# Non-Linear Regression
© 2006-2008 Samuel L. Baker

The linear least squares method that you have been using fits a straight line or a flat plane to a bunch of data points. Sometimes the true relationship that you want to model is curved, rather than flat. For example, if something is growing exponentially, which means growing at a steady rate, the relationship between X and Y is curve, like that shown to the right.

To fit something like this, you need non-linear regression. Often, you can adapt linear least squares to do this. The method is to create new variables from your data. The new variables are non-linear functions of the variables in your data. If you construct your new variables properly, the curved function of your original variables can be expressed as a linear function of your new variables.

We will be discussing two popular non-linear equation forms that are amenable to this technique:
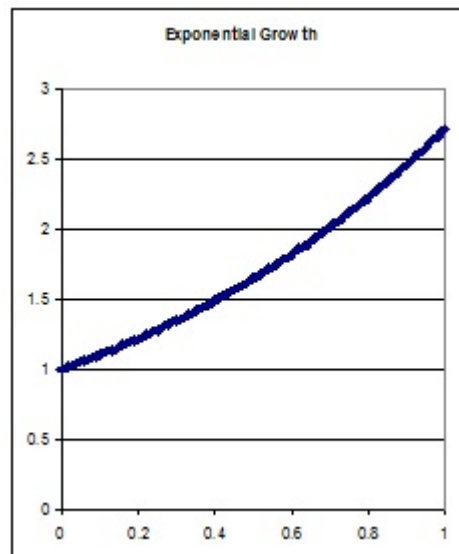
Equation	Interpretation	Linear Form
$Y = Ae^{bX}u$	Y is growing (or shrinking) at a constant relative rate of b.	$\ln(Y) = \ln(A) + bX + \ln(u)$
$Y = AX^{b}u$	The elasticity of Y with respect to X is a constant, b.	$\ln(Y) = \ln(A) + b\cdot\ln(X) + \ln(u)$

Consider the first equation, $Y = Ae^{bX}u$. This is an exponential growth equation. b is the growth rate. u is a random error term.

If you take the logarithm of both sides of that equation, you get $\ln(Y) = \ln(A) + bX + \ln(u)$. This equation has logarithms in it, but they relate in a linear way. It is in the form y=a+bX+error, except that y, a, and the error are logarithms. (A review of logarithms and e is coming up on the next page. Bear with me until then.)

This means that if you create a new variable, the base-e logarithm of Y, written as $\ln(Y)$, you can use the regular least squares method to fit $\ln(Y) = \ln(A) + bX + \ln(u)$. That is a way of fitting the curve $Y = Ae^{bX}$ to your data. (We will go over this again after the review of logarithms.)

Similarly, consider the second equation, $Y = AX^{b}u$. This is a constant-elasticity equation (more on why I call it that later), typically used for demand curves. Take the logarithm of both sides of that equation and you get $\ln(Y) = \ln(A) + b\cdot\ln(X) + \ln(u)$.

For this equation, if you create the variable $\ln(Y)$ and also a variable for the base-e logarithm of X, written

as ln(X), you can use the regular least squares method to fit the curve $Y = AX^b$ to your data.

Before we go further into how to use these new equations, we had better review logarithms and e.

Logarithms defined:

If $a^b = c$ then $b = \log_a c$.

$b = \log_a c$ is read, "b is the log to the base a of c."

I remember the logarithm definition by remembering that:

The log of 1 is 0, regardless of the base.
Anything to the 0 power is 1.

If I can remember those two facts, then I can arbitrarily pick 10 as my base and write:
$10^0 = 1$ and $\log_{10} 1 = 0$.
From that I get $a^b = c$ and $\log_a c = b$, by substituting "a" for 10, "b" for 0, and "c" for 1.

