14

NONLINEAR REGRESSION MODELS

The major emphasis of this book is on linear regression models, that is, models that are linear in the parameters and/or models that can be transformed so that they are linear in the parameters. On occasions, however, for theoretical or empirical reasons we have to consider models that are nonlinear in the parameters.[1] In this chapter we take a look at such models and study their special features.

14.1 INTRINSICALLY LINEAR AND INTRINSICALLY NONLINEAR REGRESSION MODELS

When we started our discussion of linear regression models in Chapter 2, we stated that our concern in this book is basically with models that are linear in the parameters; they may or may not be linear in the variables. If you refer to Table 2.3, you will see that a model that is linear in the parameters as well as the variables is a linear regression model and so is a model that is linear in the parameters but nonlinear in the variables. On the other hand, if a model is nonlinear in the parameters it is a nonlinear (in-the-parameter) regression model whether the variables of such a model are linear or not.

[1]We noted in Chap. 4 that under the assumption of normally distributed error term, the OLS estimators are not only BLUE but are BUE (best unbiased estimator) in the entire class of estimators, linear or not. But if we drop the assumption of normality, as Davidson and MacKinnon note, it is possible to obtain nonlinear and/or biased estimators that may perform better than the OLS estimators. See Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics,* Oxford University Press, New York, 1993, p. 161.

However, one has to be careful here, for some models may look nonlinear in the parameters but are **inherently** or **intrinsically** linear because with suitable transformation they can be made linear-in-the-parameter regression models. But if such models cannot be linearized in the parameters, they are called **intrinsically nonlinear regression models.** *From now on when we talk about a nonlinear regression model, we mean that it is intrinsically nonlinear.* For brevity, we will call them **NLRM.**

To drive home the distinction between the two, let us revisit exercises 2.6 and 2.7. In exercise 2.6, Models **a, b, c,** and **e** are linear regression models because they are all linear in the parameters. Model **d** is a mixed bag, for β_2 is linear but not $\ln \beta_1$. But if we let $\alpha = \ln \beta_1$, then this model is linear in α and β_2.

In exercise 2.7, Models **d** and **e** are intrinsically nonlinear because there is no simple way to linearize them. Model **c** is obviously a linear regression model. What about Models **a** and **b**? Taking the logarithms on both sides of **a,** we obtain $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$, which is linear in the parameters. Hence Model **a** is *intrinsically* a linear regression model. Model **b** is an example of the **logistic (probability) distribution function,** and we will study this in Chapter 15. On the surface, it seems that this is a nonlinear regression model. But a simple mathematical trick will render it a linear regression model, namely,

$$\ln\left(\frac{1 - Y_i}{Y_i}\right) = \beta_1 + \beta_2 X_i + u_i \tag{14.1.1}$$

Therefore, Model **b** is intrinsically linear. We will see the utility of models like (14.1.1) in the next chapter.

Consider now the famous **Cobb–Douglas (C–D) production function.** Letting Y = output, X_2 = labor input, and X_3 = capital input, we will write this function in three different ways:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \tag{14.1.2}$$

or,

$$\ln Y_i = \alpha + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \tag{14.1.2a}$$

where $\alpha = \ln \beta_1$. Thus in this format the C–D function is intrinsically linear.

Now consider this version of the C–D function:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} u_i \tag{14.1.3}$$

or,

$$\ln Y_i = \alpha + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \ln u_i \tag{14.1.3a}$$

where $\alpha = \ln \beta_1$. This model too is linear in the parameters.

But now consider the following version of the C–D function:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} + u_i \tag{14.1.4}$$

As we just noted, C–D versions (14.1.2*a*) and (14.1.3*a*) are intrinsically linear (in the parameter) regression models, but there is no way to transform (14.1.4) so that the transformed model can be made linear in the parameters.[2] Therefore, (14.1.4) is intrinsically a nonlinear regression model.

