

THE USE OF PIECEWISE LINEAR MODELS
TO PREDICT HYDROELECTRIC LOAD FOR
MANITOBA HYDRO

BY
MARC SILVA

A Practicum Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

©April, 2002



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-76869-4

**THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION PAGE**

**THE USE OF PIECEWISE LINEAR MODELS TO PREDICT
HYDROELECTRIC LOAD FOR MANITOBA HYDRO**

BY

Marc Silva

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University
of Manitoba in partial fulfillment of the requirements of the degree**

of

MASTER OF SCIENCE

MARC SILVA © 2002

Permission has been granted to the Library of The University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilm Inc. to publish an abstract of this thesis/practicum.

The author reserves other publication rights, and neither this thesis/practicum nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Abstract

Techniques for forecasting hourly load and monthly peak load have been developed and used at Manitoba Hydro for many years. One part of the methodology involves temperature-adjusting the historical load data. This process involves fitting piecewise linear models to the relationship between temperature and load. In this practicum we investigate models with one or two knots, or turning points, with the locations of the knots being either predetermined or estimated from the data. If the knots are predetermined, then the standard theory for linear models can be utilized. However, if the knots are estimated from the data, then this involves fitting nonlinear models. In this latter case, iterative procedures such as "Dud" may be used, and approximate confidence intervals and regions for unknown parameters may be constructed, based on linear approximations to the nonlinear problem. In the problems that are investigated it is shown that the approximate intervals and regions provide good approximations to the likelihood intervals based on the profile sum of squares functions. This is due to the fact that there is a small degree of nonlinearity in the model-data set combinations.

For the data sets considered, the two-knot model with the knots estimated from the data, appears to provide the best fit. It is also shown that the addition of variables such as day, year, and month increase the predictive power of the model. Future research in this area will focus on the utilization of these additional variables in random coefficient, piecewise linear models, in order to provide a better understanding of the influence of noise factors, such as temperature, on load forecasts.

Acknowledgements

I would first like to thank my supervisor Dr. John F. Brewster for introducing me to this project, and for all his help and support since day one. I would also like to thank the members of my committee, Dr. Tom Berry and Dr. Dean Slonowsky, for taking time out of their busy schedules and helping with the proof reading of the practicum and offering their suggestions and opinions. Finally, I would like to thank Manitoba Hydro with awarding a research grant to the Institute of Industrial Mathematical Sciences at the University of Manitoba. Part of the grant was used as funding for my practicum. The financial support was very much appreciated.

Contents

1 Introduction and Overview	1
1.1 Preamble	1
1.2 Scope of the Practicum	2
1.3 Piecewise Linear Models for Load-Temperature Relationships	3
2 Fitting Piecewise Linear Models With Predetermined Knots—The Linear Case	7
2.1 Introduction	7
2.2 Linear Regression in General	8
2.3 Least Squares Estimation and Inference	12
2.4 Fitting Piecewise Linear Models with One Predetermined Knot to Manitoba Hydro Data	15
2.4.1 The Model	15
2.4.2 Least Squares Inference	17
2.4.3 Assessing The Model	20
2.5 Fitting Piecewise Linear Models with Two Predetermined Knots to Manitoba Hydro Data	22
2.5.1 The Model	23
2.5.2 Least Squares Inference	23
2.5.3 Assessing The Model	27

2.6	Conclusion	28
3	Fitting Piecewise Linear Models With Unknown Knots—The Non-linear Case	30
3.1	Introduction	30
3.2	Nonlinear Least Squares	31
3.3	Iterative Methods for Finding Least Squares Estimates in Nonlinear Regression	33
3.3.1	The Gauss-Newton Procedure	34
3.3.2	The Grid Search Procedure	38
3.3.3	The Dud (doesn't use derivatives) Procedure	39
3.4	Fitting Piecewise Linear Models with One Unknown Knot to Manitoba Hydro Data	43
3.4.1	The Model	43
3.4.2	Using the Grid Search Method	44
3.4.3	Using the Dud Method	48
3.5	Fitting Piecewise Linear Models with Two Unknown Knots to Manitoba Hydro Data	54
3.5.1	The Model	54
3.5.2	Using the Grid Search Method	55
3.5.3	Using the Dud Method	60
3.6	Conclusion	65
4	Further Aspects	67
4.1	Introduction	67
4.2	Comparing the Models	68
4.2.1	The Linear Models	68
4.2.2	The Nonlinear Models	69

4.2.3	Linear vs. Nonlinear Models	71
4.3	The Profile Likelihood Intervals Based on the Sum of Squares Function	73
4.3.1	The Piecewise Linear Model With One Unknown Knot	73
4.3.2	The Nonlinear Model With Two Knots	75
4.4	Advantages and Disadvantages of Linear Least Squares	78
4.4.1	The Advantages	78
4.4.2	The Disadvantages	78
4.5	Advantages and Disadvantages of Nonlinear Least Squares	80
4.5.1	The Advantages	80
4.5.2	The Disadvantages	82
4.6	Conclusion	86
5	Modifying The Model	88
5.1	Introduction	88
5.2	The Time of Day Affect	89
5.3	The Day of the Week Affect	91
5.4	The Seasonal Affect	95
5.5	The Year Affect	98
5.6	Conclusion	102
6	Conclusion	103
A	The Data	106
B	SAS Code	108
B.1	PROC REG code for linear models	108
B.2	PROC NLIN code for nonlinear models	110
	Bibliography	113

List of Figures

1.1	Plot of load vs. time from April 1, 1992 to March 31, 1993, taken at 6pm.	4
1.2	Plot of temperature vs. time from April 1, 1992 to March 31, 1993, taken at 6pm.	4
1.3	Plot of load vs. temperature from April 1, 1992 to March 31, 1993, taken at 6pm.	5
2.1	Plot of load vs. temperature for all data at 6pm.	16
2.2	Piecewise linear fit with one predetermined knot for 6pm data.	17
2.3	SAS output for piecewise linear fit with one predetermined knot for 6pm data.	18
2.4	Plot of residuals vs. temperature based on piecewise linear fit with one predetermined knot.	21
2.5	Normal probability plot of residuals.	21
2.6	Plot of load vs. temperature for all data at 6pm.	22
2.7	Piecewise linear fit with two predetermined knots for data at 6pm.	24
2.8	SAS output for piecewise linear fit with two predetermined knots for data at 6pm.	25
2.9	Plot of residuals vs. temperature based on piecewise linear fit with two predetermined knots.	27
2.10	Normal probability plot of residuals.	28

3.1	Plot of $S^*(\hat{\theta}_3)$ vs. θ_3	45
3.2	SAS output for piecewise linear fit using 17.09°C as the unknown knot.	46
3.3	Piecewise fit using Grid Search method with 17.09 as the unknown knot.	46
3.4	SAS output showing the iterations from the Dud method.	50
3.5	SAS output for nonlinear fit using the Dud method.	51
3.6	Nonlinear fit for piecewise linear model with one unknown knot.	52
3.7	Plot of residuals vs. temperature for piecewise linear model with one unknown knot.	53
3.8	Normal probability plot of residuals.	53
3.9	3-Dimensional plot of $S^*(\hat{\theta}_3, \hat{\theta}_4)$ vs. θ_3 and θ_4	56
3.10	SAS output for piecewise linear fit using 14.25°C and 20.70°C as the unknown knots.	57
3.11	Piecewise fit using Grid Search method with 14.25 and 20.70 as the unknown knots.	58
3.12	SAS output showing iterations from the Dud method.	61
3.13	SAS output for nonlinear fit using the Dud method.	63
3.14	Nonlinear fit for piecewise linear model with two unknown knots.	64
3.15	Plot of residuals vs. temperature for piecewise linear model with two unknown knots.	64
3.16	Normal probability plot of residuals	65
4.1	Sum of squares function for Hydro data for nonlinear one knot model. The solid straight line corresponds to the profile likelihood interval.	74
4.2	Plot of sum of squares function, $S'(\hat{\theta}_3)$ vs. θ_3 . The solid straight line corresponds to the profile likelihood interval.	76
4.3	Plot of sum of squares function, $S'(\hat{\theta}_4)$ vs. θ_4 . The solid straight line corresponds to the likelihood interval.	77
4.4	Sum of squares contours for two knot nonlinear model.	81

4.5	Proc NLIN output for BOD data.	84
4.6	Sum of squares contours for the BOD data.	85
4.7	Sum of squares function for BOD data for θ_1	86
5.1	Plots of θ_1 , θ_2 , θ_3 and θ_4 each versus hour of day.	90
5.2	Plot of residuals vs. temperature for piecewise linear fit with two unknown knots.	92
5.3	Plot of residuals vs. day of the week for original piecewise linear model with two unknown knots.	92
5.4	PROC NLIN output for Hydro data using nonlinear model with added day component.	94
5.5	Plot of residuals vs. day of the week after adjusting nonlinear model for day variable.	95
5.6	Plot of residuals vs. month for our original piecewise linear model with two unknown knots.	96
5.7	PROC NLIN output for Hydro data using nonlinear model with added season component.	97
5.8	Plot of residuals vs. month after adjusting nonlinear model for seasonal variable.	98
5.9	Plot of residuals vs. year for our original nonlinear piecewise model with two unknown knots.	99
5.10	PROC NLIN output for Hydro data using nonlinear model with added year component.	100
5.11	Plot of residuals vs. year after adjusting nonlinear model for year variable.	102

List of Tables

4.1	The extra sum of squares analysis for nested models.	70
4.2	The extra sum of squares analysis for the 4 and 5 parameter piecewise model fitted to the Manitoba Hydro data.	70
4.3	The extra sum of squares analysis for the 3 and 4 parameter piecewise model fitted to the Manitoba Hydro data.	71
4.4	The extra sum of squares analysis for the 3 and 5 parameter piecewise model fitted to the Manitoba Hydro data.	72
4.5	BOD versus time.	83
5.1	θ_i versus hour of day.	90
5.2	The extra sum of squares analysis for the 5 and 8 parameter piecewise model fitted to the Manitoba Hydro data.	94
5.3	The extra sum of squares analysis for the 5 and 8 parameter piecewise model fitted to the Manitoba Hydro data.	97
5.4	The extra sum of squares analysis for the 10 and 18 parameter piecewise model fitted to the Manitoba Hydro data.	101
A.1	Partial data set collected by Manitoba Hydro	106

Chapter 1

Introduction and Overview

1.1 Preamble

For electrical companies, such as Manitoba Hydro, it is very important to be able to accurately forecast the amount of power that they will need to produce at a particular point in time. Capacity in excess of that required by domestic customers is becoming increasingly more valuable, as excess capacity can be sold to other utilities. The value of excess capacity is dependent on when it is sold, and accurate prediction of anticipated demand is critical in maximizing the value of any excess. Effective cost-benefit analysis requires reliable assessments of the uncertainties associated with predictions of excess capacity, and achieving a good understanding of the influence of temperature and other noise factors on predictions is particularly crucial. Estimates of future load requirements affect many management decisions. Areas that are affected include capital investment in generation, transmission and distribution, purchasing decisions on fuel, and tariffs and revenue calculations [10].

Techniques for forecasting hourly load and monthly peak loads have been developed and used at Manitoba Hydro for many years. During this time the models have undergone a number of changes, and the process has evolved. In the past the tradi-

tional focus has been on forecasting the annual peak, which occurs during the winter months. Now, however, there are more opportunities for export sales in the summer months and, as a consequence, there is increased interest in the ability of the current models to forecast summer monthly peaks. Manitoba Hydro has decided that it is an appropriate time to examine the statistical properties of the current forecasting methods, to see whether refinements should be introduced.

As part of this review, the Research and Development Program of Manitoba Hydro awarded a research grant to the Institute of Industrial Mathematical Sciences at the University of Manitoba for a project titled “Robust Modelling and Prediction of the Demand for Electrical Power”. The primary focus of the research is on achieving a better understanding of the influence of noise factors on load forecasts. This practicum forms a part of this research.

1.2 Scope of the Practicum

The methods that Manitoba Hydro uses to forecast demand are complex. The models use factors such as historical load data, historical temperature data and type of customer, among others. Moreover, different assumptions are made concerning future economic, demographic and market conditions, in order to reflect the sensitivity of demand to each of these factors.

In this practicum we will not be examining Manitoba Hydro’s forecasting methods in detail. Rather, we will look at one aspect of the methodology, which is related to the need to temperature-adjust the historical load data in some manner.

The methodology that Manitoba Hydro uses for temperature-adjusting the load data involves fitting piecewise linear models to load-temperature relationships. In this practicum our primary focus will be on an examination of the nonlinear regression methods that are utilized for this purpose. An important issue is the extent to which

linear approximation methods can be used for statistical inferences in such problems.

The research in this practicum will play an important role in the larger research project mentioned above. The intention in the larger project is to use piecewise linear models for load-temperature relationships, as in this practicum, but to also incorporate random coefficients into the models, in order to capture variability across other factors. These piecewise linear, random coefficient, models will then be used in conjunction with historical weather data to examine different weather scenarios, in order to get an idea of the potential load extremes. This will help to produce reliable estimates of the uncertainties associated with the forecasts.

1.3 Piecewise Linear Models for Load-Temperature Relationships

In order to understand the need to temperature-adjust the historical load data, let us look at a few different plots. Figure 1.1 plots load versus time for the period of April 1, 1992 to March 31, 1993, recorded at 6 pm. Now let us compare this plot with a plot of temperature versus time for the same time period, given in Figure 1.2.

It is obvious by comparing the two plots that there is some sort of relationship between load and temperature. Figure 1.3 plots load versus temperature for the same time period. We can see that the load is high for lower and higher temperatures. There seems to be a “V” shape to the plot.

In this practicum the focus is on fitting piecewise linear models to data such as that in Figure 1.3. The models considered will have one or two knots, or turning points. We will tackle several problems concerning these models. Should we use one knot or two knots? Should we fix the knots at certain predetermined temperatures, or should we estimate the location of the knots from the data? How do we attach precision to the resulting estimates? For our purposes we will look at all 6 pm data

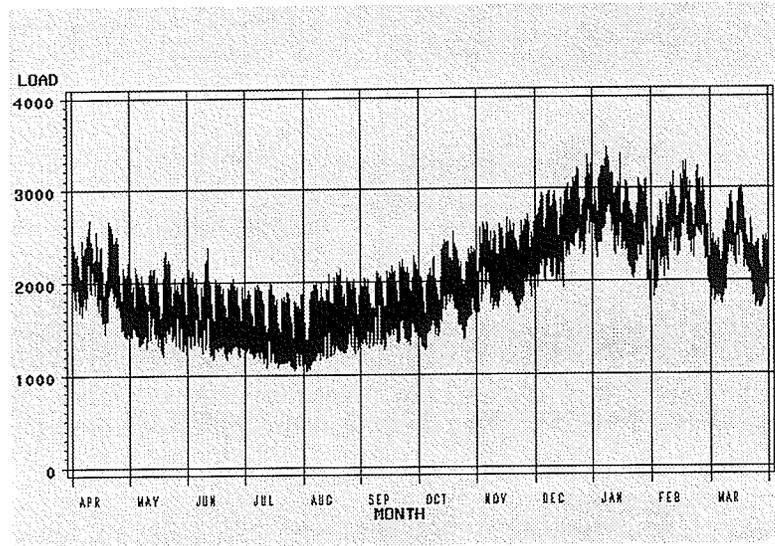


Figure 1.1: Plot of load vs. time from April 1, 1992 to March 31, 1993, taken at 6pm.

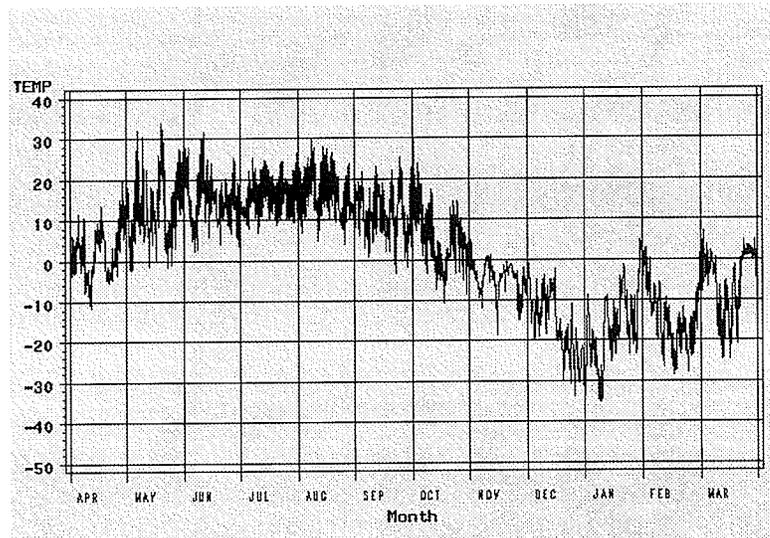


Figure 1.2: Plot of temperature vs. time from April 1, 1992 to March 31, 1993, taken at 6pm.

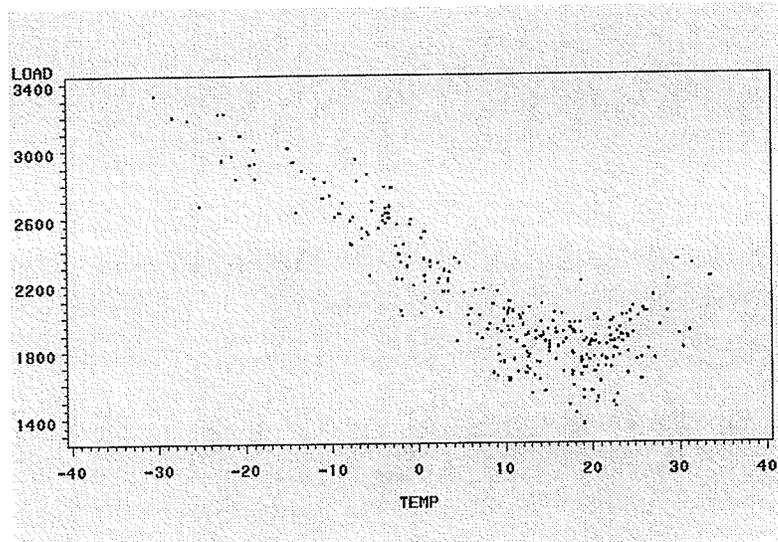


Figure 1.3: Plot of load vs. temperature from April 1, 1992 to March 31, 1993, taken at 6pm.

from 1992 to the present and examine each of these questions in turn. A portion of the data set is shown in Appendix A, along with an explanation of the variables used.

When we assume that the knots are fixed we have a “linear” model, i.e., a model that is linear in the unknown parameters. In this linear case, statistical procedures for estimating parameters, and for constructing confidence intervals and confidence regions for the unknown parameters are standard and well-known. In chapter 2 we introduce the linear case, and fit one and two knot models to the 6 pm data using the least-squares criterion. Confidence intervals and regions for unknown parameters are constructed, and residual plots are examined in order to assess the adequacy of the models.

When the knots are estimated from the data, however, we have a “nonlinear” model, because the model is nonlinear in the unknown parameters. In this nonlinear case, statistical procedures are also available. However, these procedures are typically based on linear approximations to the nonlinear models and such approximations may or may not work well. So one of the questions to be addressed in this practicum is the adequacy of the inference procedures based on these linear approximations.

Another issue in the nonlinear case concerns the use of the “Dud” (doesn’t use derivatives) iterative procedure that is typically used for fitting piecewise linear models with unknown knots. Approximate confidence intervals and confidence regions can be constructed in conjunction with the Dud method, but again the adequacy of the approximations needs to be addressed.

In chapter 3 we introduce the nonlinear regression models, and use iterative nonlinear least-squares procedures to fit these models to the 6 pm data. Approximate confidence intervals and confidence regions are also constructed. In chapter 4 these approximate intervals are compared to more appropriate likelihood intervals and regions based on the profile sums of square functions. In chapter 4 we also discuss the advantages and disadvantages of the linear and nonlinear cases.

In chapter 5 we briefly look at modifications to the models that incorporate other variables such as day of the week, year and month, in an additive fashion. The intention is not to choose a final model, but rather to see which of the additional variables might be important in subsequent research on random coefficient models, as mentioned above.

Finally in chapter 6 we give some concluding remarks and discuss some of the further work that could be done.

Chapter 2

Fitting Piecewise Linear Models With Predetermined Knots—The Linear Case

2.1 Introduction

One of our main goals is to study the relationship between load and temperature in the Manitoba Hydro data. In this chapter we use piecewise linear models with predetermined knots to describe the relationship between load and temperature. With the knots being fixed, we have a “linear” model. What we mean by “linear” is that the models are linear in the parameters. For this linear case, well-known statistical procedures to construct confidence intervals and confidence regions for the unknown parameters are used. In section 2.2 linear regression is introduced. We look at how general piecewise linear models with one or two predetermined knots can be fit to the data. In section 2.3 we introduce the least squares procedure, which is the method most widely used to choose appropriate parameter estimates for regression models. We also show how to construct confidence intervals for our parameters.

In section 2.4 we fit a piecewise linear model with one fixed knot to 6pm data of load versus temperature. We then study the residuals to determine if our model assumptions are met. Then in section 2.5 we fit a piecewise linear model with two fixed knots to the same data. Again, we study the residuals and see if the assumptions for our two knot model are valid.

2.2 Linear Regression in General

We use regression analysis to predict one quantitative variable based on its relationship with another quantitative variable [7]. A simple example of a regression model is the straight line relationship between a pair of variables. The model can be written as follows,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (2.1)$$

where y_i is the value of the response variable,

β_0 and β_1 are unknown parameters,

x_i is a known independent variable,

ε_i is a random error term,

$i = 1, \dots, n$,

n =number of observations.

We assume $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$, where σ^2 denotes the variance in the model.

We may write model 2.1 in matrix form as follows,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.2)$$

where \mathbf{y} is an $n \times 1$ vector of responses and has the form

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

\mathbf{X} is $n \times 2$ and is often called the derivative matrix [1]. As we will see later the derivative matrix is very important when it comes to inference in linear regression.

For model 2.1 the derivative matrix has the following form

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

$\boldsymbol{\beta}$ is a 2×1 parameter vector. For model 2.1 it has the form

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Finally $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors. It has the form

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

As before, we assume $E(\boldsymbol{\varepsilon}) = 0$ and $Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$, where \mathbf{I} is the identity matrix.

We say that model 2.1 is linear in the parameters. This means that the derivatives with respect to any of the parameters are independent of all of the parameters [1].

Linear models take many forms. One form that is of particular interest to us in this paper is what is known as a piecewise linear model. We encounter piecewise linear models when the regression of y on x follows a particular linear relation in some range of x , but follows a different linear relation elsewhere. The linear relation we are interested in begins with a negative slope, reaches some fixed point, or knot, we denote as a , at which point the slope of the regression line changes to being positive. The model for that piecewise linear regression with fixed knot is as follows,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - a) z_a + \varepsilon_i, \quad (2.3)$$

where

$$z_a(x_i) = \begin{cases} 1, & \text{if } x_i > a, \\ 0, & \text{otherwise.} \end{cases}$$

For notational purposes, in the rest of the practicum, we will suppress the x_i in $z_a(x_i)$ and remember that z_a is a function of x_i .

Writing model 2.3 in the matrix form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ we obtain

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & (x_1 - a)z_a \\ 1 & x_2 & (x_2 - a)z_a \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n - a)z_a \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \quad (2.4)$$

In 2.3 and 2.4, z_a is an indicator variable, also known as a heaviside step function, which forces the change in slope in the model. In combination with the form of the model, z_a also insures that the function is continuous at $x_i = a$.

When $x_i > a$, $z_a = 1$ and model 2.3 becomes

$$y_i = (\beta_0 - a\beta_2) + (\beta_1 + \beta_2)x_i + \varepsilon_i, \quad (2.5)$$

on the other hand, when $x_i \leq a$, $z_a = 0$ and we obtain

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (2.6)$$

which is our simple linear regression model 2.1.

