Pakistan Journal of Social Sciences (PJSS)
Vol. 31, No. 2 (December 2011), pp. 423-432

Nonparametric Regression Estimation for Nonlinear Systems: A Case Study of Sigmoidal Growths

Atif Akbar
Department of Statistics,
Bahauddin Zakariya University, Multan, Pakistan
E-mail: atifakbar@bzu.edu.pk

Muhammad Aman Ullah
Department of Statistics,
Bahauddin Zakariya University, Multan, Pakistan
E-mail: aman_stat@bzu.edu.pk

Abstract

Sigmoidal growths are well approximated by the non-linear sigmoidal growth models including Richards (1959) Morgan et al (1975), Davies and Ku (1977) and Muller et al (2006) among many others. This article deals with the comparison of the nonparametric regression with the non-linear regression models in order to locate the better approximation for sigmoidal growths. To unwind the standard assumptions, nonparametric regression estimation is used for data analysis, which enables us to look at the data more flexibly, uncovering structure in the data that might missed otherwise. To consider noisy data, observed at certain time or design points, an effort has been made to approximate the relationships without defining the parametric functional form, and to use the function that bears the ability to deal the situation with mild assumptions. The assumptions of smoothness and simplicity are made to make up for usual assumptions of regression models. Kernel estimation and k-nearest neighbor estimation (Hardle, 1990; Takezawa, 2006) is used and the analysis shows the evidence in favor of nonparametric regression estimation (k-nearest neighbor estimation) with the comparison of non-linear sigmoidal growth models.

Keywords: Intrinsic non-linearity; Kernel estimation; Non-linear models; Nonparametric regression; Parameter effect non-linearity; Smoothing; Sigmoidal growth models

I. Introduction

Regression analysis plays an imperative role in Statistics because of its most potent and commonly used techniques. It is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data that is used to deal with problems of locating suitable models to represent relationship between a response variable and a set of explanatory variables based on data collected from a series of experiments. Regression models are used to represent existing data and also to predict new observations. Two types of such models are prevailing, linear and nonlinear. Usually nonlinear models provide better approximation to switch real world phenomena. Different nonlinear models have been established including, logistic, gompert and Weibull, see Weibull (1951), Richards (1959), Holiday (1960), Farazdaghi and Harris (1968), Morgan et al (1975), Davies and Ku (1977) and Muller et al (2006), etc.

There exist many systems in which the representation of data form a curve that starts from a point and increase steadily and then at a greater rate than at the early stage and then relatively in lesser rate and slowly get to an asymptote. These curves are referred as S-shaped or sigmoidal growth curves. The applications of such relationships found in agriculture, biology, economics and engineering. These growth curves include logistic, gompertz, Weibull type and can be found with Weibull (1951), Richards (1959), Morgan et al (1975), etc. For detailed discussion of nonlinear models, intrinsic nonlinearity, parameter effect nonlinearity, curvature measures and their estimation methods, see Ratkowsky (1983) and Seber and Wild (1989).

There is substantial literature regarding the efforts made in the field of nonlinear regression. To measure the nonlinearity also to right of entry the adequacy of regression models with their estimation, certain work is available (see e.g., Beale, 1960; Guttman and Meeter, 1965). Bates and Wates (1980) presented measure of nonlinearity about the geometric behavior of the curvature. They have found two components of nonlinearity i.e., intrinsic nonlinearity (*IN*) and parameter effect (*PE*) nonlinearity. Bates and Wates (1980) examined the work of Beale (1960) and Box (1971) and showed that Beale's measure generally tend to underestimate the true nonlinearity, but the bias measure of Box is closely related to the parameter effect nonlinearity.

Birch (1999) presented a sigmoidal growth model based on logistic model, known as generalized logistic sigmoid growth equation. He claimed that the proposed model is unique to deal growth at small sizes or low densities are expected to be approximately exponential, also the maximum slope of the growth curve can be at any value. He established a comparison between his proposed model and Richard model and shown that the proposed model described the sequences of points much closely to the curves form which they are generated also it always gave closer fits and more correct estimates of the characteristics of the conventional sigmoid curves and the Richard model. He also concluded that the proposed model is more apposite when the exponential is assumed for initial growth, when estimates of maximum relative growth are required, or for generic growth simulations.

