WinDat Thematic Network

www.windat.org

Windows as Renewable Energy Sources for Europe
Window Energy Data Network

Project supported by DG for Energy and Transport
of the European Commission

contract NNE5-2000-122

Solving non-linear data fitting problems using DataFit 8.0

Frank Pedersen

Technical University of Denmark
Department of Civil Engineering
Brovej, Building 118
DK-2800 Kgs. Lyngby

WinDat document N2.10 - Public **March 2004**

This document was produced within WinDat, with active contribution by the members of this European Thematic Network. See www.windat.org for more information.

Contact:

On this report:

The authors:
Email: Frank Pedersen fp@byg.dtu.dk

On Thematic Network WinDat:

WinDat coordinator: Dick van Dijk, TNO Building and Construction Research, Delft, The Netherlands
Email: H.vanDijk@bouw.tno.nl

Research funded in part by
THE EUROPEAN COMMISSION
in the framework of the Community activities
in the Field of the specific programme for RTD and
demonstration on Energy, Environment and Sustainable
Development

BYG DTU

Frank Pedersen

Solving non-linear data fitting problems using DataFit 8.0

TECHNICAL
UNIVERSITY
OF DENMARK

BYG · DTU
marts 2004

Index

1 Introduction

This note concerns solving non-linear data fitting problems of the type described in [1], using the application DataFit 8.0. The note is outlined as follows: The data fitting problem is described in this section. Section 2 concerns the regression model used for representing the data. In section 3, some general information about DataFit is provided. Section 4 concerns practical aspects of the use of DataFit, i.e. how to define a new project (section 4.1), how to implement the model (section 4.2), how to solve the data fitting problem (section 4.3), and finally, how to evaluate the model, once the optimal model parameters have been estimated (section 4.4).

The purpose of the method described in [1] is to enable the user to express the linear thermal transmittance Ψ of the interaction of window frames and glazing, as a function of the spacer profile L-value, the glass thickness d, and the center U_g-value of the glazing unit.

It is assumed that m detailed calculations of Ψ are available, for various combinations of L, d and U_g. The available dataset is denoted

$$\left\{\left(x_1,\Psi_1\right),\dots,\left(x_m,\Psi_m\right)\right\}, \tag{1.1}$$

where $x_i=\left[L_i,d_i,U_{g,i}\right]^{\mathrm{T}}$, $i=1,\dots,m$ is the independent data, and Ψ_i, $i=1,\dots,m$ is the dependent data. The detailed calculations of Ψ are represented using a regression model

$$\hat{\Psi}:\mathbb{R}^n\times\mathbb{R}^3\to\mathbb{R}, \tag{1.2}$$

where the aim is to find a set of coefficients, or model parameters $b^*\in\mathbb{R}^n$, such that

$$\hat{\Psi}\left(b^*,x_i\right)\simeq\Psi_i,\ i=1,\dots,m. \tag{1.3}$$

The model parameters b^* are estimated by solving the following non-linear data fitting problem:

$$b^*=\arg\min_{b\in\mathbb{R}^n}R\left(b\right), \tag{1.4}$$

where

$$R\left(b\right)=\left\|r\left(b\right)\right\|_q, \tag{1.5}$$

and where q is either 1, 2 or ∞. The implications of this choice will not be considered in this note. The vector function $r:\mathbb{R}^n\to\mathbb{R}^m$ is the residual between the model and the data:

$$r_i\left(b\right)=\hat{\Psi}\left(b,x_i\right)-\Psi_i,\ i=1,\dots,m. \tag{1.6}$$

If the model (1.2) is intended for predicting data at previously unknown points, it is advisable to ensure that the model is "conservative", i.e. it does not predict better values than the known observations. Since the model (1.2) is intended for representing Ψ, small values are better than large ones. A conservative parameter estimate in this case therefore means estimating a set of model parameters, where the values obtained with the model are larger than the observations, i.e. all residuals must be positive in the solution to (1.4).

This means that the model parameters b^* must be estimated by solving the following data fitting problem with general constraints:

$$b^* = \arg\min_{b\in\mathbb{R}^n} R(b) \text{ subject to } c(b) \geq 0, \tag{1.7}$$

where $c:\mathbb{R}^n \to \mathbb{R}^l$ is the constraint function. In (1.7), the statement $c(b) \geq 0$ is true if and only if the statements $c_i(b) \geq 0$, $i = 1,\ldots,l$ are true. Using the following constraint function:

$$c(b) = r(b), \tag{1.8}$$

therefore provides a conservative parameter estimate.

