Nonparametric Trend Analysis

C.T. Jose
Central Plantation Crops Research Institute
Regional Station, Vittal-574 243
Karnataka, INDIA
Email:ctjos@yahoo.com

Abstract

The use of nonparametric techniques has a long tradition in time series analysis. Running mean, a very simple type of smoother has been used since the early 1900's for determining trends in time series. The increased data availability and the explosion of computing power have made it possible to use a wide range of other modern nonparametric techniques in time series analysis recently. The nonparametric estimation of trend and growth rate has been discussed in this paper. Kernel weighted linear regression method has been proposed to estimate the trend and growth rate nonparametrically. The method is applied to the data of area and production of arecanut in India.

1. Introduction

In recent years nonparametric techniques for functional estimation have become increasingly popular as tools for data analysis. These techniques impose only few assumptions about the shape of the function and therefore the nonparametric methods are more flexible than the usual parametric approaches. Smoothing techniques are commonly used to estimate the functon nonparametrically. The use of nonparametric techniques has a long tradition in time series analysis. As early as the late 19th century Schuster (1898) introduced the periodogram which may be regarded as the origin of spectral analysis. Running mean, a very simple type of smoother have been used since the early 1900's for determining trends in time series (Whittaker, 1923). The increased data availability and the explosion of computing power have made it possible to use a wide range of other modern nonparametric techniques in time series analysis recently. The nonparametric estimation of trend and growth rate has been discussed in this paper. The trend analysis of area and production of arecanut in India have been carried out based on the proposed procedure.

2. Model Settings and Estimators

2.1 Trend:

The time series model considered is of the form

$$y_i=m(t_i)+\varepsilon_i,\ t_i=i/n,\ i=1,\dots,n \tag{1}$$

Where y_i, is the observation of the i^{th} time interval, m(.) is the trend function, which is assumed to be smooth and ε_i are random errors with mean zero and finite variance. The kernel weighted linear regression smoother (Fan, 1992) has been used to estimate the trend function nonparametrically. Under the assumption that the second derivative of m(t) exists, the value of the local linear regression smoother is the solution of a_0 to the following weighted least squares problem.

$$\sum_{i=1}^{n}\left[y_i-a_0-a_1(t_i-t)\right]^2 K\left(\frac{t-t_i}{h}\right) \tag{2}$$

where K is a bounded symmetric kernel density function and h is the bandwidth. Let $\hat{a}_0$ and $\hat{a}_1$ be the solution to the weighted least squares problem (2). The estimate of the trend function m(t) is given by

$$\hat{m}(t)=\hat{a}_0=\frac{\sum_{j=1}^{n} w_{tj}y_j}{\sum_{j=1}^{n} w_{tj}} \tag{3}$$

where, $$w_{tj}=K\left(\frac{t-t_j}{h}\right)\left\{\sum_{i=1}^{n}K\left(\frac{t-t_i}{h}\right)(t-t_i)^2-(t-t_j)\sum_{i=1}^{n}K\left(\frac{t-t_i}{h}\right)(t-t_i)\right\}$$

The slope m'(t) of m(t) at t has been considered as the simple growth rate at the time point t. The estimate of m'(t) is given by

$$\hat{m}'(t) = \hat{a}_1 = \frac{\sum_{j=1}^{n} w'_{tj}\, y_j}{\sum_{j=1}^{n} w'_{tj}} \tag{4}$$

where, $$w'_{tj} = K\left(\frac{t-t_j}{h}\right)\left\{(t-t_j)\sum_{i=1}^{n} K\left(\frac{t-t_i}{h}\right) - \sum_{i=1}^{n} K\left(\frac{t-t_i}{h}\right)(t-t_i)\right\}$$

Under some regularity conditions on the bandwidth, the consistency and asymptotic normality of the abve estimators can be proved (Jose, 1999 and Jose and Ismail, 2000). In the above method, we assume that the trend function m is smooth. Estimation of m in the case of discontinuities in m and/or its derivative is discussed by various authors (Jose and Ismail, 1997&1999; Loader, 1996; Muller, 1992).

