## Non-linear Models



Ass. Prof. Andriy Stavytskyy

## References

- Greene W. H., Econometric Analysis, 7th ed., Prentice Hall, 2011.
- Murray M. P. (2005) Econometrics: A Modern Introduction. Prentice Hall.
- Stock James H., Mark W. Watson (2010) Introduction to Econometrics (3rd Edition)


## Agenda

- Examples of non-linear models
- Estimation of Non-Linear Regression

"What should we do, if we fail to find an appropriate model that satisfy all tests?"



## Examples of non-linear models



## Inherently Linear Models

Non-linear models that can be expressed in linear form
-Can be estimated by least square in linear form

- Require data transformation


## Logarithmic Transformation

$$
\mathrm{P}=\beta_{0}+\beta_{1} \ln x_{1}+\beta_{2} \ln x_{2}+\varepsilon
$$

$$
\beta_{1}>0
$$

$\beta_{1}<0$

## Square-Root Transformation

$$
Y_{i}=\beta_{0}+\beta_{1} \sqrt{X_{1 i}}+\beta_{2} \sqrt{X_{2 i}+\varepsilon_{i}}
$$

## Reciprocal Transformation

$$
Y_{i}=\beta_{0}+\beta_{1} \frac{1}{X_{1 i}}+\beta_{2} \frac{1}{X_{2 i}}+\varepsilon_{i}
$$

## V Asymptote

$\beta_{1}<0$

## $\beta_{1}>0$

## Exponential Transformation

 $Y_{i}=e^{\beta_{0}+\beta_{1} X_{1 i}+\beta_{2} X_{2 i}} \varepsilon_{i}$
## Y <br> $\beta_{1}>0$

## $\beta_{1}<0$

## Mathematical transformations: problems

- Outdated!
- Data become transformed, but errors don't
- Violates linear regression assumptions


## Same Data Transformation



- Nonlinear Regression
- Although usually inappropriate to analyze transformed data, it is often helpful to display data after a linear transformation for ease of interpretation

- Scratched linear transformation
- Note how the transformation amplified and distorted the scatter, and the line doesn't fit the maximum values


## Main differences between NLR and LR

- It is not possible to directly derive an equation to compute the best-fit values for your data
- Requires computationally intensive, iterative approach with matrix algebra


## Nonlinear vs. Linear Model Fit



## Linear vs. Nonlinear - 1

- In this context, linear and nonlinear are used with reference to the parameters, not the variables
- For example,

$$
\begin{aligned}
& Y=A+B X \\
& Y=A+B X^{3} \\
& Y=A e^{X}
\end{aligned}
$$

are all linear (in the parameters $A$ and $B$ )

## Linear vs. Nonlinear - 2

- All of these are non-linear:

$$
\begin{aligned}
& Y=A+B^{2} X \\
& Y=A+X^{3} B \\
& Y=A e^{B X}
\end{aligned}
$$

(but can be reparameterized)

- Big Difference:
- Linear regression: can be calculated analytically
- Optimization: starting parameter values need to be provided!


## Getting Started

- Your Data
- Equation (from literature or prior theoretical work)
- Initial/Starting parameter values for the equation
- If initial parameters sufficiently close to optimal values, convergence will occur within a few steps
- Ensure that procedure reached global minimum rather than local minimum
- Important to choose reasonable starting values


## Popular non-linear regression models

- Exponential model:
- Power model:

$$
\left(y=a e^{b x}\right)
$$

$$
\left(y=a x^{b}\right)
$$

- Saturation growth model:
- Polynomial model: $\quad\left(y=a_{0}+a_{1} x+\ldots+a_{m} x^{m}\right)$



## Example: The Mechanistic Growth Model - 1

- many models cannot be transformed into a linear model

$$
Y=\alpha\left(1-\beta e^{-k x}\right)+\varepsilon
$$

> or (ignoring $\varepsilon$ ) "rate of increase in $Y$ "

$$
\frac{d Y}{d x}=k(\alpha-Y)
$$

## Example: The Mechanistic Growth Model - 2

Mechanistic Growth Model


## Example: The Logistic Growth Model - 1

$$
Y=\frac{\alpha}{1+\beta e^{-k x}}+\varepsilon
$$

- or (ignoring $\varepsilon$ ) "rate of increase in $Y$ "

$$
\frac{d Y}{d x}=\frac{k Y(\alpha-Y)}{\alpha}
$$

## Example: The Logistic Growth Model - 2

## Logistic Growth Model



## Example: The Gompertz Growth Model - 1

$$
Y=\alpha e^{-\beta e^{-k x}}+\varepsilon
$$

- or (ignoring e) "rate of increase in $Y$ "