You can't take the log of 0 or of a negative number. The c in $\log_a c$ must be a positive number.

e ($\approx$ 2.718282) is the base of the "natural logarithms" ($\log_e$ is written "ln").
Natural logarithms are used for continuous growth rates.
Something growing at a 100% annual rate, compounded continuously, will grow to e times its original size in one year. (The diagram on the preceding page shows a 100% growth rate.)
Something growing at an r annual rate, compounded continuously, will grow to e^r times its original size in one year.

In Excel, the function =EXP(A1) raises e the power of whatever number is in cell A1. =LN(A1) takes the natural $\log_e$ of whatever is in cell A1.

Some exponential algebra rules

$e^{a+b} = e^a e^b$
$e^{ab} = (e^a)^b$
$e^{\ln(a)} = a$
$e^{b \cdot \ln(a)} = a^b$
$e^0 = 1$

Some corresponding logarithm algebra rules

$\ln(ab) = \ln(a) + \ln(b)$
$\ln(a^b) = b \cdot \ln(a)$
$\ln(e^a) = a$

$\ln(1) = 0$

Non-linear models that can be made linear using logarithms:

Now we can get back to the curved models introduced at the beginning of this section and give fuller explanations.

Constant growth or decay equation $Y = Ae^{bX}u$

Suppose you're modeling a process that you think has continuous growth at a constant growth rate. This is like the diagram on page 1. You want to estimate the growth rate from your data.

$Y = Ae^{bX}u$ is your equation.

X is often a measure of time. X goes up by 1 for each second, hour, day, year, or whatever time unit you are using. Sometimes, X is a measure of dosage.
A is the starting amount, the amount expected when X is 0.
b is the relative growth rate, with continuous compounding.
u is the random error. Notice that we multiply by the error, rather than adding or subtracting the error as in standard linear regression. u has a mean of 1 and is always bigger than 0. u is never 0 or negative.

Use this equation if:

The analogy with steady growth or decay makes sense.
Y cannot be 0 or negative.
X can be positive, 0, or negative.

The interpretation of b in $Y = Ae^{bX}u$:

b is the parameter in which you are most interested, usually. Your estimate of b is your estimate of the *relative* change in Y associated with a unit change in X.

Mathematically, if X goes up by 1, Y is multiplied by e^b. That is because $Ae^{b(X+1)}$ equals Ae^{bX+b}, which equals $Ae^{bX}e^b$, which is Y is multiplied by e^b. That may not look like "relative change," but it is, if you are using continuous compounding.

People who use these models usually make a simpler statement about b. They say that if X goes up by 1, Y goes up by 100b percent. They mean that if X goes up by 1, Y is multiplied by (1+b). For example, if your estimate for b is 0.02, then it is said that if X goes up by 1, Y goes up by 2%, because Y is multiplied by 1.02. (1 + .02 = 1.02.)

The 1+b idea is fine so long as b is near 0. For small values of b, e^b is close to 1+b. Multiplying by 1+b is close to the same as multiplying by e^b. For example, if b is 0.05 and X increases by 1, Y is multiplied by e^5, which is about 1.051, so Y really increases by 5.1%. That is pretty close to 5%. Most things in the financial world and the public health world grow pretty slowly, so the 1+b or b-percent idea is OK. The downloadable file about time series has more details on this.

How to make the growth/decay equation linear:

To convert $Y = Ae^{bX}u$ to a linear equation, take the natural log of both sides:

$\ln(Y) = \ln(Ae^{bX}u)$ Apply the rules above and you get:

$\ln(Y) = \ln(A) + bX + \ln(u)$

To implement this, create a new variable $y = \ln(Y)$. (The Y in the original equation is "big Y." The new variable is "little y.")

Also, define v as equaling $\ln(u)$ and a as equaling $\ln(A)$.

Subsitute y for $\ln(Y)$, a for $\ln(A)$, and v for $\ln(u)$ in $\ln(Y) = \ln(A) + bX + \ln(u)$ and you get:

$$y = a + bX + v$$

This is a linear equation! It may seem like cheating to turn a curve equation into a straight line equation this way, but that is how it's done.