2 The regression model

This section concerns a simple regression model, intended for representing the detailed calculations of Ψ.

The model is based on algebraic expressions, chosen such that the tendency of the data presented in [2], Figures 7 and 8 is captured. The linear thermal transmittance Ψ depends on L in a way that resembles a power model:

$$\hat{\Psi}_L = b_1 \cdot x_1^{b_2}. \tag{1.9}$$

The following linear correction term is added, in order to give a more precise fit:

$$c_L = b_3 + b_4 \cdot x_1. \tag{1.10}$$

Inserting a pane with different thickness d can be done either by keeping the inner or the outer distance between the panes fixed. Keeping the inner distance fixed may result in an altered frame construction, depending on the width of the weather strips. Keeping the outer distance fixed results in an altered edge construction. In order to accommodate the dependency of d, the following quadratic correction term is added:

$$c_d = b_5 \cdot x_2 + b_6 \cdot x_2^2. \tag{1.11}$$

Finally, in order to accommodate the dependency of the conductance U_g of the center of the glazing unit, the following linear correction term is added:

$$c_{U_g} = b_7 \cdot x_3 . \tag{1.12}$$

The regression model used for representing Ψ thus becomes:

$$\begin{aligned} \hat{\Psi}(b,x) &= \hat{\Psi}_L + c_L + c_d + c_{U_g} \\ &= b_1 \cdot x_1^{b_2} + b_3 + b_4 \cdot x_1 + b_5 \cdot x_2 + b_6 \cdot x_2^2 + b_7 \cdot x_3 \end{aligned} \tag{1.13}$$

where

$$b = [b_1, \ldots, b_7]^T , \tag{1.14}$$

are the model parameters, and where

$$x = [x_1, x_2, x_3]^T , \tag{1.15}$$

are the sample parameters.

3 An introduction to DataFit 8.0

3.1 General information

DataFit provides possibilities for fitting a wide range of non-linear models to a given data set. Version 8.0 provides 298 built-in 2D models, and 242 3D models. The application also enables the user to define models. By default, derivatives of user-defined models are obtained using finite differences, but if desired, the user can implement analytical derivatives.

The Levenberg-Marquardt method is used when solving non-linear regression problems, and Singular Value Decomposition is used when solving linear regression problems.

Once a solution estimate is obtained, the application provides sensitivity information, and other statistical information about the solution.

The data can depend on up to 20 independent parameters, and regression models can depend on up to 100 model parameters.

Homepage: www.oakdaleengr.com.

3.2 License

A 30-day evaluation copy can be downloaded from the homepage, free of charge. A license costs US$ 229 when downloading the software from the Internet.

3.3 Remarks

The current version only supports unconstrained weighted least-squares estimates, where the weighting factors for the data points are the reciprocal of the standard deviation. Performing a conservative parameter estimate by solving (1.7), is therefore not possible. Minimizing the maximum residual, i.e. using $q = \infty$ in (1.5), is also not possible.

The current version does not provide means for checking user-defined derivatives of the residual function (1.6) (e.g. comparing them with finite difference approximations).

It is not possible to make (Ψ, L) plots for various combinations of d and U, as the ones used in [1].

In the documentation for the software, data with $n-1$ independent parameters is referred to as n-D regression, since it requires n axes when plotting the data.

4 Solving non-linear data fitting problems using DataFit 8.0

This section concerns the use of DataFit 8.0 for solving non-linear data fitting problems. In order to solve the data fitting problem described in [1], the user is required to perform the following tasks:

1. Define a new project, which is described in section 4.1.
2. Define a regression model, which is described in section 4.2.
3. Solve the data fitting problem, which is described in section 4.3.
4. Evaluate the model, which is described in section 4.4.

In DataFit, the sample parameters are denoted $x_1, \ldots, x_3$. For the model parameters, any symbol other than $x_1, \ldots, x_3$ and y can be used.

4.1 Defining a new project

First the user is required to define a new project. This is done by starting DataFit, and choosing the item "New" in the menu "File". This opens up the "New Project" dialog box, shown in Figure 4.1. The user is required to define the number of independent variables, in this case 3.

The next step is to enter the data set (1.1). Double-clicking the column header gives a dialog box, which can be used for entering a description of the data in the column.