Another well-known but intrinsically nonlinear function is the **constant elasticity of substitution (CES)** production function of which the Cobb–Douglas production is a special case. The CES production takes the following form:

$$Y_i = A\left[\delta K_i^{-\beta} + (1-\delta)L_i^{-\beta}\right]^{-1/\beta} \tag{14.1.5}$$

where Y = output, K = capital input, L = labor input, A = scale parameter, δ = distribution parameter $(0 < \delta < 1)$, and β = substitution parameter $(\beta \geq -1)$.[3] No matter in what form you enter the stochastic error term u_i in this production function, there is no way to make it a linear (in parameter) regression model. It is intrinsically a nonlinear regression model.

14.2 ESTIMATION OF LINEAR AND NONLINEAR REGRESSION MODELS

To see the difference in estimating linear and nonlinear regression models, consider the following two models:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{14.2.1}$$

$$Y_i = \beta_1 e^{\beta_2 X_i} + u_i \tag{14.2.2}$$

By now you know that (14.2.1) is a linear regression model, whereas (14.2.2) is a nonlinear regression model. Regression (14.2.2) is known as the **exponential regression model** and is often used to measure the growth of a variable, such as population, GDP, or money supply.

[2] If you try to log-transform the model, it will not work because $\ln(A + B) \neq \ln A + \ln B$.

[3] For properties of the CES production function, see Michael D. Intriligator, Ronald Bodkin, and Cheng Hsiao, *Econometric Models, Techniques, and Applications,* 2d ed., Prentice Hall, 1996, pp. 294–295.

Suppose we consider estimating the parameters of the two models by OLS. In OLS we minimize the residual sum of squares (RSS), which for model (14.2.1) is:

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{14.2.3}$$

where as usual $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimators of the true β's. Differentiating the preceding expression with respect to the two unknowns, we obtain the **normal equations** shown in (3.1.4) and (3.1.5). Solving these equations simultaneously, we obtain the OLS estimators given in Eqs. (3.1.6) and (3.1.7). Observe very carefully that in these equations the unknowns (β's) are on the left-hand side and the knowns (X and Y) are on the right-hand side. As a result we get explicit solutions of the two unknowns in terms of our data.

Now see what happens if we try to minimize the RSS of (14.2.2). As shown in Appendix 14A, Section 14A.1, the normal equations corresponding to (3.1.4) and (3.1.5) are as follows:

$$\sum Y_i e^{\hat{\beta}_2 X_i} = \beta_1 e^{2\hat{\beta}_2 X_i} \tag{14.2.4}$$

$$\sum Y_i X_i e^{\hat{\beta}_2 X_i} = \hat{\beta}_1 \sum X_i e^{2\hat{\beta}_2 X_i} \tag{14.2.5}$$

Unlike the normal equations in the case of the linear regression model, the normal equations for nonlinear regression have the unknowns (the $\hat{\beta}$'s) both on the left- and right-hand sides of the equations. As a consequence, we *cannot obtain explicit solutions* of the unknowns in terms of the known quantities. To put it differently, the unknowns are expressed in terms of themselves and the data! Therefore, although we can apply the method of least squares to estimate the parameters of the nonlinear regression models, we cannot obtain explicit solutions of the unknowns. Incidentally, OLS applied to a nonlinear regression model is called **nonlinear least squares (NLLS).** So, what is the solution? We take this question up next.

14.3 ESTIMATING NONLINEAR REGRESSION MODELS: THE TRIAL-AND-ERROR METHOD

To set the stage, let us consider a concrete example. The data in Table 14.1 relates to the management fees that a leading mutual fund in the United States pays to its investment advisors to manage its assets. The fees paid depend on the net asset value of the fund. As you can see, the higher the net asset value of the fund, the lower are the advisory fees, which can be seen clearly from Figure 14.1.

To see how the exponential regression model in (14.2.2) fits the data given in Table 14.1, we can proceed by trial and error. Suppose we assume that

TABLE 14.1 ADVISORY FEES CHARGED AND ASSET SIZE

	Fee, %	Asset*
1	0.520	0.5
2	0.508	5.0
3	0.484	10
4	0.46	15
5	0.4398	20
6	0.4238	25
7	0.4115	30
8	0.402	35
9	0.3944	40
10	0.388	45
11	0.3825	55
12	0.3738	60

*Asset represents net asset value, billions of dollars.

