Part III

Nonlinear Regression
Chapter 13

Introduction to Nonlinear Regression

We look at nonlinear regression models.

13.1 Linear and Nonlinear Regression Models

We compare linear regression models,

\[ Y_i = f(X_i, \beta) + \epsilon_i \]

\[ = X_i^T \beta + \epsilon_i \]

\[ = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i \]

with nonlinear regression models,

\[ Y_i = f(X_i, \gamma) + \epsilon_i \]

where \( f \) is a nonlinear function of the parameters \( \gamma \) and

\[ X_i = \begin{bmatrix} X_{i1} \\ \vdots \\ X_{i,q} \end{bmatrix}, \quad \gamma = \begin{bmatrix} \gamma_0 \\ \vdots \\ \gamma_{p-1} \end{bmatrix} \]

In both the linear and nonlinear cases, the error terms \( \epsilon_i \) are often (but not always) independent normal random variables with constant variance. The expected value in the linear case is

\[ E\{Y\} = f(X_i, \beta) = X_i^T \beta \]

and in the nonlinear case,

\[ E\{Y\} = f(X_i, \gamma) \]
Exercise 13.1 (Linear and Nonlinear Regression Models) Identify whether the following regression models are linear, intrinsically linear (nonlinear, but transformed easily into linear) or nonlinear.

1. \( Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \)
   This regression model is **linear / intrinsically linear / nonlinear**

2. \( Y_i = \beta_0 + \beta_1 \sqrt{X_{i1}} + \varepsilon_i \)
   This regression model is **linear / intrinsically linear / nonlinear** because
   \[
   Y_i = \beta_0 + \beta_1 \sqrt{X_{i1}} + \varepsilon_i = \beta_0 + \beta_1 X'_{i1} + \varepsilon_i
   \]

3. \( \ln Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \)
   This regression model is **linear / intrinsically linear / nonlinear** because
   \[
   \ln Y_i = \beta_0 + \beta_1 \sqrt{X_{i1}} + \varepsilon_i = \ln Y'_i = \ln \left( \beta_0 + \beta_1 X'_{i1} + \varepsilon_i \right)
   \]
   where \( Y'_i = \ln Y_i \)

4. \( Y_i = \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \varepsilon_i \)
   This regression model\(^1\) is **linear / intrinsically linear / nonlinear**

5. \( f(X_{i1}; \beta) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \)
   This regression model is **linear / intrinsically linear / nonlinear**

6. \( f(X_{i1}; \gamma) = \gamma_0 [\exp(\gamma_1 X_i)] \)
   This regression model is **linear / intrinsically linear / nonlinear** because
   \[
   f(X_{i1}; \gamma) = \gamma_0 [\exp(\gamma_1 X_i)] = \ln f(X_{i1}; \gamma) = \ln \gamma_0 [\exp(\gamma_1 X_i)] = \ln \gamma_0 + \ln [\exp(\gamma_1 X_i)] = Y'_i = \ln \gamma_0 + \gamma_1 X_i
   \]
   where \( Y'_i = \ln Y_i \) and \( Y'_0 = \ln \gamma_0 \)

\(^1\)Traditionally, the “\( \gamma \)” parameters are reserved for nonlinear regression models and the “\( \beta \)” parameters are reserved for linear regression models. But, of course, traditions are made to be broken.
7. \( f(X_i, \gamma) = \gamma_0 + \frac{\gamma_1}{\gamma_2 \gamma_3} X_i \)
   This regression model is **linear** / intrinsically linear / **nonlinear** even though
   \[
   f(X_i, \gamma) = \gamma_0 + \frac{\gamma_1}{\gamma_2 \gamma_3} X_i \\
   Y_i = \gamma_0 + \gamma'_1 X_i
   \]
   where \( \gamma'_1 = \frac{\gamma_1}{\gamma_2 \gamma_3} \) where three parameters have been condensed into one.