$$
\frac{d Y}{d x}=k Y \ln \left(\frac{\alpha}{Y}\right)
$$

## Example: The Gompertz Growth Model - 2

## Gompertz Growth Model



## Estimation of Non-Linear Regression



## Non-Linear Regression

Previously we have fitted, by least squares, the General Linear model which were of the type:

$$
\mathrm{Y}=\beta_{0}+\beta_{1} \mathrm{X}_{1}+\beta_{2} \mathrm{X}_{2}+\ldots+\beta_{\mathrm{k}} \mathrm{X}_{\mathrm{k}}+\varepsilon
$$

- When we are led to a model of nonlinear form, we would usually prefer to fit such a model whenever possible, rather than to fit an alternative, perhaps, less realistic, linear model.


## Non-Linear Regression

- This model will generally be of the form:
$\mathrm{Y}=\mathrm{f}\left(\mathrm{X}_{1}, \mathrm{X}_{2}, \ldots, \mathrm{X}_{\mathrm{k}-1} \mid \beta_{0}, \beta_{1}, \beta_{2}, \ldots, \beta_{\mathrm{q}}\right)+\varepsilon$ where the function (expression) f is known except for the $q$ unknown parameters $\beta_{0}, \beta_{1}, \beta_{2}, \ldots, \beta_{\mathrm{q}}$.



## Non-Linear Regression

- Suppose that we have collected data on the Y ,

$$
\left(y_{1}, y_{2}, \ldots y_{n}\right)
$$

- corresponding to n sets of values of the independent variables $X_{1}, X_{2}, \ldots$ and $X_{k-1}$

$$
\begin{gathered}
\left(x_{11}, x_{21}, \ldots, x_{k-1}, 1\right), \\
\left(x_{12}, x_{22}, \ldots, x_{k-1}, 2\right), \\
\ldots \\
\left(x_{1 n}, x_{2 n}, \ldots, x_{k-1}, n\right) .
\end{gathered}
$$

## Estimation

- For a set of possible values $\boldsymbol{\beta}_{1}, \boldsymbol{\beta}_{2}, \ldots, \boldsymbol{\beta}_{\mathrm{k}-1}$ of the parameters, a measure of how well these values fit the model is the residual sum of squares function

$$
\begin{gathered}
S\left(\beta_{0}, \beta_{1}, \beta_{2}, \ldots, \beta_{k-1}\right)=\sum_{i=1}^{n}\left(y_{i}-\hat{y}_{i}\right)^{2}= \\
=\sum_{i=1}^{n}\left[y_{i}-f\left(x_{1 i}, x_{2 i}, \ldots, x_{k-1 i} \mid \beta_{0}, \beta_{1}, \beta_{2}, \ldots, \beta_{q}\right)\right]^{2}
\end{gathered}
$$

where $\quad \hat{y}_{i}=f\left(x_{1 i}, x_{2 i}, \ldots, x_{k-1 i} \mid \beta_{0}, \beta_{1}, \beta_{2}, \ldots, \beta_{q}\right)$
is the predicted value of the response variable $y_{i}$ from the values of the $k$ independent variables $\mathrm{x}_{1 \mathrm{i}}, \mathrm{x}_{2 \mathrm{i}}, \ldots, \mathrm{x}_{\mathrm{k}-1, \mathrm{i}}$ using the model and the values of the parameters.

## Example-1

- We have collected data on two variables $(\mathrm{Y}, \mathrm{X}):\left(\mathrm{x}_{1}, \mathrm{y}_{1}\right)\left(\mathrm{x}_{2}, \mathrm{y}_{2}\right)\left(\mathrm{x}_{3}, \mathrm{y}_{3}\right), \ldots,\left(\mathrm{x}_{\mathrm{n}}, \mathrm{y}_{\mathrm{n}}\right)$
- The model we will consider is

$$
\mathrm{Y}=\beta_{1} \mathrm{e}^{-\mathrm{X}}+\beta_{2}+\varepsilon
$$



## Example - 2

- The predicted value of $y_{i}$ using the model is:

$$
\hat{y}_{i}=\beta_{1} e^{-x_{i}}+\beta_{2}
$$

The Residual Sum of Squares is:
$S\left(\beta_{1}, \beta_{2}\right)=\sum_{i=1}^{n}\left(y_{i}-\hat{y}_{i}\right)^{2}=\sum_{i=1}^{n}\left(y_{i}-\left[\beta_{1} e^{-x_{i}}+\beta_{2}\right]\right)^{2}$
The least squares estimates of $\beta 1$ and $\beta 2$ satisfy
$S\left(\hat{\beta}_{1}, \hat{\beta}_{2}\right)=\min _{\beta_{1}, \beta_{2}} S\left(\beta_{1}, \beta_{2}\right)=\min _{\beta_{1}, \beta_{2}} \sum_{i=1}^{n}\left(y_{i}-\left[\beta_{1} e^{-x_{i}}+\beta_{2}\right]\right)^{2}$