We can also extend 2.3 to a piecewise linear model with two or more fixed knots. For example, the slope of our regression line may change twice at predetermined knots of a and b . The model follows a negative slope up to a , then changes to a horizontal relation with slope equal to zero. It then reaches b , where a positive slope exists from there. Our model then has the form

$$y_i = \beta_0 + \beta_1 x_i - \beta_1(x_i - a)z_a + \beta_2(x_i - b)z_b + \varepsilon_i, \quad (2.7)$$

where

$$z_a = \begin{cases} 1, & \text{if } x_i > a, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$z_b = \begin{cases} 1, & \text{if } x_i > b, \\ 0, & \text{otherwise.} \end{cases}$$

Or, again writing in matrix form, we obtain

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 - (x_1 - a)z_a & (x_1 - b)z_b \\ 1 & x_2 - (x_2 - a)z_a & (x_2 - b)z_b \\ \vdots & \vdots & \vdots \\ 1 & x_n - (x_n - a)z_a & (x_n - b)z_b \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \quad (2.8)$$

From 2.7 we notice that when $x_i \leq a$, both z_a and z_b are zero. The model then becomes

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (2.9)$$

which is the simple linear regression model given in 2.1. When $a < x_i \leq b$, $z_a = 1$ and $z_b = 0$, and 2.7 becomes

$$y_i = \beta_0 + \beta_1 a + \varepsilon_i, \quad (2.10)$$

which gives us the horizontal portion of the model. Finally when $x_i > b$, both z_a and z_b are equal to one, and we obtain

$$y_i = \beta_0 + \beta_1 a + \beta_2 x_i - \beta_2 b + \varepsilon_i, \quad (2.11)$$

which forces the relationship between y and x to have a positive slope.

This section briefly introduced linear regression analysis and how we can build models to describe relationships between two quantitative variables. In particular we looked at piecewise linear models in which changes in slope in the relationship exist. Models 2.3 and 2.7 will be used when we look at the relationship between load and temperature in the Manitoba Hydro problem.

2.3 Least Squares Estimation and Inference

The goal of regression analysis is to use the data to estimate the unknown regression parameters. We need a logical method to find these parameters. One method that is used more extensively than any other estimation procedure is the method of least squares [5].

For the basic linear model given by 2.1, the expected value of y_i given x is

$$E(y_i) = \beta_0 + \beta_1 x_i. \quad (2.12)$$

The method of least squares considers the deviation of y_i from its expected value,

$$y_i - (\beta_0 + \beta_1 x_i). \quad (2.13)$$

We need to approximate $E(y_i)$ in some logical manner. Imagine $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators for β_0 and β_1 respectively. We can then replace the model parameters in 2.13 with these estimators to obtain

$$y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i). \quad (2.14)$$

Result 2.14 is called the i^{th} residual, or error, denoted by e_i . If the linear model used to fit the relationship between y and x is accurate then these random errors should be as small as possible.

The method of least squares is used to choose estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ to force the residuals to be as small as possible. In other words, the idea behind the least squares method is to choose estimators in such a way that the sum of squares of the residuals given by

$$S(\hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2, \quad (2.15)$$

is minimized [6]. To achieve this, one can do long drawn out manual calculations which become almost impossible once the number of unknown parameters increases beyond two or three. These calculations involve taking partial derivatives with respect to each of the parameters, then setting each partial derivative equal to zero and finally solving for the unknown parameters. These tedious procedures are avoided by using software packages such as SAS to calculate the parameter estimates.

In matrix terms, $\hat{\beta}$ is the value of β which minimizes $S(\hat{\beta})$ [6]. $\hat{\beta}$ is the least squares estimate and can be written as

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2.16)$$

Under normal assumptions we can show that $\hat{\beta}$ is normally distributed with $E(\hat{\beta}) = \beta$ and $Var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ [6]. From these results a $1 - \alpha$ marginal confidence interval for the parameter β_i , $i = 0, 1$, is given by

$$\hat{\beta}_i \pm se(\hat{\beta}_i)t(n - p; \alpha/2), \quad (2.17)$$

where $t(n - p; \alpha/2)$ is the upper $\alpha/2$ quantile for the Student's t distribution with $n - p$ degrees of freedom. The standard error of the parameter estimate is given by

$$se(\hat{\beta}_i) = \sqrt{s^2(\mathbf{X}'\mathbf{X})_{ii}^{-1}}, \quad (2.18)$$

with $(\mathbf{X}'\mathbf{X})_{ii}^{-1}$ equal to the i^{th} diagonal term of the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix and

$$s^2 = \frac{S(\hat{\beta})}{n - p}, \quad (2.19)$$

which is an estimator for σ^2 [6].

We may also create joint confidence regions for β . A $1 - \alpha$ joint confidence region for β is an ellipsoid given by

$$(\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta})' \leq ps^2F(p, n - p; \alpha), \quad (2.20)$$

where $F(p, n - p; \alpha)$ is the upper α quantile for the Fisher's F distribution with p and $n - p$ degrees of freedom [6].

Inference is an important aspect of linear regression. Creating confidence intervals

and confidence regions helps us make conclusions about the behaviour of our unknown parameters. The method of least squares was introduced in this section as a way to choose the best possible parameter estimates. We also looked at how confidence intervals and confidence regions are created. In the next couple of sections, we will use these notions in the Manitoba Hydro data to examine the relationship between load and temperature. Specifically we will see how piecewise linear models can be used to fit models to the data.

2.4 Fitting Piecewise Linear Models with One Predetermined Knot to Manitoba Hydro Data

In section 2.2 we briefly looked at piecewise linear models. These models are used to fit data that show changes in slope at a certain point we called a knot. We will test if this model is appropriate for the Manitoba Hydro load versus temperature relationship.

2.4.1 The Model

Figure 2.1 shows us a plot of load versus temperature for data collected everyday at 6pm. When we examine the data, we see that two slopes in the plot exist. There is what we call a “heating slope”, where a negative relationship between load and temperature exists. We call this the “heating slope” because at those temperatures consumers are using hydroelectric power to heat their homes or businesses. This negative slope reaches a certain point and then changes to a positive slope in the relationship between load and temperature, which we call the “cooling slope”. We name this the “cooling slope” because over this temperature range most consumers are using hydroelectric power to cool their homes or businesses.

For now, we will consider the knot where the slope changes to be predetermined.

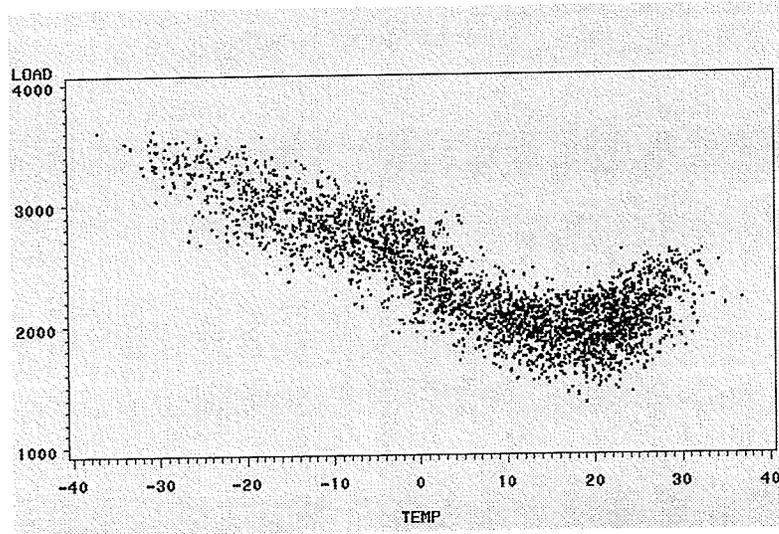


Figure 2.1: Plot of load vs. temperature for all data at 6pm.

Manitoba Hydro has suggested using 18°C as the knot. For our purposes, in this chapter, we will use Manitoba Hydro's suggestion. Later we will consider the problem where the location of the knot is unknown, and has to be estimated from the data.

We modify model 2.3 by replacing a with our predetermined knot of 18°C . Model 2.3 then can be stated as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 18) z_{18} + \varepsilon_i, \quad (2.21)$$

where

$$z_{18} = \begin{cases} 1, & \text{if } x_i > 18, \\ 0, & \text{otherwise.} \end{cases}$$

In this model y_i is the i^{th} value of load and x_i is the known i^{th} value of temperature. We can call β_0 a height parameter. The magnitude of the "heating slope" is given by β_1 , and β_2 is the magnitude of the "cooling slope". Here the derivative matrix, \mathbf{X} , has

the form

$$\begin{bmatrix} 1 & x_1 & (x_1 - 18)z_{18} \\ 1 & x_2 & (x_2 - 18)z_{18} \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n - 18)z_{18} \end{bmatrix}. \quad (2.22)$$

2.4.2 Least Squares Inference

We will use the least squares procedure to choose appropriate parameter estimates which minimize the residual sum of squares. That is, choose $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ that minimize

$$S(\hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 (x_i - 18)z_{18})^2. \quad (2.23)$$

That means we need to calculate $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Using PROC REG in SAS [11] a piecewise fit to load versus temperature data is obtained. Part of the SAS code used is given in Appendix B. Figure 2.2 shows us the 6pm data along with the solid line that represents the piecewise linear fit with one knot at 18°C.

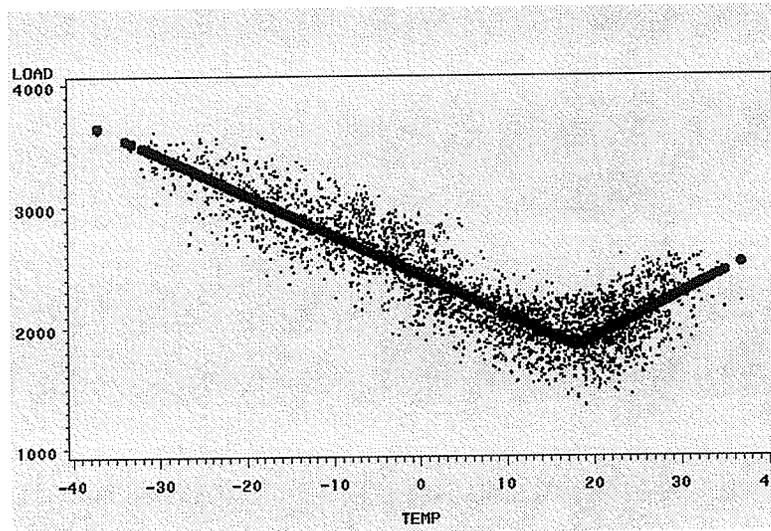


Figure 2.2: Piecewise linear fit with one predetermined knot for 6pm data.

From the PROC REG output in Figure 2.3 we see that the R^2 for the piecewise fit is .7944, which indicates a reasonably good fit. The parameter estimates and

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	
Model	2	499849833	249924917	6051.96	
Error	3133	129382000	41297		
Corrected Total	3135	629231833			

Source	Pr > F
Model	<.0001

	Root MSE	203.21545	R-Square	0.7944
Dependent Mean	2329.46918	Adj R-Sq	0.7942	
Coeff Var	8.72372			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
B0	1	2433.76322	4.12523	589.97	<.0001
B1	1	-32.85522	0.31076	-105.72	<.0001
B2	1	69.27470	1.51908	45.60	<.0001

Figure 2.3: SAS output for piecewise linear fit with one predetermined knot for 6pm data.

their respective standard errors are also obtained from the output. We have $\hat{\beta} = (2433.76, 32.86, 69.27)'$, $s^2 = 41297$ and the degrees of freedom, $df = 3133$. The standard errors of the parameter estimates are $se(\hat{\beta}_0) = 4.13$, $se(\hat{\beta}_1) = 0.31$ and $se(\hat{\beta}_2) = 1.52$. We see by the individual t values, that all parameters are significant and should be all included in the model. The fitted model then can be written as follows,

$$\hat{y} = 2433.76 - 32.86x + 69.27(x - 18)z_{18}, \quad (2.24)$$

where

$$z_{18} = \begin{cases} 1, & \text{if } x > 18, \\ 0, & \text{otherwise.} \end{cases}$$

Using the derivative matrix 2.22 we can also calculate the $\mathbf{X}'\mathbf{X}$ matrix, and the

$(\mathbf{X}'\mathbf{X})^{-1}$ matrix. They are as follows,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3136 & 20913.3 & 5196.9 \\ 20913.3 & 880072.29 & 133150.99 \\ 5196.9 & 133150.99 & 39606.79 \end{bmatrix},$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} .000412 & -.00000328 & -.0000430 \\ -.00000328 & .00000234 & -.00000743 \\ -.0000430 & -.00000743 & .0000559 \end{bmatrix}.$$

From the information given in the SAS output we may now construct marginal 95% confidence intervals for our parameter estimates. We use 2.17 with $t(3133; .025) = 1.96$. The 95% marginal confidence interval for β_0 is

$$2433.76 \pm (4.13)(1.96)$$

or

$$2425.68 \leq \beta_0 \leq 2441.85,$$

the 95% marginal confidence interval for β_1 is

$$-32.86 \pm (0.31)(1.96)$$

or

$$-33.47 \leq \beta_1 \leq -32.25$$

and the 95% marginal confidence interval for β_2 is

$$69.27 \pm (1.52)(1.96)$$

or

$$66.29 \leq \beta_2 \leq 72.25.$$

We can also construct a joint 95% confidence region for our three parameters. We use 2.20 with $p = 3$, $s^2 = 41297$ and $F(3, 3133; .05) = 2.61$. The joint 95% confidence region is an ellipsoid given by the following

$$\begin{aligned} & 3136(\beta_0 - 2433.76)^2 + 880072.29(\beta_1 + 32.86)^2 + 39606.79(\beta_2 - 69.27)^2 + \\ & 418266(\beta_0 - 2433.76)(\beta_1 + 32.86) + 10393.8(\beta_0 - 2433.76)(\beta_2 - 69.27) \\ & \quad + 266301.98(\beta_1 + 32.86)(\beta_2 - 69.27) \\ & = (3)(41297)(2.61) = 323355.51. \end{aligned}$$

2.4.3 Assessing The Model

To see if our model is appropriate we can examine the residuals. Plotting the residuals against the predictor variable can tell us whether a linear regression function is appropriate for the data being analyzed [7]. The plot of the residuals versus temperature given in Figure 2.4 shows a nice random pattern with no apparent deviations from this pattern. There is no evidence to suggest that our model is inappropriate here.

Caution must be taken though. The relationship between load and temperature is definitely far from perfect. This suggests that there might be other variables contributing to the variation in load. Variables such as year, day and month are assumed to be adding to the variation in load. Later we will touch upon this problem and see if we can adjust our models to account for these additional variables.

Another assumption we could verify is the normality of the error terms. This can be done using a normal probability plot. In this plot each residual is plotted against its expected value under normality [7]. A plot that is linear suggests agreement with

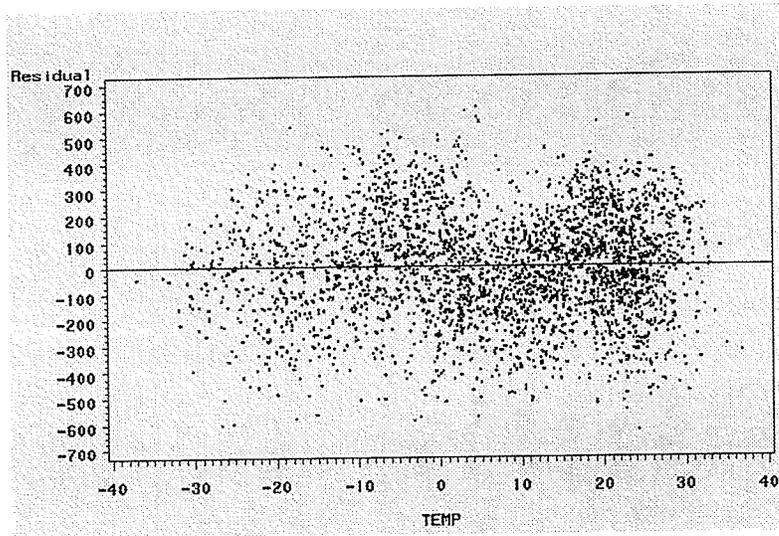


Figure 2.4: Plot of residuals vs. temperature based on piecewise linear fit with one predetermined knot.

normality, but a plot that departs substantially from linearity suggests non-normality in the error terms. The normal probability plot for our data is given in Figure 2.5.

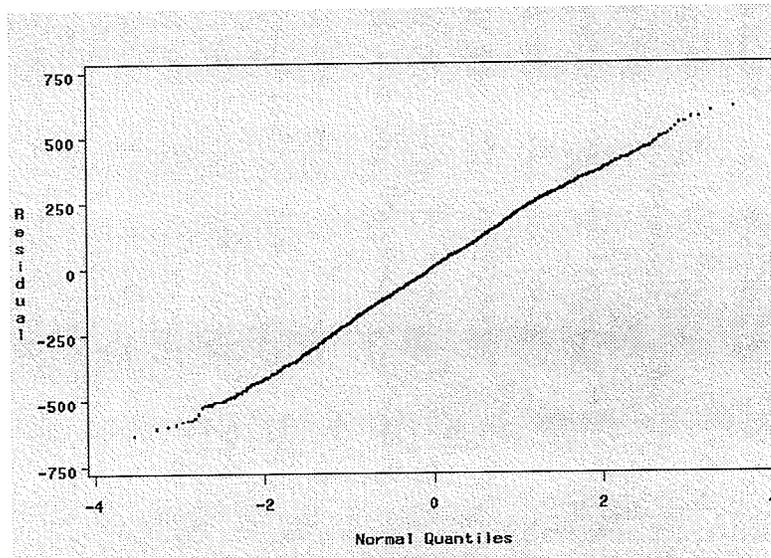


Figure 2.5: Normal probability plot of residuals.

We can see that there are small departures in the tails, but nothing to concern us. In general the plot is very nearly linear. Therefore the assumption of normality in the error terms is supported by the plot. By these two plots we can say that there is no

evidence that our piecewise model is incorrect.

One problem which we will tackle later is should we use a knot at a fixed temperature of 18°C or should we estimate the knot from the data itself? It is safe to assume that the knot will be unknown and possibly different from 18°C at 6pm. Also, we could assume that the knot changes depending on the time of day and using 18°C for all these models could be inappropriate. Later, we will see how nonlinear modeling can be used to tackle this problem. In the next section we look at the piecewise linear model again, this time using two fixed knots.

2.5 Fitting Piecewise Linear Models with Two Predetermined Knots to Manitoba Hydro Data

In this section we will extend our discussion of piecewise linear models. Specifically, instead of one knot we will fit a piecewise linear model with two predetermined knots. Here we reintroduce Figure 2.1 as Figure 2.6 and notice that a different fit to the data is possible. We still have our “heating slope” and “cooling slope”, but we can

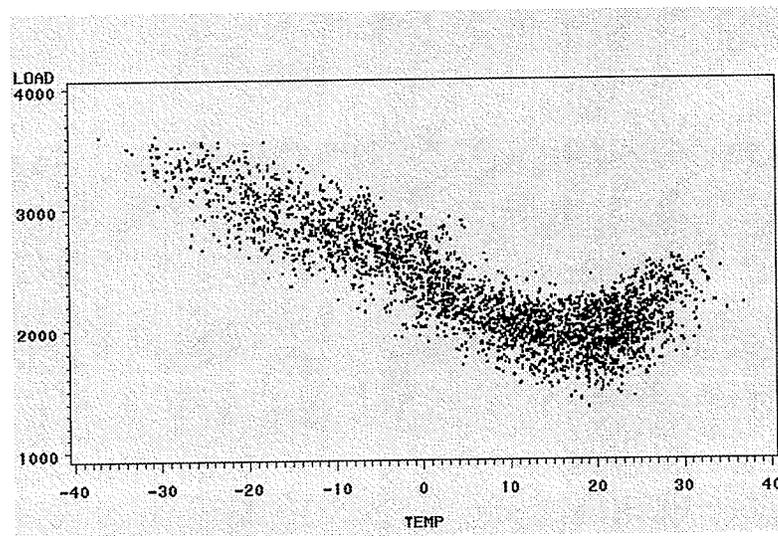


Figure 2.6: Plot of load vs. temperature for all data at 6pm.

also see that there may be a constant relation between approximately 10°C and 20°C .

Therefore it is possible to fit a piecewise linear model with knots at 10°C and 20°C with a horizontal portion between the two knots.

2.5.1 The Model

Model 2.7 is modified by replacing a and b by 10°C and 20°C respectively. The model is then

$$y_i = \beta_0 + \beta_1 x_i - \beta_1(x_i - 10)z_{10} + \beta_2(x_i - 20)z_{20} + \varepsilon_i, \quad (2.25)$$

where

$$z_{10} = \begin{cases} 1, & \text{if } x_i > 10, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$z_{20} = \begin{cases} 1, & \text{if } x_i > 20, \\ 0, & \text{otherwise.} \end{cases}$$

The derivative matrix, \mathbf{X} , has the form

$$\begin{bmatrix} 1 & x_1 - (x_1 - 10)z_{10} & (x_1 - 20)z_{20} \\ 1 & x_2 - (x_2 - 10)z_{10} & (x_2 - 20)z_{20} \\ \vdots & \vdots & \vdots \\ 1 & x_n - (x_n - 10)z_{10} & (x_n - 20)z_{20} \end{bmatrix}. \quad (2.26)$$

2.5.2 Least Squares Inference

The procedure to find the parameter estimates is the same as before. We use the method of least squares to choose $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ that minimizes

$$S(\hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i + \hat{\beta}_1(x_i - 10)z_{10} - \hat{\beta}_2(x_i - 20)z_{20})^2. \quad (2.27)$$

Again, this is done by calculating $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Figure 2.7 shows the piecewise linear fit, obtained from SAS, with knots at 10°C and 20°C. Again, a portion of the SAS code used is given in Appendix B. The output

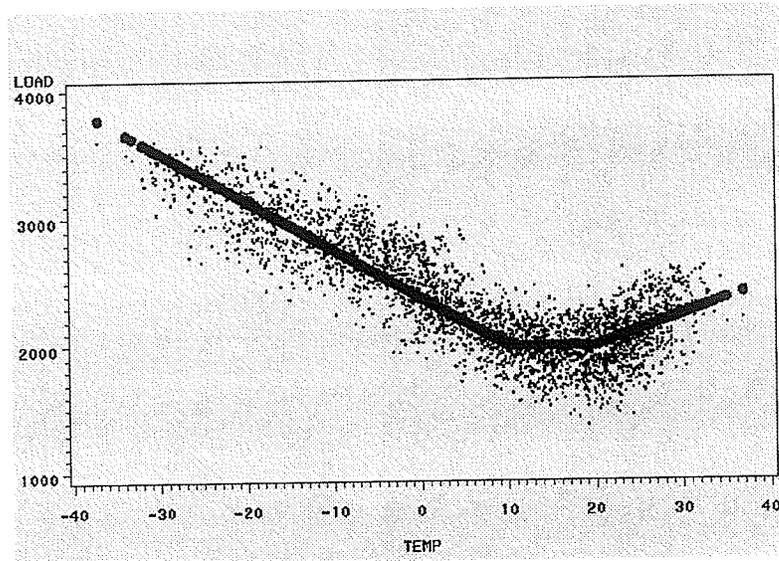


Figure 2.7: Piecewise linear fit with two predetermined knots for data at 6pm.

from PROC REG in Figure 2.8 gives us an R^2 value for this model to be .7841 indicating a reasonably good fit. We notice, though, that the fit using one knot gave us a higher R^2 and lower sum of squares. This could be due to the fact that for the piecewise fit with two knots we guessed at what the two knots should be without any prior information. In the next chapter, when we estimate the knots from the data, we will compare the two types of piecewise models again. It will probably give us a better indication of which model fits the data better. The parameter estimates given by the output are $\hat{\beta} = (2372.60, -37.99, 24.32)'$. The standard errors are $se(\hat{\beta}_0) = 4.08$, $se(\hat{\beta}_1) = 0.36$ and $se(\hat{\beta}_2) = 1.62$. We also have $s^2 = 43371$ and $df = 3133$. All three parameter estimates are significant. Using these parameter estimates, the fitted piecewise model with two fixed knots is written as follows,

$$\hat{y} = 2372.60 - 37.99x + 37.99(x - 10)z_{10} + 24.32(x - 20)z_{20}, \quad (2.28)$$

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	2	493350189	246675094	5687.55
Error	3133	135881644	43371	
Corrected Total	3135	629231833		

Source	Pr > F
Model	<.0001

	Root MSE	208.25728	R-Square	0.7841
Dependent Mean	2329.45918	Adj R-Sq	0.7839	
Coeff Var	8.94016			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
B0	1	2372.59501	4.08136	581.32	<.0001
B1	1	-37.98688	0.36379	-104.42	<.0001
B2	1	24.32439	1.62190	15.00	<.0001

Figure 2.8: SAS output for piecewise linear fit with two predetermined knots for data at 6pm.

where

$$z_{10} = \begin{cases} 1, & \text{if } x > 10, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$z_{20} = \begin{cases} 1, & \text{if } x > 20, \\ 0, & \text{otherwise.} \end{cases}$$

We also obtain the $\mathbf{X}'\mathbf{X}$ and $(\mathbf{X}'\mathbf{X})^{-1}$ matrices using 2.26, they are