Yin et al. (2003) proposed a sigmoid function called as the beta growth function and compared it with logistic, Richards, Gompertz and Weibull growth models and two expolinear models. They shown that the beta model is flexible in describing various asymmetrical sigmoid patterns which is the property of Richards model, its parameters are numerically stable in statistical model which is the characteristic of Gompertz and logistic model, it predicts zero mass at time zero, that Weibull bears, but it also deal with various initial conditions that can be easily obtained additionally it provides reasonable estimates of final quantity and duration of a growth process. They claimed that their proposed model is unique for dealing with determinate growth, and is more suitable than other functions for establishing in process-based crop simulation models that are used to represent the characteristics of the system.

Muller et al. (2006) presented a new approach to describe and estimate organ area indices (AI). They shown that the proposed model provide reliable predictions of time of organ AI and it found better in the sense that non-asymptotic pattern at minimum and maximum, high flexibility, unique interpretability and numeric stability of all the parameters and simplicity of parameterization based on measurement data.

Smoothing regression curve and obtaining smooth curve for density is an issue of deep interest. The initial performance concerning smoothing in time series is seen in Macauley (1931). The number of nonparametric regression methods is developed that are used to obtain the estimate of unknown regression function in a smooth manner (Härdle, 1990).

Priestly and Chao (1970, 1972) considered the problem of estimating the unknown regression function *g(x)* given observations at a fixed set of points. Their estimate is nonparametric in the sense that *g(x)* is restricted only by certain smoothing requirements. The weights of moving averages of sample realizations are based on a class of kernels suggested by Rosenblatt (1956) and Parzon (1962). To extend the work of Priestly and Chao (1972), Benedetti (1977) reviewed the results of Priestly and Chao (1970, 1972) and also considered further properties, i.e. asymptotic normality, convergence of *g(x)* and choices of kernel. Benedetti (1977) provided the optimal class of kernels in the sense of optimizing an asymptotic expression for the mean square error. Cleveland (1979) and Cleveland and Devlin (1988) provided the robust version of nonparametric regression. Hardle and Mammen (1993) provided the grounds for the comparison between parametric and nonparametric regression and offered that squared deviation may be used to identify the appropriateness of parametric or nonparametric fit. Furthermore they proposed the wild bootstrap method, following Wu (1986), for comparison after showing the failure of standard bootstrap method, see also Akbar et el. (2010) for comparison of nonlinear and nonparametric regression towards yield-density relationships. For detailed discussion of nonparametric regression see Hardle (1990), Takezawa (2006) among many others.

Stone (1977) defined the consistency of nonparametric regression estimators and also provided the regularity conditions upon which the consistency is attained. He defined the consistency criteria for nonparametric regression estimators and established the consistency for many nonparametric regression estimators including nearest neighbor and local polynomial estimators. The applications of these results are also studied in density and Bayes risk estimators.

Müller (1987) has studied theoretical relationships of locally weighted regression and has shown that locally weighted regression is asymptotically equivalent to the kernel estimators and discussed that these estimators are approximately equal in finite sample applications. He also discussed that there is close correspondence between the orders of polynomial used in locally weighted regression and the order of the kernel used in the kernel regression. The correspondence between the shapes and the weight function is also observed there in.

k-NN nonparametric estimator was initially proposed by Royall (1966) and The consistency of *k*-nearest neighbor regression estimator is calculated by Devroye et al. (1994) and have shown that, in probability, almost sure and complete convergence are equivalent for bounded regression variate. They also provided the strong uniform consistency for *k*-nearest neighbor regression estimator under regularity conditions.

Priestly and Chao (1972) proposed a nonparametric regression estimator for both equally and not equally spaced design points. They also derived approximate expressions for mean and variance of their proposed estimator and also showed that their proposed estimator is consistent. They considered different weight schemes and while handling the nonparametric regression, they have inferred that the weight function used in the regression estimator has less critical effect on the estimator of unknown function but the selection of bandwidth has a serious effect on the properties and performance of the estimators of unknown functions.

II. Parametric Estimation of Sigmoidal Growths

Four Data Sets are used in the current study, taken from Ratkowsky (1983). The data sets are Pasture Regrwoth, Onion Bulbs Plus Tops, Cocumber Cotyledons and Been Root Cells. Following are the nonlinear models used in the study and compared with nonparametric regression fits.

$$Y = \alpha \exp(-\exp(\beta - \gamma X)) \quad (1)$$

$$Y = \frac{\alpha}{1 + \exp(\beta - \gamma X)} \quad (2)$$

$$Y = \frac{\alpha}{[1 + \exp(\beta - \gamma X)]^{1/\delta}} \quad (3)$$

$$Y = \frac{\beta\gamma + \alpha X^{\delta}}{\gamma + X^{\delta}} \quad (4)$$

$$Y = \alpha - \beta \exp(-\gamma X^{\delta}) \quad (5)$$

Equation 1 - 5 are Gompertz, Logistic, Richards (1959), Morgan-Mercer-Flodin (MMF), and Weibull (1951) type model respectively.

