

www.statsoft.com

- Products
- Solutions
- Buy
- Trials
- Support

TextbookGeneralized Additive Models

What can we help you find?

Search

Looking for info about statistics?

We wrote the book on it.
And you can read it for free!

- Elementary Concepts
- Statistics Glossary
- Basic Statistics
- ANOVA / MANOVA
- Association Rules
- Boosting Trees
- Canonical Analysis
- CHAID Analysis
- C & R Trees
- Classification Trees
- Cluster Analysis
- Correspondence Analysis
- Data Mining Techniques
- Discriminant Analysis
- Distribution Fitting
- Experimental Design
- Factor Analysis
- General Discrim. Analysis
- General Linear Models
- Generalized Additive Mod.
- Generalized Linear Mod.
- General Regression Mod.
- Graphical Techniques
- Ind.Components Analysis
- Linear Regression
- Log-Linear Analysis
- MARSplines
- Machine Learning
- Multidimensional Scaling
- Neural Networks
- Nonlinear Estimation
- Nonparametric Statistics
- Partial Least Squares
- Power Analysis
- Process Analysis
- Quality Control Charts
- Reliability / Item Analysis
- SEPATH (Structural eq.)
- Survival Analysis
- Text Mining
- Time Series / Forecasting

Generalized Additive Models (GAM)

- Additive Models
- Generalized Linear Models
- Distributions and Link Functions
- Generalized Additive Models
- Estimating the Nonparametric Function of Predictors via Scatterplot Smoothers
- A Specific Example: The Generalized Additive Logistic Model
- Fitting Generalized Additive Models
- Interpreting the Results
- Degrees of Freedom
- A Word of Caution









The methods available in *Generalized Additive Models* are implementations of techniques developed and popularized by Hastie and Tibshirani (1990). A detailed description of these and related techniques, the algorithms used to fit these models, and discussions of recent research in this area of statistical modeling can also be found in Schimek (2000).

Additive Models

The methods described in this section represent a generalization of [multiple regression](#) (which is a special case of [general linear models](#)). Specifically, in linear regression, a linear least-squares fit is computed for a set of predictor or X variables, to predict a dependent Y variable. The well known linear regression equation with m predictors, to predict a dependent variable Y, can be stated as:

$$Y = b_0 + b_1 \cdot X_1 + \dots + b_m \cdot X_m$$

Where Y stands for the (predicted values of the) dependent variable, X_1 through X_m represent the m values for the predictor variables, and b_0 , and b_1 through b_m are the regression coefficients estimated by multiple regression. A generalization of the multiple regression model would be to maintain the additive nature of the model, but to replace the simple terms of the linear equation $b_1 \cdot X_i$ with $f_i(X_i)$ where f_i is a non-parametric function of the predictor X_i . In other words, instead

-  [Variance Components](#)
-  [Statistical Advisor](#)
-  [Distribution Tables](#)
-  [References Cited](#)
-  [Send Comments](#)
-  [Business Solutions](#)
-  [Free Resources](#)
-  [About Textbook](#)

of a single coefficient for each variable (additive term) in the model, in additive models an unspecified (non-parametric) function is estimated for each predictor, to achieve the best prediction of the dependent variable values.

Generalized Linear Models

To summarize the basic idea, the [generalized linear model](#) differs from the [general linear model](#) (of which multiple regression is a special case) in two major respects: First, the distribution of the dependent or response variable can be (explicitly) non-normal, and does not have to be continuous, e.g., it can be [binomial](#); second, the dependent variable values are predicted from a linear combination of predictor variables, which are "connected" to the dependent variable via a [link function](#). The general linear model for a single dependent variable can be considered a special case of the generalized linear model: In the general linear model the dependent variable values are expected to follow the [normal distribution](#), and the link function is a simple identity function (i.e., the linear combination of values for the predictor variables is not transformed).

To illustrate, in the general linear model a response variable Y is linearly associated with values on the X variables while the relationship in the generalized linear model is assumed to be

$$Y = g(b_0 + b_1 X_1 + \dots + b_m X_m)$$

where $g(\dots)$ is a function. Formally, the inverse function of $g(\dots)$, say $g_i(\dots)$, is called the link function; so that:

$$g_i(\mu Y) = b_0 + b_1 X_1 + \dots + b_m X_m$$

where μY stands for the expected value of Y .