Doing the analysis this way is not exactly the same as fitting a curve to points using least squares. This method this method minimizes the sum of the squares of $\ln(Y) - (\ln(A) + bX)$, not the sum of the squares of $Y - Ae^{bX}$. In our equation, the error term is something we multiply by rather than add. Our equation is $Y=Ae^{bX}u$, rather than $Y = Ae^{bX}+u$.

We make our equation $Y=Ae^{bX}u$, rather than $Y = Ae^{bX}+u$, so that when we take the log of both sides we get an error, which we call v, the logarithm of u in the original equation, that behaves like errors are supposed to behave for linear regression. For example, for linear regression we have to assume that the error's expected value is 0. To make this work, we assume that the expected value of u is 1. That makes the expected value of $\ln(u)$ equal zero, because $\ln(1) = 0$. In $Y=Ae^{bX}u$, u is never 0 or negative, but v can take on positive or negative values, because if u is less than 1, $v=\ln(u)$ is less than 0.

When you transform $Y=Ae^{bX}u$ to $y = a + bX + v$ and do a least squares regression, here is how you interpret the coefficients:
$\hat{b}$ is your estimate of the growth rate. $\hat{a}$ is your estimate of $\ln(A)$. To get an estimate of A, calcualte e to the a-hat power.

Getting a prediction for Y from your prediction for y

After you calculate a prediction for little y, you have to convert it to big Y.

Your prediction for $Y = e^{\text{Your prediction for y}}$. In Excel, use the =EXP() function.

An example

The first two columns are your data in a dose-response experiment. The third column is calculated from the second by taking the natural logarithm.

Dose	Bacteria	LnBacteria
1	35500	10.4773
2	21100	9.95703
3	19700	9.88837
4	16600	9.71716
5	14200	9.561
6	10600	9.26861
7	10400	9.24956
8	6000	8.69951
9	5600	8.63052
10	3800	8.24276
11	3600	8.18869
12	3200	8.07091
13	2100	7.64969
14	1900	7.54961
15	1500	7.31322

Your regression results are:

```
LnBacteria is the dependent variable, with a mean of 8.8309284

    Variable Coefficient    Std Error T-statistic      P-Value
    Intercept    10.57833    0.0597781   176.95997    0
        Dose   -0.2184253    0.0065747  -33.222015    0
```

You have found that each dose reduces the number of bacteria by about 22%.

```
Prediction, showing how it is calculated:
Variable      Coefficient * Your Value =  Product
Intercept        10.57833           1.0   10.57833
Dose           -0.2184253          16.0  -3.4948041
           Total is the prediction:   7.0835264
90% conf. interval is    6.861792 to   7.3052607
```

Your prediction for the number of bacteria if the dose is 16 is $e^{7.0835264} = 1192$. In Excel, calculate =EXP(7.0835264). The 90% confidence interval for that prediction is from $e^{6.861792}$ to $e^{7.3052607}$, which is about 955 to 1488.

Constant elasticity equation $Y=AX^b u$

Another non-linear equation that is commonly used is the constant elasticity model. Applications include supply, demand, cost, and production functions.

$Y = AX^b u$ is your equation

X is some continuous variable that's always bigger than 0.
A determines the scale.
b is the elasticity of Y with respect to X (more on elasticity below).
u , the error, has a mean of 1 and is always bigger than 0.

Graph of $Y=X^2 u$
u is log-normally distributed with a mean of 1.

Use this equation if:

•   X and Y are always positive. X and Y can not be 0 or negative.
•   You suspect, or wish to test for, diminishing ($|b|<1$) or increasing ($|b|>1$) returns to scale
•   If it is reasonable that the percentage change in Y caused by a 1% change in X is the same at any level of X (or any level of Z, if you are doing a multiple regression).   That is why this is called a "constant elasticity" equation.