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3136 & 5772.6 & 3453.7 \\ 5772.6 & 380966.5 & 34537 \\ 3453.7 & 34537 & 22435.17 \end{bmatrix},$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} .000384 & -.000000534 & -.000058 \\ -.000000534 & .0000305 & -.0000462 \\ -.000058 & -.00000462 & .0000607 \end{bmatrix}.$$

Again we can construct the 95% marginal confidence intervals for our parameters. Once more we use 2.17 with $t(3133; .025) = 1.96$. The 95% marginal confidence interval for β_0 is

$$2372.60 \pm (4.08)(1.96)$$

or

$$2364.60 \leq \beta_0 \leq 2380.60,$$

the 95% marginal confidence interval for β_1 is

$$-37.99 \pm (0.36)(1.96)$$

or

$$-38.70 \leq \beta_1 \leq -37.28,$$

and the 95% marginal confidence interval for β_2 is

$$24.32 \pm (1.62)(1.96)$$

or

$$21.14 \leq \beta_2 \leq 27.50.$$

We can also construct the 95% joint confidence region for our three parameters. Using 2.20 with $p = 3$, $s^2 = 43371$ and $F(3, 3133; .05) = 2.61$. The 95% confidence

ellipsoid for our parameters is given by the following

$$\begin{aligned}
 &3136(\beta_0 - 2372.60)^2 + 380966.5(\beta_1 + 37.99)^2 + 22435.17(\beta_2 - 24.32)^2 + \\
 &11545.2(\beta_0 - 2372.60)(\beta_1 + 37.99) + 6907.4(\beta_0 - 2372.60)(\beta_2 - 24.32) \\
 &\quad + 69074(\beta_1 + 37.99)(\beta_2 - 24.32) \\
 &= (3)(43371)(2.61) = 339594.93.
 \end{aligned}$$

2.5.3 Assessing The Model

The next step is to look at the residual plot to see if it shows us any unusual patterns. We take a look at the plot of the residuals versus temperature as given in Figure 2.9. The residual plot does not show us any unusual pattern. There is a nice random

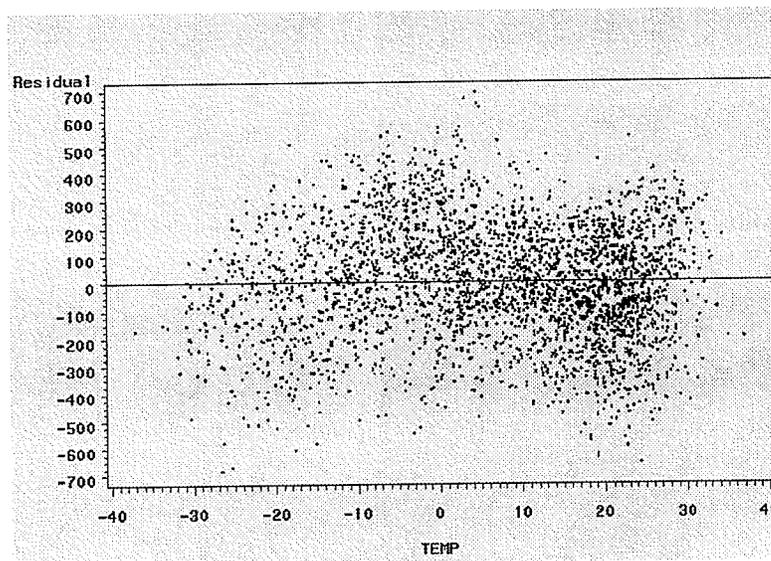


Figure 2.9: Plot of residuals vs. temperature based on piecewise linear fit with two predetermined knots.

scatter which suggests that our model is appropriate to use. As we stated with the piecewise model with one knot, caution has to be taken because, as we will discuss later, there are other variables that are contributing to the variation in load.

We also have the normal probability plot of the error terms in Figure 2.10. The

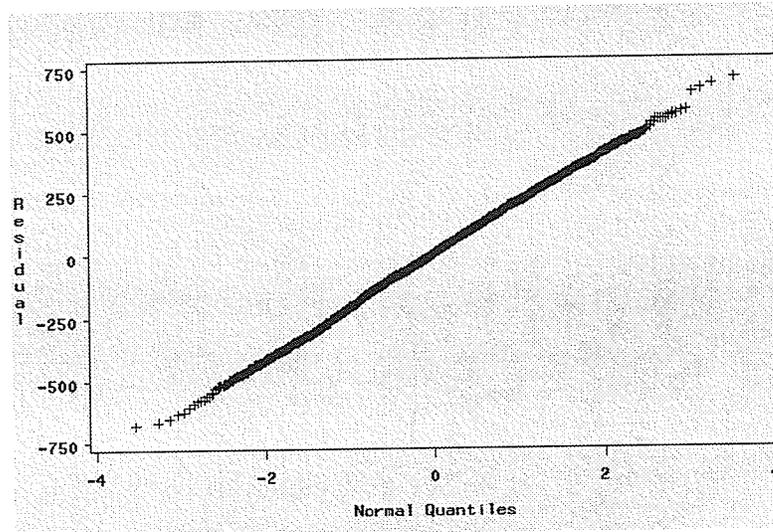


Figure 2.10: Normal probability plot of residuals.

plot shows us a linear pattern with no significant deviations from this linearity. This suggests that the assumption of normality in the error terms is valid. By looking at the residuals we can conclude that there is no evidence that our model is incorrect.

2.6 Conclusion

In this chapter we introduced linear regression. In particular we talked about fitting piecewise linear models with either one or two predetermined knots. We introduced least squares inference and showed how marginal and joint confidence intervals are constructed. Here, we assumed that the knots were fixed predetermined points. We took all data collected at 6pm and fit piecewise linear models. The piecewise models fit the data reasonably well. But we saw that the relationship between load and temperature is not perfect. There is still a lot of variation that is not accounted for. We suggested that other variables are contributing to the variation. We may suggest such variables as daytype, year and month. Later we will look at this topic in more detail. We will see if we can build up a model, that incorporates all these predictor variables, to improve our fit to the data.

Manitoba Hydro suggested using 18°C as the knot in the one knot model. As was mentioned earlier our goal is to choose parameter estimates that minimize the residual sum of squares. It would make sense then to check to see if estimating the location of the knot from the data instead of using 18°C would minimize our sum of squares even further and therefore giving us a better linear fit. Also, in the two knot problem we do not have prior information on where the two knots should be located. We took an educated guess when choosing 10°C and 20°C . Therefore, estimating the knots from the data may be our only choice. In the next chapter we will assume that the knots are unknown and have to be estimated from the data. We will introduce nonlinear regression and use this type of regression to fit piecewise linear models to the Manitoba Hydro Data using estimated knots.

Chapter 3

Fitting Piecewise Linear Models With Unknown Knots—The Nonlinear Case

3.1 Introduction

In chapter 2 we studied the relationship between load and temperature. We took all data collected at 6pm and saw how we can use linear regression to predict load based on its relationship with temperature. We determined that piecewise models could be used to fit the data. Piecewise linear models with one and two predetermined knots were fit to the data. As an example, for the piecewise model with one knot we used 18°C as our fixed knot. That model had the form

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 18) z_{18} + \varepsilon_i, \quad (3.1)$$

where z_{18} is an indicator variable defined in section 2.4.

In this chapter we look at the situation where we do not know what our knot is, and we must estimate it from the data. In fact the location of this unknown knot

becomes another unknown parameter in our model. Model 3.1 is then written as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - \beta_3) z_{\beta_3} + \varepsilon_i, \quad (3.2)$$

where

$$z_{\beta_3} = \begin{cases} 1, & \text{if } x_i > \beta_3, \\ 0, & \text{otherwise.} \end{cases}$$

The fixed knot of 18°C has been replaced by another unknown parameter, β_3 .

We will see that this is not a linear fitting problem anymore but is in fact a nonlinear fitting problem. New techniques to find the estimates in our models will be introduced. Again, as in chapter 2, we will look at the piecewise linear model with one knot and two knots using the same 6pm data. This time we will need to estimate the knots from the data.

In section 3.2 we introduce the notion of nonlinear least squares. We see how the basic idea behind nonlinear least squares is similar to that of linear least squares. The goal is to choose parameter estimates that minimize the residual sum of squares. In section 3.3 we introduce three iterative methods that are used in nonlinear least squares. Specifically, we examine the Gauss-Newton method, the Grid Search method, and the Dud method. In section 3.4 we tackle the one knot piecewise linear problem again. This time we assume that the knot is unknown and therefore must use the iterative methods. In section 3.5 we look at the two knot piecewise linear problem, assuming unknown knots. In sections 3.4 and 3.5 we again examine the residuals and determine if our model assumptions are valid.

3.2 Nonlinear Least Squares

The idea behind the least squares procedure in nonlinear regression is the same as that used in linear regression. The least squares principle is used to estimate the

parameters in nonlinear models just as in the linear model case [9]. Let us consider the general model

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i, \quad (3.3)$$

where $f(x_i, \boldsymbol{\theta})$ is a nonlinear function,

$\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_{p-1})'$ is a vector containing p parameters,

$i = 1, \dots, n$,

$n > p$,

and

$$x_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

We use $\boldsymbol{\theta}$ instead of $\boldsymbol{\beta}$ to remind us that the function here is nonlinear in the parameters. That is why we say this is a nonlinear model. To illustrate what we mean by nonlinear in $\boldsymbol{\theta}$, let us look at our one knot nonlinear piecewise model 3.2. In that model,

$$f(x_i, \boldsymbol{\theta}) = \theta_0 + \theta_1 x_i + \theta_2 (x_i - \theta_3) z_{\theta_3}. \quad (3.4)$$

When we take the partial derivative with respect to θ_2 , for example, we obtain

$$\frac{\partial f}{\partial \theta_2} = (x_i - \theta_3) z_{\theta_3}. \quad (3.5)$$

This derivative with respect to θ_2 depends on another parameter θ_3 . So, we say that model 3.2 is nonlinear in the parameters. We can not compute estimates by elementary linear algebra, as we can in linear regression [5]. In nonlinear regression, iterative methods are used to find the parameter estimates [9].

We call $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_{p-1})'$ the estimator of $\boldsymbol{\theta}$ that minimizes

$$S(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n (y_i - f(x_i, \hat{\boldsymbol{\theta}}))^2. \quad (3.6)$$

Therefore we see that the general idea of least squares is still used. What changes is that we are using iterative methods to find the parameter estimates. These iterative methods require starting values for the parameters, which we label $\boldsymbol{\theta}^0 = (\theta_0^0, \theta_1^0, \dots, \theta_{p-1}^0)'$. These starting values are substituted for $\boldsymbol{\theta}$ to compute the residual sum of squares and compute adjustments to $\boldsymbol{\theta}^0$ that will reduce the residual sum of squares. Hopefully, these adjustments will move $\boldsymbol{\theta}^0$ closer to the least squares solution. The new estimates of the parameters are then used to repeat the process until a sufficiently small adjustment is being made at each step. A criteria that is often used is that the relative change in the sum of squares is very small in at least five successive iterations, usually less than 10^{-8} . When this happens, the process is said to have converged to the solution [12]. In the next section we introduce three of these iterative methods.

3.3 Iterative Methods for Finding Least Squares Estimates in Nonlinear Regression

In the next three subsections we introduce three iterative methods that can be used to find parameter estimates in nonlinear regression. There are several methods available, but here we focus on the Grid Search method, the Dud method and the Gauss-Newton method. The first two methods mentioned will be used for our Hydro data. The third is the method most widely used in nonlinear regression and so we feel it should be discussed.

3.3.1 The Gauss-Newton Procedure

The method most often used for finding the least squares estimator $\hat{\theta}$ in a nonlinear model is the Gauss-Newton method, also called the linearization method [12]. Even though we will not be using this method for our data, we feel that the Gauss-Newton method builds the basis for all other iterative nonlinear procedures. That is why we introduce it here first.

For this iterative procedure we need starting value estimates for the parameters, denoted by θ^0 . Remember, the goal of least squares is to choose parameter estimates that minimizes 3.6. To find the value of θ that minimizes 3.6, we first expand the nonlinear function in 3.3 in a Taylor expansion about the starting values θ^0 and keep only linear terms. That is, $f(x_i, \theta)$ is replaced with

$$f(x_i, \theta) \cong f(x_i, \theta^0) + \sum_{j=0}^{p-1} \left(\frac{\partial f(x_i, \theta^0)}{\partial \theta_j} \right) (\theta_j - \theta_j^0), \quad (3.7)$$

where θ_j^0 is the j^{th} component of the vector θ^0 .

This is simply a linear approximation to $f(x_i, \theta)$ in a region near the starting values θ^0 [9].

We can rewrite 3.7 as

$$f(x_i, \theta) \cong f_i^0 + \sum_{j=0}^{p-1} V_{ij}^0 \beta_j^0, \quad (3.8)$$

where

$$f_i^0 = f(x_i, \theta^0), \quad (3.9)$$

$$\beta_j^0 = \theta_j - \theta_j^0, \quad (3.10)$$

$$V_{ij}^0 = \left(\frac{\partial f(x_i, \theta)}{\partial \theta_j} \Big|_{\theta=\theta^0} \right). \quad (3.11)$$

Therefore we can write out the approximation to the nonlinear regression model 3.3

as

$$y_i \cong f_i^0 + \sum_{j=0}^{p-1} V_{ij}^0 \beta_j^0 + \varepsilon_i. \quad (3.12)$$

We can move the f_i^0 term to the left and denote $y_i - f_i^0$ as y_i^0 , the residual. Finally, we obtain the following linear regression model approximation

$$y_i^0 \cong \sum_{j=0}^{p-1} V_{ij}^0 \beta_j^0 + \varepsilon_i. \quad (3.13)$$

The partial derivatives, V_{ij}^0 , act as explanatory variables in a linear model. The coefficient, β_j^0 , is the difference between the true regression parameter and the starting value.

We can write 3.13 in matrix form as

$$\mathbf{y}^0 = \mathbf{V}^0 \boldsymbol{\beta}^0 + \boldsymbol{\varepsilon}, \quad (3.14)$$

where

$$\mathbf{y}^0 = \begin{bmatrix} y_1 - f_1^0 \\ y_2 - f_2^0 \\ \vdots \\ y_n - f_n^0 \end{bmatrix},$$
$$\mathbf{V}^0 = \begin{bmatrix} V_{10}^0 & \cdots & V_{1p-1}^0 \\ \vdots & & \vdots \\ V_{n0}^0 & \cdots & V_{np-1}^0 \end{bmatrix},$$
$$\boldsymbol{\beta}^0 = \begin{bmatrix} \beta_0^0 \\ \vdots \\ \beta_{p-1}^0 \end{bmatrix},$$

and

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

3.14 has the same form as our linear regression model 2.2, but with the \mathbf{V} matrix taking the place of the \mathbf{X} matrix. So it is possible to estimate the parameters $\boldsymbol{\beta}^0$ by ordinary least squares by solving

$$\hat{\boldsymbol{\beta}}^0 = (\mathbf{V}'\mathbf{V}^0)^{-1}\mathbf{V}'\mathbf{y}^0. \quad (3.15)$$

We then use these estimates to revise our regression coefficients θ_j^1 by using

$$\theta_j^1 = \theta_j^0 + \hat{\beta}_j^0, \quad (3.16)$$

where θ_j^1 is the revised estimate of θ_j at the end of the first iteration.

Therefore linear least squares is used on the linearized model 3.13 to estimate the shift in the parameters, or the amount to adjust the starting values. From this we obtain new values for the parameters. We linearize the model about these new values and linear least squares is applied again to find a second set of adjustments. This procedure is repeated until a desired degree of convergence is obtained. Convergence implies that the residual sum of squares and parameter estimates are no longer changing [12]. The convergence criteria used by SAS [11] is the following

$$\frac{|S(\boldsymbol{\theta}^k) - S(\boldsymbol{\theta}^{k-1})|}{S(\boldsymbol{\theta}^{k-1})} \leq 10^{-8} \quad (3.17)$$

for five successive iterations. In 3.17 k represents the current iteration. In other words, 3.17 is saying that convergence has been met when the relative change in the sum of squares on five successive iterations is very small.

One has to be careful because the Gauss-Newton procedure will not converge for some functions without the use of some sort of step shortening procedure to decrease $S(\hat{\theta})$. In fact the Gauss-Newton procedure may produce an increase in the residual sum of squares [1]. If an increase in the sum of squares is observed, we introduce a step factor λ , and calculate

$$\theta^k = \theta^{k-1} + \lambda\beta^{k-1}, \quad (3.18)$$

where k is the current iteration and λ is chosen to ensure that

$$S(\theta^k) < S(\theta^{k-1}). \quad (3.19)$$

PROC NLIN in SAS [11] uses step-halving to find the step size λ . The method starts by setting $\lambda = 1$ and then halves it until 3.19 is satisfied.

Inference

If the sample size is large enough and ε is normally distributed we can use the Large Sample Theory and assume that $\hat{\theta}$ is approximately normally distributed with $E(\hat{\theta}) = \theta$ and $Var(\hat{\theta}) = \sigma^2(\mathbf{V}'\mathbf{V})^{-1}$ [7].

We may now construct approximate inference regions and intervals for our parameters in the nonlinear model. Recall that in the linear case a $1 - \alpha$ confidence region for the parameters is given by

$$(\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta}) \leq ps^2F(p, n - p; \alpha). \quad (3.20)$$

In nonlinear regression we replace \mathbf{X} by the derivative matrix \mathbf{V} , where \mathbf{V} is evaluated at the least squares parameter estimates. Therefore a $1 - \alpha$ approximate joint

confidence region for $\boldsymbol{\theta}$ is given by

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{V}' \mathbf{V} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq ps^2 F(p, n - p; \alpha). \quad (3.21)$$

A $1 - \alpha$ approximate marginal confidence interval for the parameter θ_i , $i = 0, \dots, p - 1$, is given by

$$\hat{\theta}_i \pm se(\hat{\theta}_i) t(n - p; \alpha/2), \quad (3.22)$$

where $se(\hat{\theta}_i) = \sqrt{s^2 (\mathbf{V}' \mathbf{V})_{ii}^{-1}}$, with $(\mathbf{V}' \mathbf{V})_{ii}^{-1}$ equal to the i^{th} diagonal term of the $(\mathbf{V}' \mathbf{V})^{-1}$ matrix and $s^2 = S(\hat{\boldsymbol{\theta}})/(n - p)$.

3.3.2 The Grid Search Procedure

The Grid Search method is conceptually the simplest method used to find least squares parameter estimates. In this method we search over a grid of all possible values of the parameters for the combination of values that gives the smallest residual sum of squares. As we will show with our Hydro examples, we can vary the knot (or knots) from a model, perform linear least squares and calculate $S(\hat{\boldsymbol{\theta}})$ for each combination of knot (or knots). We can then plot $S(\hat{\boldsymbol{\theta}})$ versus the parameter (or parameters) of interest. In this method we are finding the minimum sum of squares over all the other coordinates for each point on the grid. The minimum point of the curve or surface will be the point where the residual sum of squares is minimized. The parameters that correspond to this minimum point will be our parameter estimates. It is important that our grid is fine enough, so that precise results can be obtained. For more than two parameters making plots becomes impossible, so this Grid Search method becomes difficult.

For our piecewise linear models we can use the Grid Search method by exploiting the conditional linearity seen in the models. For example, looking at model 3.2, we notice that θ_0 , θ_1 , and θ_2 are conditionally linear. When we fix θ_3 , we may obtain

estimates for the other three parameters using linear least squares. Even though we are performing linear least squares in this method, we can consider this a nonlinear problem for our Hydro data. This is because the knots are unknown to begin with. So, the way the Grid Search procedure works in our case is that we vary the knot (or knots) over a grid of possible values and calculate the residual sum of squares for each value. At each possible knot we are calculating estimates for the other parameters in the model. The knot (or knots) that gives us the smallest sum of squares will be our estimate for the true knot (or knots), along with the estimates for the other parameters that correspond to this minimization.

This method is the easiest to perform but it is not the most efficient. This method can be used to provide good starting values for other methods. Later, we will use the Grid Search method on our Manitoba Hydro data and see what parameter estimates this method comes up with. We will see if this method will lower our residual sum of squares from the ordinary linear regression that we used in chapter 2.

3.3.3 The Dud (doesn't use derivatives) Procedure

Often, it may be inconvenient or impossible to provide the derivatives required by nonlinear least squares procedures such as the Gauss-Newton method. Therefore, there is a need for a derivative-free method. One method, known as Dud (doesn't use derivatives), was introduced by Ralston and Jenrich [8]. This derivative-free method is similar to the Gauss-Newton method, except that the derivatives are estimated from the history of iterations rather than supplied analytically [8]. Dud works well as long as there is a lot of data [9].

Once again, let us consider the least squares fitting problem. We want to choose $\hat{\theta}$ as to minimize

$$S(\hat{\theta}) = \sum_{i=1}^n (y_i - f(x_i, \hat{\theta}))^2. \quad (3.23)$$

As we saw with the Gauss-Newton procedure, $f(x_i, \boldsymbol{\theta})$ is approximated by a first order Taylor expansion about the current value of the parameter vector $\boldsymbol{\theta}$ and then uses linear least squares to obtain a new value of $\boldsymbol{\theta}$. On the other hand Dud takes the starting values and displaces each parameter in turn by 10% to produce a further set of $p+1$ parameter vectors [8]. This step is known as the initialization process. The next step is the iteration process. Dud uses a function which equals $f(x_i, \boldsymbol{\theta})$ at $p+1$ previous values of the parameter vector to approximate $f(x_i, \boldsymbol{\theta})$ at each iteration.

At the k^{th} step, Dud does the following. Let $\hat{\boldsymbol{\theta}}_1^k, \hat{\boldsymbol{\theta}}_2^k, \dots, \hat{\boldsymbol{\theta}}_{p+1}^k$ be parameter estimates from previous iterations. We can approximate $f(x_i, \boldsymbol{\theta})$ by a linear function $g(x_i, \boldsymbol{\theta})$ which is equal to $f(x_i, \boldsymbol{\theta})$ at the $p+1$ points. We then find $\boldsymbol{\theta}_{\text{new}}$ as the point which minimizes the distance between y and $g(x_i, \boldsymbol{\theta})$. Let us assume that $\boldsymbol{\theta}_{p+1}$ is this such point.

The linear approximation is written as a function of $\hat{\alpha}$, where $\boldsymbol{\theta}_{\text{new}}$ at each iteration is given by

$$\hat{\boldsymbol{\theta}}_{\text{new}} = \boldsymbol{\theta}_{p+1} + \mathbf{T}\hat{\alpha}, \quad (3.24)$$

where $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \dots, \hat{\boldsymbol{\theta}}_{p+1}$ are the estimates from previous iterations. The i^{th} column of the matrix \mathbf{T} is equal to $\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{p+1}$, where $i = 1, \dots, p$. Therefore the Dud linear approximation is written as

$$g(x_i, \boldsymbol{\theta}) = f(x_i, \boldsymbol{\theta}_{p+1}) + \mathbf{H}\hat{\alpha}, \quad (3.25)$$

where the i^{th} column of \mathbf{H} is equal to $f(x_i, \hat{\boldsymbol{\theta}}_i) - f(x_i, \hat{\boldsymbol{\theta}}_{p+1})$, $i = 1, \dots, p$. The iteration consists of minimizing

$$S(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n (y_i - g(x_i, \hat{\boldsymbol{\theta}}))^2. \quad (3.26)$$

Finally we solve for $\hat{\alpha}$ using

$$\hat{\alpha} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'(y - f(x_i, \hat{\boldsymbol{\theta}}_{p+1})), \quad (3.27)$$

and $\hat{\boldsymbol{\theta}}_{\text{new}}$ is calculated using 3.24.