## Techniques for Estimating the Parameters of a Nonlinear

## System

- In some nonlinear problems it is convenient to determine equations (the Normal Equations) for the least squares estimates and develop an iterative technique for solving them.
- We shall mention three of these:
- Steepest descent,
- Linearization,
- Marquardt's procedure.


## Steepest Descent - 1

The steepest descent method focuses on determining the values of regression coefficients that minimize the sum of squares function, S.

- The basic idea is to determine from an initial point, $\quad \beta_{0}, \beta_{1}^{0}, \beta_{2}^{0}, \ldots, \beta_{k}^{0}$, and the tangent plane to $\mathrm{S}\left(\beta_{0}, \beta_{1}, \beta_{2}, \ldots, \beta_{\mathrm{k}}\right)$ at this point, the vector along which the function $\mathrm{S}\left(\beta_{0}, \beta_{1}, \beta_{2}, \ldots\right.$ , $\beta_{\mathrm{k}}$ ) will be decreasing at the fastest rate.
- The method of steepest descent than moves from this initial point along the direction of steepest descent until the value of $\mathrm{S}\left(\beta_{0}, \beta_{1}, \beta_{2}\right.$,
$\ldots, \beta_{\mathrm{k}}$ ) stops decreasing.


## Steepest Descent - 2

- While, theoretically, the steepest descent method will converge, it may do so in practice with agonizing slowness after some rapid initial progress.
- Slow convergence is particularly likely when the $S$ contours are attenuated and banana-shaped.


## Steepest Descent - 3

- A further disadvantage of the steepest descent method is that it is not scale invariant.
- The indicated direction of movement changes if the scales of the variables are changed, unless all are changed by the same factor.
- The steepest descent method will work satisfactorily for many nonlinear problems, especially if modifications are made to the basic technique.


## Steepest Descent: Graph

Steepest
descent path


Initial guess

## Linearization

The linearization (or Taylor series) method uses the results of linear least squares in a succession of stages.

- For initial values for the parameters the linearization method approximates $S$ with a linear function using a Taylor series expansion about the point and curtailing the expansion at the first derivatives.

The procedure is then repeated again until the successive approximations converge to hopefully at the least squares estimates.

## Linearization: Graph - 1

Contours of RSS for linear approximation

## Linearization: Graph - 2

Contours of RSS for linear approximation

## Linearization: Drawbacks

- It may converge very slowly; that is, a very large number of iterations may be required before the solution stabilizes.
- It may oscillate widely, continually reversing direction, and often increasing, as well as decreasing the sum of squares.
- It may not converge at all, and even diverge, so that the sum of squares increases iteration after iteration without bound.


## Levenberg-Marquardt algorithm (damped least-squares method)

- Provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function.
- The LMA interpolates between the Gauss-Newton algorithm and the method of gradient descent.
- The LMA finds a solution in many cases even if it starts very far off the final minimum. For well-behaved functions and reasonable starting parameters, the LMA tends to be a bit slower than other algorithms.
- The LMA is a very popular curve-fitting algorithm used in many software applications for solving generic curvefitting problems.
- However, the LMA finds only a local minimum, not a global minimum.

Review


## Mathematical transformations: problems

- Outdated!
- Data become transformed, but errors don't
- Violates linear regression assumptions



## Linear vs. Nonlinear

- All of these are non-linear:
- $Y=A+B^{2} X$
- $Y=A+X^{3} B$
- $Y=A e^{B X}$
(but can be reparameterized)
- Big Difference:
- Linear regression: can be calculated analytically
- Optimization: starting parameter values need to be provided!


## Techniques for Estimating the Parameters of a Nonlinear

## System

- In some nonlinear problems it is convenient to determine equations (the Normal Equations) for the least squares estimates and develop an iterative technique for solving them.
- We shall mention three of these:
- Steepest descent,
- Linearization,
- Marquardt's procedure.

Questions?