Chice of bandwidth: The choice of bandwidth is of great importance in nonparametic regression estimation. The cross-validation or the leave-one-out method is commonly used to estimate the optimum bandwidth. The cross validaton score, CV(h) is given by

$$CV(h) = n^{-1}\sum\left[y_j - \hat{m}_{h,j}(t_j)\right]^2$$

where $\hat{m}_{h,j}(t_j)$ is the estimate of $m(t_j)$ without using the jth observation. Then the optmum bandwidth is the value of h, which minimizes CV(h). Note that, the ordinary cross-validation tends to select a bandwidth many order of magnitude too small, if the data are highly positvely correlated. Hart (1994), proposed a method called time series cross-validation for selecting the bandwidth and Jose (1999) proposed a method based on the residual analysis to obtain the optimum bandwidth under the correlated error situations.

2.2 Relative growth rate:

Consider a series of independent random impulses occuring in the order $\xi_1, \xi_2, \ldots, \xi_n$, the effect of ξ_t being to increase the momentary size of the time series observations y_t to y_{t+1} by the relation

$$y_{t+1} = y_t + \xi_t y_t$$
$$= (1+\xi_t) y_t$$
$$= \eta_t y_t$$

Or $$z_t = \log(\eta_t) = \log(y_{t+1}) - \log(y_t)$$

In growh rate studies we usually assume that η_t follows a lognormal distribution, then z_t can be written as

$$z_t = g(t) + \varepsilon_t \tag{5}$$

where $E(z_t)=g(t)$ is assumed to be smooth and ε_t are normally distributed with mean 0 and variance $\sigma^2<\infty$. The function g can be estimated in a similar way as that of the trend functon estimate provided in Section 2.1.

$$\hat{g}(t) = \frac{\sum_{j=1}^{n} w_{tj} z_j}{\sum_{j=1}^{n} w_{tj}}$$

where, $$w_{tj} = K\left(\frac{t-t_j}{h}\right)\left\{\sum_{i=1}^{n} K\left(\frac{t-t_i}{h}\right)(t-t_i)^2 - (t-t_j)\sum_{i=1}^{n} K\left(\frac{t-t_i}{h}\right)(t-t_i)\right\}$$

To obtain an estimate of the variance of $\hat{g}(t)$, first we have to get an estimate of σ^2 which is given by

$$\hat{\sigma}^2 = \frac{1}{2(n-1-2q)}\sum_{i=2+q}^{n-q} \psi_i$$

where, ψ_i, $i=1,2,\ldots,n$ are the rearranged $(z_i-z_{i-1})^2$ and ψ_i are in ascending order. The number q is determined by the expermenter. Then

$$V\left(\hat{g}(t)\right) = \hat{\sigma}^2 \frac{\sum_{j=1}^{n} w_{tj}^{\,2}}{\left(\sum_{j=1}^{n} w_{tj}\right)^2}$$

Since $\hat{g}(t)$ is normally distributed, the confidence interval for $\hat{g}(t)$ can be worked out.

The estimate of the relative growth rate at time t is given by

$$\hat{\xi}_t = Anti\log\left(\hat{g}(t)\right) - 1$$

Note that in the parametric compound growth rate analysis, the function g in (5) is assumed to be a constant and the estimate of the compound growth rate is obtained by the method of least squares. In the present study we assume that the relative growth rate is a smooth function and obtained the estimate of the relative growth rate function nonparametricaly.

3. Trend Analysis of Area and Production of Arecanut in India

The proposed method is applied to the yearly data of area and production of arecanut in India given by the Directorate of Economics and Statistics from 1967-68 to 1997-98 to study the trend and growth rate. The estimated trend functions of area and production of arecanut in India obtained using the proposed method (Section 2.1) are given in Fig.1 and Fig.2 respectively. The simple growth rate functons (slope of the trend function) are given in Fig.3. The relative growth rate fuctions of area and production of arecanut obtained using the proposed method (Section 2.2) is given in Fig.4. The trend functions of area and production fitted extremly well to the data (Fig. 1&2.). The simple growth rate of area is decreasing upto the year 1979 and then it is increasing smoothly, whereas in the case of production, the simple growth rate is more or less same upto 1975 and afterwards it is increasing smoothly (Fig.3). The relative growth rate of both area and production declined initially from 1967 and showed an increase after 1980 (Fig.4).

