue button will cause Solver to carry on searching for a solution.
Convergence	As fitting proceeds, Solver compares the most recent solution (for our purposes this would be the value of *SSR*) with previous solutions. If the fractional reduction in the solution over five iterations is less than the value in the Convergence box, Solver reports that optimisation is complete. If this value is made very small (say, 10^{-6}) Solver will continue iterating (and hence take longer to complete) than if that number is larger (say, 10^{-2}).
Derivatives: Forward or Central	The partial derivatives of the function in the target cell with respect to the parameter estimates are found by the method of finite differences. It is possible to perturb the estimates 'forward' from a particular point or to perturb the estimates forward and backward from the point in order to obtain a better estimates of the partial derivatives. Both methods of determining the partial derivatives produce the same final results for the examples described in this chapter.
Multi-start	An important feature introduced in Excel 2010 that was not available in previous versions of Excel, is the ability of Solver to automatically choose many sets of starting values for the parameters. Excel performs a fit using each set of starting values in an effort to find the global minimum in *SSR*. It is necessary to specify the bounds of the parameter estimates (i.e. the maximum and minimum of each estimate) before Solver is able to execute its Multi-start feature. The bounds are entered by specifying the constraints for the parameters which is done in the main Solver Parameters dialog box. The wider the bounds that are specified, the longer it will take Solver to find the global minimum. Though this option does increase the probability of finding the global minimum, it does not guarantee it.

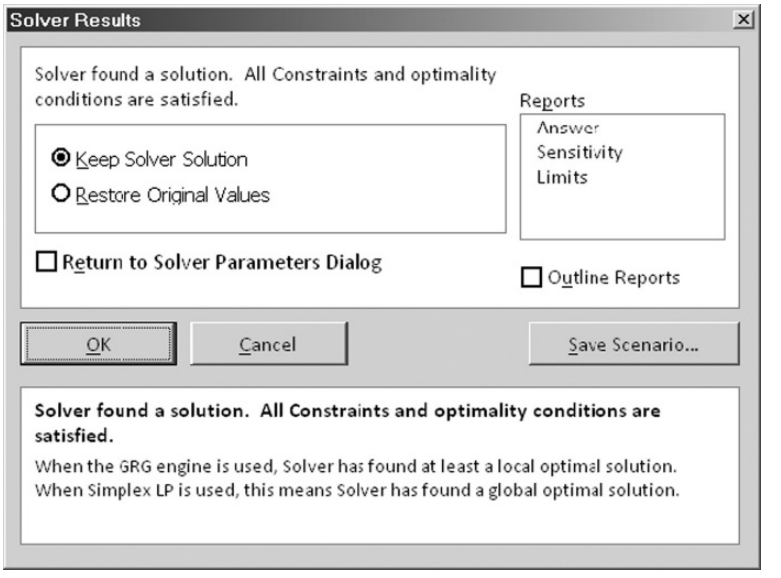


Figure 8.15. Solver Results dialog box.

8.5.2 Solver Results

Once Solver completes optimisation, it displays the Solver Results dialog box shown in figure 8.15.

Clicking on OK will retain the solution found by Solver (i.e. the starting parameters are permanently replaced by the final parameter estimates). At this stage Excel is able to present three reports: Answer, Sensitivity and Limits. Of the three reports, the Answer report is the most useful as it gives the starting values of the parameter estimates and the associated *SSR*. The report also displays the final parameter estimates and the final *SSR*, allowing for easy comparison with the original values. An Answer report is shown in figure 8.16.

8.6 Review

This chapter considered fitting equations to data using the technique of non-linear least squares. In particular, the use of the Solver tool within Excel was introduced and employed for non-linear least squares fitting.

In order to quote a coverage interval for parameters it is necessary to be able to determine standard uncertainties in the estimates made of any parameters appearing in an equation. As Solver does not provide standard

Microsoft Excel 14.0 Answer Report
Worksheet: [chapter8_Figure8_4.xlsx]Sheet1
Report Created: 29/04/2010 9:01:15 PM
Result: Solver has converged to the current solution. All Constraints are satisfied.
Solver Engine
Engine: GRG Nonlinear
Solution Time: 0.42 Seconds.
Iterations: 10 Subproblems: 0
Solver Options
Max Time 100 sec, Iterations 100, Precision 0.000001
Convergence 0.0001, Population Size 100, Random Seed 0, Derivatives Forward, Require Bounds
Max Subproblems Unlimited, Max Integer Sols Unlimited, Integer Tolerance 5%, Solve Without Integer Constraints

Objective Cell ()

Cell	Name	Original Value	Final Value
D:	SSR (y-yhat)^2 (°C2)	2.295668212	0.292634971