Distributions and Link Functions

Generalized Additive Models allows you to choose from a wide variety of distributions for the dependent variable, and [link functions](#) for the effects of the predictor variables on the dependent variable (see McCullagh and Nelder, 1989; Hastie and Tibshirani, 1990; see also [GLZ Introductory Overview - Computational Approach](#) for a discussion of link functions and distributions):

[Normal](#), [Gamma](#), and [Poisson](#) distributions:

Log link: $f(z) = \log(z)$

Inverse link: $f(z) = 1/z$

Identity link: $f(z) = z$

[Binomial](#) distributions:

Logit link: $f(z) = \log(z/(1-z))$

Generalized Additive Models

We can combine the notion of [additive models](#) with [generalized linear models](#), to derive the notion of generalized additive models, as:

$$g_i(\mu Y) = S_i(f_i(X_i))$$

In other words, the purpose of generalized additive models is to maximize the quality of prediction of a dependent variable Y from various distributions, by estimating unspecific (non-parametric) functions of the predictor variables which are "connected" to the dependent variable via a link function.

[To index](#)

Estimating the Nonparametric Function of Predictors via Scatterplot Smoothers

A unique aspect of [generalized additive models](#) are the non-parametric functions f_j of the predictor variables X_j . Specifically, instead of some kind of simple or complex parametric functions, Hastie and Tibshirani (1990) discuss various general [scatterplot smoothers](#) that can be applied to the X variable values, with the target criterion to maximize the quality of prediction of the (transformed) Y variable values. One such scatterplot smoother is the [cubic smoothing splines](#) smoother, which generally produces a smooth generalization of the relationship between the two variables in the scatterplot. Computational details regarding this smoother can be found in Hastie and Tibshirani (1990; see also Schimek, 2000).

To summarize, instead of estimating single parameters (like the regression weights in [multiple regression](#)), in generalized additive models, we find a general unspecific (non-parametric) function that relates the predicted (transformed) Y values to the predictor values.

A Specific Example: The Generalized Additive Logistic Model

Let us consider a specific example of the generalized additive models: A generalization of the logistic (logit) model for binary dependent variable values. As also described in detail in the context of [Nonlinear Estimation](#) and [Generalized Linear/Nonlinear Models](#), the logistic regression model for binary responses can be written as follows:

$$y = \exp(b_0 + b_1 \cdot x_1 + \dots + b_m \cdot x_m) / \{1 + \exp(b_0 + b_1 \cdot x_1 + \dots + b_m \cdot x_m)\}$$

Note that the distribution of the dependent variable is assumed to be binomial, i.e., the response variable can only assume the values 0 or 1 (e.g., in a market research study, the purchasing decision would be binomial: The customer either did or did not make a particular purchase). We can apply the logistic link function to the probability p (ranging between 0 and 1) so that:

$$p' = \log \{p/(1-p)\}$$

By applying the logistic link function, we can now rewrite the model as:

$$p' = b_0 + b_1 \cdot X_1 + \dots + b_m \cdot X_m$$

Finally, we substitute the simple single-parameter additive terms to derive the generalized additive logistic model:

$$p' = b_0 + f_1(X_1) + \dots + f_m(X_m)$$

An example application of this model can be found in Hastie and Tibshirani (1990).

Fitting Generalized Additive Models

Detailed descriptions of how [generalized additive models](#) are fit to data can be found in Hastie and Tibshirani (1990), as well as Schimek (2000, p. 300). In general there are two separate iterative operations involved in the algorithm, which are usually labeled the *outer* and *inner* loop. The purpose of the outer loop is to maximize the overall fit of the model, by minimizing the overall likelihood of the data given the model (similar to the [maximum likelihood](#) estimation procedures as described in, for example, the context of Nonlinear Estimation). The purpose of the inner loop is to refine the [scatterplot smoother](#), which is the [cubic splines smoother](#). The smoothing is performed with respect to the [partial residuals](#); i.e., for every predictor k , the weighted cubic spline fit is found that best represents the relationship between variable k and