For estimation of the above mentioned nonlinear models, Gauss-Newton Method is adopted (see Ratkowsky, 1983 and Draper and Smith, 1998, for more details). Bates and Wates (1980) and Box (1971) selected the suitable models on the basis of least standard error, high R^2, insignificant IN, PE and bias.

Following Box (1971) and Bates and Wates (1980), the IN and PE measures are established as well as bias calculation is carried out for model selection. To test the significance of the IN and PE the relation $\frac{1}{2\sqrt{F}}$ is used, where F is the critical value obtained from the F-distribution table. Parameter estimates for above mentioned models with four data sets can be found in Table 4.1 page 65 provided by Ratkowsky (1983). The squared deviations for above mentioned models for four data sets are given in Table 1 - 4.

III. Nonparametric Estimation of Sigmoidal Growths

For nonparametric regression estimation two approaches are used for curve fits, i.e. Kernal smoothing and nearest neighborhood (NN) estimates as presented under.
The relationship between different variables represented in a form of regression model and the general nonparametric regression model is;

$$Y = m(x) + U \qquad (6)$$

where *Y* is response variable, *m*(*x*) is unknown, nebulous and assumed smooth function of explanatory variable *x*, also $m(x) = E(Y|X{=}x)$ with $E(Y) < \infty$ and *U* is the residual. An estimate of *m*(*x*) is found in terms of weighted sums in many estimation approaches. The error term having zero mean, constant variance and normality.

By using sample data, the nonparametric regression estimation is the estimation of unknown regression function, *m*(*x*), by considering the population mean of *Y* when *X* = *x*. Generally, the nonparametric regression estimators are the weighted mean of y_i as given by;

$$\hat{m}(x) = \sum_{i=1}^{n} w_{ni}(x) yi \qquad (7)$$

where, $w_{ni}(x) = w_{ni}(x_i, x)$ is the weight allocated to the *i*th observation y_i. The allocation of weight depends upon the distance of x_i from the point *x*. Typically, large weight is allocated to the small distances and small weights for large distances to make balance the effect of large distant values. These weights are called probability weights with following properties (Hardle, 1990).

(i) $w_{ni}(x) \geq 0$ and

(ii) $\sum w_{ni}(x) = 1$.

a) *Nadaraya-Watson* (*NW*) *Kernel Estimator*

Nadaraya (1964) and Watson (1964) proposed kernel weights $w_{ni}(x)$ for Eq. (7) above, and the book-length treatment concerning theoretical and applied aspects is given in Härdle (1990).
The relation for $w_{ni}(x)$ is given by;

$$w_{ni}(x) = K\left(\frac{x_i - x}{h_n}\right) \Big/ \hat{f}(x),$$

with $\hat{f}(x) = \sum_{i=1}^{n} K\left(\frac{x_i - x}{h_n}\right)$

and then the resultant estimator $\hat{m}(x)$ is given by;

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h_n}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h_n}\right)} \quad (8)$$

where $K(.)$ is a kernel and is a continuous bounded function which sums to 1 with a bandwidth $h_n \to 0$ as $n \to \infty$. The bandwidth (h_n) or smoothing parameter encounters the roughness of the regression line while simultaneous minimizing the variance and bias usually referred as bias-variance tradeoff. Many methods are available to calculate bandwidth (see Rice, 1984; Silverman, 1986 for details).

b) *k- Nearest Neighbour* (*NN*) *Estimator*

Royall (1966) proposed *k*-nearest neighbour estimator while taking into account the Euclidian distance between design points and its *k*-th neighbour among *x*'s, i.e. h_n having $h_n \to 0$ as $n \to \infty$. The resultant estimator for $w_{ni}(x)$ is given by;

$$w_{ni}(x) = K\left(\frac{x_i - x}{k_n}\right) \Big/ \hat{f}(x), \quad (9)$$

$$\hat{f}(x) = \sum_{j=1}^{n} K\left(\frac{x_j - x}{k_n}\right)$$

the k_n is the Euclidian distance between x and its k-th neighbor among the x_i and k_n satisfies $k_n \to 0$ as $n \to \infty$. The k is the bandwidth and has the same role as in Nadaraya-Watson estimator.