The procedure is repeated using this new value of $\boldsymbol{\theta}$. The procedure is repeated continuously until convergence is met. As in the Gauss-Newton method we use the criteria given in 3.17, where the relative change in the sum of squares on successive iterations is small. As in the Gauss-Newton method $\boldsymbol{\theta}_{\text{new}}$ may produce a larger sum of squares than the previous estimate, $\boldsymbol{\theta}_{\text{old}}$. Therefore a step factor, λ , is introduced to determine a better point. Using λ we calculate

$$\boldsymbol{\theta}_{\text{trial}} = \boldsymbol{\theta}_{\text{new}}\lambda + \boldsymbol{\theta}_{\text{old}}(1 - \lambda) \quad (3.28)$$

If $\boldsymbol{\theta}_{\text{trial}}$ gives us a reduction in the sum of squares, then we will use it instead of $\boldsymbol{\theta}_{\text{new}}$. Ralston and Jenrich [8] allow the step factor to be negative, by choosing λ from a sequence of values $1, 1/2, -1/4, 1/8, -1/16, \dots$

At convergence the approximate covariance matrix is given by $s^2\mathbf{T}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{T}'$. With the Dud method the linear approximation standard errors may not be reliable [1]. Using this covariance matrix we can write out the confidence intervals and regions for the unknown parameters under the Dud method. A $1 - \alpha$ approximate confidence region for $\boldsymbol{\theta}$ is given by

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{T}(\mathbf{H}'\mathbf{H})\mathbf{T}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq ps^2F(p, n - p; \alpha). \quad (3.29)$$

A $1 - \alpha$ approximate marginal confidence interval for θ_i , $i = 0, \dots, p - 1$, is given by

$$\hat{\theta}_i \pm se(\hat{\theta}_i)t(n - p; \alpha/2), \quad (3.30)$$

where $se(\hat{\theta}_i) = \sqrt{s^2(\mathbf{T}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{T}')_{ii}^{-1}}$ with $(\mathbf{T}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{T}')_{ii}^{-1}$ equal to the i^{th} diagonal term of the $(\mathbf{T}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{T}')_{ii}^{-1}$ matrix and $s^2 = S(\hat{\boldsymbol{\theta}})/(n - p)$.

We have seen that the Dud method builds upon the Gauss-Newton method. The difference between the two methods is that the Dud method computes numerical estimates of the derivatives. This is due to the fact that for some models the required derivatives are inconvenient or impossible to find. PROC NLIN in SAS [11] does nonlinear least squares using methods such as Dud and Gauss-Newton, among others. Later we will use the Dud method on our Manitoba Hydro data. We will see that with our piecewise models with unknown knots we have to use derivative-free methods to compute least squares estimates. This is due to the fact that some of the partial derivatives in those models are very difficult to find. We will compare the estimates from the Dud method with both the Grid Search method and the usual linear least squares methods of chapter 2. Ralston and Jenrich [8] showed that the Dud method compares very well with other derivative-free and derivative-requiring methods.

Now that we have introduced nonlinear least squares and some of the available algorithms, let us go ahead and examine the Manitoba Hydro data once again. In the next couple of sections we will reintroduce our piecewise linear models. This time we will assume that our knots are unknown and must be estimated from the data. We will use the nonlinear techniques just introduced to come up with parameter estimates.

3.4 Fitting Piecewise Linear Models with One Unknown Knot to Manitoba Hydro Data

3.4.1 The Model

In section 2.4 we looked at how we can fit a linear piecewise model with one knot to Manitoba Hydro data. There we assumed that our knot was known to be 18°C. The model for that case with predetermined knot is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 18) z_{18} + \varepsilon_i, \quad (3.31)$$

where z_{18} is defined as in section 2.4. One may argue that we should not fix our knot at 18°C, but in fact try to estimate this knot from the data available. It is safe to assume that the knot may vary when we look at different combinations of data. Data from different times of day, different years, etc. will probably have different knots. Even the data from 6pm could possibly have a different knot that minimizes the residual sum of squares even further than what we observed in section 2.4. Therefore we modify our model by replacing the known knot by another parameter. Using θ_i instead of β_i , the model then becomes

$$y_i = \theta_0 + \theta_1 x_i + \theta_2 (x_i - \theta_3) z_{\theta_3} + \varepsilon_i, \quad (3.32)$$

where

$$z_{\theta_3} = \begin{cases} 1, & \text{if } x_i > \theta_3, \\ 0, & \text{otherwise.} \end{cases}$$

As we mentioned earlier this is not a “linear” model anymore, but a “nonlinear” model. From 3.5 we saw that the partial derivative with respect to θ_2 is dependent on θ_3 . We say that this model is nonlinear in the parameters. Now, to obtain

parameter estimates we must use a nonlinear least squares method just like the ones we introduced in the last section. We will start off with the Grid Search method to find the parameter estimates that minimize $S(\hat{\theta})$. Then we will use the Dud method and compare the two methods. We will also see if these methods lower our $S(\hat{\theta})$ given by linear least squares in chapter 2.

3.4.2 Using the Grid Search Method

In chapter 2 we set our knot at 18°C and calculated the residual sum of squares. We did not ask the question of what would happen to the sum of squares if we chose different knots. Be it in the linear or nonlinear scenario, the goal of least squares is to choose parameter estimates that will minimize $S(\hat{\theta})$. So, what we will do here is vary θ_3 in 3.32, which is the parameter that corresponds to our unknown knot, and see which knot will give us the lowest sum of squares. We are taking advantage of the conditional linearity of the model. In other words, we will pick a knot from the region of possible values, perform linear least squares with PROC REG and obtain the corresponding residual sum of squares and parameter estimates. As a reminder, we are not fixing any of the other three parameters, we are letting them vary also. We will repeat this for different knots. The knot that minimizes $S(\hat{\theta})$ will be our estimate for θ_3 along with the other estimates of θ_0 , θ_1 , and θ_2 from PROC REG that correspond to this minimization. We will define this sum of squares as

$$S^*(\hat{\theta}_3) = \min_{\theta_0, \theta_1, \theta_2} S(\hat{\theta}) \quad (3.33)$$

Figure 3.1 plots $S^*(\hat{\theta}_3)$ vs. different values of θ_3 . This function that we have plotted is known as the likelihood profile sum of squares function. We will use this function again in chapter 4 to calculate likelihood profile intervals. The plot shows that the residual sum of squares is minimized when the knot, or θ_3 is equal to approximately

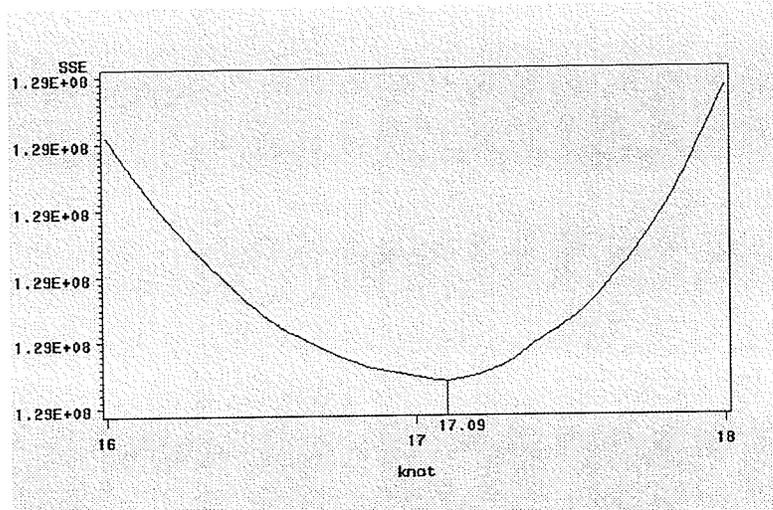


Figure 3.1: Plot of $S^*(\hat{\theta}_3)$ vs. θ_3 .

17.09°C. We see that estimating the knot from the data has given us a lower sum of squares than using 18°C as our knot. Linear least squares using PROC REG was run on the piecewise linear model with 17.09°C as our knot. The output from SAS is given in Figure 3.2. We can see that the R^2 for the piecewise model with one knot has increased to .7951 from .7944 given in section 2.4. The mean square error is also slightly lower, 41156 compared to 41297. The change is not much but none the less has given us a better fit. Figure 3.3 shows us the piecewise fit with 17.09 as our knot. The parameter estimates for our new model are $\hat{\theta} = (2427.29, -33.56, 64.72)$. These estimates are all significant and are slightly different from the results of section 2.4. Their respective standard errors are also given in the output as $se(\hat{\theta}_0) = 4.16$, $se(\hat{\theta}_1) = 0.32$, and $se(\hat{\theta}_2) = 1.41$. We also have $s^2 = 41156$ and $df = 3133$.

Replacing the parameters in model 3.32 with the estimates gives us the following

$$\hat{y} = 2427.29 - 33.56x + 64.72(x - 17.09)z_{17.09}, \quad (3.34)$$

where

$$z_{17.09} = \begin{cases} 1, & \text{if } x > 17.09, \\ 0, & \text{otherwise.} \end{cases}$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	
Model	2	500291275	250145637	6078.04	
Error	3133	128940558	41156		
Corrected Total	3135	629231833			

Source	Pr > F
Model	<.0001

Root MSE	202.86848	R-Square	0.7951
Dependent Mean	2329.45918	Adj R-Sq	0.7950
Coeff Var	8.70882		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
B0	1	2427.29420	4.15952	583.55	<.0001
B1	1	-33.56002	0.32010	-104.84	<.0001
B2	1	64.71877	1.41313	45.80	<.0001

Figure 3.2: SAS output for piecewise linear fit using 17.09°C as the unknown knot.

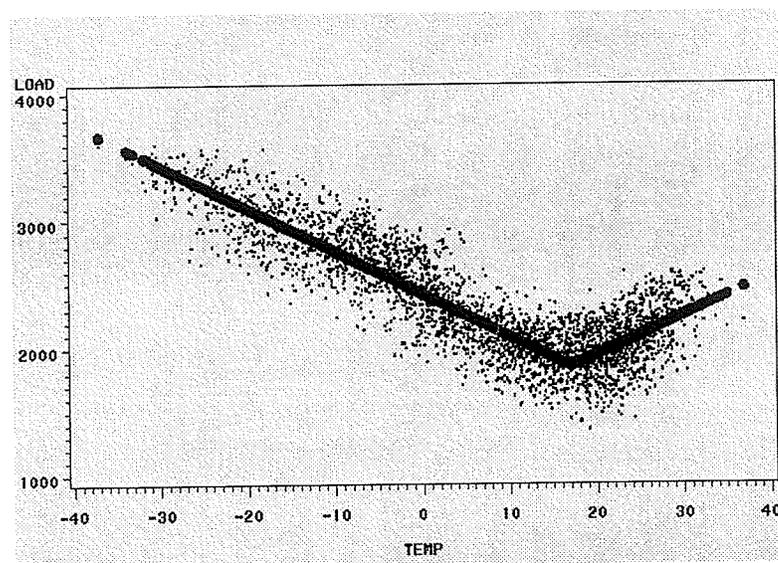


Figure 3.3: Piecewise fit using Grid Search method with 17.09 as the unknown knot.

For this model the derivative matrix has the following form

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & (x_1 - 17.09)z_{17.09} \\ 1 & x_2 & (x_2 - 17.09)z_{17.09} \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n - 17.09)z_{17.09} \end{bmatrix}. \quad (3.35)$$

Using 3.35 the $\mathbf{X}'\mathbf{X}$ and $(\mathbf{X}'\mathbf{X})^{-1}$ matrices are respectively

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3136 & 20913.3 & 6103.95 \\ 20913.3 & 880072.29 & 154198.77 \\ 6103.95 & 154198.77 & 49882.27 \end{bmatrix}, \quad (3.36)$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} .000420 & -.00000213 & -.0000450 \\ -.00000213 & .00000249 & -.00000743 \\ -.000045 & .00000743 & .000049 \end{bmatrix}. \quad (3.37)$$

Once more we can construct 95% confidence intervals for our parameter estimates using 2.17 with $t(3133; .025) = 1.96$. A 95% marginal confidence interval for θ_0 is

$$2419.14 \leq \theta_0 \leq 2435.44,$$

a 95% marginal confidence interval for θ_1 is

$$-34.19 \leq \theta_1 \leq -32.93,$$

and a 95% marginal confidence interval for θ_2 is

$$61.96 \leq \theta_2 \leq 67.48.$$

Using 2.20 and 3.36 we can construct a 95% joint confidence region for θ . We have $p = 3$, $s^2 = 41156$ and $F(3, 3133; .05) = 2.61$. The 95% joint confidence region for θ is given by

$$\begin{aligned} & 3136(\theta_0 - 2427.29)^2 + 880072.29(\theta_1 + 33.56)^2 + 49882.27(\theta_2 - 64.72)^2 + \\ & 41826.6(\theta_0 - 2427.29)(\theta_1 + 33.56) + 12207.9(\theta_0 - 2427.29)(\theta_2 - 64.72) \\ & + 308397.54(\theta_1 + 33.56)(\theta_2 - 64.72) \\ & = (3)(41156)(2.61) = 322251.48. \end{aligned}$$

We can say that using 18°C as our knot was a good estimate to use. On the other hand, estimating the knot from the data has given us the approximate point where the sum of squares is at its lowest.

We have shown how the Grid Search method can be used to find parameter estimates in a nonlinear problem. People can argue that this method seems too simple to come up with the right values and that a more mathematical algorithm should be used instead. As suggested before, the Grid Search method is not efficient but gives very good, if not perfect, results. Next, we will use the Dud method on the same data and see if it gives us the same results as the Grid Search method.

3.4.3 Using the Dud Method

The first thing one may ask is why are we using a derivative-free method. Can we not use a more reliable Gauss-Newton method instead? To answer this, let us rewrite model 3.32,

$$y_i = \theta_0 + \theta_1 x_i + \theta_2 (x_i - \theta_3) z_{\theta_3} + \varepsilon_i, \quad (3.38)$$

where z_{θ_3} is the indicator variable that was previously defined. We mentioned earlier that we must use derivative-free methods, such as Dud, when it is impossible or

difficult to provide derivatives from the model. This is the case we have here with model 3.38. The problem arises if we try to take the partial derivative with respect to θ_3 . We see that this derivative is very difficult to calculate. This is due to the fact that θ_3 is present in the indicator variable, z_{θ_3} . If you recall,

$$z_{\theta_3} = \begin{cases} 1, & \text{if } x_i > \theta_3, \\ 0, & \text{otherwise.} \end{cases}$$

This is known as a heaviside step function. The derivative of this step function with respect to θ_3 is very hard to find. To use nonlinear least squares, the most convenient choice we have is to use a derivative-free procedure, such as Dud. That is what we do here.

The Dud method requires that we have some starting values for our four unknown parameters in model 3.38. Good starting values to use are the parameter estimates and knot of 18°C given from section 2.4. Therefore our starting values will be $\theta^0 = (2433.76, -32.86, 69.27, 18)'$. Figure 3.4 is the output from PROC NLIN in SAS, it shows us the iterations that the Dud method went through to get our least squares parameter estimates. Part of the SAS code used is given in Appendix B. The first five steps are the initialization phase that we talked about earlier. As we can see from the output it took 16 iterations for $S(\hat{\theta})$ to converge. The parameter estimates for our nonlinear model from Dud are $\hat{\theta} = (2427.30, -33.56, 64.73, 17.09)'$. The approximate standard errors are $se(\hat{\theta}_0) = 4.44$, $se(\hat{\theta}_1) = 0.36$, $se(\hat{\theta}_2) = 1.78$ and $se(\hat{\theta}_3) = 0.22$. Our fitted model is then written as

$$\hat{y} = 2427.30 - 33.56x + 64.73(x - 17.09)z_{17.09}, \quad (3.39)$$

DUD Initialization					Sum of
DUD	theta0	theta1	theta2	theta3	Squares
-5	2433.8	-32.8600	69.2700	18.0000	1.2938E8
-4	2677.1	-32.8600	69.2700	18.0000	3.1507E8
-3	2433.8	-36.1460	69.2700	18.0000	1.3892E8
-2	2433.8	-32.8600	76.1970	18.0000	1.3127E8
-1	2433.8	-32.8600	69.2700	19.8000	1.4322E8
Iterative Phase					Sum of
Iter	theta0	theta1	theta2	theta3	Squares
0	2433.8	-32.8600	69.2700	18.0000	1.2938E8
1	2432.6	-32.9842	67.7701	17.7675	1.2918E8
2	2431.8	-33.0251	67.9766	17.7609	1.2917E8
3	2429.5	-33.3153	66.0356	17.4000	1.29E8
4	2429.7	-33.4915	66.3604	17.3204	1.2898E8
5	2429.5	-33.4716	66.3225	17.3192	1.2898E8
6	2429.2	-33.4003	66.1761	17.3052	1.2898E8
7	2428.5	-33.4259	66.3202	17.3048	1.2898E8
8	2427.7	-33.5123	65.0110	17.1543	1.2894E8
9	2427.6	-33.5285	64.8833	17.1322	1.2894E8
10	2427.6	-33.5750	64.9670	17.1093	1.2894E8
11	2427.5	-33.5614	64.9420	17.1061	1.2894E8
12	2427.5	-33.5604	64.9367	17.1063	1.2894E8
13	2427.3	-33.5555	64.7483	17.0963	1.2894E8
14	2427.3	-33.5559	64.7438	17.0956	1.2894E8
15	2427.3	-33.5605	64.7298	17.0914	1.2894E8
16	2427.3	-33.5589	64.7282	17.0912	1.2894E8

NOTE: Convergence criterion met.

Figure 3.4: SAS output showing the iterations from the Dud method.

where

$$z_{17.09} = \begin{cases} 1, & \text{if } x > 17.09, \\ 0, & \text{otherwise.} \end{cases}$$

Figure 3.5 shows us the rest of the SAS output and gives us more information about the fit. The approximate 95% marginal confidence intervals are also given. They were calculated using 3.30. The approximate 95% marginal interval for θ_0 is

$$2418.6 \leq \theta_0 \leq 2436.0,$$

the approximate 95% marginal interval for θ_1 is

$$-34.26 \leq \theta_1 \leq -32.85,$$

the approximate 95% marginal interval for θ_2 is

$$61.24 \leq \theta_2 \leq 68.22,$$

and the approximate 95% marginal interval for θ_3 is

$$16.66 \leq \theta_3 \leq 17.53.$$

We take caution here because these intervals are based on a linear approximation and may not be absolutely accurate. We notice a couple of things from this output. First of all the residual sum of squares for this nonlinear fit is lower than that given by the linear fit using 18°C as the knot. The other thing we notice is that the 95% confidence interval given for θ_3 , which represents our knot, does not cover 18. This suggests that the knot we used in the linear case is an overestimate of the actual knot. Estimating the knot from the data gives us a more accurate result for the knot. We

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	4	1.752E10	4.3794E9	106376	<.0001
Residual	3132	1.2894E8	41168.8		
Uncorrected Total	3136	1.765E10			
Corrected Total	3135	6.2923E8			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
theta0	2427.3	4.4401	2418.6	2436.0
theta1	-33.5589	0.3594	-34.2636	-32.8542
theta2	64.7282	1.7808	61.2365	68.2200
theta3	17.0912	0.2244	16.6513	17.5311

Approximate Correlation Matrix				
	theta0	theta1	theta2	theta3
theta0	1.0000000	0.1038328	-0.0209555	0.3495245
theta1	0.1038328	1.0000000	-0.2018197	0.4548101
theta2	-0.0209555	-0.2018197	1.0000000	0.6076275
theta3	0.3495245	0.4548101	0.6076275	1.0000000

Figure 3.5: SAS output for nonlinear fit using the Dud method.

also show the linear fit in Figure 3.6. The solid line shows us our nonlinear fit to the

data with the slope change at 17.09°C .

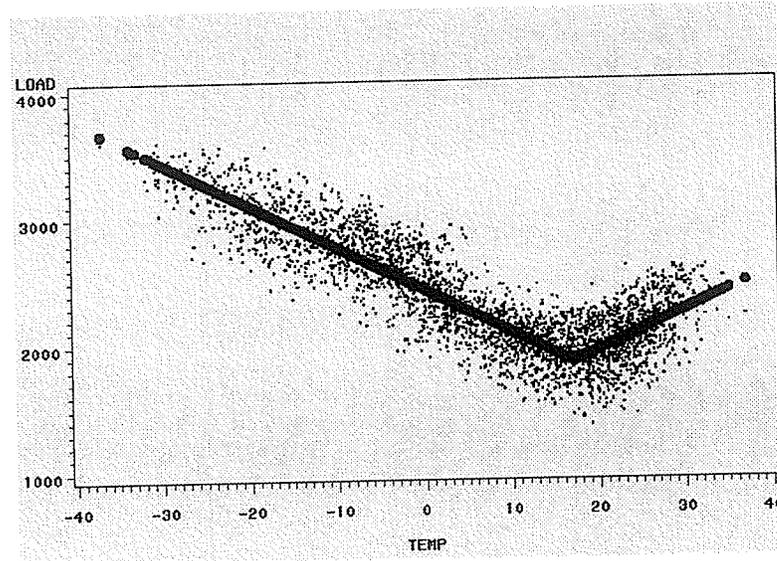


Figure 3.6: Nonlinear fit for piecewise linear model with one unknown knot.

We notice that the parameter estimates given by the Dud method are equal to the parameter estimates given by the Grid Search method. Given the reliability of the Grid Search method we are safe to say that the Dud method has given us the correct results. One last thing we should point out is that the marginal intervals constructed from the Dud method are wider than the intervals constructed with the Grid Search method and the linear methods of section 2.4. This is because the Dud method uses a linear approximation to construct the intervals, whereas in the other methods we are using linear least squares.

As we did with the linear case, we examine the residuals. Figure 3.7 shows us the plot of residuals versus temperature. The plot shows us a nice random pattern suggesting that there is no evidence that our model is incorrect. We also have the normal probability plot in Figure 3.8. We see that we have nice linear pattern with no deviations indicating that the error terms are following a normal distribution. We may also want to check to see whether any of the parameters are excessively highly correlated. This is important to check since high correlations among parameters

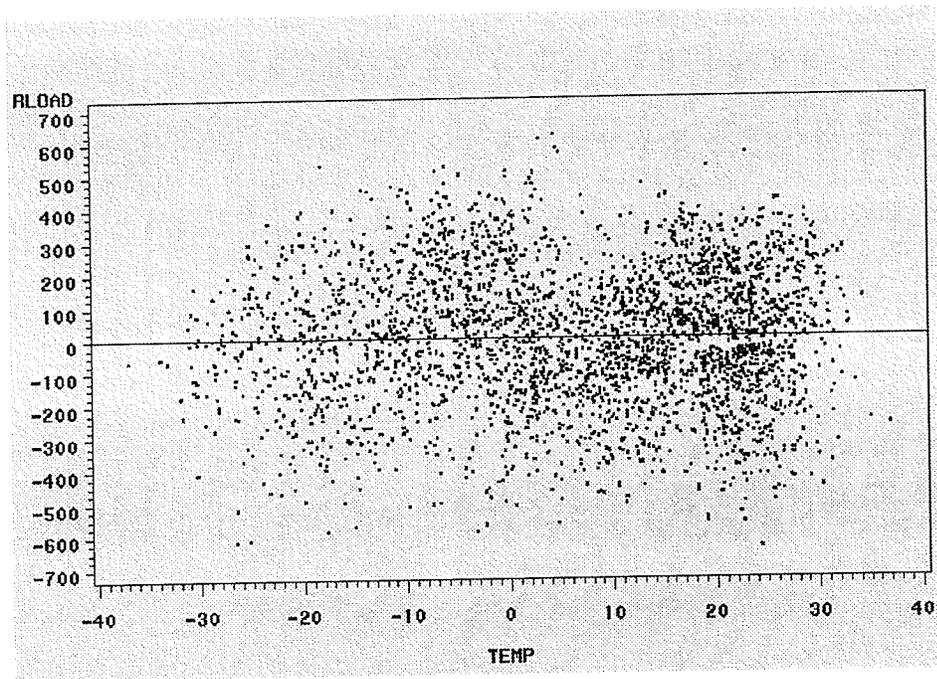


Figure 3.7: Plot of residuals vs. temperature for piecewise linear model with one unknown knot.

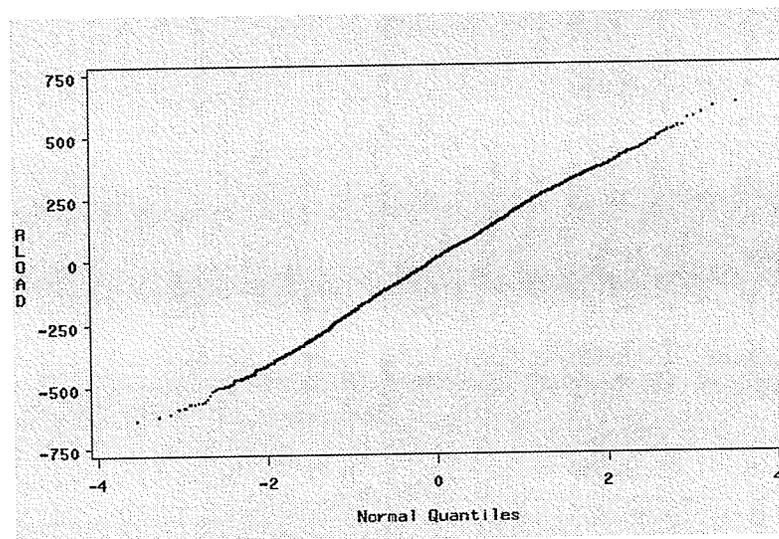


Figure 3.8: Normal probability plot of residuals.

may indicate that we have too many parameters in our model. We can do this by examining the approximate parameter correlation matrix given in the SAS output of Figure 3.5. Bates and Watts [1] suggest that correlations above 0.99 should be investigated. We see by our output that we do not seem to have a problem with high correlation among parameters.