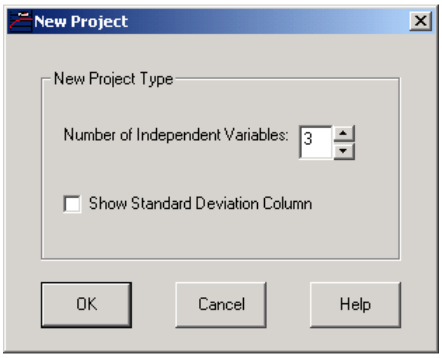

Figure 4.1. The "New Project" dialog box. In this case 3 independent variables are needed.

Once the data has been entered, the main window looks like Figure 4.2. The data is the same as the one used in [1].

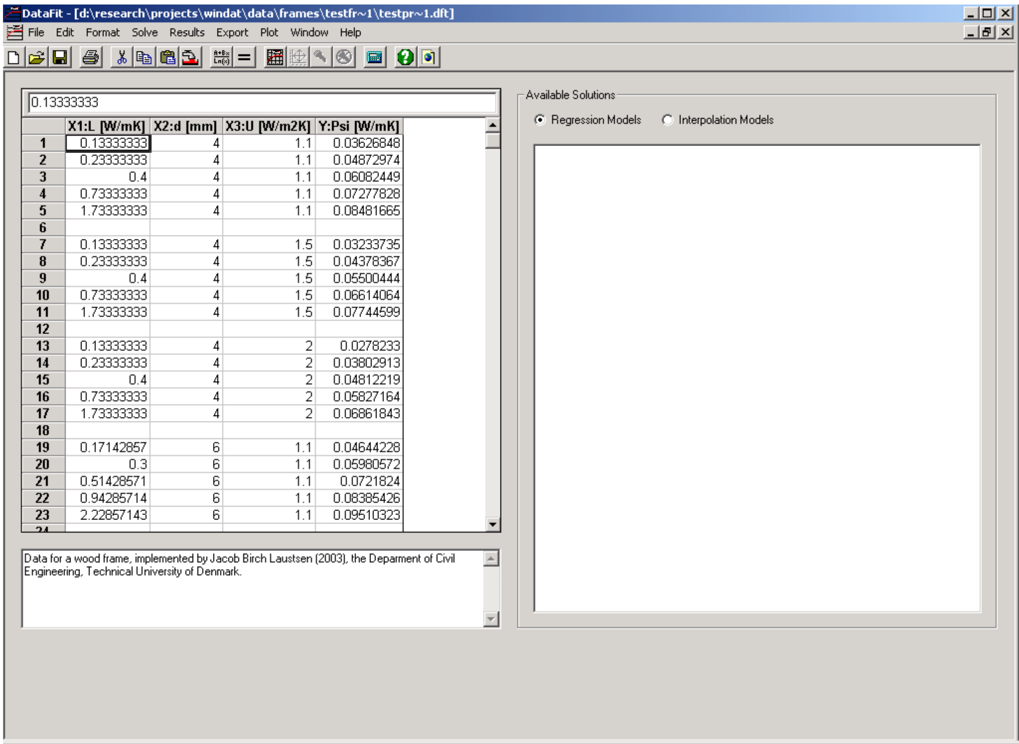

Figure 4.2. The main window after the data set has been entered.

4.2 Implementing the model

In order to define a regression model in DatatFit, the following steps are needed:

1. In the menu "Solve", select the item "Define User Model...". This opens the "User Defined Models" window, shown in Figure 4.3.

2. Click the "New..." button. This opens the "Model Editor" window, shown in Figure 4.4.

3. To define a model, the user has to provide a model identification string, which can be entered in the text field labeled "Model ID", and the expression for the model, which can be entered in the text field labeled "Model Defini-

tion". In Figure 4.4, a string representing (1.13) is entered. In the text field labeled "Description", the user can enter a description of the model (optional). Once the model is defined, click "OK" to accept the changes.