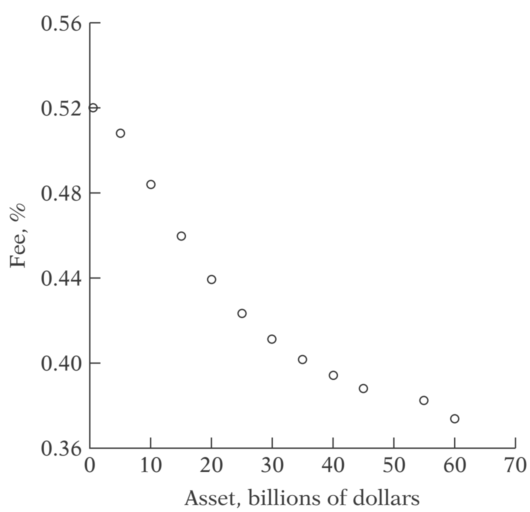

FIGURE 14.1 Relationship of advisory fees to fund assets.

initially $\beta_1 = 0.45$ and $\beta_2 = 0.01$. These are pure guesses, sometimes based on prior experience or prior empirical work or obtained by just fitting a linear regression model even though it may not be appropriate. At this stage do not worry about how these values are obtained.

Since we know the values of β_1 and β_2, we can write (14.2.2) as:

$$u_i = Y_i - \beta_1 e^{\beta_2 X_i} = Y_i - 0.45e^{0.01X_i} \quad \textbf{(14.3.1)}$$

Therefore,

$$\sum u_i^2 = \sum (Y_i - 0.45e^{0.01X_i})^2 \tag{14.3.2}$$

Since Y, X, β_1, and β_2 are known, we can easily find the *error sum of squares* in (14.3.2).[4] Remember that in OLS our objective is to find those values of the unknown parameters that will make the error sum of squares as small as possible. This will happen if the estimated Y values from the model are as close as possible to the actual Y values. With the given values, we obtain $\sum u_i^2 = 0.3044$. But how do we know that this is the least possible error sum of squares that we can obtain? What happens if you choose another value for β_1 and β_2, say, 0.50 and −0.01, respectively? Repeating the procedure just laid down, we find that we now obtain $\sum u_i^2 = 0.0073$. Obviously, this error sum of squares is much smaller than the one obtained before, namely, 0.3044. But how do we know that we have reached the lowest possible error sum of squares, for by choosing yet another set of values for the β's, we will obtain yet another error sum of squares?

As you can see, such a trial-and-error, or **iterative,** process can be easily implemented. And if one has infinite time and infinite patience, the trial-and-error process *may* ultimately produce values of β_1 and β_2 that may guarantee the lowest possible error sum of squares. But you might ask, how did we go from ($\beta_1 = 0.45$; $\beta_2 = 0.01$) to ($\beta_1 = 0.50$; $\beta_2 = -0.1$)? Clearly, we need some kind of *algorithm* that will tell us how we go from one set of values of the unknowns to another set before we stop. Fortunately such algorithms are available, and we discuss them in the next section.

14.4 APPROACHES TO ESTIMATING NONLINEAR REGRESSION MODELS

There are several approaches, or algorithms, to NLRMs: (1) direct search or trial and error, (2) direct optimization, and (3) iterative linearization.[5]

Direct Search or Trial-and-Error or Derivative-Free Method

In the previous section we showed how this method works. Although intuitively appealing because it does not require the use of calculus methods as the other methods do, this method is generally not used. *First,* if an NLRM

[4]Note that we call $\sum u_i^2$ the error sum of squares and not the usual residual sum of squares because the values of the parameters are assumed to be known.