In this section we have shown how nonlinear regression could be used to fit a piecewise linear model when the knot is unknown. The residual sum of squares are lower when we estimate the knot from the data compared to what we obtained in section 2.4. The residual plots showed us the same patterns as we witnessed with the linear case in chapter 2. In the next section we look at the two knot problem again, this time using nonlinear regression.

3.5 Fitting Piecewise Linear Models with Two Unknown Knots to Manitoba Hydro Data

3.5.1 The Model

We introduced the piecewise linear model with two knots in section 2.5. There we assumed that the two knots were predetermined. We arbitrarily chose 10°C and 20°C as the knots. The linear model in that case was

$$y_i = \beta_0 + \beta_1 x_i - \beta_1(x_i - 10)z_{10} + \beta_2(x_i - 20)z_{20} + \varepsilon_i, \quad (3.40)$$

where z_{10} and z_{20} are as defined in section 2.5.

In this chapter we will look at the nonlinear case. We are going to assume that the two knots are unknown and must be estimated from the data. Therefore we introduce

two new parameters into our model. Model 3.40 becomes the following,

$$y_i = \theta_0 + \theta_1 x_i - \theta_1(x_i - \theta_3)z_{\theta_3} + \theta_2(x_i - \theta_4)z_{\theta_4} + \varepsilon_i, \quad (3.41)$$

where

$$z_{\theta_3} = \begin{cases} 1, & \text{if } x_i > \theta_3, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$z_{\theta_4} = \begin{cases} 1, & \text{if } x_i > \theta_4, \\ 0, & \text{otherwise.} \end{cases}$$

As in the one knot case, this is not a “linear” problem anymore. This model is said to be nonlinear in the parameters. For example, the derivative with respect to θ_1 is dependent on θ_3 . Again, we will need to use a nonlinear iterative method to come up with the parameter estimates that minimizes $S(\hat{\theta})$. We will first use the Grid Search method to come up with the estimates and then use the Dud method. We will compare our findings with the linear case and also see if using two knots gives us an improvement over that of using just one knot.

3.5.2 Using the Grid Search Method

In subsection 3.4.2 we showed how the Grid Search method can give us accurate results for finding parameter estimates in the one knot problem. We will repeat the method here for the two knot case. Remember, the goal is to choose parameter estimates that will minimize the residual sum of squares. As we did in the one knot case, we take advantage of the fact that our model is conditionally linear. What we will do here is vary the two unknown knots, θ_3 and θ_4 , over a 2-dimensional grid of possible values, while at the same time letting the other parameters in the model vary also, and calculating $S(\hat{\theta})$ by linear least squares at each combination of knots in this

grid using PROC REG. The combination of knots that minimizes the residual sum of squares will be our estimates for θ_3 and θ_4 . The other parameter estimates will be the values of θ_0, θ_1 and θ_2 that correspond to this minimization of $S(\hat{\theta})$. We define this minimizing sum of squares as

$$S^*(\hat{\theta}_3, \hat{\theta}_4) = \min_{\theta_0, \theta_1, \theta_2} S(\hat{\theta}) \quad (3.42)$$

Figure 3.9 shows us the plot of $S^*(\hat{\theta}_3, \hat{\theta}_4)$ vs. θ_3 and θ_4 . We may call this the profile likelihood surface. This plot will come in to use in chapter 4 when we produce the profile contour plots of θ_3 and θ_4 . We notice from the plot that the surface is minimized when $\theta_3 = 14.25$ and $\theta_4 = 20.70$. We notice that both knots, especially θ_3 , are different from what we used in chapter 2. We can then run PROC REG using 14.25°C and 20.70°C as our knots. The output from PROC REG is given in Figure 3.10. From the output we see that $R^2 = .7972$, which is higher than the R^2 given by both the Grid Search method for one knot and the piecewise linear fit using known knots of 10°C and 20°C . The linear fit is plotted in Figure 3.11. Also, the residual sum of squares is lower here compared to the two other cases.

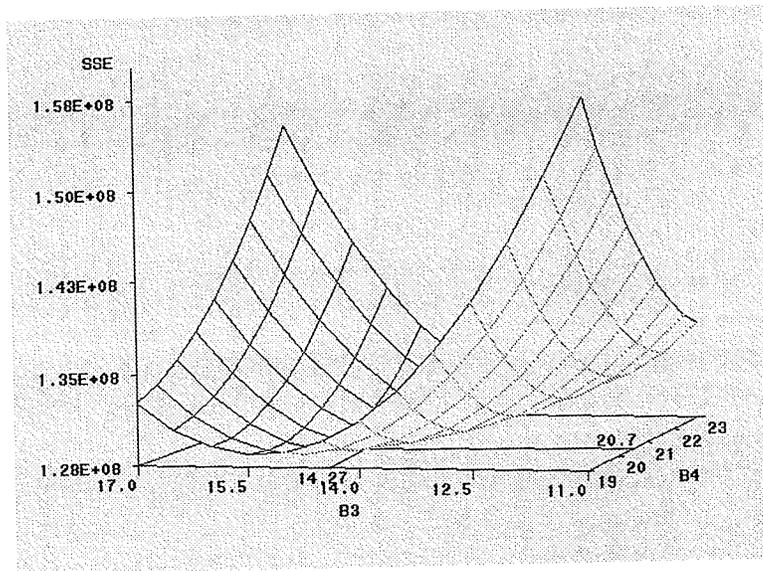


Figure 3.9: 3-Dimensional plot of $S^*(\hat{\theta}_3, \hat{\theta}_4)$ vs. θ_3 and θ_4 .

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	
Model	2	501630831	250815415	6158.30	
Error	3133	127601002	40728		
Corrected Total	3135	629231833			

Source	Pr > F
Model	<.0001

Root MSE	201.81193	R-Square	0.7972
Dependent Mean	2329.45918	Adj R-Sq	0.7971
Coeff Var	8.66347		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
B0	1	2421.19154	3.95534	612.13	<.0001
B1	1	-34.17581	0.31285	-109.24	<.0001
B2	1	37.38596	1.75433	21.31	<.0001

Figure 3.10: SAS output for piecewise linear fit using 14.25°C and 20.70°C as the unknown knots.

The parameter estimates are $\hat{\theta} = (2421.19, -34.18, 37.39)'$. The standard errors are $se(\hat{\theta}_0) = 3.96$, $se(\hat{\theta}_1) = 0.31$ and $se(\hat{\theta}_2) = 1.75$. We also have $s^2 = 40728$ and $df = 3133$. Again, the parameter estimates are statistically significant. Using these parameter estimates along with the estimates for the knots in model 3.41 the fitted model becomes

$$\hat{y} = 2421.19 - 34.18x + 34.18(x - 14.25)z_{14.25} + 37.39(x - 20.70)z_{20.70}, \quad (3.43)$$

where

$$z_{14.25} = \begin{cases} 1, & \text{if } x > 14.25, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$z_{20.70} = \begin{cases} 1, & \text{if } x > 20.70, \\ 0, & \text{otherwise.} \end{cases}$$

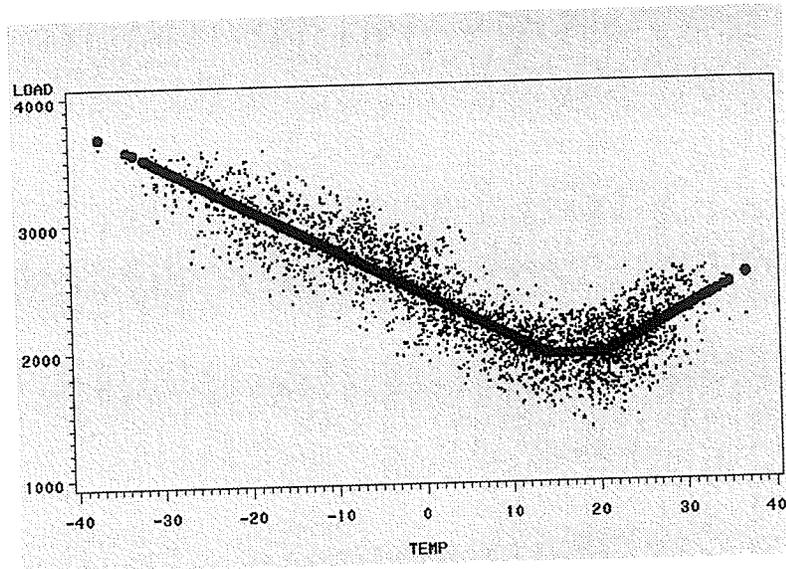


Figure 3.11: Piecewise fit using Grid Search method with 14.25 and 20.70 as the unknown knots.

We see that estimating the knots from the data gives us very accurate results. This is especially true when we don't have information on what the knots should be or if the knots vary when we use different data sets.

The derivative matrix, \mathbf{X} , for this model has the following form

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 - (x_1 - 14.25)z_{14.25} & (x_1 - 20.70)z_{20.70} \\ 1 & x_2 - (x_2 - 14.25)z_{14.25} & (x_2 - 20.70)z_{20.70} \\ \vdots & \vdots & \vdots \\ 1 & x_n - (x_n - 14.25)z_{14.25} & (x_n - 20.70)z_{20.70} \end{bmatrix}. \quad (3.44)$$

From 3.44 we obtain

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3136 & 11620.85 & 2931.7 \\ 11620.85 & 521889.17 & 41776.73 \\ 2931.7 & 41776.73 & 17970.13 \end{bmatrix}, \quad (3.45)$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} .00384 & -.0000044 & -.000053 \\ -.0000044 & .0000024 & -.0000049 \\ -.000053 & -.0000049 & .000076 \end{bmatrix}. \quad (3.46)$$

Using 2.17, with $t(3133; .025) = 1.96$, we can construct the 95% marginal confidence intervals for the parameters. A 95% marginal confidence interval for θ_0 is

$$2413.43 \leq \theta_0 \leq 2428.95,$$

a 95% marginal confidence interval for θ_1 is

$$-34.79 \leq \theta_1 \leq -33.57,$$

and a 95% marginal confidence interval for θ_2 is

$$33.96 \leq \theta_2 \leq 40.82.$$

We can also construct a 95% joint confidence region for the parameters. Using 2.20 and 3.45 with $p = 3$, $s^2 = 40728$ and $F(3, 3133; .05) = 2.61$, the 95% joint confidence region is

$$\begin{aligned} & 3136(\theta_0 - 2421.19)^2 + 521889.17(\theta_1 + 34.18)^2 + 17970.13(\theta_2 - 37.39)^2 + \\ & 23241.7(\theta_0 - 2421.19)(\theta_1 + 34.18) + 5863.4(\theta_0 - 2421.19)(\theta_2 - 37.39) \\ & + 83553.46(\theta_1 + 34.18)(\theta_2 - 37.39) \\ & = (3)(40728)(2.61) = 318900.24. \end{aligned}$$

We are confident that the Grid Search method has given us accurate parameter estimates. To verify this, we will next use the Dud method to come up with the

parameter estimates. We will see if these two methods agree.

3.5.3 Using the Dud Method

First of all we rewrite model 3.41 here,

$$y_i = \theta_0 + \theta_1 x_i - \theta_1(x_i - \theta_3)z_{\theta_3} + \theta_2(x_i - \theta_4)z_{\theta_4} + \varepsilon_i, \quad (3.47)$$

where z_{θ_3} and z_{θ_4} are the indicator variables that have been defined. For the same reasons we had in section 3.4.3 it is difficult to find the derivatives in model 3.47. This is because of the indicator variables, z_{θ_3} and z_{θ_4} . These are step functions in which derivatives, with respect to θ_3 and θ_4 respectively, are very hard to find. Therefore, we will use the derivative-free method, Dud, to find the parameter estimates.

The first thing we need are starting values for the parameters. We will use the parameter estimates and knots given by the linear fit in section 2.5. Therefore the starting values are $\theta^0 = (2372.60, -37.99, 24.32, 10, 20)'$. The output from PROC NLIN in SAS is given in Figure 3.12. Part of the SAS code used to produce the output is given in Appendix B. The output first shows us the six steps of the initialization phase. We notice that the first step of the iterative phase starts off with $\theta_3 = 11.00$ instead of 10.00. Notice that the initialization phase has chosen the parameter vector that gives the smallest sum of squares. The Dud procedure went through 24 iterations to obtain the parameter estimates that minimized $S(\hat{\theta})$. The parameter estimates for model 3.47 are $\hat{\theta} = (2421.30, -34.16, 37.29, 14.26, 20.70)'$. The knots estimated from the data are definitely different from the knots we assumed known with the linear model. The fitted model is then written as

$$\begin{aligned} \hat{y} = & 2421.30 - 34.16x + 34.16(x - 14.26)z_{14.26} \\ & + 37.29(x - 20.70)z_{20.70}, \end{aligned} \quad (3.48)$$

DUD Initialization						Sum of
DUD	theta0	theta1	theta2	theta3	theta4	Squares
-6	2372.6	-37.9900	24.3200	10.0000	20.0000	1.353E8
-5	2609.9	-37.9900	24.3200	10.0000	20.0000	3.1061E8
-4	2372.6	-41.7890	24.3200	10.0000	20.0000	1.4084E8
-3	2372.6	-37.9900	26.7520	10.0000	20.0000	1.3543E8
-2	2372.6	-37.9900	24.3200	11.0000	20.0000	1.3344E8
-1	2372.6	-37.9900	24.3200	10.0000	22.0000	1.3595E8

Iterative Phase						Sum of
Iter	theta1	theta1	theta2	theta3	theta4	Squares
0	2372.6	-37.9900	24.3200	11.0000	20.0000	1.3344E8
1	2429.8	-33.4550	38.5355	13.9339	21.9053	1.2835E8
2	2426.9	-33.6143	38.0831	14.6522	20.0948	1.2822E8
3	2426.6	-33.6110	38.0797	14.6621	20.1400	1.2813E8
4	2427.0	-33.8216	38.9506	14.6446	20.8392	1.2768E8
5	2427.1	-33.7987	38.7466	14.6574	20.8116	1.2768E8
6	2422.5	-34.0067	37.9872	14.4301	20.7284	1.2761E8
7	2421.5	-34.0863	37.3730	14.3128	20.7065	1.276E8
8	2421.5	-34.0931	37.3158	14.3009	20.7044	1.276E8
9	2421.8	-34.1308	37.1379	14.2942	20.6829	1.276E8
10	2421.6	-34.1209	37.2354	14.3087	20.6699	1.276E8
11	2421.6	-34.1298	37.3233	14.2983	20.6885	1.276E8
12	2421.5	-34.1564	37.5354	14.2793	20.7173	1.276E8
13	2421.3	-34.1684	37.5069	14.2714	20.7148	1.276E8
14	2421.4	-34.1614	37.2962	14.2786	20.6933	1.276E8
15	2421.4	-34.1637	37.2812	14.2755	20.6908	1.276E8
16	2421.3	-34.1617	37.2865	14.2713	20.6957	1.276E8
17	2421.3	-34.1609	37.2918	14.2696	20.6977	1.276E8
18	2421.3	-34.1579	37.2960	14.2695	20.6974	1.276E8
19	2421.3	-34.1562	37.2988	14.2709	20.6976	1.276E8
20	2421.3	-34.1577	37.3317	14.2682	20.7025	1.276E8
21	2421.3	-34.1629	37.2822	14.2605	20.6957	1.276E8
22	2421.3	-34.1619	37.2821	14.2591	20.6984	1.276E8
23	2421.3	-34.1616	37.2861	14.2588	20.6996	1.276E8
24	2421.3	-34.1624	37.2851	14.2585	20.6997	1.276E8

NOTE: Convergence criterion met.

Figure 3.12: SAS output showing iterations from the Dud method.

where

$$z_{14.26} = \begin{cases} 1, & \text{if } x > 14.26, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$z_{20.70} = \begin{cases} 1, & \text{if } x > 20.70, \\ 0, & \text{otherwise.} \end{cases}$$

We notice that there is a slight difference in the estimates found here with the estimates from the Grid Search procedure. As we have mentioned before, the Grid Search method is not efficient. Clearly, the grid has to be fine enough to give us accurate precision in the estimates [12]. Our Grid Search produced very good estimates, but perhaps we needed a finer grid to increase our precision.

Figure 3.13 shows us the rest of the SAS output. Given in the output are the parameter estimates along with their respective approximate standard errors. We notice that the residual sum of squares for this fit is lower than any other fit we have seen so far. This indicates that fitting a piecewise linear model with two unknown knots to the Hydro data may give us a better fit. The approximate standard errors are $se(\hat{\theta}_0) = 4.77$, $se(\hat{\theta}_1) = .40$, $se(\hat{\theta}_2) = 2.60$, $se(\hat{\theta}_3) = .36$ and $se(\hat{\theta}_4) = .42$. The approximate 95% marginal confidence intervals for the parameters are also given in the output. They were calculated using 3.26. The approximate 95% marginal interval for θ_0 is

$$2412 \leq \theta_0 \leq 2430.7,$$

the approximate 95% marginal interval for θ_1 is

$$-34.94 \leq \theta_1 \leq -33.38,$$

the approximate 95% marginal interval for θ_2 is

$$32.20 \leq \theta_2 \leq 42.37,$$

the approximate 95% marginal interval for θ_3 is

$$13.56 \leq \theta_3 \leq 14.96,$$

and the approximate 95% marginal interval for θ_4 is

$$19.87 \leq \theta_4 \leq 21.53.$$

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	5	1.752E10	3.5038E9	85972.9	<.0001
Residual	3131	1.276E8	40754.1		
Uncorrected Total	3136	1.765E10			
Corrected Total	3135	6.2923E8			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
theta0	2421.3	4.7674	2412.0	2430.7
theta1	-34.1624	0.3985	-34.9438	-33.3809
theta2	37.2851	2.5914	32.2041	42.3662
theta3	14.2585	0.3586	13.5553	14.9616
theta4	20.6997	0.4229	19.8704	21.5289

Approximate Correlation Matrix					
	theta0	theta1	theta2	theta3	theta4
theta0	1.0000000	0.2522757	0.0004895	0.5060848	0.0004711
theta1	0.2522757	1.0000000	0.0032593	0.5616752	0.0031774
theta2	0.0004895	0.0032593	1.0000000	0.0052170	0.6672235
theta3	0.5060848	0.5616752	0.0052170	1.0000000	-0.4168366
theta4	0.0004711	0.0031774	0.6672235	-0.4168366	1.0000000

Figure 3.13: SAS output for nonlinear fit using the Dud method.

The nonlinear fit to the data is given in Figure 3.14. The solid line in the plot indicates the fit with the slopes changing at 14.26°C and 20.70°C.

As we have done before, let us examine the residuals. The plot of residuals versus temperature, shown in Figure 3.15, has the same random pattern that we have seen

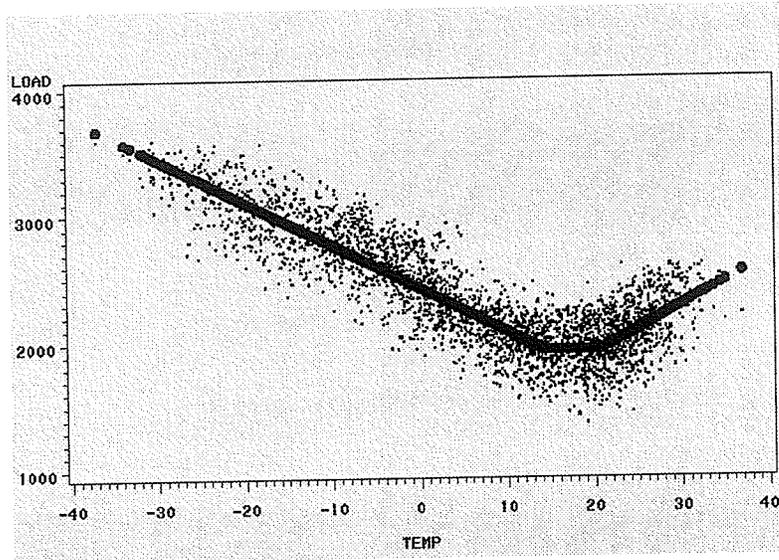


Figure 3.14: Nonlinear fit for piecewise linear model with two unknown knots.

before. There is no evidence to indicate that our model is incorrect. The normal probability plot of the residuals is shown in Figure 3.16. From the plot we can say that the error terms are following a normal distribution. We also see by the correlation matrix given in Figure 3.13, that we do not have a problem with high correlation among parameters.

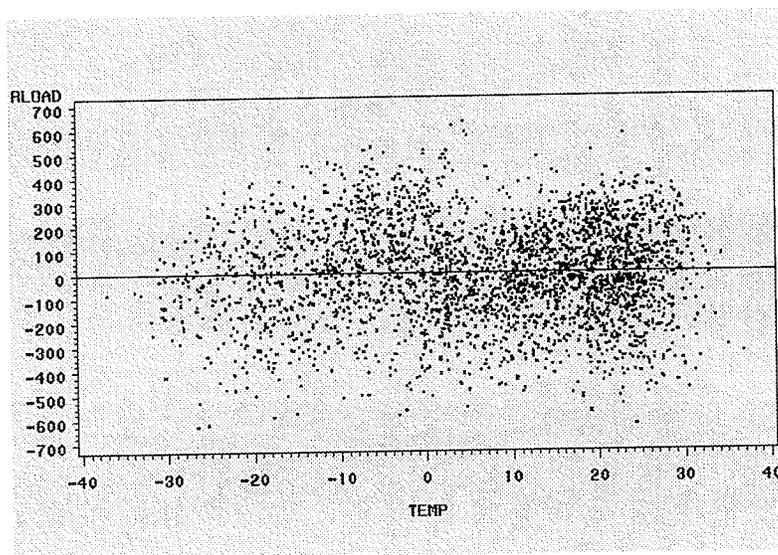


Figure 3.15: Plot of residuals vs. temperature for piecewise linear model with two unknown knots.

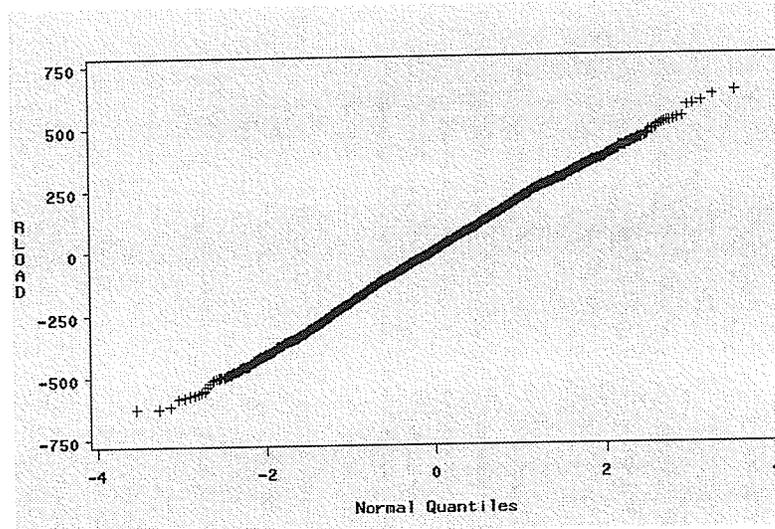


Figure 3.16: Normal probability plot of residuals

We tackled the two knot problem again in this section. This time we assumed the knots were unknown, and therefore used nonlinear regression to estimate the parameters. We found that the piecewise linear model with two unknown knots gave us the lowest residual sum of squares thus far. The residual plot showed us the same random pattern that we had noticed with the other fits. It seems that adding the extra knot to the model improved our fit to the data.

3.6 Conclusion

In this chapter we have examined our piecewise linear models a bit further. This time we assumed that the knots in the model were unknown. We showed how this now became a nonlinear problem and how iterative methods needed to be used to find the parameter estimates. We introduced nonlinear least squares and three of the iterative methods that are often used to find parameter estimates. One particular method, the Dud method, was shown to be an appropriate method to use for the Manitoba Hydro data.

In the piecewise model with one knot we saw that a more precise value of the

knot was found. In chapter 2 we used 18°C for our knot, but it turned out that the data estimated the knot to be 17.09°C instead. The iterative methods gave us a more accurate estimate of the knot. The same goes for the two knot model. The knots we used in the linear case were different from the knots that were estimated from the data. This shows that the iterative methods are an efficient way to produce appropriate estimates for the knots when we are not sure ourselves what the knots should be. Obviously, the knot will vary when looking at different data sets. If, instead of 6pm data, we examined 6am data then the knot would probably be at a different location. Nonlinear least squares uses the data to estimate where the knot is.