A constant elasticity model can never give a negative prediction. If a negative prediction would make sense, you do not want this model. If a negative prediction would not make sense, this model may be what you want.

The interpretation of b, the elasticity:

b is the percentage change in Y that you get when X changes by 1%. Economists call b the elasticity of Y with respect to X.
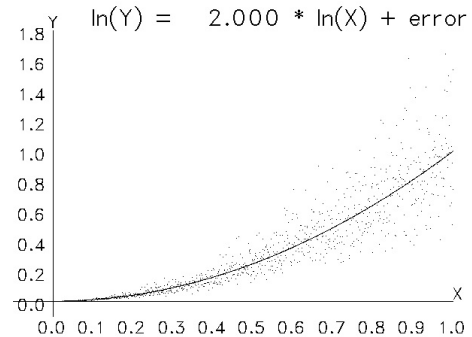
If b is positive, you see a pattern like in the diagram above.

If b is negative in this equation, the pattern is like a hyperbola. The points get closer and closer to the X axis as you go the right. The points get closer and closer to the Y axis and they get higher and higher as X goes to the left and gets near 0.
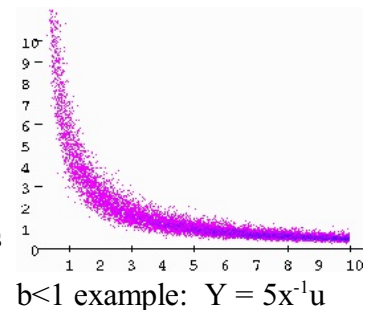
If you do a multiple regression version of this, with the equation $Y = A \cdot X^b \cdot Z^c \cdot u$, then b is the percentage change in Y that you get when X changes by 1% and Z does not change. Economists say b and c are elasticities of Y with respect to X and Z.

$b<1$ example: $Y = 5x^{-1}u$

How to make this linear:

If you take the log of both sides of $Y = A \cdot X^b \cdot u$, you get:

$$\ln(Y) = \ln(A) + b \cdot \ln(X) + \ln(u)$$

To make this linear, create two new variables, $y = \ln(Y)$ and $x = \ln(X)$. (If you have more than one X variable, take the logs of all of them.) Let a be $\ln(A)$ and v be $\ln(u)$, and our equation becomes nice and linear:

$$y = a + bx + v$$

As with the growth model, to use least squares, we must assume that v behaves like errors are supposed to behave for linear regression. One of those assumptions is that v's expected value is 0. That is why we assume that the mean of u is 1. That fits because $\ln(1) = 0$. u is never 0 or negative, but v can take on positive or negative values, because if u is less than 1, $v = \ln(u)$ is less than 0.

The multiple regression version of the constant elasticity model works like this:

$Y = A \cdot X^b \cdot Z^c \cdot u$ is the equation.

Take the log of both sides to get:

$\ln(Y) = \ln(A) + b \cdot \ln(X) + c \cdot \ln(Z) + \ln(u)$

which you estimate as

$y \quad = a + b \cdot x + c \cdot z + v$

where $y = \ln(Y)$, $a = \ln(A)$, $x = \ln(X)$, etc.

This equation lets you look for diminishing or increases returns to scale. If $|b+c| < 1$, you have diminishing returns. Doubling X and Z makes Y go up less than twofold. If $|b+c| > 1$, you have increasing returns. Doubling X and Z makes Y more than double.

Sometimes a linear model shows heteroskedasticity -- apparently unequal variances of the errors -- that takes the form of the residuals fanning out or in as you look down the residual plot. This constant elasticity form can correct that.

Getting a prediction for Y from your prediction for y

As with the growth-decay model, when you calculate a prediction for y, you have to convert it to Y. You do this by raising e to the y power, because $Y = e^y$. In Excel, use the =EXP() function.