[5]The following discussion leans heavily on these sources: Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts,* 4th ed., McGraw-Hill, 1998, Chap. 10; Norman R. Draper and Harry Smith, *Applied Regression Analysis,* 3d ed., John Wiley & Sons, 1998, Chap. 24; Arthur S. Goldberger, *A Course in Econometrics,* Harvard University Press, 1991, Chap. 29; Russell Davidson and James MacKinnon, op. cit., pp. 201–207; John Fox, *Applied Regression Analysis, Linear Models, and Related Methods,* Sage Publications, 1997, pp. 393–400; and Ronald Gallant, *Nonlinear Statistical Models,* John Wiley and Sons, 1987.

involves several parameters, the method becomes very cumbersome and computationally expensive. For example, if an NLRM involves 5 parameters and 25 alternative values for each parameter are considered, you will have to compute the error sum of squares $(25)^5 = 9{,}765{,}625$ times! *Second,* there is no guarantee that the final set of parameter values you have selected will necessarily give you the absolute minimum error sum of squares. In the language of calculus, you may obtain a local and not an absolute minimum. In fact, no method guarantees a global minimum.

Direct Optimization

In direct optimization we differentiate the error sum of squares with respect to each unknown coefficient, or parameter, set the resulting equation to zero, and solve the resulting normal equations simultaneously. We have already seen this in Eqs. (14.2.4) and (14.2.5). But as you can see from these equations, they cannot be solved explicitly or *analytically*. Some iterative routine is therefore called for. One routine is called the **method of steepest descent.** We will not discuss the technical details of this method as they are somewhat involved, but the reader can find the details in the references. Like the method of trial and error, the method of steepest descent also involves selecting initial trial values of the unknown parameters but then it proceeds more systematically than the hit-or-miss or trial-and-error method. One disadvantage of this method is that it may converge to the final values of the parameters extremely slowly.

Iterative Linearization Method

In this method we linearize a nonlinear equation around some initial values of the parameters. The linearized equation is then estimated by OLS and the initially chosen values are adjusted. These adjusted values are used to *relinearize* the model, and again we estimate it by OLS and readjust the estimated values. This process is continued until there is no substantial change in the estimated values from the last couple of iterations. The main technique used in linearizing a nonlinear equation is the **Taylor series expansion** from calculus. Rudimentary details of this method are given in Appendix 14A, Section 14A.2. Estimating NLRM using Taylor series expansion is systematized in two algorithms, known as the **Gauss–Newton iterative method** and the **Newton–Raphson iterative method.** Since one or both of these methods are now incorporated in several computer packages, and since a discussion of their technical details will take us far beyond the scope of this book, there is no need to dwell on them here.[6] In the next section we discuss some examples using these methods.

[6]There is another method that is sometimes used, called the **Marquard method,** which is a compromise between the method of steepest descent and the linearization (or Taylor series) method. The interested reader may consult the references for the details of this method.

14.5 ILLUSTRATIVE EXAMPLES

EXAMPLE 14.1

MUTUAL FUND ADVISORY FEES

Refer to the data given in Table 14.1 and the NLRM (14.2.2). Using the Eviews 4 nonlinear regression routine, which uses the linearization method,[7] we obtained the following regression results; the coefficients, their standard errors, and their *t* values are given in a *tabular form:*

Variable	Coefficient	Std. error	*t* value	*p* value
Intercept	0.5089	0.0074	68.2246	0.0000
Asset	−0.0059	0.00048	−12.3150	0.0000

$$R^2 = 0.9385 \qquad d = 0.3493$$

From these results, we can write the estimated model as:

$$\widehat{\text{Fee}}_i = 0.5089\,\text{Asset}^{-0.0059} \tag{14.5.1}$$

Before we discuss these results, it may be noted that if you do not supply the initial values of the parameters to start the linearization process, Eviews will do it on its own. It took Eviews five iterations to obtain the results shown in (14.5.1). However, you can supply your own initial values to start the process. To demonstrate, we chose the initial value of $\beta_1 = 0.45$ and $\beta_2 = 0.01$. We obtained the same results as in (14.5.1) but it took eight iterations. *It is important to note that fewer iterations will be required if your initial values are not very far from the final values.* In some cases you can choose the initial values of the parameters by simply running an OLS regression of the regressand on the regressor(s), simply ignoring the nonlinearities. For instance, using the data in Table 14.1, if you were to regress fee on assets, the OLS estimate of β_1 is 0.5028 and that of β_2 is −0.002, which are much closer to the final values given in (14.5.1). (For the technical details, see Appendix 14A, Section 14A.3.)