Again, as in chapter 2 we took a close look at the residuals. As before, the same patterns were evident in the residual plots and normal probability plots. Something that we touched upon earlier, is that other variables may be adding to the variation we see in load. There is definitely a need to address these variables in our models. In chapter 5 we will briefly look at this problem and see what possible adjustments we can make to our model.

Chapter 4

Further Aspects

4.1 Introduction

In chapters 2 and 3 we introduced four piecewise linear models. In chapter 2 we introduced two piecewise models that were linear in the parameters, which we will call piecewise linear models. Then in chapter 3 we introduced two piecewise models that were nonlinear in the parameters, we will call these piecewise nonlinear models. These models were used to describe the relationship between load and temperature in our Manitoba Hydro data. One of our goals is to try to determine which of these four models may be the best one to use. In section 4.2 we compare our four models to each other, by using the extra sum of squares test, or just by comparing the R^2 and mean square errors. In section 4.3 we discuss how profile likelihood intervals based on the sum of squares are constructed. We then compare our DUD linear approximation intervals to these profile likelihood intervals. We will discuss how adequate or inadequate these approximations are. In section 4.4 we briefly discuss the advantages and disadvantages of linear least squares. Then, in section 4.5, we discuss the advantages and disadvantages of nonlinear least squares. Through an example we will discuss how linear approximations sometimes are very inadequate in

certain nonlinear problems.

4.2 Comparing the Models

4.2.1 The Linear Models

In chapter 2 we looked at two possible linear models that we could use to fit the Manitoba Hydro data. It is of course of interest for us to compare these models. The two models we looked at in that chapter are known as non-nested models [9]. Non-nested models are models that are not reduced versions of each other, where both models have the same number of parameters. With non-nested models we would like to select the model that gives the best fit. The first model we looked at was the one knot linear model,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 18) z_{18} + \varepsilon_i. \quad (4.1)$$

The other model was the two knot linear model,

$$y_i = \beta_0 + \beta_1 x_i - \beta_1 (x_i - 10) z_{10} + \beta_2 (x_i - 20) z_{20} + \varepsilon_i. \quad (4.2)$$

If we were to choose one of these models, what criteria should we use to choose the more appropriate model? There are a few things we could look at to compare our two models. We will use the suggestion of Myers [5] and compare the R^2 value and the residual mean square, s^2 . For model 4.1 the R^2 value for the fitted model was .7944 compared to an R^2 value of .7841 for model 4.2. The R^2 is a good measure of the capability of the model to fit the data. On this criterion we would suggest choosing model 4.1.

The other measure that we can look at to compare the models, is the magnitude of the residual mean square, s^2 . Generally the model with the smallest residual mean square should be chosen [1]. When we fit model 4.1 to the data, $s^2 = 41297$. We

compare this to the fit of model 4.2, which gave us $s^2 = 43371$. From this criterion, again we would choose model 4.1.

Model 4.1, the one knot linear model, seems to be the model we would choose here. One thing to note here is that with model 4.2 we guessed at what our two knots should be, without any prior knowledge. We saw, in chapter 3, the nonlinear least squares procedure estimated the two knots from the data, and these knots were very different from the knots we guessed. This, of course, could bias our decision to what the more appropriate linear model here is.

4.2.2 The Nonlinear Models

In chapter 3 we looked at two possible nonlinear models that could be used to fit our load versus temperature data. The first model we looked at was the piecewise model with one unknown knot. It had the following form,

$$y_i = \theta_0 + \theta_1 x_i + \theta_2 (x_i - \theta_3) z_{\theta_3} + \varepsilon_i. \quad (4.3)$$

The other model was the piecewise model with two unknown knots. It had the form,

$$y_i = \theta_0 + \theta_1 x_i - \theta_1 (x_i - \theta_3) z_{\theta_3} + \theta_2 (x_i - \theta_4) z_{\theta_4} + \varepsilon_i. \quad (4.4)$$

In contrast to what we had in chapter 2, these two models are nested models. Model 4.4 is known as our full model. Model 4.3 is known as our partial (or reduced) model, because it has a reduced number of parameters. We want to find the simplest model which adequately fits our data. To compare the two models we use a likelihood ratio test [1]. Let S denote the residual sum of squares, df the corresponding degrees of freedom, and k the number of parameters. Let subscripts f and p denote full and partial (or reduced) models, and let subscript e denote extra. The calculations are

summarized in Table 4.1 We then compare the F-ratio to $F(df_e, df_f; \alpha)$ and accept

Source	DF	Sum of Squares	Mean Square	F-ratio
Extra	$df_e = k_f - k_p$	$S_e = S_p - S_f$	$s_e^2 = S_e/df_e$	$F = s_e^2/s_f^2$
Full model	$df_f = n - k_f$	S_f	$s_f^2 = S_f/df_f$	
Partial model	$df_p = n - k_p$	S_p		

Table 4.1: The extra sum of squares analysis for nested models.

the reduced model if the calculated F-ratio is lower than the table value.

In our problem we are testing the hypothesis

$$H_o : \text{use reduced 4 parameter model}$$

vs. the alternative,

$$H_a : \text{use full 5 parameter model.}$$

Using the information from our SAS outputs we can fill out Table 4.1. The results are shown in Table 4.2. The F-ratio is 32.87, and we compare this to $F(1, 3131; .05) =$

Source	DF	Sum of Squares	Mean Square	F-ratio
Extra	1	1339594.5	1339594.5	32.87
5-parameter	3131	127601097.1	40754.1	
4-parameter	3132	128940681.6		

Table 4.2: The extra sum of squares analysis for the 4 and 5 parameter piecewise model fitted to the Manitoba Hydro data.

3.84. The F-ratio is much larger than 3.84, so the p-value is very small. The likelihood test shows that we should add the extra parameter into the model. Therefore, model 4.4 may be the more appropriate model to use for our data. We note that for nonlinear models the F-ratio will not have an exact F-distribution. But, Bates and Watts [1], suggest that the approximation is good enough to give us good results.

4.2.3 Linear vs. Nonlinear Models

Finally, we can make two more comparisons. In all, we introduced four different models, two linear models and two nonlinear models. We would like to compare each linear model with its nonlinear counterpart. We have suggested before that estimating the knots from the data instead of setting fixed knots may give us a better fit. Now let us formally test that assumption.

First of all let us compare our one knot models. To repeat, the linear one knot model has the form

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 18) z_{18} + \varepsilon_i, \quad (4.5)$$

and the nonlinear one knot model has the form

$$y_i = \theta_0 + \theta_1 x_i + \theta_2 (x_i - \theta_3) z_{\theta_3} + \varepsilon_i, \quad (4.6)$$

These two models are nested. Model 4.6, with four parameters, builds upon model 4.5, which has three parameters. Therefore, we use the extra sum of squares technique, to see if adding the extra parameter into the model is appropriate. That is, test

$$H_o : \text{use reduced 3 parameter model}$$

vs.

$$H_a : \text{use full 4 parameter model.}$$

The extra sum of squares results are shown in Table 4.3. The F-ratio is 10.72 and

Source	DF	Sum of Squares	Mean Square	F-ratio
Extra	1	441318.4	441318.4	10.72
4-parameter	3132	128940681.6	41168.8	
3-parameter	3133	129382000		

Table 4.3: The extra sum of squares analysis for the 3 and 4 parameter piecewise model fitted to the Manitoba Hydro data.

we compare this to $F(1, 3132; .05) = 3.84$. The F-ratio is larger than 3.84 and the p-value is approximately .001. So we can conclude that adding the extra parameter into the model is appropriate. Therefore, model 4.6 seems to be the model we should choose here.

Finally, we can compare our two knot piecewise models. In chapter 2 we introduced the two knot linear model

$$y_i = \beta_0 + \beta_1 x_i - \beta_1(x_i - 10)z_{10} + \beta_2(x_i - 20)z_{20} + \varepsilon_i. \quad (4.7)$$

Then, in chapter 3, we built on model 4.7 and introduced the two knot nonlinear model

$$y_i = \theta_0 + \theta_1 x_i - \theta_1(x_i - \theta_3)z_{\theta_3} + \theta_2(x_i - \theta_4)z_{\theta_4} + \varepsilon_i. \quad (4.8)$$

Again, these two models are nested, therefore we will use the extra sum of squares technique to compare them. Model 4.7 has three parameters, and model 4.8 has five parameters. We want to see if adding the two extra parameters into the model is appropriate. We are testing,

H_o : use reduced three parameter model.

vs. the alternative,

H_a : use full five parameter model.

The extra sum of squares results are shown in Table 4.4. The F-ratio=101.59

Source	DF	Sum of Squares	Mean Square	F-ratio
Extra	2	8280556.9	4140278.45	101.59
5-parameter	3131	127601087.1	40754.1	
3-parameter	3133	135881644		

Table 4.4: The extra sum of squares analysis for the 3 and 5 parameter piecewise model fitted to the Manitoba Hydro data.

and we compare this to $F(2, 3131; .05) = 3.00$. The F-ratio is much larger than 3.00 and the p-value is very small. Therefore, we can conclude that adding the extra two parameters into the model is appropriate. The two knot nonlinear model is considered the more appropriate model to use for our data.

4.3 The Profile Likelihood Intervals Based on the Sum of Squares Function

4.3.1 The Piecewise Linear Model With One Unknown Knot

In section 3.4 we introduced the piecewise linear model with one unknown knot. In that case there were four parameters that we needed to estimate. But, the parameter that is of most importance to us is θ_3 , the knot. We obtained a point estimate of 17.09 and we also constructed an approximate 95% marginal confidence interval for θ_3 . This interval was

$$16.66 \leq \theta_3 \leq 17.53.$$

This approximate marginal interval was constructed using

$$\hat{\theta}_i \pm se(\hat{\theta}_i)t(n-p; \alpha/2), \quad (4.9)$$

where $i = 0, \dots, p-1$.

How does this approximate interval, from Dud, compare with a profile likelihood interval based on $S(\boldsymbol{\theta})$? In section 3.4 we constructed the profile sum of squares function, $S^*(\hat{\theta}_3)$ and plotted it in Figure 3.1. We use this plot here to construct the profile likelihood interval for θ_3 . Just as a reminder of how this plot was constructed, we varied θ_3 over a grid of possible values. At each θ_3 we fit a linear model and found the minimum sum of squares over all the other coordinates. The plot shown

in Figure 4.1 is the view of the sum of squares function along the θ_3 axis. We notice that $S^*(\hat{\theta}_3)$ has an almost symmetric parabolic shape. This tells us that there is only a small degree of nonlinearity in the model-data set combination [1]. We also see that the minimum of the sum of squares function corresponds to $\theta_3 = 17.09$, our least square estimate for θ_3 .

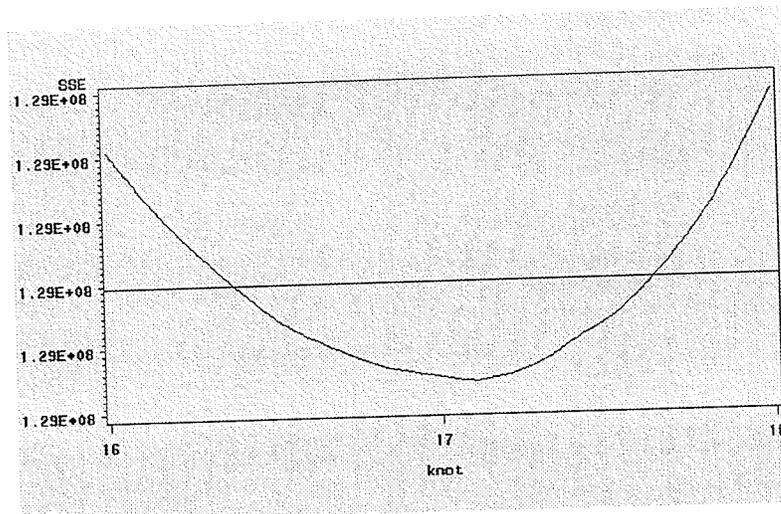


Figure 4.1: Sum of squares function for Hydro data for nonlinear one knot model. The solid straight line corresponds to the profile likelihood interval.

To associate a confidence level with the sum of squares function, we let the marginal likelihood interval be all values of θ such that

$$S(\theta) \leq S^*(\hat{\theta}_3) \left(1 + \frac{1}{n-p} F(1, n-p; .05)\right) \quad (4.10)$$

where $S^*(\hat{\theta}_3)$ is the sum of squares associated with the least squares estimates [1]. We superimpose the line given by 4.10 to the plot in Figure 4.1 as the solid straight line. The solution to the points where the straight line crosses the sum of squares function will correspond to our 95% profile likelihood interval. The profile likelihood interval turns out to be

$$16.38 \leq \theta_3 \leq 17.64.$$

This interval is wider than the approximate interval constructed from Dud. The

Dud method did give us a good approximation to the profile likelihood interval. It is true that when the degree of nonlinearity is small in a model-data set combination, the linear approximation intervals and likelihood intervals will be very close, see The Rumford 5 Example in Bates and Watts [1]. Later we will discuss an example where the degree of nonlinearity is very large and the linear approximations are not very good.

4.3.2 The Nonlinear Model With Two Knots

In chapter 3 we also introduced the piecewise linear model with two unknown knots. That model had five parameters, but the parameters that are of interest to us are θ_3 and θ_4 , the two knots. In chapter 3 we constructed approximate 95% marginal intervals. The intervals were as follows,

$$13.56 \leq \theta_3 \leq 14.96,$$

and

$$19.87 \leq \theta_4 \leq 21.53.$$

As we did with the one knot nonlinear model we want to compare these 95% Dud intervals with the profile likelihood intervals. First of all we define

$$S'(\hat{\theta}_3) = \min_{\theta_0, \theta_1, \theta_2, \theta_4} S(\theta). \quad (4.11)$$

We can then plot $S'(\hat{\theta}_3)$ versus θ_3 . Figure 4.2 shows us this sum of squares function. This plot is constructed by varying θ_3 over a grid of possible values. At each possible value, we run a least squares fit on the model to calculate the minimizing sum of squares over all coordinates. Note that everytime we run the least squares fit we are fitting a nonlinear model, because θ_4 is still included in the model. Again, the

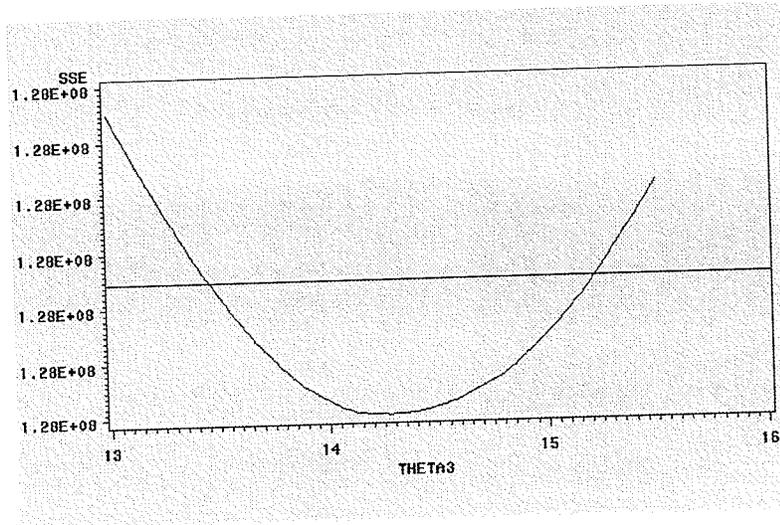


Figure 4.2: Plot of sum of squares function, $S'(\hat{\theta}_3)$ vs. θ_3 . The solid straight line corresponds to the profile likelihood interval.

sum of squares function has a nice parabolic shape, again we note that it is almost symmetric. The plot also shows the straight line corresponding to 4.10, but replacing $S^*(\hat{\theta}_3)$ with $S'(\hat{\theta}_3)$. The solutions to the intersection of the lines gives us our profile likelihood interval for θ_3 as

$$13.45 \leq \theta_3 \leq 15.22.$$

As before, the likelihood interval is wider, but Dud has produced a good approximation.

Finally, we look at the plot of the sum of squares function versus θ_4 . This particular sum of squares function is defined as follows,

$$S'(\hat{\theta}_4) = \min_{\theta_0, \theta_1, \theta_2, \theta_3} S(\theta). \tag{4.12}$$

The plot is given in Figure 4.3. Similar to the previous plot, this plot is constructed by varying θ_4 over a grid of possible values. At each value of θ_4 we run a nonlinear least squares fit to calculate the minimizing sum of squares over all coordinates. The least squares method is nonlinear because θ_3 is making the model be nonlinear at

each fitting. Again, we see the same parabolic shape to the sum of squares function.

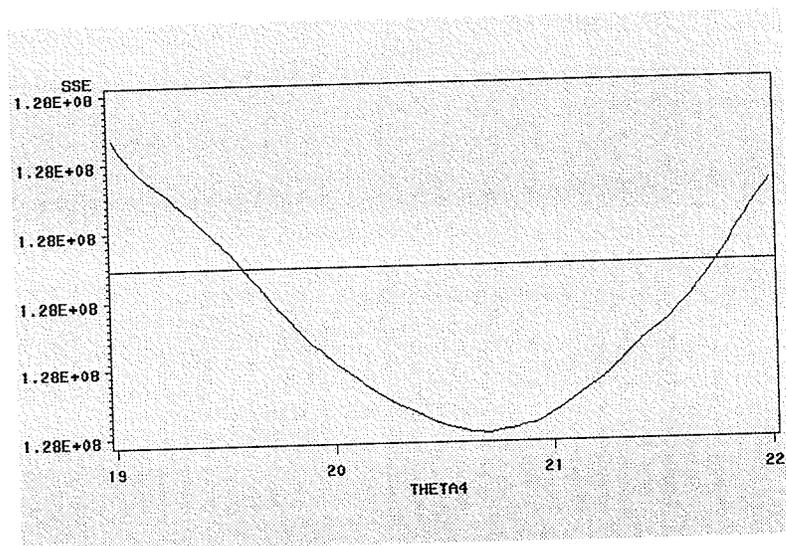


Figure 4.3: Plot of sum of squares function, $S'(\hat{\theta}_4)$ vs. θ_4 . The solid straight line corresponds to the likelihood interval.

The straight line in the plot corresponds to 4.10, but with $S'(\hat{\theta}_4)$ replacing $S^*(\hat{\theta}_3)$. The solution to the intersection of the straight line and $S'(\hat{\theta}_4)$ gives us the profile likelihood interval as

$$19.59 \leq \theta_4 \leq 21.74.$$

Once again, the likelihood intervals are wider, but Dud has produced a good approximation.

From this section we have seen that as long as the degree of nonlinearity is small, the Dud procedure will give good linear approximations. This is true of any nonlinear method [1]. In the next section we take a look at advantages and disadvantages of both the linear and nonlinear least squares methods. We will look at a disadvantage of nonlinear least squares when the degree of nonlinearity is high. This will be shown through an example.

4.4 Advantages and Disadvantages of Linear Least Squares

4.4.1 The Advantages

Linear least squares regression has earned its place as the primary tool for process modeling [7]. Although there are types of data that are better described by functions that are nonlinear in the parameters, many processes in science and engineering are well described by linear models. This can either be because the processes are linear or because, over short ranges of data, any process can be described to a high degree of accuracy by a linear model. The estimates of the unknown parameters obtained from linear least squares regression are unbiased and have minimum variance among all unbiased linear estimators. Linear least squares makes very efficient use of the data[5]. Good results can be obtained with relatively small data sets. The theory associated with linear regression is well understood and allows for construction of different types of easy to understand statistical intervals and regions for inference.

With our Manitoba Hydro data, the linear models were good models to use. This is true under the assumption that we know what our knots are and they are fixed. With predetermined knots we have simple piecewise linear models that described the relationship between load and temperature very well. Appropriate confidence intervals and regions were constructed, so that we could make proper conclusions about the unknown parameters.

4.4.2 The Disadvantages

The main disadvantages of linear least squares are limitations in the shapes that linear models can take over long ranges, poor extrapolation properties and sensitivity to outliers [5]. Linear models curve relatively slowly, so for nonlinear processes it

becomes increasingly difficult to find a linear model that fits the data well as the range of the data increases. As the explanatory variable increases to infinity, the predicted response from the linear model will also always go to positive or negative infinity. This means that linear models may not be effective for extrapolating the results of a process for which data cannot be collected in the region of interest. One may say that extrapolation is dangerous regardless of the model type, but linear models are especially bad choices for processes which it is known that the response does not approach infinity.

While the method of least squares often gives optimal estimates of the unknown parameters, it is very sensitive to unusual data points found in the data set. One or two outliers can sometimes seriously skew the results of a least squares analysis. That is why studying residual plots to examine potential outliers is important.

For our piecewise linear models we did not have the problem of outliers. So our results were very accurate. The problem we did have is that we were assuming fixed knots and using the least squares procedure to calculate parameter estimates. We witnessed in chapter 3 that estimating the knots from the data gave us reliable locations for knots. These knots were different from the knots we had used in chapter 2. In turn, using nonlinear methods and adding more parameters to our models gave us improved models for fitting the data. A nonlinear function seems to be more appropriate for our Manitoba Hydro problem.

4.5 Advantages and Disadvantages of Nonlinear Least Squares

4.5.1 The Advantages

The biggest advantage of nonlinear least squares regression over many other techniques is the broad range of functions that can be fit. Although many scientific and engineering processes can be described well using linear models, there are many other processes that are inherently nonlinear.

Being a "least sum of squares" procedure, nonlinear least squares shares some of the same advantages (and disadvantages) that linear least squares regression has over other methods. One shared advantage is the efficient use of the data. Nonlinear regression can produce good estimates of the unknown parameters in the model with relatively small datasets. Another advantage that nonlinear least squares shares with linear least squares is a fairly well developed theory for computing confidence intervals. In most cases the intervals produced by nonlinear regression are only approximately correct, but these intervals still work very well in practice.

In chapter 3 we used the Dud method to compute our parameter estimates. The Dud method is very useful when it is not possible to provide the derivatives required for other nonlinear procedures. With Dud, numerical estimates of the derivatives are computed from observed shifts in \hat{y} as the values of the θ_j are changed. As long as the data set is large enough, the Dud method will work just as well as any other iterative method.

In our Hydro problem, we introduced more parameters into our linear models that corresponded to our unknown knots. From there the relationship between load and temperature could only be explained by a nonlinear function. The Dud method seemed to give us very sensible results. In section 4.3 we compared the approximate

intervals from Dud with the profile likelihood intervals based on $S(\boldsymbol{\theta})$. We concluded that the Dud intervals were very good approximations to the profile likelihood intervals. This is due to the fact that there was a small degree of nonlinearity in the model-data set combination. By plotting the sum of squares function, $S(\boldsymbol{\theta})$, versus the parameters of interest we can make conclusions about the degree of nonlinearity. In all three plots of section 4.3, there was a symmetric parabolic shape to the sum of squares function. This indicated to us a small degree of nonlinearity.

Another way we could have made conclusions about the degree of nonlinearity in our two knot model was to plot the sum of squares contours. In Figure 4.4 we have done just that. In section 4.3 we plotted the sum of squares functions in one dimension, for the contours we are looking at the function in two dimensions and in a sense “slicing” across at the desired levels. Recall in chapter 3 that we plotted the $S^*(\hat{\theta}_3, \hat{\theta}_4)$ function versus θ_3 and θ_4 . To produce the contours we “slice” this function at our desired levels. We have plotted our two knots versus each other and produced the sum of squares contours. The plot gives us the 50%, 95%, and 99% contours. We

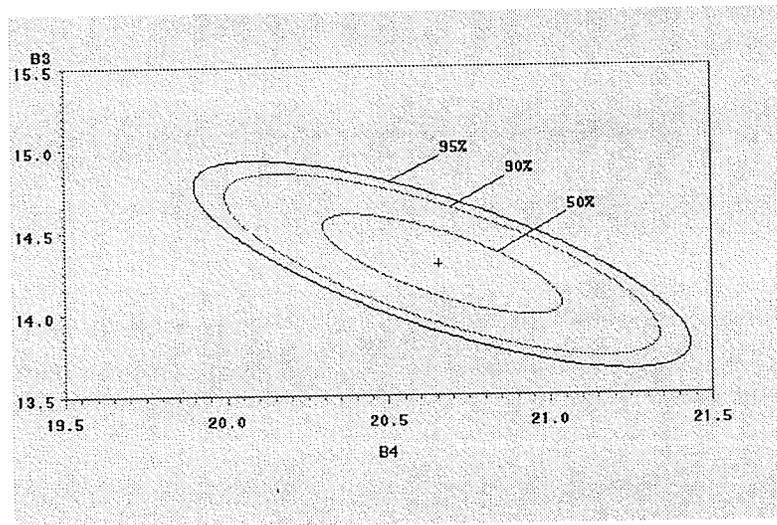


Figure 4.4: Sum of squares contours for two knot nonlinear model.

notice that these contours have a very nice elliptical shape to them. We come to the same conclusion as we did in section 4.3, the degree of nonlinearity is very small. The

Dud results are very good approximations to the likelihood results. In general, this will be true with any problem. If sum of squares contour plots show nice elliptical shapes, then the degree of nonlinearity is small. If the degree of nonlinearity is small, then the approximations given by nonlinear least squares will be very good.