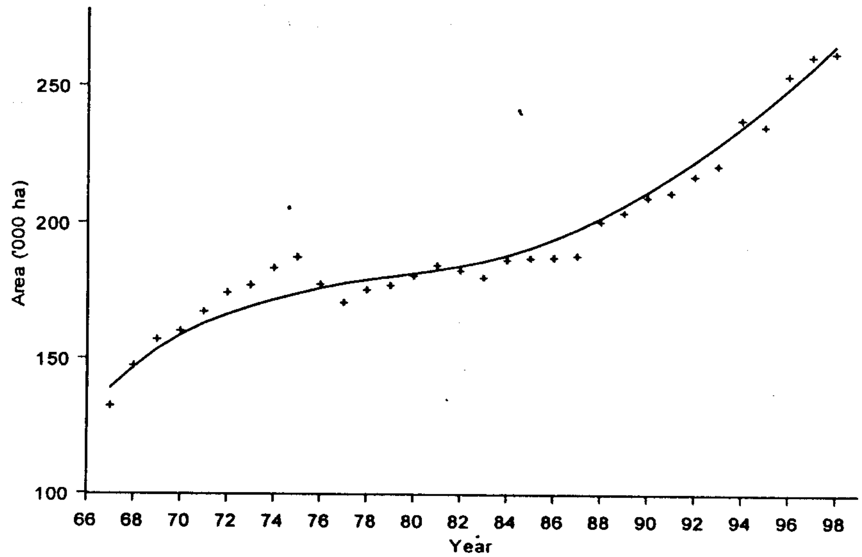

Fig. 1. The plus signs denote the area of arecanut in India given by *the Directrate of Econmics and Statistics* and the smooth curve is the estimated trend function.

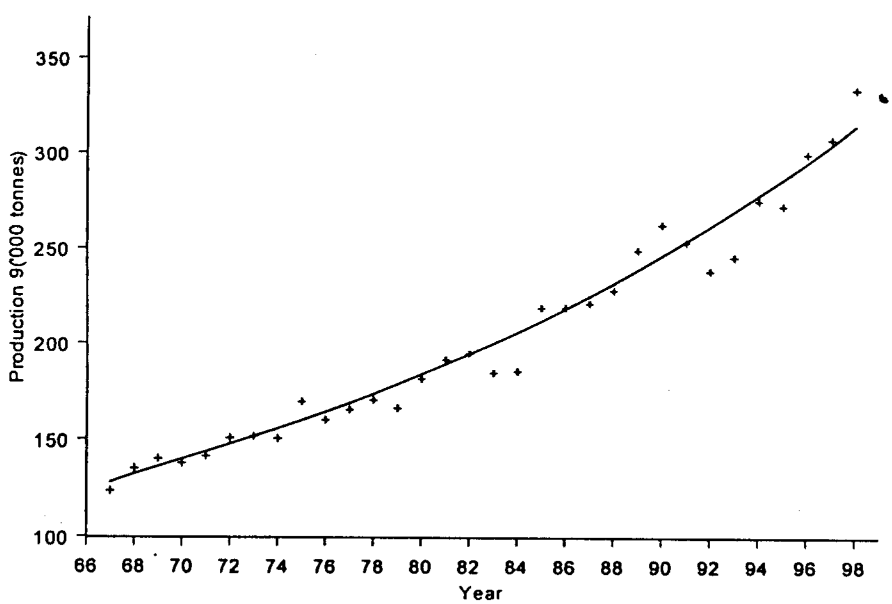

Fig. 2. The plus signs denote the production of arecanut in India given by *the Directrate of Econmics and Statistics* and the smooth curve is the estimated trend function.

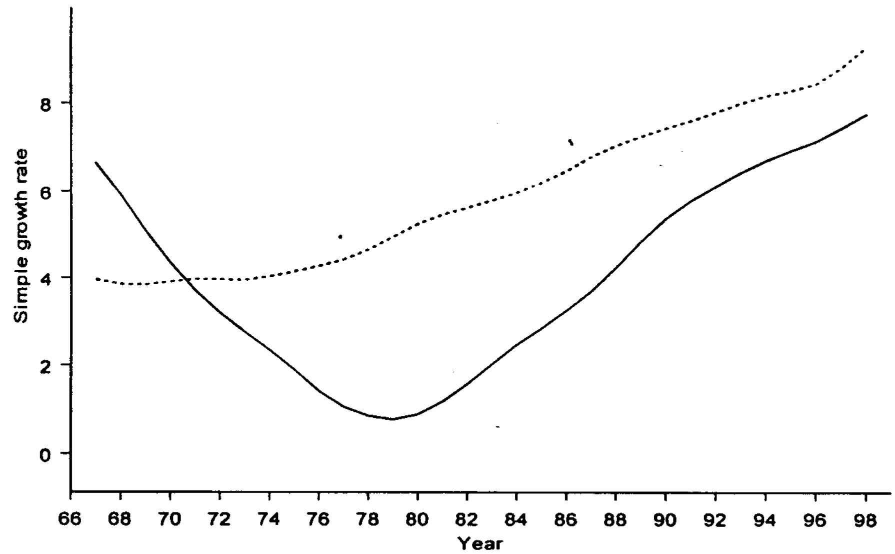

Fig.3. The simple growth rate of area (solid line) and production (dotted line) of arecanut in India

