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,

$$\begin{aligned} Y_i &= f(\mathbf{X}_i, \beta) + \varepsilon_i \\ &= \mathbf{X}_i'\beta + \varepsilon_i \\ &= \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \end{aligned}$$

with nonlinear regression models,

$$Y_i = f(\mathbf{X}_i, \gamma) + \varepsilon_i$$

where f is a nonlinear function of the parameters γ and

$$\underset{\mathbf{q\times 1}}{\mathbf{X_i}} = \begin{bmatrix} X_{i1} \\ \vdots \\ X_{iq} \end{bmatrix} \quad \underset{\mathbf{p\times 1}}{\gamma} = \begin{bmatrix} \gamma_0 \\ \vdots \\ \gamma_{p-1} \end{bmatrix}$$

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

$$E\{Y\} = f(\mathbf{X}_i, \beta) = \mathbf{X}_i'\beta$$

and in the nonlinear case,

$$E\{Y\} = f(\mathbf{X}_i, \gamma)$$

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

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 \sqrt{X_{i1}} + \varepsilon_i \\ &= \beta_0 + \beta_1 X'_{i1} + \varepsilon_i \end{aligned}$$

3. $\ln Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$
   This regression model is **linear** / **intrinsically linear** / **nonlinear** because

$$\begin{aligned} \ln Y_i &= \beta_0 + \beta_1 \sqrt{X_{i1}} + \varepsilon_i \\ Y'_i &= \beta_0 + \beta_1 X'_{i1} + \varepsilon_i \end{aligned}$$

   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. $f(\mathbf{X}_i, \beta) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$
   This regression model is **linear** / **intrinsically linear** / **nonlinear**

6. $f(\mathbf{X}_i, \gamma) = \gamma_0[\exp(\gamma_1 X)]$
   This regression model is **linear** / **intrinsically linear** / **nonlinear** because

$$\begin{aligned} f(\mathbf{X}_i, \gamma) &= \gamma_0[\exp(\gamma_1 X_i)] \\ \ln f(\mathbf{X}_i, \gamma) &= \ln[\gamma_0[\exp(\gamma_1 X_i)]] \\ Y'_i &= \ln \gamma_0 + \ln[\exp(\gamma_1 X_i)] \\ Y'_i &= \gamma'_0 + \gamma_1 X_i \end{aligned}$$

   where $Y'_i = \ln Y_i$ and $\gamma'_0 = \ln \gamma_0$

[1] Traditionally, the "γ_i" parameters are reserved for *non*linear regression models and the "β_i" parameters are reserved for linear regression models. But, of course, traditions are made to be broken.

7. $f(\mathbf{X}_i, \gamma) = \gamma_0 + \frac{\gamma_1}{\gamma_2\gamma_3} X_i$
   This regression model is **linear / intrinsically linear / nonlinear** even though

$$\begin{aligned} f(\mathbf{X}_i, \gamma) &= \gamma_0 + \frac{\gamma_1}{\gamma_2\gamma_3} X_i \\ Y_i &= \gamma_0 + \gamma_1' X_i \end{aligned}$$

   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

SAS programs:
att12-13-3-read-bounded, att12-13-3-read-logistic,

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

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

and involves minimizing the least squares criterion[2]

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

with respect to the linear regression parameters β_0, β_1, $\ldots$, β_{p-1}, and so gives the following (analytically–derived) estimators,

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

[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, ε, 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 *non*linear regression model,

$$Y_i = f(\mathbf{X}_i, \gamma) + \varepsilon_i$$

involves minimizing the least squares criterion

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

with respect to the nonlinear regression parameters, $\gamma_0, \gamma_1, \ldots, \gamma_{p-1}$ 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)

illumination, X	1	2	3	4	5	6	7	8	9	10
ability to read, Y	70	70	75	88	91	94	100	92	90	85

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

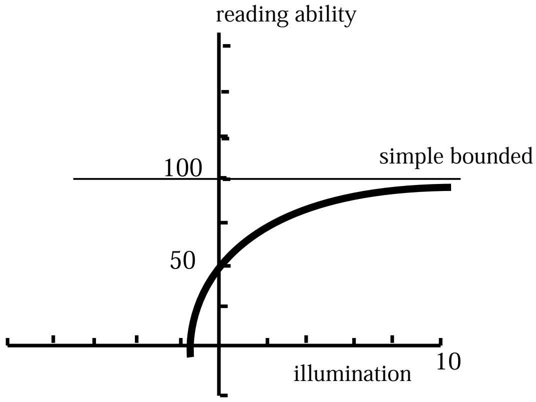
   $$Y_i = \gamma_2(1-\exp(-\gamma_1 X_i)) + \gamma_0 \exp(-\gamma_1 X_i) = 100(1-\exp(-X_i)) + 50\exp(-X_i)$$

   In other words, the γ_0, γ_1 and γ_2 parameters are initially estimated by (choose one)

   i. $g_0^{(0)} = 100$, $g_1^{(0)} = -1$, $g_2^{(0)} = 1$
   ii. $g_0^{(0)} = 50$, $g_1^{(0)} = 1$, $g_2^{(0)} = 100$
   iii. $g_0^{(0)} = 50$, $g_1^{(0)} = -1$, $g_2^{(0)} = 100$

   (b) *Initial Estimates, Part 2*
   A sketch of the simple bounded function where
   $g_0^{(0)} = 50$, $g_1^{(0)} = 1$, $g_2^{(0)} = 100$
   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.