Variable Cells

Cell	Name	Original Value	Final Value	Integer
G:	a from solver	25.5	24.98112622	Contin
G:	b from solver	5.5	6.387885919	Contin
G:	c from solver	-0.12	-0.093674867	Contin

Constraints

NONE

Figure 8.16. Answer report created by Excel.

uncertainties, this chapter describes a way in which these may be determined using Excel. An advantage of employing Excel is that aspects of fitting by non-linear least squares which are normally hidden from view when using a conventional computer based statistics package can be made visible with Excel. This approach leads to an insight into the process of fitting using non-linear-least squares which goes beyond the act of entering numbers into a statistics package and waiting for the fruits of the analysis to emerge.

Some general issues relating to fitting by non-linear least squares have been discussed, such as the existence of local minima in *SSR* and the means by which good starting values may be established in advance of fitting.

End of chapter problems

(1) Standard addition analysis is routinely used to establish the composition of a sample. In order to establish the concentration of Fe^{3+} in water, solutions

Table 8.6. *Data for problem (1).*

Concentration (ppm), x	Absorbance (arbitrary units), y
0	0.240
5.55	0.437
11.10	0.621
16.65	0.809
22.20	1.009

containing known concentrations of Fe^{3+} were added to water samples.[34] The absorbance of each solution, y, was determined for each concentration of added solution, x. The absorbance/concentration data are shown in table 8.6.

The relationship between absorbance, y, and concentration, x, of added solution may be written

$$y = B(x - x_C), \tag{8.45}$$

where B is the slope of the line of y versus x, and x_C is the intercept on the x axis which represents the concentration of Fe^{3+} in the water before additions are made.

Use non-linear least squares to fit equation 8.45 to the data in table 8.6. Determine

(i) best estimates of B and x_C;
(ii) standard uncertainties in best estimates of B and x_C.

(2) Another way to analyse the data in table 8.6 is to write

$$y = A + Bx. \tag{8.46}$$

Here A is the intercept on the y axis at $x = 0$, and B is the slope. The intercept on the x axis, x_C (found by setting $y = 0$ in equation 8.46), is given by

$$x_C = \frac{-A}{B}. \tag{8.47}$$

Use linear least squares to fit equation 8.46 to the data in table 8.6. Determine,

(i) best estimates of A, B and x_C;
(ii) standard uncertainties in the best estimates of A, B and x_C.

Note that the errors in the best estimate of slope and intercept in equation 8.47 are correlated and so the normal 'propagation of uncertainties' method is not valid when calculating x_C.

[34] This problem is adapted from Skoog and Leary (1992).

Table 8.7. *Data for problem (3).*

t (s)	$V(t)$ (mL)
145	4.0
314	7.6
638	12.2
901	15.6
1228	18.6
1691	21.6
2163	24.0
2464	24.8

(3) In a study of first order kinetics, the volume of titrant required, $V(t)$, to reach the end point of a reaction is measured as a function of time, t (see table 8.7). The data obtained are shown in table 8.7.[35]

The relationship between V and t can be written

$$V(t) = V_\infty - (V_\infty - V_0)\exp(-kt), \tag{8.48}$$

where k is the rate constant. V_∞ and V_0 are also constants.

(i) Find approximate values[36] for V_∞ , V_0 and k.
(ii) Use Solver to establish best estimates of V_∞ , V_0 and k.
(iii) Determine standard uncertainties in the estimates of V_∞ , V_0 and k.
(iv) Determine the 95% coverage intervals for V_∞ , V_0 and k.

(4) Table 8.8 contains data obtained from a simulation of a chemical reaction in which noise of constant variance has been added to the data.[37]

Assuming that the relationship between concentration, C, and time, t, can be written[38]

$$C = \frac{C_0}{1 + C_0 kt}, \tag{8.49}$$

where C_0 is the concentration at $t = 0$ and k is the second order rate constant.

(i) Find approximate values for C_0 and k.
(ii) Fit equation 8.49 to the data in table 8.8 to obtain best estimates for C_0 and k.
(iii) Determine standard uncertainties in the best estimates.

[35] See Denton (2000).

[36] Hint: draw a graph of $V(t)$ versus t. Equation 8.48 indicates that, as t gets large, $V(t)$ nears V_∞. V_0 is the value of $V(t)$ when $t = 0$ and this may be estimated from the graph. Using the estimates of V_∞ and V_0, equation 8.48 can be written in a form that allows for linearisation from which a starting value for k can be found (see section 8.3.2 for a similar problem).

[37] See Zielinski and Allendoerfer (1997).

[38] The assumption is made that a second order kinetics model can represent the reaction.

