A Robust Non Parametric Approach to Fitting a Quadratic Model

Samprit Chatterjee
New York University
44 West 4th Street
New York,USA
schatter@stern.nyu.edu

Introduction

In this paper we present a nonparametric approach to quadratic fitting. The method is an extension to a method proposed by Theil (1950). Theil's method for simple regression involves basically calculating the regression coefficients for all possible pairs of points. We extend this method for quadratic regression. We present two methods, a nonparametric approach, and a least squares approach which calculates all possible regressions for all possible triplets of points. Calculating the standard error of estimates involves the application of dependent U-statistics.

Model and Method

We have n independent observation (y_i, x_i),and our model for the data is

$$y_i = \boldsymbol{a} + \boldsymbol{b}x_i + g x_i^2 + \boldsymbol{e}_i$$

The estimate for g is given in two stages. First we get $\hat{g}_{ijk}$ which is given by

$$\hat{g}_{ijk} = \frac{1}{x_k - x_j}\left\{\frac{y_k - y_i}{x_k - x_i} - \frac{y_j - y_i}{x_j - x_i}\right\}$$

Then we take the median of the $\hat{g}_{ijk}$, $\hat{g} = median\left(g_{ijk}\right)$. We get estimates for the other parameters similarly.

$$\hat{\boldsymbol{b}}_{ij} = \frac{y_j - y_i}{x_j - x_i} - \hat{g}\left(x_j + x_i\right)$$

$$\hat{\boldsymbol{b}} = median\left(\boldsymbol{b}_{ij}\right)$$

$$\hat{\boldsymbol{a}}_i = \hat{Y}_i - \hat{\boldsymbol{b}}x_i - g x_i^2$$

$$\hat{\boldsymbol{a}} = median\{\boldsymbol{a}_i\}$$

To find the variance of these estimates we have to use the results for the variance of the median of correlated observations. This is difficult and we will present the results at the Conference.

An alternative set of estimates can also be derived. The least squares estimates of the three parameters can be given in a close form. This is possible because the matrix to be inverted is of a special form- a Vandermonde matrix-which has a closed form inverse. We take the median of all the generated estimates as our

estimator.

The estimates will be robust and will not require any probabilistic assumptions. We present a numerical example to illustrate the method. The method proposed here can be used as an alternative to local quadratic smoothing.

As must be apparent that no exact small sample estimates of the variance of these estimators can be obtained. We will provide large sample asymptotic variances. These can be used for constructing confidence intervals, and statistical inference.

The variance of the quadratic coefficient $\hat{g}$ is given by

$$\operatorname{var}\left(\hat{l}\right) ; \frac{C}{A^2}$$

where A = $Ave\left(\dfrac{1}{s_{ijk}}\right)$ B = $Ave\left\{\arcsin r_{ijk,ilm}\right\}$

and

$$C = \frac{9}{n}\left[\frac{1}{4} + \frac{1}{2p}B\right]$$

Acknowledgement

I would like to thank Prof. Ingram Olkin for bringing this problem to my attention. I have also benefited from discussions with Profs. Pranab Sen, and S.R.S.Varadhan.

RESUME

Samprit Chatterjee is professor of statistics at New York University. His books on regression analysis have been translated into German, Japanese, and Korean.