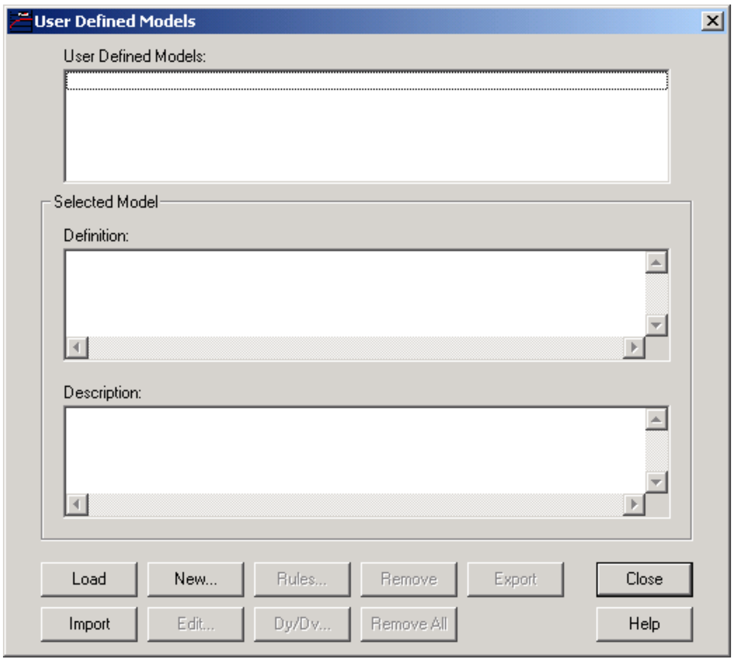

Figure 4.3. The "User Defined Models" window. This is used for adding/removing regression models to/from the current project.

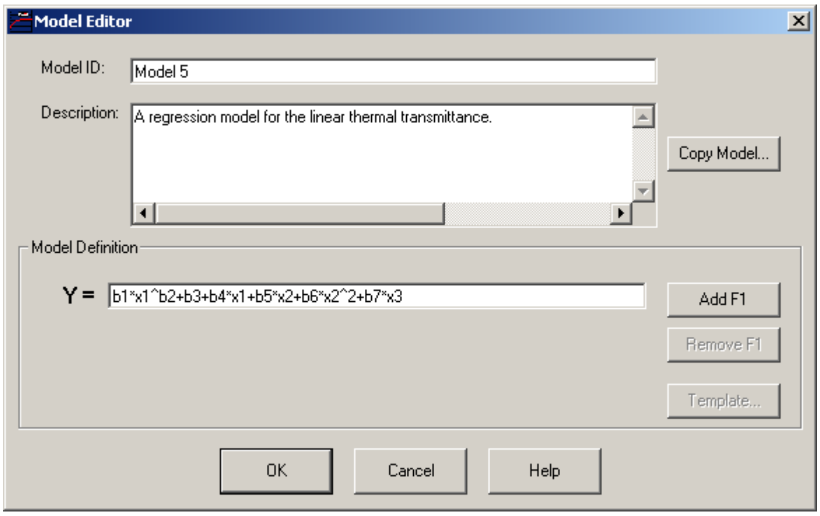

Figure 4.4. The "Model Editor" window. This is used for creating or editing user defined regression models.

4.3 Solving the data fitting problem

When a dataset is entered, and a regression model is defined, a weighted lest squares solution for the data fitting problem (1.4) can be estimated. Solving non-linear optimization problems is done in an iterative way, by solving a sequence of approximated sub-problems. The user therefore has to provide the following information:

1. An initial estimate for the model parameters (a "starting point"), used for initializing the algorithm.
2. Tolerance level for the final solution estimate, which determines when the algorithm terminates.
3. Maximum number of iterations, which is a safeguard feature, that ensures that the algorithm always terminates, regardless of the numerical properties of the regression model.

The starting point is defined by opening the "User Defined Model" window, using the "Define User Model..." item in the "Solve" menu. Clicking the button labeled "Rules...", opens the "User Defined Model Rules" window, shown in Figure 4.5.

The starting point can be defined using a number of rules, where the simplest one is to define it as a fixed value. All other rules define the starting point relative to a value derived from the data, for instance the largest or smallest dependent data, the average dependent data, etc.

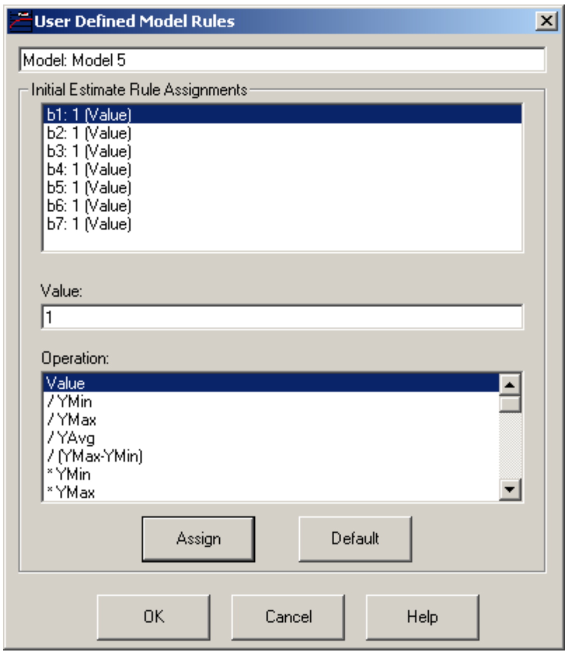

Figure 4.5. The "User Defined Model Rules" window, which is used for defining starting points.