Table 8.8. *Simulated data drawn from Zielinski and Allendoerfer (1997).*

Time, t (s)	Concentration, C (mol/L)
0	0.01000
20000	0.00862
40000	0.00780
60000	0.00687
80000	0.00648
100000	0.00595
120000	0.00536
140000	0.00507
160000	0.00517
180000	0.00450
200000	0.00482
220000	0.00414
240000	0.00359
260000	0.00354
280000	0.00324
300000	0.00333
320000	0.00309
340000	0.00285
360000	0.00349
380000	0.00273
400000	0.00271

(5) Table 8.9 gives the temperature dependence of the energy gap of high purity crystalline silicon. The variation of energy gap with temperature can be represented by the equation

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{\beta + T}, \tag{8.50}$$

where $E_g(0)$ is the energy gap at absolute zero and α and β are constants.

Fit equation 8.50 to the data in table 8.9 to find best estimates of $E_g(0)$, α and β as well as standard uncertainties in the estimates. Use starting values, 1.1, 0.0004, and 600 respectively for estimates of $E_g(0)$, α and β .

(6) In an experiment to study phytoestrogens in Soya beans, an HPLC system was calibrated using known concentrations of the phytoestrogen, biochanin. Table 8.10 contains data of the area under the chromatograph absorption peak

Table 8.9. *Energy gap versus temperature data.*

T (K)	$E_g(T)$ (eV)
20	1.1696
40	1.1686
60	1.1675
80	1.1657
100	1.1639
120	1.1608
140	1.1579
160	1.1546
180	1.1513
200	1.1474
220	1.1436
240	1.1392
260	1.1346
280	1.1294
300	1.1247
320	1.1196
340	1.1141
360	1.1087
380	1.1028
400	1.0970
420	1.0908
440	1.0849
460	1.0786
480	1.0723
500	1.0660
520	1.0595

as a function of biochanin concentration.[39] A comparison is to be made of two equations fitted to the data in table 8.10. The equations are

$$y = A + Bx \tag{8.51}$$

and

$$y = A + Bx^c. \tag{8.52}$$

[39] Some of these data were published by Kirkup and Mulholland (2004).

Table 8.10. *HPLC data for biochanin.*

Conc. (x) (mg/L)	Area (y) (arbitrary units)
0.158	0.121342
0.158	0.121109
0.315	0.403550
0.315	0.415226
0.315	0.399678
0.631	1.839583
0.631	1.835114
0.631	1.835915
1.261	3.840554
1.261	3.846146
1.261	3.825760
2.522	8.523561
2.522	8.539992
2.522	8.485319
5.045	16.80701
5.045	16.69860
5.045	16.68172
10.09	34.06871
10.09	33.91678
10.09	33.70727

Assuming an unweighted fit is appropriate, fit equations 8.51 and 8.52 to the data in table 8.10.

For each equation fitted to the data, calculate the

(i) best estimates of parameters;
(ii) standard uncertainties in estimates;
(iii) sum of squares of residuals (SSR);
(iv) corrected Akaike's information criterion.
(v) Draw graphs of residuals versus concentration for each equation fitted to the data.

Which equation better fits the data?

(7) The relationship between critical current, I_c, and temperature, T, for a high-temperature superconductor can be written

$$I_c = 1.74A\left(1 - \frac{T}{T_c}\right)^{1/2} \tanh\left[0.435B\frac{T_c}{T}\left(1 - \frac{T}{T_c}\right)^{1/2}\right], \tag{8.53}$$

Table 8.11. *Critical current versus temperature data for a high temperature superconductor with critical temperature of 90.1 K.*

T (K)	I_c (mA)
5	5212
10	5373
15	5203
20	4987
25	4686
30	4594
35	4245
40	4091
45	3861
50	3785
55	3533
60	3199
65	2903
70	2611
75	2279
80	1831
85	1098
90	29

where A and B are constants and T_c is the critical temperature of the superconductor. The following data for critical current and temperature were obtained for a high-temperature superconductor with a T_c equal to 90.1 K.

(i) Draw a graph of I_c versus T.
(ii) Show that, as T tends to zero, equation 8.53 simplifies to $I_c = 1.74A$. Use this, and your graph, to estimate a starting value for A.
(iii) Estimate a starting value[40] for B.
(iv) Fit equation 8.53 to the data in table 8.11, to obtain best estimates for the parameters A and B.

[40] By choosing a data pair from table 8.11 and inserting the pair into equation 8.53 (as well as the estimate for A) that equation can be solved for B. This will give you an approximate value for B which is good enough to use with Solver.