The squared deviation for the above mentioned nonparametric estimation procedures are given in Table 1-4 for four data sets of sigmoidal growths.

Table 1: Comparison of Squared Deviations for Pasture Yield Data

Squared Deviations for Parametric and Nonparametric Approaches						
Parametric Models					Nonparametric Approaches	
Gompertz	*Logistic*	*Richards*	*MMF*	*Weibull Type*	*Kernel*	*NN*
9445.676	9342.816	15768.71	9355.972	9325.459	5543.296	0.000

Table 2: Comparison of Squared Deviations for Onion Bulbs Plus Tops Data

Squared Deviations for Parametric and Nonparametric Approaches						
Parametric Models					Nonparametric Approaches	
Gompertz	*Logistic*	*Richards*	*MMF*	*Weibull Type*	*Kernel*	*NN*
3764146	8930.712062	236114.5	11165.44	7834.526	18794.17	5391.151

Table 3: Comparison of Squared Deviations for Cucumber Cotyledons Yield Data

Squared Deviations for Parametric and Nonparametric Approaches						
Parametric Models					**Nonparametric Approaches**	
Gompertz	***Logistic***	***Richards***	***MMF***	***Weibull Type***	***Kernel***	***NN***
0.371689	0.211932	0.301072	0.0238	0.133839	0.457964	0.0000

Table 4: Comparison of Squared Deviations for Been Root Cells Yield Data

Squared Deviations for Parametric and Nonparametric Approaches						
Parametric Models					**Nonparametric Approaches**	
Gompertz	***Logistic***	***Richards***	***MMF***	***Weibull Type***	***Kernel***	***NN***
12.59052	6.210316	610.582	6.36995	5.446859483	15.37105	3.965548

IV. Discussion and Conclusion

The main purpose of the article is to evaluate the performance of the parametric and nonparametric regression estimation techniques. Moreover, nonparametric regression estimation methods, i.e. kernel estimation and nearest neighbor estimation are also compared. Certain work has been done but we choose different add ups with parametric and nonparametric regression estimation and observed the upshots of these linkages.

Initially we considered the nonlinear data and applied the nonlinear modeling and then estimation of these nonlinear models as suggested and given in Ratkowsky (1983). We than applied the nonparametric kernel estimation and nearest neighbor estimation to these data and come to the conclusions.

Thorough examination of the above tables signifies that the nonparametric regression approach is more suitable to deal with the sigmoidal growths. The *k*-nearest neighbor nonparametric estimation is more credible because it estimates the values that are very close to the actual values and in some cases (Table 1 and Table 3), the estimated values are exactly same as the actual values that results in the zero squared deviation of actual and estimated values.

Irrespective of rigid assumptions for the regression model, we used less restrictive nonparametric regression approach for the regression model and come to the conclusion that nonparametric regression estimation is suitable in the considered case of sigmoidal growths. So the nonparametric regression estimation appears more suitable in the cases discussed above for the sigmoidal growths.

References

Akbar,A., Pasha, G.R. and Aslam, M. (2010). Yield Density Rapporates: A Nonparametric regression Approach. *International Research Journal of Finance and Economics,* Vol. 43, Issue 1.

Bates, D. M. and Watts, D. G. (1980). Relative Curvature Measure of Nonlinearity. *J.R. Statist. Soc. Ser. B*, 42, pp.1-25.

Beale, E. M. L. (1960). Confidence Region in Non-Linear Estimation. *J. R. Stat. Soc. Ser. B*, Vol.22, pp.41-76.

Benedetti, J. K. (1977). On the Nonparametric Regression Estimation of Regression Functions. *J. R. Stat. Soc. Ser. B* 39 (2), 248-253.

Birch, C. P. D. (1999). A New Generalized Loistic Sigmoid Growth Equation Campred with the Richards Growth Equation. *Annals of Botany*, 83, 713 - 723.

Bleasdale, J. K. A. and Nelder, J. A. (1960). Plant Population and Crop Yield. *Nature,* 188, 342.

Box, M. J. (1971). Bias in Nonlinear Estimation. *J. R. Stat. Soc. Ser. B*, Vol.33, No.2, pp. 171-201.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *JASA*, Vol. 74, No. 368, 829-836.

Cleveland, W. S. and Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *JASA*, Vol. 83. No. 403, 596-610.