The tolerance level and the maximum number of iterations can be defined in the "Edit" menu, under the item "Preferences" ► "Solution...", which opens the "Solution Preferences" window shown in Figure 4.6. Here, the default values are used.

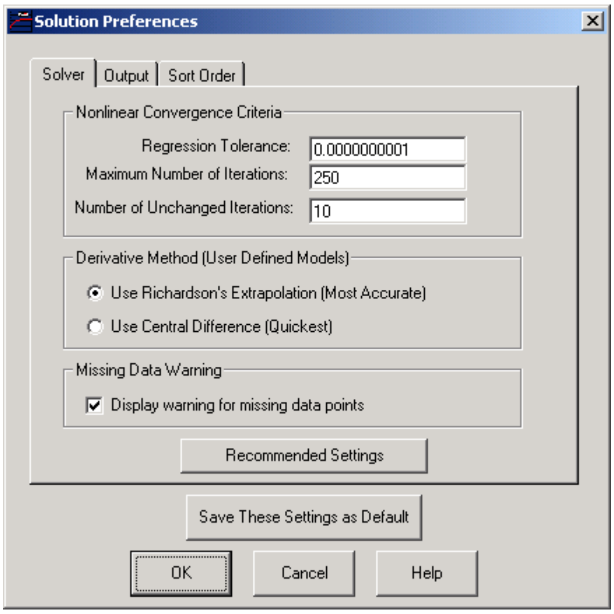

Figure 4.6. The "Solution Preferences" window, where the tolerance level and maximum number of iterations are defined.

When the starting point and solution preferences are defined, the data fitting problem is solved in the following way:

1. Open the "Solution Setup" window, shown in Figure 4.7, using the "Regression" item in the "Solve" menu. Check "Single Model" and "Nonlinear", and then click "OK".

2. This opens the "Single Model Regression Setup" window, shown in Figure 4.8. Click "Solve". This should hopefully result in a message window with the message "Solution was successful".

3. Close the message window.

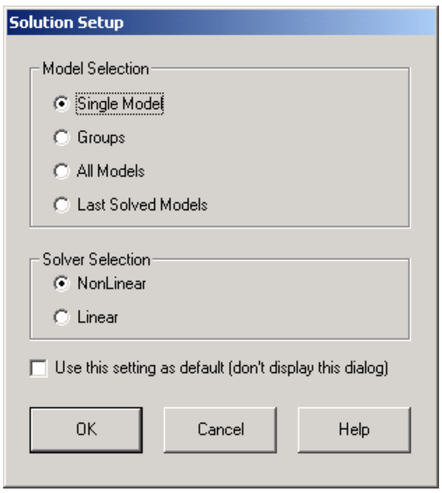

Figure 4.7. The "Solution Selection" window.

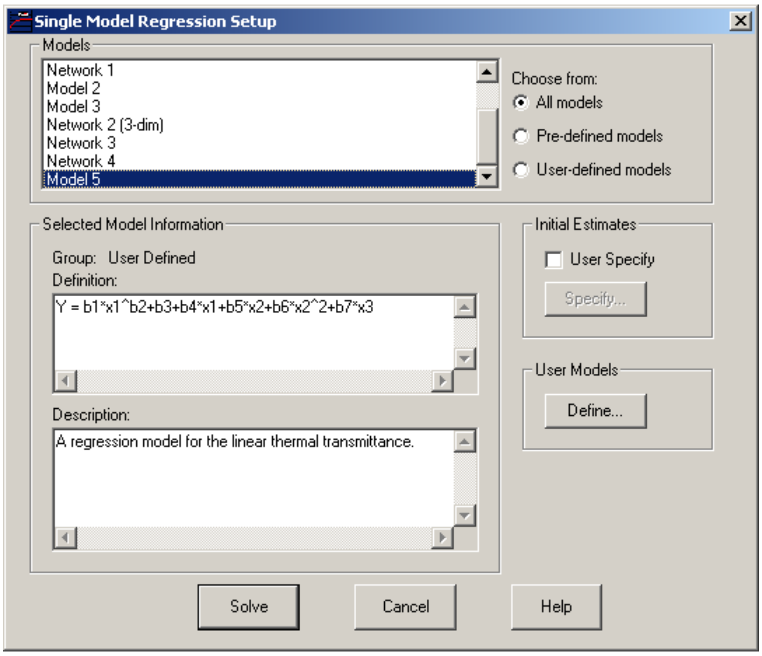

Figure 4.8. The "Single Model Regression Setup" window.