### 13.2 Example

An interesting example which uses the nonlinear exponential regression model to describe hospital data is given in this section.

### 13.3 Least Squares Estimation in Nonlinear Regression

The least squares estimation of the **linear** regression model,

\[
Y = X \beta + \varepsilon
\]

and involves minimizing the least squares criterion\(^2\)

\[
Q = \sum_{i=1}^{n} [Y_i - f(X_i, \beta)]^2
\]

with respect to the linear regression parameters \( \beta_0, \beta_1, \ldots, \beta_{p-1} \), and so gives the following (analytically-derived) estimators,

\[
b = (X'X)^{-1}X'Y
\]

\(^2\)Another estimation method is the maximum likelihood method which involves minimizing the likelihood of the distribution function of the linear regression model with respect to the parameters. If the error terms, \( \varepsilon \), are independent identically normal, then the least squares method and MLE methods give the same estimators.
In a similar way, the least squares estimation of the nonlinear regression model,

\[ Y_i = f(X_i, \gamma) + \varepsilon_i \]

involves minimizing the least squares criterion

\[ Q = \sum_{i=1}^{n} [Y_i - f(X_i, \gamma)]^2 \]

with respect to the nonlinear regression parameters, \( \gamma_0, \gamma_1, \ldots, \gamma_p \) but, often, it is not possible to analytically derive the estimators. A numerical method\(^3\) is often required. We will use the numerical Gauss–Newton method given in SAS to derive least squares estimators for various nonlinear regression models.

**Exercise 13.2 (Least Squares Estimation in Nonlinear Regression)**

<table>
<thead>
<tr>
<th>illumination, ( X )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ability to read, ( Y )</td>
<td>70</td>
<td>70</td>
<td>75</td>
<td>88</td>
<td>91</td>
<td>94</td>
<td>100</td>
<td>92</td>
<td>90</td>
<td>85</td>
</tr>
</tbody>
</table>

Fit two different nonlinear (and intrinsically linear) regressions to this data. Since there is, typically, no analytical solution to nonlinear regressions, a numerical procedure, called the Gauss–Newton method is used to derive each of the nonlinear regressions below.

1. **Simple Bounded Model**, \( Y_i = \gamma_2 + (\gamma_0 - \gamma_2) \exp(-\gamma_1 X_i) + \varepsilon_i \)
   
   (a) **Initial Estimates, Part 1**
   
   Assume the simple bounded model is initially given by
   
   \[ Y_i = \gamma_2(1-\exp(-\gamma_1 X_i))+\gamma_0 \exp(-\gamma_1 X_i) = 100(1-\exp(-X_i)) + 50 \exp(-X_i) \]
   
   In other words, the \( \gamma_0, \gamma_1 \) and \( \gamma_2 \) parameters are initially estimated by (choose one)
   
   i. \( g_0^{(0)} = 100, \quad g_1^{(0)} = -1, \quad g_2^{(0)} = 1 \)
   
   ii. \( g_0^{(0)} = 50, \quad g_1^{(0)} = 1, \quad g_2^{(0)} = 100 \)
   
   iii. \( g_0^{(0)} = 50, \quad g_1^{(0)} = -1, \quad g_2^{(0)} = 100 \)

   (b) **Initial Estimates, Part 2**
   
   A sketch of the simple bounded function where \( g_0^{(0)} = 50, \quad g_1^{(0)} = 1, \quad g_2^{(0)} = 100 \)
   
   reveals that the upper bound of this function can be found at (choose one) \( 1 / 50 / 100 \)

\(^{3}\)The text not only shows how it is not possible to analytically derive the estimators for the exponential regression model, but also gives a detailed step-by-step explanation of how to derive the estimators for this model using the Gauss–Newton numerical method.
In other words, the upper bound of the simple bounded function, parameter $\gamma_0$ which is estimated by $g_0^{(0)}$, is initially set at 100 because we know the largest observed response is at $Y = 100$ (when $X = 7$). In a similar way, the other two parameters, the intercept ($\gamma_0$) and rate of change ($\gamma_1$) of the function, are initially set to 50 and 1, respectively.