Davies, O. L. and Ku, J. Y. (1977). Re-examination of the Fitting of the Richards Growth Function. *Biometrics*, 33, 546-547.

Devroye, L., Gyorfi, L., Krzyzak, A., and Lugosi, G. (1994). On the Strong Universal Consistency of Nearest Neighbor Regression Function Estimates. Annals of Statistics, 22:1371–1385.

Farazdaghi, H. and Harris, P. M. (1968). Plant Competition and Crop Yield. *Nature,* 217, 289-290.

Gillis, P. R. and Ratkowsky, D. A. (1978). The Behaviour of the Estimators of the Parameters of Various Yield-Density Relationship. *Biometrics*, 34, 191-198.

Guttman, I. and Meeter, D. A. (1965). On Beale's Measures of Non-Linearity. *Technometrics*, Vol.7, pp. 623-637.

Hardle, W. (1990). Applied Nonparametric Regression. Cambridge University Press.

Hardle, W. and Mammen, E. (1993). Comparing Nonparametric Versus Parametric Regression Fits. *Annals of Statistics*, Vol., 21, No. 4, 1926-1947.

Hartley, H.O. (1961). The Modification of Gauss-Newton Method for the Fitting of Nonlinear Regression Function by Least Squares. *Technometrics*, Vol.3, pp. 269-280.

Holiday, R. (1960). Plant Population and Crop Yield. *Field Crop Abstr*., 13, 159-167, 247-254.

Loftsgaarden, D. O. and Quesenberry, G. P. (1965). A Nonparametric Estimate of a Multivariate Density Function. *Annals of Mathematical Statistics*, 36, 1049-1051.

Macauley, R.R. (1931). The Smoothing of Time Series. National Bureau of Economic Research, New York.

Morgan, P. H, Mercer, L. P., and Flodin, N. W. (1975). General Model for Nutritional Responses of Higher Organisms, Proc. Nat. Acad. Sci. U. S. A. 72, 4327-4331.

Muller, H. G. (1987). Weighted Local Regression and Kernel Methods for Nonparametric Curve Fitting. *JASA*, 82, 231-238.

Muller, J., Beherns, T. and Diepenbrock, W. (2006). Use of a New Sigmoid Growth Equation to Estimate Organ Area Canopy Area Index in Winter Oilseed Rape (Brassica Napus L.). *Field Crop Research*.

Nadaraya, E.A. (1964). On estimating regression. Theory Probe Appl., 9, 141-142.

Priestly, M. B. and Chao, M. T. (1970). Nonparametric Function Estimation. Department of Mathematics, Technical Report No. 21, University of Manchester.

Priestly, M. B. and Chao, M. T. (1972). Nonparametric Function Fitting. *J. R. Statist. Soc., B*, 34, 385-392.

Ratkowsky, D. (1983). Nolinear Regression Modeling. Marcel Dekker.

Rice. J (1984). Bandwidth choice for nonparametric regression. The Annals of Statistics, 12:1215-1230.

Richards, F. J. (1959). A Flexible Growth Function for Empirical Use. *J. Exp. Biol.,* 10, 290-300

Rossenblat, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.,* 27, 832-837.

Royall, R. M. (1966). A Class of Nonparametric Estimators of a Smooth Regression Function. Ph.D. Thesis, Stanford University.

Seber, G. A. F. and Wild, C. J. (1989). Nonlinear Regression. Wiley New York.

Silverman, B.W. (1985). Some aspects of the spline smoothing approach to onparametric regression curve fitting (with discussion). J. Royal Statist. Soc. B, 47, 1-52.

Stone, C.J. (1977). Consistent Nonparametric Regression. Ann. Statist., 5, 595-645.

Takezawa, K. (2006). Introduction to Nonparametric Regression. Wiley Interscience.

Watson, G.S. (1964). Smooth Regression Analysis. Sankhya, Series A, 26, 359-372.

Wiebull, W. (1951). A Statistical Distribution Function of Wide Applicability, *J. Appl. Mech.,* 18, 293-296.

Yin, X. Goudriaan, J. Lantinga, E. A. Vos, J. and Spiertz, H. J. (2003). A Flexible Function of Determinate Growth. *Annals of Botany*, 91, 361-371.

Wu, C. F. J. (1986). Jacknife, Bootstrap and other Resampling Methods in Regression Analysis (with discussion). *Annals of Statistics*, 14, 1261-1350.

(see Loftsgaarden and Quesenberry, 1965; Muller, 1987; Hardle, 1990, among many others).