4.4 Evaluating the model

In the main window, in the frame labeled "Available Solutions Sorted By RSS", there should now be a line reading "Model 5". The item labeled "Detailed..." in the "Results" menu opens the "Regression Results" window, shown in Figure 4.9.

This window provides information about the solution, amongst other number of iterations, average residual, the sum of the squares of the residuals, etc. The section labeled "Regression Variable Results" contain the estimated optimal regression model parameters, in this case:

$$b^* = \begin{bmatrix} 4.538179680 \cdot 10^{-1} \\ 5.427638824 \cdot 10^{-2} \\ 2.186555135 \\ -8.815026056 \cdot 10^{-3} \\ -1.060223109 \\ 1.063484956 \cdot 10^{-1} \\ -1.388614808 \cdot 10^{-2} \end{bmatrix} \tag{3.1}$$

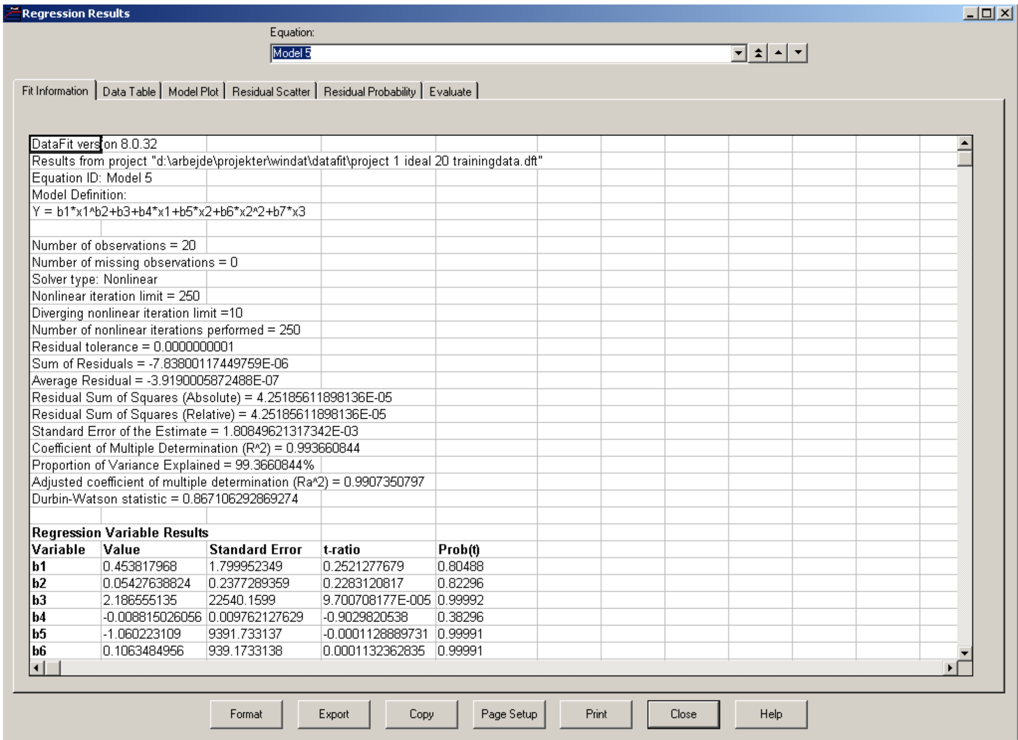

Figure 4.9. The "Regression Results" window, which provides statistical details about the solution.

5 References

[1] Frank Pedersen, Jacob Birck Laustsen and Svend Svendsen (2003, preliminary version), *A method for characterizing the thermal properties of windows frame profiles*, unpublished report, Department of Civil Engineering, Technical University of Denmark.

[2] Jacob Birck Laustsen and Svend Svendsen (2003), *WinDat WP2.3 – Edge seals, Frames and Windows - Edge constructions and frames*, unpublished report, Department of Civil Engineering, Technical University of Denmark.