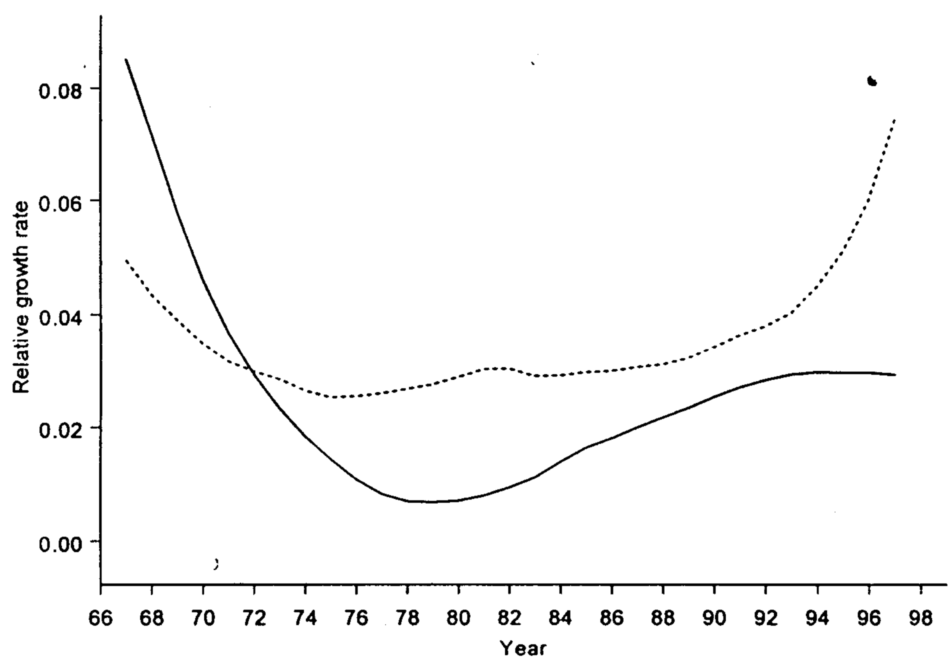

Fig.4. The relative growth rate of area (solid line) and production (dotted line) of arecanut in India

4. Conclusion.

The nonparametric approach used in this paper for trend analysis is more flexible than the usual parametric methods. In parametric approaches to study the trend and growth rate, we assume some parametric functional form and the inferences are drawing based on the estimated parameters. In the present nonarametric approach, the only assumption about the form of the trend and growth rate functions are that they are smooth. The local changes in trend and growth rate can be studied using the present study whereas, in the parametric approach , we are estimating the paramaters for an interval and that will not reflect the local changes.

References

Fan, J.(1992). Design-adaptive nonparametric regression, *Journal of the American Statistical Association*, **87**, 998-1004.

Hart, J.D.(1994). Automated kernel smoothing of dependent data y using tme series cross-validation. *Journal of the Royal Statistical Society, Series*, B, **56**, 529-542.

Jose, C.T. (1999). Nonparametric Inference on Structural Changes in Regression, *Ph.D thesis submitted to Mangalore University, Mangalore, India.*

Jose, C.T. and Ismail, B. (1997) Estimation of jump points in nonparametric regression through residual analysis, *Communications in Statistics- Theory and Methods*, **26**, 2583-2607.

Jose, C.T. and Ismail, B. (1999) Change points in regression functions, *Communications in Statistics- Theory and Methods*, **28**, 1883-1902.

Jose, C.T. and Ismail, B. (2000) Nonparametric inference on change points in regression, *Statistical Methods.* Vol.2, Issue 2. (To appear)

Loader, C.R.(1996). Change point estimation using nonparametric regression. *Annals of Statistics*, **24**, 1667-1678.

Muller H.G. (1992). Change points in nonparametric regression analysis, *Annals of Statistics*, **20**, 737-761.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena. *Ter.Mag.Atmos.Elect.*, **3**, 13-41.

Whittakar, E.T.(1923). On a new method of graduation, *Proc.Edinburgh Math. Soc.*,**41**, 63-75.