4.5.2 The Disadvantages

The major disadvantage of moving to nonlinear least squares regression from simpler modeling techniques like linear least squares is the need to use iterative procedures to compute the parameter estimates. With functions that are linear in the parameters the least squares estimates of the parameters can always be obtained analytically, while that is generally not the case with nonlinear models. However, these procedures are easily performed by software such as SAS.

The use of iterative procedures requires the user to provide starting values for the unknown parameters before the software can begin the iteration process. The starting values must be reasonably close to the yet unknown parameter estimates or the procedure will not converge. Bad starting values can also cause the software to converge to a local minimum rather than the global minimum that defines the least squares estimates.

Disadvantages shared with linear least squares procedures include a strong sensitivity to outliers. Just as in the linear least squares analysis, the presence of one or two outliers in the data can seriously affect the results of a nonlinear analysis. Unfortunately, there are fewer model validation tools for detecting outliers in nonlinear regression than there are for linear regression [5].

As we mentioned above, the Dud method works well when we have a "rich" amount of data. There have been cases with relatively small data sets where the Dud method did not appear to work as well as the derivative-requiring methods [9]. In those cases convergence either was not obtained, was not as fast, or the final estimates did not

appear to be as good.

Earlier, we talked about the advantages of nonlinear least squares. One aspect we talked about is how the nonlinear methods will give us good approximations as long as the degree of nonlinearity in the model-data set combination is small. We showed that in our Hydro example the degree of nonlinearity was indeed very small and so the approximations given by Dud were very good. By plotting the sum of squares function and contours we can tell what the degree of nonlinearity is. Sum of squares functions with symmetric parabolic shapes and sum of squares contours that are quite elliptical both indicate small degrees of nonlinearity.

Here we take a look at an example which shows us the situation where the degree of nonlinearity is very large. The linear approximation will then not be very good. The following is taken from Bates and Watts [1].

BOD example

Data on biochemical oxygen demand (BOD) was obtained. To determine the BOD, a sample of stream water was injected with organic matter, inorganic nutrients, and dissolved oxygen, and divided into BOD bottles. Each bottle was inoculated with a mixed culture of microorganisms, sealed, and incubated at constant temperature. Then the bottles were opened periodically and analyzed for dissolved oxygen concentration from which the BOD was calculated in mg/L. The results are shown in Table 4.5. A model was derived to be

Time (days)	BOD (mg/L)
1	8.3
2	10.3
3	19.0
4	16.0
5	15.6
7	19.8

Table 4.5: BOD versus time.

$$y = \theta_1(1 - e^{-\theta_2 x}), \quad (4.13)$$

where y is the predicted BOD and x is time.

Using Proc NLIN in SAS we can fit a nonlinear model to this data. Since derivatives are easily obtained for this model, the Gauss-Newton method is used. The results from SAS are given in Figure 4.5.

The NLIN Procedure				
Parameter	Estimate	Approx Std Error	Approximate 95% Confidence	
theta1	19.1425	2.4959	12.2129	26.0722
theta2	0.5311	0.2031	-0.0327	1.0949

Approximate Correlation Matrix		
	theta1	theta2
theta1	1.0000000	-0.8528007
theta2	-0.8528007	1.0000000

Figure 4.5: Proc NLIN output for BOD data.

The parameter estimates are $\hat{\theta} = (19.143, 0.5311)'$. The approximate 95% marginal confidence interval for θ_1 is

$$12.21 \leq \theta_1 \leq 26.07,$$

and the approximate 95% marginal confidence interval for θ_2 is

$$-0.03 \leq \theta_2 \leq 1.09.$$

How do these approximate intervals compare to the profile likelihood intervals? What is the degree of nonlinearity in the model-data set combination? To answer the second question let us take a look at the sum of squares contour plots. Recall in the last section we plotted the sum of squares contours for our two knot nonlinear problem. We saw that the contours were quite elliptical, indicating a small degree of nonlinearity. Therefore, the approximate intervals from Dud were very good approx-

imations to the likelihood profile intervals. The sum of squares contours for the BOD example are shown in Figure 4.6. We notice right away that the contours do not

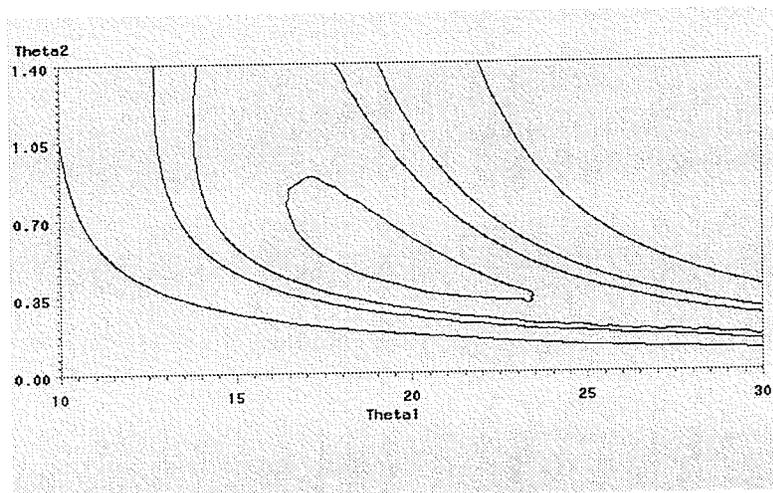


Figure 4.6: Sum of squares contours for the BOD data.

follow elliptical shapes. One can even consider these contours as having hyperbolic shapes. We would hypothesize that the Gauss-Newton linear approximations are not going to be adequate.

Let us take a look at just one of the parameters, θ_1 . By varying θ_1 over a grid of possible values and calculating the minimizing sum of squares over the coordinates for both parameters we can then plot the sum of squares function versus θ_1 . The sum of squares function has the form,

$$S^*(\hat{\theta}_1) = \min_{\theta_2} S(\theta). \quad (4.14)$$

That plot is given in Figure 4.7. We notice a few things right away. First of all the sum of squares function is definitely not parabolic. The function is asymmetric, with a very apparent steeper rise to the right of the minimum than to the left. From SAS we obtained the least squares estimate for θ_1 to be 19.143, which corresponds to the minimum point of the function in Figure 4.7. In the plot we also have the straight line corresponding to 4.10, with $S^*(\hat{\theta}_1)$ replacing $S^*(\hat{\theta}_3)$. The intersection of $S^*(\hat{\theta}_1)$

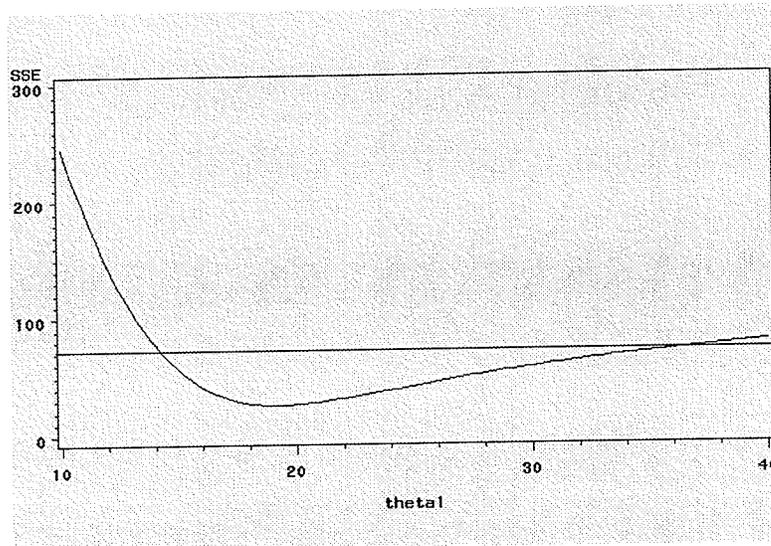


Figure 4.7: Sum of squares function for BOD data for θ_1 .

with this line gives us a 95% profile likelihood interval for θ_1 as

$$14.2 \leq \theta_1 \leq 36.8.$$

This is definitely different from the linear approximation interval given by the Gauss-Newton method. The profile likelihood interval is much wider because of the high degree of asymmetry in the function. This example shows us that in many nonlinear analyses, linear approximations provide inadequate summaries of the inferential results. This is true especially when the degree of nonlinearity is high. There are improved graphical methods for summarizing inferential results of a nonlinear analysis. Bates and Watts [1] discuss methods such as profile t plots, profile traces and profile pair sketches.

4.6 Conclusion

The purpose of this chapter was to take a look at some further considerations when it comes to linear and nonlinear least squares. We started off by comparing our four

models. We concluded that the five parameter piecewise linear model with unknown knots was the most adequate model of the four to use in our Hydro problem. We then looked at how likelihood profile intervals based on $S(\boldsymbol{\theta})$ are constructed. The sum of squares function had nice parabolic shapes and so our DUD linear approximation intervals were adequate representations of the likelihood profile intervals. Since our functions were symmetric and parabolic, the model-data set combination had a small degree of nonlinearity. Briefly we discussed the advantages and disadvantages of linear least squares. We then turned our attention to the advantages and disadvantages of nonlinear least squares. Sum of squares contours were constructed for our two knot nonlinear problem. We saw that the contours had nice elliptical shapes, which indicated a small degree of nonlinearity. Therefore, the linear approximation intervals were again good approximations to the likelihood profile intervals. Through the BOD example, we showed the case where the linear approximations are very inadequate. This was due to the large degree of nonlinearity in the model-data set combination.

Chapter 5

Modifying The Model

5.1 Introduction

Up to this point we have looked at a few different models that could be used to describe the relationship between load and temperature. What we have done so far is only taken a subset of the data. We have only taken a look at all data recorded at 6pm. We have shown that the five parameter piecewise linear model with unknown knots fit the data the best. From now on, we will only discuss this five parameter nonlinear model. We may want to ask a few questions. How do the parameters in the model change for different times of the day? What about over different days? months? years? Is there one model that we can use to incorporate all these factors? Or do we need to use several different models for different combinations? In this chapter we will try to answer some of these questions by examining the data further and fitting our model to different subsets of the data.

5.2 The Time of Day Affect

Throughout our discussion so far we have only taken a look at 6pm data. The fitted five parameter model for the data is

$$\hat{y} = 2421.19 - 34.18x + 34.18(x - 14.25)z_{14.25} + 37.39(x - 20.70)z_{20.70}, \quad (5.1)$$

where

$$z_{14.25} = \begin{cases} 1, & \text{if } x > 14.25, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$z_{20.70} = \begin{cases} 1, & \text{if } x > 20.70, \\ 0, & \text{otherwise.} \end{cases}$$

But, how does the fitted model change over the course of a day? With the temperature varying so much over the period of 24 hours, it is right to assume that the parameters will change from hour to hour. To examine this, we ran a nonlinear analysis using PROC NLIN on data from each hour of the day. Table 5.1 shows us how the five parameters change over the course of a day. By examining Table 5.1 we do see that the parameters are definitely changing from hour to hour.

To get a better idea of how much the parameters vary let us take a look at Figure 5.1. Figure 5.1 gives us four different plots. We have the plots of θ_1 , θ_2 , θ_3 and θ_4 each versus hour of day. We can see that the “heating slope”, θ_1 , varies a lot during the course of a day. We see that after midnight the slope decreases until about 8am, then from there the slope increases until it reaches a maximum at about 3pm and then begins to decrease again. It then shows a smaller peak at around 10 or 11pm. Looking at the “cooling slope”, θ_2 , we see that the variation isn’t as much, but the slope does peak at about 10am and then slowly decreases throughout the rest of the

Hour	θ_0	θ_1	θ_2	θ_3	θ_4
1	1858.4	-33.40	27.35	11.06	15.40
2	1793.6	-33.95	27.92	10.89	16.04
3	1760.2	-34.62	29.75	10.81	16.50
4	1745.6	-35.20	24.18	10.99	15.40
5	1745.8	-35.88	24.55	10.91	15.68
6	1783.5	-36.73	20.91	11.30	15.20
7	1916.6	-37.33	26.59	12.70	16.54
8	2149.3	-36.47	35.95	14.87	17.80
9	2311.5	-33.88	48.06	15.10	19.67
10	2373.8	-31.52	50.44	15.69	20.80
11	2392.7	-29.87	46.68	15.09	21.50
12	2401.6	-28.49	46.52	14.58	22.43
13	2368.7	-27.55	46.62	14.02	23.01
14	2343.4	-27.18	43.06	13.74	22.87
15	2317.2	-27.15	38.13	13.80	22.12
16	2314	-28.04	36.72	13.80	21.44
17	2369.6	-30.24	37.85	14.20	21.38
18	2421.2	-34.18	37.39	14.25	20.70
19	2387.6	-35.219	41.95	14.24	20.80
20	2353.5	-34.97	40.59	14.37	20.11
21	2315.8	-33.73	36.52	14.03	19.00
22	2255.1	-32.68	33.42	13.00	17.30
23	2139.9	-32.54	36.69	11.60	16.66
24	1974.3	-32.86	33.00	10.90	16.07

Table 5.1: θ_i versus hour of day.

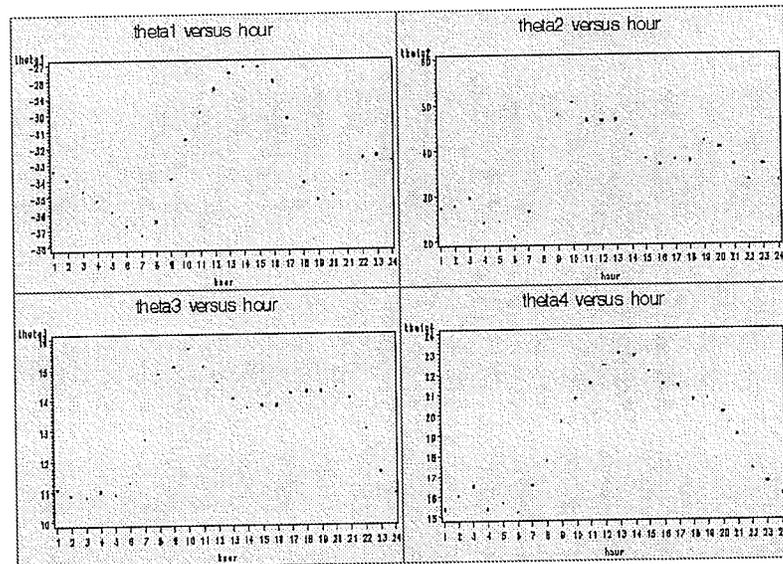


Figure 5.1: Plots of θ_1 , θ_2 , θ_3 and θ_4 each versus hour of day.

day. The first knot, θ_3 , peaks at about 10am and then decreases, remains constant during the evening hours and then decreases over night. Finally, the second knot, θ_4 , remains somewhat constant over night and then increases to a maximum at about 1pm and then steadily decreases throughout the rest of the day.

So, what does all of this mean? Do we need to have a different model for each hour of the day? Do we somehow incorporate a time of day variable into our model? The choice is up to us. Going in either direction will not give us incorrect results. As we will shortly see, there are other variables contributing to the variation in load. These variables will need to be incorporated into the model. We do not want a model that is too complicated, having too many parameters. Adding some sort of indicator variable for each hour of the day will only complicate our model even more. For now, we will choose to fit our model 24 times, once for each hour of the day.

In the next few sections we will look at three additional variables that are adding to the variation in load. These three variables are day, month and year.

5.3 The Day of the Week Affect

In section 3.5 we fit a piecewise linear model with two unknown knots to the Manitoba Hydro 6pm data. One of the things we did to check the assumptions of the model was to examine the residuals. Figure 5.2 shows us the plot of residuals versus temperature that we had originally in section 3.5.

We see from the plot that the residuals have a nice random scatter with no obvious patterns and no obvious outliers. But, we also notice that there is a lot of variation. What we will do here is examine our residuals even further. Let us break down the residuals by day of the week. Figure 5.3 shows us the plot of residuals versus day of the week. We can see that there is an apparent pattern in the residuals. The residuals seem to be constant except for a difference for Friday, Saturday and Sunday. This

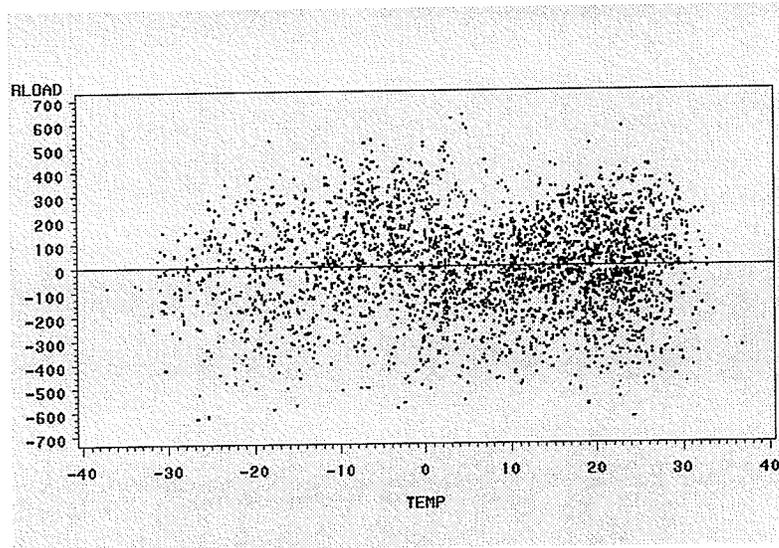


Figure 5.2: Plot of residuals vs. temperature for piecewise linear fit with two unknown knots.

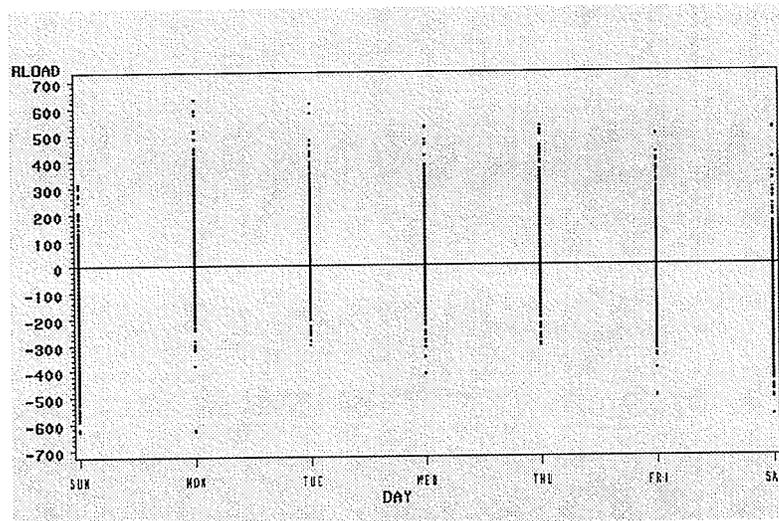


Figure 5.3: Plot of residuals vs. day of the week for original piecewise linear model with two unknown knots.

means that we may need to adjust our model by adding some sort of indicator variable for each of these three days in the model. A suggestion for the new model would be

$$y_i = \theta_0 + \theta_1 x_i - \theta_1(x_i - \theta_3)z_{\theta_3} + \theta_2(x_i - \theta_4)z_{\theta_4} + \gamma_1 \text{Friday} + \gamma_2 \text{Saturday} + \gamma_3 \text{Sunday} + \varepsilon_i, \quad (5.2)$$

where

$$z_{\theta_3} = \begin{cases} 1, & \text{if } x_i > \theta_3, \\ 0, & \text{otherwise,} \end{cases}$$

$$z_{\theta_4} = \begin{cases} 1, & \text{if } x_i > \theta_4, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$Day = \begin{cases} 1, & \text{if data point is from } Day, \\ 0, & \text{otherwise.} \end{cases}$$

If data is from a day other than Friday, Saturday or Sunday than we have our original nonlinear two knot model.

When we fit model 5.2 to the data using PROC NLIN in SAS we obtain the output in Figure 5.4. The fitted model is written as

$$y = 2507.60 - 34.33x + 34.33(x - 14.11)z_{14.11} + 36.50(x - 20.69)z_{20.69} - 215.60 \text{Friday} - 270.20 \text{Saturday} - 68.10 \text{Sunday}. \quad (5.3)$$

We notice from the SAS output that none of the approximate 95% confidence intervals cover the value zero. That means we can say that all the parameters in the model are significant and should be included in the model.

We may also perform an extra sum of squares test to see if adding the three extra

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	8	1.756E10	2.1946E9	76903.2	<.0001
Residual	3128	89265736	28537.6		
Uncorrected Total	3136	1.765E10			
Corrected Total	3135	6.2923E8			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
theta0	2507.6	4.8505	2498.1	2517.1
theta1	-34.3287	0.3356	-34.9848	-33.6686
theta2	36.4967	2.1724	32.2371	40.7563
theta3	14.1086	0.2976	13.5251	14.6920
theta4	20.6929	0.3612	19.9847	21.4011
gamma1	-215.6	8.7926	-232.8	-198.4
gamma2	-270.2	8.4554	-286.8	-253.6
gamma3	-68.1017	9.0226	-85.7927	-50.4106

Figure 5.4: PROC NLIN output for Hydro data using nonlinear model with added day component.

parameters into the model is valid. Table 5.2 shows the results of the extra sum of squares test for the eight parameter model. We see that the F-ratio is very large and

Source	DF	Sum of Squares	Mean Square	F-ratio
Extra	3	38335331.1	12778450.4	447.78
8-parameter	3128	89265736	28537.6	
5-parameter	3131	127601087.1		

Table 5.2: The extra sum of squares analysis for the 5 and 8 parameter piecewise model fitted to the Manitoba Hydro data.

therefore highly significant. Therefore it is valid to add the three extra parameters into our nonlinear model.

Let us again examine the plot of residuals versus day of the week to see if the new model has adjusted for the day factor. Figure 5.5 shows us the new plot of residuals versus day. We see that the residuals are essentially constant throughout the whole week. There are no significant deviations that we need to adjust for. Model 5.2 has taken into account the day variable.

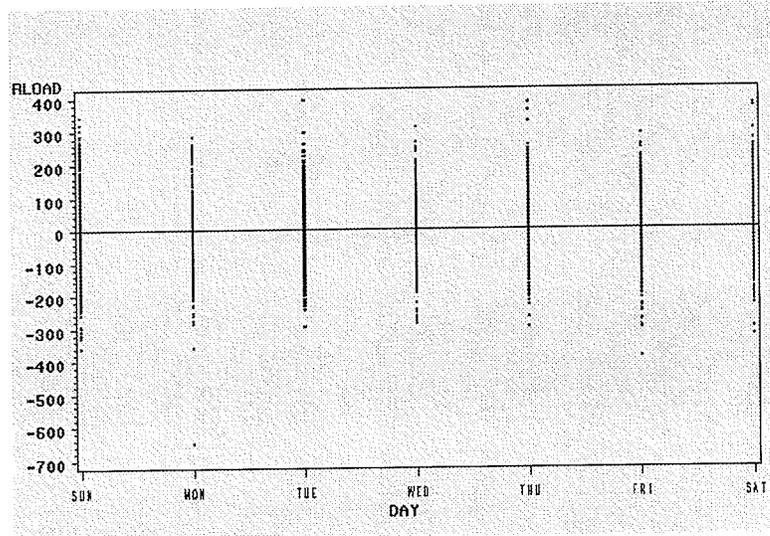


Figure 5.5: Plot of residuals vs. day of the week after adjusting nonlinear model for day variable.