An example

Here are some data:

Price	Visits	LnPrice	LnVisits
10	512	2.30259	6.23832
10	396	2.30259	5.98141
10	456	2.30259	6.12249
10	525	2.30259	6.2634
20	347	2.99573	5.84932
20	377	2.99573	5.93225
20	423	2.99573	6.04737
20	330	2.99573	5.79909
30	366	3.4012	5.90263
30	335	3.4012	5.81413
30	301	3.4012	5.70711
30	309	3.4012	5.73334
40	345	3.68888	5.84354
40	320	3.68888	5.76832
40	325	3.68888	5.78383
40	285	3.68888	5.65249

The first two columns are the actual data. The third and fourth columns are calculated as the natural logs of the first two columns. 3.6888 approximately = LN(40), for example.

```
LnVisits is the dependent variable, with a mean of 5.9024412

    Variable Coefficient    Std Error T-statistic     P-Value
    Intercept   6.8044237   0.1496568   45.466851   0
     LnPrice  -0.2912347    0.047653  -6.1115685   0.0000269
```

The estimated coefficient of LnPrice says that when the price is 1% higher, there are 0.29% fewer visits.

To predict visits if the price is $25, first calculate =EXP(25). It is about 3.218875825. Use that in the prediction form.

```
Prediction, showing how it is calculated:
Variable      Coefficient * Your Value =  Product
Intercept       6.8044237          1.0   6.8044237
LnPrice        -0.2912347    3.2188758  -0.9374482
          Total is the prediction:   5.8669755
90% conf. interval is   5.6865181 to   6.0474328
```

The prediction for the logarithm of visits is 5.8669755. The prediction for visits is =EXP(5.8669755), which is about 353. The 90% confidence interval for that prediction is from =EXP(5.6865181) to =EXP(6.0474328), which is about 295 to 423. Notice that 353 is not in the middle of this confidence

interval.

More comments on our two non-linear models

Time series often use the $\ln(Y) = \ln(A) + bX + \ln(u)$ form.
> The multiple regression version has more than one X variable. Your independent (X) variables measure time, either
>> Sequentially: 0,1,2,3,... or 1996, 1997, 1998 , ... or
>> Cyclically: seasonal dummy variables (to be explained next week)
> Notice that you do not take the logs of these X variables, just the Y variable.

> The coefficient b is the growth/decline rate, meaning that a change in X of 1 corresponds with a 100b percent change in Y, if b is small.

Cross section studies, that look for causation rather than time trends, often use the constant elasticity form,
$\ln(Y) = \ln(A) + b \cdot \ln(X) + \ln(u)$

> For this model, your independent variables have a proportional effect on your Y variable.

> You do take the logs of your X variables.

> Coefficients are elasticities, meaning that a change in X by 1% (of X) changes Y by b percent.

General comments on working with transformed equations

You must assume that the error in the **transformed** equation has the desired properties (normal distribution with mean 0).

Once you get your estimates from the transformed equation, going back to the original equation can be tricky. Some original-equation parameter estimates are biased, but consistent, if the parameter was transformed (e.g. A in the models above). Confidence intervals around predicted values are no longer symmetrical. You must get the confidence interval from the transformed equation and then transform the ends back.

Before you start with using a non-linear regression equation:

> Have a good reason for not using a linear model, such as
>> a theory of how the process that you're observing works, or
>> a pattern you see on a graph or in the residuals from a linear regression
> Determine which non-linear equation is best for your data. Can you make a sensible analogy with steady growth or decay? How about with demand or production?

To do the regression:
> Use algebra (or just read the above) to transform your original non-linear equation into one that is linear. For example, $Y = AX^b u$ transforms to
>> $\ln(Y) = \ln(A) + b \cdot \ln(X) + \ln(u).$

Create the variables you need from your data.

Use least squares, including the transformed variables in your data set.

Your predictions, and at least some of your estimated coefficients, will have to be transformed to anti-logs to get back to the parameters in the original non-linear equation. In both of these equations, the predicted value of ln(Y) must be transformed with the exp function to get a predicted value for Y.