(c) **Nonlinear Least Squares Estimated Regression**

Using the starting values $g_0^{(0)} = 50$, $g_1^{(0)} = 1$, $g_2^{(0)} = 100$

SAS gives us

$$g_0 = 52.2624, \quad g_1 = 0.4035, \quad g_2 = 93.6834$$

In other words, the best fitting nonlinear exponential model,

$$Y_i = \gamma_2(1 - \exp(-\gamma_1 X_i)) + \gamma_0 \exp(-\gamma_1 X_i)$$

is given by (circle one)

\[ Y_1 = \frac{93.2731}{1 + 0.0994 \exp(-0.5117X_i)} \]

\[ Y_1 = 93.6834(1 - \exp(-0.4035X_i)) + 52.2624 \exp(-0.4035X_i) \]

\[ Y_1 = 84.3854 + 0.0509 \exp(0.119X_i) \]

(d) **Various Residual Plots**

i. **True / False**

The nonlinear regression plot shows that the regression line is a fairly good fit to the data.

ii. **True / False**

The residual versus predicted plot indicates non–homogeneous variance in the error.

---

4Use -10, 10, 1, 0, 150, 10, 1 in your calculator.
iii. True / False
The residual versus predictor, illumination, indicates non-homogeneous variance in the error.

iv. True / False
The normal probability plot of residuals is fairly linear which indicates the error is normal.

On the basis of these residual plots, it seems the simple bounded model is not a great fitting model, but certainly, will be shown to be a better fitting model to the data than the (to-be-analyzed) exponential model.

2. Logistic Model, \( Y_i = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)} + \varepsilon_i \)

(a) Initial Estimates, Part 1
Assume the logistic model is initially given by

\[
Y_i = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)} = \frac{100}{1 + \exp(-X_i)}
\]

In other words, the \( \gamma_0, \gamma_1 \) and \( \gamma_2 \) parameters are initially estimated by (choose one)

i. \( g_0^{(0)} = 100, \ g_1^{(0)} = -1, \ g_2^{(0)} = 1 \)

ii. \( g_0^{(0)} = 100, \ g_1^{(0)} = 2, \ g_2^{(0)} = 1 \)

iii. \( g_0^{(0)} = 100, \ g_1^{(0)} = 1, \ g_2^{(0)} = 1 \)

(b) Initial Estimates, Part 2
A sketch of the logistic function where \( g_0^{(0)} = 100, \ g_1^{(0)} = 1, \ g_2^{(0)} = 1 \) reveals that the upper bound of this function can be found at (choose one) 1 / 50 / 100

![Figure 13.2 (Nonlinear Regression, Logistic Model)](image-url)
Section 3. Least Squares Estimation in Nonlinear Regression

In other words, the upper bound of the logistic function, parameter $\gamma_0$ which is estimated by $g_0^{(0)}$, is initially set at 100 because we know the largest observed response is at $Y = 100$ (when $X = 7$). In a similar way, the other two parameters, which control direction and rate of increase or decrease of the function, are initially set to one (1) each.

(c) Nonlinear Least Squares Estimated Regression

Using the starting values $g_0^{(0)} = 100$, $g_1^{(0)} = 1$, $g_2^{(0)} = 1$

SAS gives us $g_0 = 93.2731$, $g_1 = 0.6994$, $g_2 = -0.5117$

In other words, the best fitting nonlinear exponential model,

$$ Y_i = \frac{\gamma_0}{1 + \gamma_1 \exp(\gamma_2 X_i)} $$

is given by (circle one)

$Y_1 = 93.2731 \frac{93.2731}{1 + 0.6994 \exp(-0.5117 X_i)}$

$Y_1 = 73.677 \exp(0.0266 X_i)$

$Y_1 = 84.3854 + 0.0509 \exp(0.119 X_i)$

(d) Various Residual Plots

i. True / False
   The nonlinear regression plot shows that the regression line is a fairly good fit to the data.

ii. True / False
   The residual versus predicted plot indicates non-homogeneous variance in the error.

iii. True / False
   The residual versus predictor, illumination, indicates non-homogeneous variance in the error.

iv. True / False
   The normal probability plot of residuals is fairly linear which indicates the error is normal.