Now about the properties of NLLS estimators. You may recall that, in the case of linear regression models with normally distributed error terms, we were able to develop exact inference procedures (i.e., test hypotheses) using the *t*, *F*, and χ^2 tests in small as well as large samples. Unfortunately, this is not the case with NLRMs, even with normally distributed error terms. *The NLLS estimators are not normally distributed, are not unbiased, and do not have minimum variance* in finite, or small, samples. As a result, we cannot use the *t* test (to test the significance of an individual coefficient) or the *F* test (to test the overall significance of the estimated regression) because we cannot obtain an unbiased estimate of the error variance σ^2 from the estimated residuals. Furthermore, the residuals (the difference between the actual *Y* values and the estimated *Y* values from the NLRM) do not necessarily sum to zero, ESS and RSS do not necessarily add up to the TSS, and therefore $R^2 = \text{ESS}/\text{TSS}$ may not be a meaningful descriptive statistic for such models. However, we can compute R^2 as:

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} \tag{14.5.2}$$

where Y = regressand and $\hat{u}_i = Y_i - \hat{Y}_i$, where $\hat{Y}_i$ are the estimated Y values from the (fitted) NLRM.

(*Continued*)

[7]Eviews provides three options: quadratic hill climbing, Newton–Raphson, and Berndt–Hall–Hall–Hausman. The default option is quadratic hill climbing, which is a variation of the Newton–Raphson method.

EXAMPLE 14.1 (*Continued*)

Consequently, inferences about the regression parameters in nonlinear regression are usually based on large-sample theory. This theory tells us that the least-squares and maximum likelihood estimators for nonlinear regression models with normal error terms, when the sample size is large, are approximately normally distributed and almost unbiased, and have almost minimum variance. This large-sample theory also applies when the error terms are not normally distributed.[8]

In short, then, all inference procedures in NLRM are large sample, or asymptotic. Returning to Example 14.1, the *t* statistics given in (14.5.1) are meaningful only if interpreted in the large-sample context. In that sense, we can say that estimated coefficients shown in Eq. (14.5.1) are individually statistically significant. Of course, our sample in the present instance is rather small.

Returning to Eq. (14.5.1), how do we find out the rate of change of Y (= fee) with respect to X (asset size)? Using the basic rules of derivatives, the reader can see that the rate of change of Y with respect to X is:

$$\frac{dY}{dX} = \beta_1\beta_2 e^{\beta_2 X} = (-0.0059)(0.5089)e^{-0.0059X} \qquad \textbf{(14.5.3)}$$

As can be seen, the rate of change of fee depends on the value of the assets. For example, if $X = 20$ (million), the expected rate of change in the fees charged can be seen from (14.5.3) to be about −0.0031 percent. Of course, this answer will change depending on the X value used in the computation. Judged by the R^2 as computed from (14.5.2), the R^2 value of 0.9385 suggests that the chosen NLRM fits the data in Table 14.1 quite well. The estimated Durbin–Watson value of 0.3493 may suggest that there is autocorrelation or possibly model specification error. Although there are procedures to take care of these problems as well as the problem of heteroscedasticity in NLRM, we will not pursue these topics here. The interested reader may consult the references.

EXAMPLE 14.2

THE COBB–DOUGLAS PRODUCTION OF THE MEXICAN ECONOMY

Refer to the data given in exercise 14.9. These data refer to the Mexican economy for years 1955–1974. We will see if the NLRM given in (14.1.4) fits the data, noting that Y = output, X_2 = labor input, and X_3 = capital input. Using Eviews 4, we obtained the following regression results, after 32 iterations.

Variable	Coefficient	Std. error	*t* value	*p* value
Intercept	0.5292	0.2712	1.9511	0.0677
Labor	0.1810	0.1412	1.2814	0.2173
Capital	0.8827	0.0708	12.4658	0.0000

$$R^2 = 0.9942 \qquad d = 0.2899$$

(*Continued*)

[8]John Neter, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman, *Applied Regression Analysis,* 3d ed., Irwin, 1996, pp. 548–549.