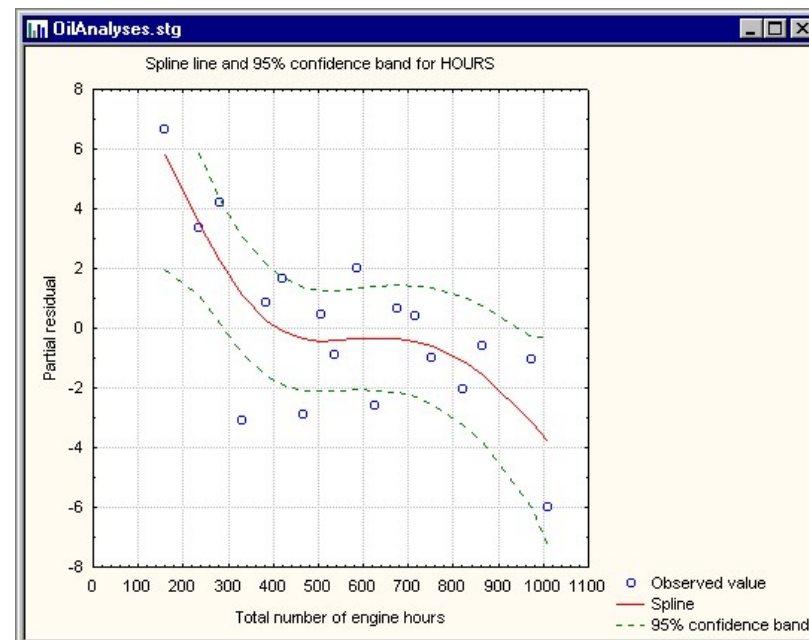
the (partial) residuals computed by removing the effect of all other j predictors ($j \neq k$). The iterative estimation procedure will terminate, when the likelihood of the data given the model can not be improved.

[To index](#)

Interpreting the Results

Many of the standard results statistics computed by *Generalized Additive Models* are similar to those customarily reported by linear or nonlinear model fitting procedures. For example, predicted and [residual](#) values for the final model can be computed, and various graphs of the residuals can be displayed to help the user identify possible [outliers](#), etc. Refer also to the description of the residual statistics computed by [Generalized Linear/Nonlinear Models](#) for details.

The main result of interest, of course, is how the predictors are related to the dependent variable. [Scatterplots](#) can be computed showing the smoothed predictor variable values plotted against the [partial residuals](#), i.e., the residuals after removing the effect of all other predictor variables.



This plot allows you to evaluate the nature of the relationship between the predictor with the residualized (adjusted) dependent variable values (see Hastie & Tibshirani, 1990; in particular formula 6.3), and hence the nature of the influence of the respective predictor in the overall model.

Degrees of Freedom

To reiterate, the [generalized additive models](#) approach replaces the simple products of (estimated) parameter values times the predictor values with a [cubic spline smoother](#) for each predictor. When estimating a single parameter value, we lose one degree of freedom, i.e., we add one degree of freedom to the overall model. It is not clear how many degrees of freedom are lost due to estimating the cubic spline smoother for each variable. Intuitively, a smoother can either be very smooth, not following the pattern of data in the scatterplot very closely, or it can be less smooth, following the pattern of the data more closely. In the most extreme case, a

simple line would be very smooth, and require us to estimate a single slope parameter, i.e., we would use one degree of freedom to fit the smoother (simple straight line); on the other hand, we could force a very "non-smooth" line to connect each actual data point, in which case we could "use-up" approximately as many degrees of freedom as there are points in the plot. Generalized Additive Models allows you to specify the degrees of freedom for the cubic spline smoother; the fewer degrees of freedom you specify, the smoother is the cubic spline fit to the partial residuals, and typically, the worse is the overall fit of the model. The issue of degrees of freedom for smoothers is discussed in detail in Hastie and Tibshirani (1990).

A word of Caution

[Generalized additive models](#) are very flexible, and can provide an excellent fit in the presence of nonlinear relationships and significant noise in the predictor variables. However, note that because of this flexibility, you must be extra cautious not to over-fit the data, i.e., apply an overly complex model (with many degrees of freedom) to data so as to produce a good fit that likely will not replicate in subsequent validation studies. Also, compare the quality of the fit obtained from *Generalized Additive Models* to the fit obtained via [Generalized Linear/Nonlinear Models](#). In other words, evaluate whether the added complexity (generality) of generalized additive models (regression smoothers) is necessary in order to obtain a satisfactory fit to the data. Often, this is not the case, and given a comparable fit of the models, the simpler generalized linear model is preferable to the more complex generalized additive model. These issues are discussed in greater detail in Hastie and Tibshirani (1990).

Another issue to keep in mind pertains to the interpretability of results obtained from [\(generalized\) linear models](#) vs. generalized additive models. Linear models are easily understood, summarized, and communicated to others (e.g., in technical reports). Moreover, parameter estimates can be used to predict or classify new cases in a simple and straightforward manner. Generalized additive models are not easily interpreted, in particular when they involve complex nonlinear effects of some or all of the predictor variables (and, of course, it is in those instances where generalized additive models may yield a better fit than generalized linear models). To reiterate, it is usually preferable to rely on a simple well understood model for predicting future cases, than on a complex model that is difficult to interpret and summarize.

[To index](#)