On the basis of these residual plots, it seems the logistic model is about as good fitting as the simple bounded model to the data.
13.4 Model Building and Diagnostics

It is often not easy to add or delete predictor variables to nonlinear regression models, in other words, to model build. However, it is possible to perform diagnostic tests to check for correlation or nonconstant error variance and to check for a lack of fit.

Large-sample theory is applicable if the nonlinear regression is “linear enough” at each point of the regression. This is true if the iterative procedure for the estimation of the nonlinear regression is quick, if various measures indicate it to be true or if a bootstrap procedure indicates it to be true.

Exercise 13.3 (Model Building and Diagnostics)

<table>
<thead>
<tr>
<th>illumination, $X$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ability to read, $Y$</td>
<td>70</td>
<td>70</td>
<td>75</td>
<td>88</td>
<td>91</td>
<td>94</td>
<td>100</td>
<td>92</td>
<td>90</td>
<td>92</td>
<td>85</td>
</tr>
</tbody>
</table>

According to the previous analysis, the logistic regression

$$Y_i = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)} + \varepsilon_i$$

is the best of the nonlinear models that were considered. Consequently, conduct a lack of fit test\(^5\) for this nonlinear regression.

1. **Logistic Model**, $Y_i = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)} + \varepsilon_i$

   (a) **Statement.**

   The statement of the test is (check none, one or more):

   i. $H_0 : E\{Y\} = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)}$ versus $H_1 : E\{Y\} > \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)}$.

   ii. $H_0 : E\{Y\} = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)}$ versus $H_1 : E\{Y\} < \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)}$.

   iii. $H_0 : E\{Y\} = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)}$ versus $H_1 : E\{Y\} \neq \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)}$.

---

\(^5\)Notice how an additional data point, $(9, 92)$, has been added to the data and so there are now two points at $X = 9$. This is necessary to conduct a lack of fit test.
(b) Test.
The test statistic\(^6\) is
\[
F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\
= \frac{SSE - SSPE}{(n - 2) - (n - c)} \div \frac{SSPE}{n - c} \\
= \frac{SSLF}{SSPE} \div \frac{c - 2}{n - c} \\
= \frac{243.2 - 2}{7} \div \frac{2}{1}
\]
(circle one) 9.075 / 17.23 / 58.57.
The critical value at \( \alpha = 0.01 \), with 7 and 1 degrees of freedom, is
(circle one) 4.83 / 5.20 / 59.28
(Use PRGM INVF ENTER 7 ENTER 1 ENTER 0.99 ENTER)

(c) Conclusion.
Since the test statistic, 17.23, is smaller than the critical value, 59.28, we
(circle one) accept / reject the null hypothesis that the regression function is
\( E(Y) = \frac{\gamma}{1 + \gamma_1 \exp(-\gamma_2 X_i)} \).

13.5 Inferences about Nonlinear Regression Parameters

SAS program: att12-13-5-read-nonlin-CI,test

For linear regression models with normal error terms, the least squares or minimum
likelihood estimators, for any given sample size, are normally distributed, are unbiased
and have minimum variance. Although this is not true for nonlinear regression models
for any sample size, it is approximately true for such models with large sample size.