Figure[4] 13.1 (Nonlinear Regression, Simple Bounded Model)

In other words, the upper bound of the simple bounded function, parameter γ_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 (γ_0) and rate of change (γ_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)
$\mathbf{Y_i} = \frac{\mathbf{93.2731}}{\mathbf{1+0.6994}\exp(-\mathbf{0.5117X_i})}$
$\mathbf{Y_i} = \mathbf{93.6834(1} - \exp(-\mathbf{0.4035X_i})) + \mathbf{52.2624}\exp(-\mathbf{0.4035X_i})$
$\mathbf{Y_i} = \mathbf{84.3854} + \mathbf{0.0509}\exp(\mathbf{0.119X_i})$

(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.

[4]Use -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 γ_0, γ_1 and γ_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**

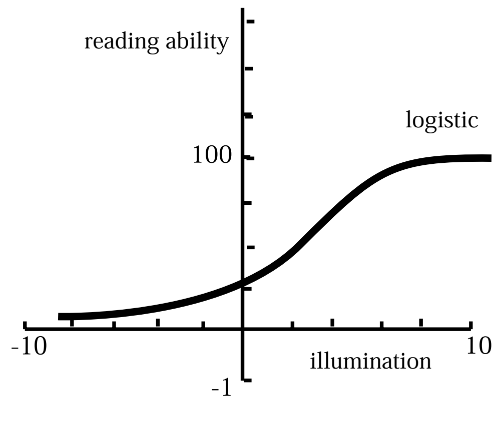


Figure 13.2 (Nonlinear Regression, Logistic Model)

In other words, the upper bound of the logistic function, parameter γ_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

$$g_0 = 93.2731, \quad g_1 = 0.6994, \quad g_2 = -0.5117$$

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)
$\mathbf{Y_i} = \frac{\mathbf{93.2731}}{\mathbf{1+0.6994}\exp(-\mathbf{0.5117X_i})}$
$\mathbf{Y_i} = \mathbf{73.677}\exp(\mathbf{0.0266X_i})$
$\mathbf{Y_i} = \mathbf{84.3854} + \mathbf{0.0509}\exp(\mathbf{0.119X_i})$

(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.

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

SAS program: att12-13-4-read-nonlin-lof

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)

illumination, X	1	2	3	4	5	6	7	8	9	9	10
ability to read, Y	70	70	75	88	91	94	100	92	90	92	85

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.

att12-13-4-read-nonlin-lof

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

$$\begin{aligned} F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\ &= \frac{SSE - SSPE}{(n-2)-(n-c)} \div \frac{SSPE}{n-c} \\ &= \frac{SSLF}{c-2} \div \frac{SSPE}{n-c} \\ &= \frac{243.2-2}{7} \div \frac{2}{1} \\ &= \end{aligned}$$

(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** / **5928**
(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, 5928, we (circle one) **accept** / **reject** the null hypothesis that the regression function is $E\{Y\} = \frac{\gamma_0}{1+\gamma_1 \exp(-\gamma_2 X_i)}$.

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 *non*linear regression models for *any* sample size, it is approximately true for such models with *large* sample size.

Specifically, if ε_i are independent $N(0, \sigma^2)$ in the nonlinear regression model

$$Y_i = f(\mathbf{X}_i, \mathbf{g}) + \varepsilon_i;$$

for large n,

$$\begin{aligned} \mathbf{E}\{\mathbf{g}\} &= \gamma \\ \mathbf{s^2}\{\mathbf{g}\} &= MSE(\mathbf{D}'\mathbf{D})^{-1} \end{aligned}$$

[6]From SAS,
$SSE = 243.2$ from the "reduced" nonlinear model, $Y_i = \frac{\gamma_0}{1+\gamma_1 \exp(-\gamma_2 X_i)} + \varepsilon_j$,
and $SSPE = 2$ from the 'full" ANOVA model, $Y_{ij} = \mu_j + \varepsilon_{ij}$.

where $\mathbf{D}$ is the matrix of partial derivatives with respect to the parameters γ evaluated at $\mathbf{g}(0)$.

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)

illumination, X	1	2	3	4	5	6	7	8	9	10
ability to read, Y	70	70	75	88	91	94	100	92	90	85

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 the 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 γ_0 is given by

$$\begin{aligned} g_0 \pm t(1-\alpha/2; n-p)s\{g_0\} &= 93.2731 \pm t(1-0.10/2; 10-3)4.0022 \\ &= 93.2731 \pm 1.8946(4.0022) = \end{aligned}$$

   (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 γ_0 (knowing that it is one of three ($m = 3$) parameters, γ_0, γ_1 and γ_2) is given by

$$\begin{aligned} g_0 \pm t(1-\alpha/2m; n-p)s\{g_0\} &= 93.2731 \pm t(1-0.10/2(3); 10-3)4.0022 \\ &= 93.2731 \pm 2.642(4.0022) = \end{aligned}$$

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

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

(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

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

(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