5.4 The Seasonal Affect

Another variable that we assume may be adding to the variation is a seasonal or monthly variable. With the temperature changing so much from month to month and season to season, it is right to assume that we may need to address this in our model. Figure 5.6 shows us the plot of residuals versus month after fitting our original five parameter nonlinear model. There is definitely a pattern in the residuals. In particular we see that the residuals are higher than average for the months of November, December and January, and the residuals are lower than average for the months of March, April, and May. The residuals for the rest of the months seem to be constant. We suggest modifying model 5.2 by adding what we will call a “winter” variable (November, December, January) and a “spring” variable (March, April, May). The model is then as follows

$$y_i = \theta_0 + \theta_1 x_i - \theta_1(x_i - \theta_3)z_{\theta_3} + \theta_2(x_i - \theta_4)z_{\theta_4} \quad (5.4)$$

$$+ \gamma_1 \text{ Friday} + \gamma_2 \text{ Saturday} + \gamma_3 \text{ Sunday}$$

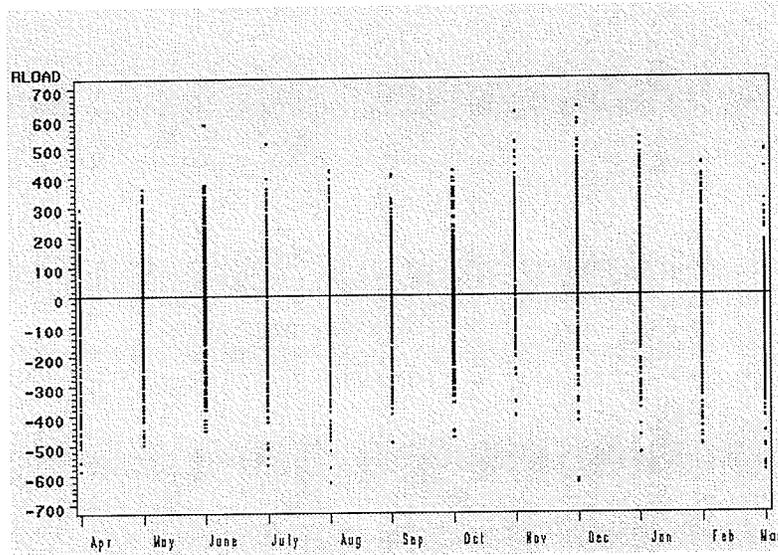


Figure 5.6: Plot of residuals vs. month for our original piecewise linear model with two unknown knots.

$$+\alpha_1 \text{ Winter} + \alpha_2 \text{ Spring} + \varepsilon_i,$$

where the conditions for z_{θ_3} , z_{θ_4} , and Day are still the same as in model 5.2. There is the extra condition,

$$Season = \begin{cases} 1, & \text{if data point is from } Season, \\ 0, & \text{otherwise.} \end{cases}$$

Using PROC NLIN in SAS we fit model 5.4 to our Hydro data. The output from SAS is given in Figure 5.7. The fitted model is written as

$$\begin{aligned} y = & 2454.8 - 28.29x + 28.29(x - 14.80)z_{14.80} & (5.5) \\ & + 36.14(x - 20.81)z_{20.81} - 221.1 \text{ Friday} \\ & - 274.20 \text{ Saturday} - 68.79 \text{ Sunday} \\ & + 229.60 \text{ Winter} - 48.68 \text{ Spring.} \end{aligned}$$

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	10	1.758E10	1.7578E9	80746.5	<.0001
Residual	3126	68052246	21769.8		
Uncorrected Total	3136	1.765E10			
Corrected Total	3135	6.2923E8			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
theta0	2454.8	5.5475	2443.9	2465.7
theta1	-28.2911	0.3439	-28.9653	-27.6169
theta2	36.1411	1.9136	32.3889	39.8933
theta3	14.8001	0.3256	14.1616	15.4386
theta4	20.8102	0.3256	20.1719	21.4485
alpha1	229.6	8.6058	212.7	246.5
alpha2	-49.6815	3.6493	-56.8369	-42.5262
gamma1	-221.1	7.6027	-236.0	-206.2
gamma2	-274.2	7.3863	-288.7	-259.8
gamma3	-68.7915	7.7689	-84.0244	-53.5586

Figure 5.7: PROC NLIN output for Hydro data using nonlinear model with added season component.

According to the approximate confidence intervals in the output, we can say that all parameters are significant and should be included in the model.

Again, we will conduct an extra sum of squares test. We will see if adding the two extra parameters to model 5.2 is a valid thing to do. Table 5.3 shows us the results of the extra sum of squares test. As in the previous section we have a very large F-ratio,

Source	DF	Sum of Squares	Mean Square	F-ratio
Extra	2	21213490	10606745	487.22
10-parameter	3126	68052246	21769.9	
8-parameter	3128	89265736		

Table 5.3: The extra sum of squares analysis for the 5 and 8 parameter piecewise model fitted to the Manitoba Hydro data.

which is highly statistically significant. Therefore, we conclude that adding the two extra parameters into the model is appropriate.

We can now plot the residuals versus month after fitting model 5.4 to the data. Figure 5.8 shows us the plot of residuals versus month. There is a nice random pattern, with no particular month showing any unusually high or low values for residuals. We

can see that fitting model 5.4 has eliminated the problem we had with the pattern in the residuals as seen in Figure 5.6.

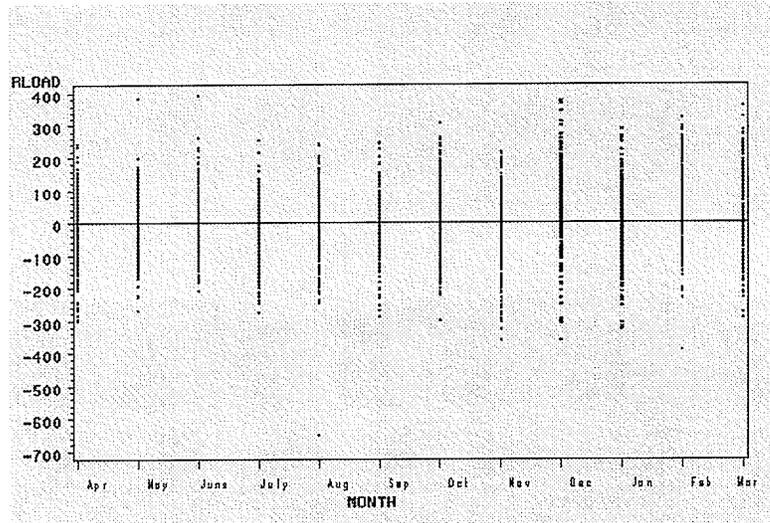


Figure 5.8: Plot of residuals vs. month after adjusting nonlinear model for seasonal variable.

5.5 The Year Affect

The Manitoba Hydro data that we have been using is data collected since 1992 until the year 2000. One of the other variables we considered was year. We assume that load may vary from year to year. Figure 5.9 is the plot of residuals versus year after fitting our original nonlinear five parameter model. By examining the plot we can see that there is an upward trend in the residuals. One reason for the trend could be that in recent years people are buying more air conditioning units and therefore the demand for electricity has gone up. It seems that we need to year adjust our model to take into account the trend. Therefore we modify model 5.4 by adding a year component. The new model is as follows

$$y_i = \theta_1 x_i - \theta_1 (x_i - \theta_3) z_{\theta_3} + \theta_2 (x_i - \theta_4) z_{\theta_4} + \gamma_1 \text{Friday} + \gamma_2 \text{Saturday} + \gamma_3 \text{Sunday} \quad (5.6)$$

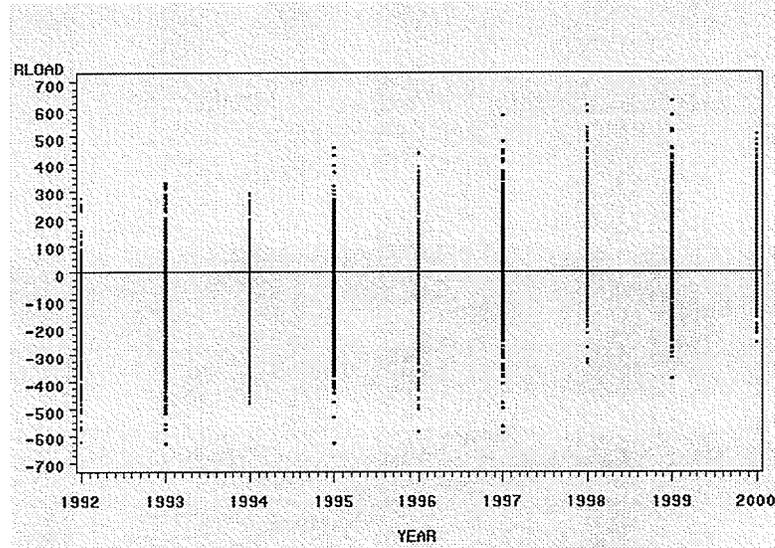


Figure 5.9: Plot of residuals vs. year for our original nonlinear piecewise model with two unknown knots.

$$\begin{aligned}
 & +\alpha_1 \text{ Winter} + \alpha_2 \text{ Spring} \\
 & +\phi_1 1992 + \phi_2 1993 + \phi_3 1994 + \phi_4 1995 + \phi_5 1996 \\
 & +\phi_6 1997 + \phi_7 1998 + \phi_8 1999 + \phi_9 2000 + \varepsilon_i,
 \end{aligned}$$

where the conditions for z_{θ_3} , z_{θ_4} , *Day*, and *Season* are as given in 5.4 and

$$\text{Year} = \begin{cases} 1, & \text{if data point is from } Year, \\ 0, & \text{otherwise.} \end{cases}$$

We notice that we do not need the “height” parameter, θ_0 in our model anymore. When we ran PROC NLIN with θ_0 in the model, the output did not give a standard error for θ_0 . There was also some sort of interaction that was seen between θ_0 and ϕ_8 . For these reasons we decided to eliminate θ_0 from our model.

We then fit model 5.6 to our data using PROC NLIN. The output from SAS is shown in Figure 5.10. All parameters seem to be significant according to the

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	18	1.761E10	9.7837E8	85608.7	<.0001
Residual	3118	35633878	11428.4		
Uncorrected Total	3136	1.765E10			
Corrected Total	3135	6.2923E8			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
theta1	-28.4549	0.2542	-28.9532	-27.9565
theta2	35.3339	1.4428	32.5050	38.1628
theta3	14.6367	0.2315	14.1827	15.0907
theta4	21.0829	0.2401	20.6120	21.5538
alpha1	232.3	6.4771	219.6	245.0
alpha2	-55.8448	4.9437	-65.5381	-46.1515
gamma1	-221.0	5.5549	-231.9	-210.1
gamma2	-274.4	5.3518	-284.9	-263.9
gamma3	-67.9720	5.7096	-79.1671	-56.7768
phi1	2296.0	7.5645	2281.1	2310.8
phi2	2345.7	6.7043	2332.5	2358.8
phi3	2377.0	6.8387	2363.6	2390.4
phi4	2394.2	6.7782	2380.9	2407.4
phi5	2445.6	6.7299	2432.4	2458.8
phi6	2494.5	6.7988	2481.2	2507.8
phi7	2594.2	6.8893	2580.7	2607.7
phi8	2550.9	6.8165	2537.6	2564.3
phi9	2602.6	7.2144	2588.4	2616.7

Figure 5.10: PROC NLIN output for Hydro data using nonlinear model with added year component.

approximate confidence intervals. The fitted model is written as follows

$$\begin{aligned}
 y = & -28.46x + 28.46(x - 14.64)z_{14.64} & (5.7) \\
 & +35.33(x - 21.08)z_{21.08} - 221.00 \text{ Friday} - 274.40 \text{ Saturday} \\
 & 67.97 \text{ Sunday} + 232.30 \text{ Winter} - 55.85 \text{ Spring} \\
 & +2296.00 * 1992 + 2345.70 * 1993 + 2377.00 * 1994 \\
 & +2394.20 * 1995 + 2445.60 * 1996 + 2494.50 * 1997 \\
 & +2594.20 * 1998 + 2550.9 * 1999 + 2602.6 * 2000.
 \end{aligned}$$

Model 5.6 has eighteen parameters. Let us perform the extra sum of squares test to see if adding the extra parameters is valid. Table 5.4 shows us the extra sum of squares results. As we had seen in the previous two sections, the F-ratio for the test is

Source	DF	Sum of Squares	Mean Square	F-ratio
Extra	8	21213490	32418368	354.58
18-parameter	3118	35633878	11428.4	
10-parameter	3126	68052246		

Table 5.4: The extra sum of squares analysis for the 10 and 18 parameter piecewise model fitted to the Manitoba Hydro data.

large and statistically significant. This tells us that adding the eight extra parameters into the model was appropriate.

Let us now look at the plot of the residuals versus year after fitting model 5.6 to the data. Figure 5.11 shows us the plot. We see that the upward trend that was present in Figure 5.9 is no longer apparent. Model 5.6 has taken into account the trend we noticed in the residuals.

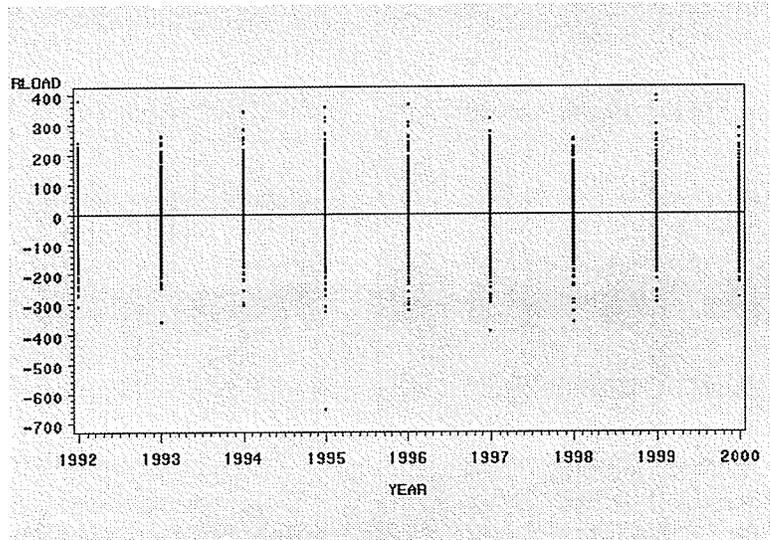


Figure 5.11: Plot of residuals vs. year after adjusting nonlinear model for year variable.

5.6 Conclusion

In this chapter we examined the data a little further. We investigated how variables such as year, month and day of the week affect our model. From residual plots we noticed trends and patterns that we needed to take into account. Using the 6pm data we created model 5.6 that could be used to best describe the relationship between load and temperature. In section 5.2 we noticed how the parameters change over the course of a day, from hour to hour. We suggest that one should fit model 5.6 for whatever hour of the day the investigator is interested in.

What we have done in this chapter is just one suggestion for a model to fully describe the relationship between load and temperature. There are other ways one can tackle the problem. For example, instead of the year component we could have a time of day component in model 5.6 and then fit a different model for each year. Engle *et al.* [3] came up with a model similar to ours but they added a holiday component, to account for holidays such as New Year's, Christmas, etc.

Chapter 6

Conclusion

This study has developed regression models for forecasting peak hourly load. In chapter 1 we discussed that one of the main goals of this practicum was the notion that we need to somehow temperature-adjust the historical load data. We introduced piecewise linear models with both predetermined and unknown knots that could be used to describe the relationship between load and temperature. Before introducing the models we gave the background of linear and nonlinear least squares. This background information helped us understand how parameter estimates are calculated and how confidence regions and intervals are constructed.

Another focus of this practicum was to examine the nonlinear methods that are utilized in the fitting of piecewise linear models. The Dud method was discussed as the nonlinear method of choice in our problem and is the method that Manitoba Hydro has used for many years. The Dud method gave us very accurate approximations to likelihood results. We concluded that a piecewise linear model with two unknown knots would give us the best fit to our Hydro data. This conclusion was made after conducting extra sum of squares tests on the models. For further discussions on the Dud method, see Ralston and Jenrich [8].

We did see that nonlinear least squares does not always give us accurate results.

Through the BOD example from Bates and Watts [1], we saw how the linear approximations were not always good in giving us good inferential results. The approximations were very much different than the likelihood results. Further work could be done here to examine such methods as profile t plots, profile traces, and profile pair sketches as alternatives to the linear approximations. But with our Hydro problem we did not see any problem with the Dud approximations.

In chapter 5 we briefly discussed how other variables are contributing to the variation in load. We examined how fitted curves change over years, months, days of the week, and over the course of a day. We constructed a model that incorporated these variables. This model was seen to be very good in describing the relationship. More work could be done here. One could see if there are better models that could be used to fit the data. If the parameters are changing significantly over different combinations of years, months, and days, then we can use “random coefficient models” to describe the relationships. For the larger research project, one may look at how these piecewise random coefficient models in conjunction with historical weather data can give an idea of potential load extremes. Also, there may be other parametric models that would be easier to interpret.

A lot of work has been done in the field of forecasting load. As we’ve mentioned at the beginning, Manitoba Hydro focuses a lot of their efforts on this topic. In literature, Mckean and Winger [4] and Engle *et al.* [2] look at the problem through econometric models. Engle *et al.* [3] produce models similar to ours but incorporate influences such as holidays and average loads. Skinner [13] uses a day variable, noon effective temperature variable, and what he calls a “sunset” variable in his model.

As was mentioned in the beginning, we have only touched upon one aspect of the whole Manitoba Hydro problem. The way our models change over time is currently being looked at. We want to see if there is one model that best describes the way the parameters are changing over time. There is the subject of accurate forecasting

and attaching precision to forecasts. Another aspect being looked at is the need to quantify the effects of noise through measures of robustness. Hopefully, the results of this project will give Manitoba Hydro a better understanding of the influence of noise factors on load forecasts.

As you can see there is a lot of research being done in this area. As mentioned before, accurate forecasts are very important for companies like Manitoba Hydro. Many important decisions have to be made based on these forecasts. An error in a forecast could cost a company like Manitoba Hydro a lot of money.

Appendix A

The Data

In our discussions we have used data collected by Manitoba Hydro since April of 1992. Many variables were measured, but Table A.1 shows a subset of the data containing only the variables that were of importance to us. WODATE is of course the date the

WODATE	HOUR	TEMP (°C)	NET (MW)	CURT (MW)	LOAD (MW)	DAYTYPE
17NOV99	18	0.5	2745	0	2745	WED
17NOV99	19	0.5	2687	0	2687	WED
17NOV99	20	0.4	2602	42	2644	WED
17NOV99	21	0	2529	48	2577	WED
17NOV99	22	-0.5	2497	7	2504	WED
17NOV99	23	-1.7	2384	1	2385	WED
17NOV99	24	-1.6	2148	1	2149	WED
18NOV99	1	-2	2038	0	2038	THU
18NOV99	2	-2.4	2009	0	2009	THU
18NOV99	3	-2.7	1967	0	1967	THU

Table A.1: Partial data set collected by Manitoba Hydro

data was taken. HOUR is the hour of the day recorded in standard hour units. TEMP is the average Winnipeg temperature taken over the corresponding hour in degrees celsius. For example in the first row of Table A.1 the average temperature between 6pm and 7pm was 0.5°C. NET is the maximum hourly demand of electricity required to meet the needs of Manitoba customers, measured in megawatts (MW). CURT is

the load that could be curtailed on short notice, also measured in megawatts. LOAD, which is the variable we are most interested in, is NET+CURT, and is measured in megawatts. Finally DAYTYPE is which day of the week the data point comes from.

Appendix B

SAS Code

In this appendix we introduce some of the SAS code used to fit the linear and nonlinear models.

B.1 PROC REG code for linear models

In chapter 2 we fit piecewise linear models to the Manitoba Hydro data using SAS. We first fit a piecewise linear model with one knot. The following code was used to fit this particular one knot model.

```
libname test8 v612 'c:\WINDOWS\desktop\Marc';
```

```
data test8.net6;
```

```
    set test4.std18;
```

```
    load = net + curt;
```

```
    if (temp > 18) then z18=1; else z18=0;
```

```

beta1=temp;
beta2=(temp-18)*z18;

run;

proc reg data=test4.net5;
    model load=beta1 beta2;
    output out=two P=Pload R=Rload;
run;

```

As you can see we have a statement that creates the indicator variable, z_{18} . The PROC REG statement needs a model statement. By using the output statement we created a data set called "two" which has the predicted values and the residuals.

The second type of linear model we fit using SAS was the two knot piecewise model. The following code was used.

```

libname test9 v612 'c:\WINDOWS\Desktop\Marc';

data test9.net7;
set test9.hour18;

if (temp>10) then z10=1; else z10=0;
if (temp>20) then z20=1; else z20=0;

beta1=temp;
beta2=(temp-20)*z20;

if (temp>10) then do;

```

```
beta1=(temp-(temp-10))*z10;
```

```
end;
```

```
run;
```

```
proc reg data=test9.net7;
```

```
model load=beta1 beta2;
```

```
output out=two P=Pload R=Rload;
```

```
run;
```

Again, notice the statements used to create the two indicator variables.

B.2 PROC NLIN code for nonlinear models

Nonlinear piecewise models were fit using SAS, in chapter 3. The following code was used to fit the nonlinear piecewise model with one knot.

```
libname nlin v612 'c:\WINDOWS\Desktop\Marc';
```

```
data nlin.net8;
```

```
set nlin.hour18;
```

```
run;
```

```
proc nlin data=nlin.net8 method=DUD;
```

```
parms b0=2433.8 b1=-32.86 b2=69.27 b3=18;
```

```
if (temp < b3) then do;
```

```
model load = b0 + b1*temp;
```

```
end;
```

```
else do;
```

```

model load = b0 + b1*temp + b2*(temp-b3);
end;
output out=two P=Pload R=Rload;
run;

```

With the PROC NLIN statement we need to indicate with the method option which iterative method we wish to use. The parms statement lists the parameters and assigns them starting values. Also notice the “if” and “else” statements that are needed because of the knot in the model.

The second nonlinear model that we fit was the two knot piecewise model. The following SAS code was used in that case.

```

libname nlin4 v612 'c:\WINDOWS\Desktop\Marc';

data nlin4.net20;
set nlin4.hour18;
run;

proc nlin data=nlin4.net20 method=DUD;
parms b0=2372.60 b1=-37.99 b2=24.32 b3=10 b4=20;
if (temp < b3) then do;
model load = b0 + b1*temp;
end;
if ((temp>b3) and (temp<b4)) then do;
model load = b0 + b1*temp - b1*(temp-b3);
end;
if (temp>b4) then do;
model load = b0 + b1*temp - b1*(temp-b3) + b2*(temp-b4);
end;

```

```
output out=two P=Pload R=Rload;  
run;
```

Again notice the parms statements along with their starting values. Also notice the conditional statements used to incorporate the two knots in the model.

For both the PROC REG and PROC NLIN procedures there are many options that can be used. See the SAS user's guide [11] for details on options available.

Bibliography

- [1] Bates, D. M. and D. G. Watts (1988). *Nonlinear Regression Analysis and Its Applications*. New York: John Wiley & Sons.
- [2] Engle, R. F.; C. W. J. Granger; J. Rice; and A. Weiss (1986). Semiparametric Estimates of the Relation Between Weather and Electricity Sales. *Journal of the American Statistical Association*, 81, 310-320.
- [3] Engle, R. F.; C. Mustafa; and J. Rice (1992). Modelling Peak Electricity Demand. *Journal of Forecasting*, 11, 241-251.
- [4] McKean, J. R.; and W. D. Winger (1992). Simultaneous Equation Estimates of Electricity Demand for the Rural South: Revenue Projection when Prices are Administered. *Journal of Forecasting*, 11, 225-240.
- [5] Myers, R. H. (1990). *Classical and Modern Regression with Applications*. Boston: Duxbury.
- [6] Myers, R. H. and J. S. Milton (1991). *A First Course in the Theory of Linear and Statistical Models*. Boston: John Wiley & Sons.
- [7] Neter, J.; M. H. Kutner; C. J. Nachtsheim; and W. Wasserman (1996). *Applied Linear Statistical Models*. Chicago: Irwin.
- [8] Ralston, M. L. and R. I. Jenrich (1978). DUD, A Derivative-Free Algorithm for Nonlinear Least Squares. *Technometrics*, 20, 7-14.

- [9] Rawlings, J. O. (1988). *Applied Regression Analysis: A Research Tool*. Pacific Grove, California: Wadsworth & Brooks.
- [10] Rhys, J. M. W. (1984). Techniques for Forecasting Electricity Demand. *Statistician*, 33, 23-33.
- [11] SAS Institute Inc. (1990). *SAS User's Guide: Statistics Version 6*. Cary, North Carolina: SAS Institute Inc.
- [12] Seber, G.A.F. and C.J. Wild (1989). *Nonlinear Regression*. New York: John Wiley & Sons.
- [13] Skinner, N. H. (1984). Load Research and Its Application to Electricity Demand Forecasting. *Statistician*, 33, 65-73.