Specifically, if \( \varepsilon_i \) are independent \( N(0, \sigma^2) \) in the nonlinear regression model
\[ Y_i = f(X_i, g) + \varepsilon_i; \]
for large \( n \),
\[
E\{g\} = \gamma \\
s^2\{g\} = MSE(D' D)^{-1}
\]
\(^6\)From SAS,
\( SSE = 243.2 \) from the "reduced" nonlinear model, \( Y_i = \frac{\gamma}{1 + \gamma_1 \exp(-\gamma_2 X_i)} + \varepsilon_i \),
and \( SSPE = 2 \) from the "full" ANOVA model, \( Y_{ij} = \mu_j + \varepsilon_{ij} \).
where $D$ is the matrix of partial derivatives with respect to the parameters $\gamma$ evaluated at $g(0)$.

Large-sample theory is applicable if the nonlinear regression is “linear enough” at each point of the regression. This is true if the iterative procedure for the estimation of the nonlinear regression is quick, if various measures indicate it to be true or if a bootstrap procedure indicates it to be true.

**Exercise 13.4 (Inferences about Nonlinear Regression Parameters)**

<table>
<thead>
<tr>
<th>illumination, $X$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ability to read, $Y$</td>
<td>70</td>
<td>70</td>
<td>75</td>
<td>88</td>
<td>91</td>
<td>94</td>
<td>100</td>
<td>92</td>
<td>90</td>
<td>85</td>
</tr>
</tbody>
</table>

According to the previous analysis, the logistic regression

$$Y_i = \frac{\gamma_0}{1 + \gamma_1 \exp(-\gamma_2 X_i)} + \varepsilon_i$$

is the best of the nonlinear models that were considered. Consequently, determine various intervals and conduct tests for this nonlinear regression. Assume large sample results hold in this case.

att12-13-4-read-nonlin-CI,test

1. **90% Confidence Interval**
   From SAS, the 90% confidence interval for $\gamma_0$ is given by
   $$g_0 \pm t(1 - \alpha/2; n - p)s\{g_0\} = 93.2731 \pm t(1 - 0.10/2; 10 - 3)4.0022 = 93.2731 \pm 1.8946(4.0022) =$$

   (circle one) (85.7, 100.9) / (88.7, 98.9) / (90.7, 94.9).

2. **90% Bonferroni Simultaneous Confidence Interval**
   The 90% Bonferroni simultaneous interval for $\gamma_0$ (knowing that it is one of three ($m = 3$) parameters, $\gamma_0$, $\gamma_1$ and $\gamma_2$) is given by
   $$g_0 \pm t(1 - \alpha/2m; n - p)s\{g_0\} = 93.2731 \pm t(1 - 0.10/2(3); 10 - 3)4.0022 = 93.2731 \pm 2.642(4.0022) =$$

   (circle one) (85.7, 100.9) / (82.7, 103.9) / (90.7, 94.9).

3. **Test of Single $\gamma_k$**
   Test if $\gamma_0 \neq 93$ at $\alpha = 0.05$. 

att12-13-4-read-nonlin-CI,test
(a) **Statement.**

The statement of the test, in this case, is (circle one)

i. $H_0 : \gamma_0 = 93$ versus $H_a : \gamma_0 < 93$

ii. $H_0 : \gamma_0 = 93$ versus $H_a : \gamma_0 > 93$

iii. $H_0 : \gamma_0 = 93$ versus $H_a : \gamma_0 \neq 93$

(b) **Test.**

The standardized test statistic of $g_1 = 0.0266$ is

$$t \text{ test statistic} = \frac{g_0 - \gamma_{10}}{s\{g_0\}} = \frac{93.2731 - 93}{4.0022} =$$

(circle one) **0.042 / 0.068 / 0.152.**

The standardized critical value at $\alpha = 0.05$ is

(circle one) **-2.31 / -1.76 / 2.32**

(Use PRGM ENTER INVT ENTER 8 ENTER 0.025 ENTER)

(c) **Conclusion.**

Since the test statistic, 0.068, is smaller than the critical value, 2.31, we (circle one) **accept / reject** the null guess of $\gamma_0 = 93$.

### 13.6 Learning Curve Example