EXAMPLE 14.2 (*Continued*)

Therefore, the estimated Cobb–Douglas function is:

$$\widehat{\text{GDP}}_t = 0.5292\text{Labor}_t^{0.1810}\text{Capital}_t^{0.8827} \qquad \textbf{(14.5.2)}$$

Interpreted asymptotically, the equation shows that only the coefficient of the capital input is significant in this model. In exercise 14.9 you are asked to compare these results with those obtained from the multiplicative Cobb–Douglas production function as given in (14.1.2).

EXAMPLE 14.3

GROWTH OF U.S. POPULATION, 1970–1999

The table in exercise 14.8 gives you data on total U.S. population for the period 1970–1999. A **logistic growth model** of the following type is often used to measure the growth of a population:

$$Y_t = \frac{\beta_1}{1 + e^{(\beta_2 + \beta_3 t)}} + u_t \qquad \textbf{(14.5.4)}$$

where Y = population; t = time, measured chronologically; and the β's are the parameters. Notice an interesting thing about this model. Although there are only two variables, population and time, there are three unknowns, which shows that in a NLRM there can be more parameters than variables.

Sample: 1970–1999
Included observations: 30
Convergence achieved after one iteration

	Coefficient	Std. error	*t* statistic	*p* value
β_1	1432.738	508.0113	2.8202	0.0089
β_2	1.7986	0.4124	4.3613	0.0002
β_3	−0.0117	0.0008	−14.0658	0.0000

$$R^2 = 0.9997 \qquad d = 0.3345$$

The estimated model, therefore, is:

$$\hat{Y}_t = \frac{1432.739}{1 + e^{1.7986 - 0.0117t}} \qquad \textbf{(14.5.5)}$$

Since we have a reasonably large sample, asymptotically all the estimated coefficients are statistically significant. The low Durbin–Watson statistic suggests that the error term is probably autocorrelated. In exercise 14.8 you are asked to compare the preceding model with the semilog model: $\ln Y_t = \beta_1 + \beta_2 \text{time} + u_t$ and compute the underlying growth rate of population for both models.

14.6 SUMMARY AND CONCLUSIONS

The main points discussed in this chapter can be summarized as follows:

1. Although linear regression models predominate theory and practice, there are occasions where nonlinear-in-the-parameter regression models (NLRM) are useful.

2. The mathematics underlying linear regression models is comparatively simple in that one can obtain explicit, or analytical, solutions of the coefficients of such models. The small-sample and large-sample theory of inference of such models is well established.

3. In contrast, for intrinsically nonlinear regression models, parameter values cannot be obtained explicitly. They have to be estimated numerically, that is, by iterative procedures.

4. There are several methods of obtaining estimates of NLRMs, such as (1) trial and error, (2) nonlinear least squares (NLLS), and (3) linearization through Taylor series expansion.

5. Computer packages now have built-in routines, such as Gauss–Newton, Newton–Raphson, and Marquard. These are all iterative routines.

6. NLLS estimators do not possess optimal properties in finite samples, but in large samples they do have such properties. Therefore, the results of NLLS in small samples must be interpreted carefully.

7. Autocorrelation, heteroscedasticity, and model specification problems can plague NLRM, as they do linear regression models.

8. We illustrated the NLLS with several examples. With the ready availability of user-friendly software packages, estimation of NLRM should no longer be a mystery. Therefore, the reader should not shy away from such models whenever theoretical or practical reasons dictate their use. As a matter of fact, if you refer to exercise 12.10, you will see from Eq. (1) that there is intrinsically a nonlinear regression model that should be estimated as such.

EXERCISES

Questions

14.1. What is meant by intrinsically linear and intrinsically nonlinear regression models? Give some examples.

14.2. Since the error term in the Cobb–Douglas production function can be entered multiplicatively or additively, how would you decide between the two?

14.3. What is the difference between OLS and nonlinear least-squares (NLLS) estimation?

14.4. The relationship between pressure and temperature in saturated steam can be expressed as[9]:

$$Y = \beta_1 (10)^{\beta_2 t/(\gamma + t)} + u_t$$

[9]Adapted from Draper and Smith, op. cit., p. 554.