Table 8.12. *Signal output from sensor as a function of electrical conductivity.*

σ (mS/cm)	V (volts)
1.504	6.77
2.370	7.24
4.088	7.61
7.465	7.92
10.764	8.06
13.987	8.14
14.781	8.15
17.132	8.19
24.658	8.27
31.700	8.31
38.256	8.34

(v) Determine the standard uncertainties in best estimates.
(vi) Determine the 95% coverage interval for A and B.

(8) A sensor developed to measure the electrical conductivity of salt solutions is calibrated using solutions of sodium chloride of known conductivity, σ. Table 8.12 contains data of signal output, V, of the sensor as a function of conductivity. Assume that the relationship between V and σ is

$$V = V_s + k[1 - \exp(\sigma^{\alpha})], \tag{8.54}$$

where V_s, k and α are constants.

Use unweighted non-linear least squares to determine best estimates of the constants and standard uncertainties in the best estimates.

(9) In a study of the propagation of an electromagnetic wave through a porous solid, the variation of relative permittivity, ε_r, of a solid was measured as a function of moisture content, v_w (expressed as a fraction). Table 8.13 contains the data obtained in the experiment.[41]

Assume the relationship between ε_r and v_w can be written

$$\varepsilon_r = v_w^2(\sqrt{\varepsilon_w} - \sqrt{\varepsilon_m})^2 + 2v_w(\sqrt{\varepsilon_w} - \sqrt{\varepsilon_m})\sqrt{\varepsilon_m} + \varepsilon_m, \tag{8.55}$$

where ε_w is relative permittivity of water, and ε_m is the relative permittivity of the (dry) porous material.

[41] Francois Malan 2002 (private communication).

Table 8.13. *Variation of relative permittivity with moisture content.*

v_w	ε_r
0.128	8.52
0.116	7.95
0.100	7.65
0.095	7.55
0.077	7.08
0.065	6.82
0.056	6.55
0.047	6.42
0.035	5.97
0.031	5.81
0.025	5.69
0.022	5.55
0.017	5.38
0.013	5.26
0.004	5.08

Use (unweighted) non-linear least squares to fit equation 8.55 to the data in table 8.13 and hence obtain best estimates of ε_w and ε_m and standard uncertainties in the best estimates.

(10) A common toughness test on steel is the Charpy V-notch test.[42] The energy that is absorbed by the specimen is measured during fracture over a range of temperatures. Table 8.14 shows the variation of absorbed energy, Y (in joules) as a function of temperature, T (in kelvins), for samples of heat-treated steel. An equation that is often fitted to the data is given by

$$Y = A + B\tanh\left[\frac{T - T_o}{C}\right], \tag{8.56}$$

where A, B, C and T_o are parameters that may be estimated using non-linear least squares.

(i) Draw an x–y graph of the data in table 8.14.
(ii) Determine approximate values for A, B, C and T_o.
(iii) Use non-linear least squares to find best estimates of A, B, C and T_o.
(iv) Determine the standard uncertainties in A, B, C and T_o.

[42] See Mathur, Needleman and Tvergaard (1994).

Table 8.14. *Variation of absorbed energy with temperature for heat treated steel.*

T (K)	Y (J)
270	11.8
280	6.6
290	16.3
300	17.5
310	22.7
320	22.7
330	32.7
340	66.2
350	90.7
360	98.8
370	114.8
380	113.4
390	113.1
400	110.1
410	114.8
420	108.4

(11) Unweighted least squares requires the minimisation of *SSR* given by

$$SSR = \sum (y_i - \hat{y}_i)^2. \tag{8.57}$$

A technique sometimes adopted when optimising parameters in optical design situations is to minimise *S4R*, where

$$S4R = \sum (y_i - \hat{y}_i)^4. \tag{8.58}$$

Carry out a simulation to compare parameter estimates obtained when equations 8.57 and 8.58 are used to fit an equation of the form $y = a + bx$ to simulated data. More specifically do the following.

(i) Use the function $y = 2.1 - 0.4x$ to generate y values for $x = 1, 2, 3$, etc., up to $x = 20$.
(ii) Add normally distributed noise of mean equal to zero and standard deviation of 0.5 to the values generated in part (i).
(iii) Find best estimates of a and b by minimising *SSR* and *S4R* as given by equations 8.57 and 8.58. (Use Solver to minimise *SSR* and *S4R*.)

(iv) Repeat steps (ii) and (iii) until 50 sets of parameter estimates have been obtained using equations 8.57 and 8.58.
(v) Is there any significant difference between the parameter estimates obtained when minimising *SSR* and *S4R*?[43]
(vi) Is there any significant difference between the variance in the parameter estimates when minimising *SSR* and *S4R*?

[43] Return to parts (v) and (vi) of this question after reading chapter 9.