Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computational Statistics & Data Analysis 49 (2005) 591 – 609
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

An adaptive algorithm for least squares piecewise monotonic data fitting☆

E. Vassiliou, I.C. Demetriou*

Department of Economics, University of Athens, 8 Pesmazoglou Street, Athens 105 59, Greece

Available online 21 June 2004

Abstract

The number of peaks and troughs of measurements of smooth function values can be unacceptably larger than the number of turning points of the function, when the measurements are too rough. It is proposed to make the least sum of squares change to the data subject to a limit on the number of sign changes of their first divided differences, but usually a suitable value of this limit is not known in advance. It is shown how to obtain automatically an adequate value for it. A test is included that attempts to distinguish between genuine trends and data errors. Specifically, if there are trends, then the monotonic sections of a tentative approximation are increased by one, otherwise this approximation seems to meet the trends and the calculation terminates. The numerical work required per iteration, beyond the second one, is quadratic in the number of data. Details for establishing the underlying algorithm are specified, numerical results from a simulation are included and the test is compared to some well-known residual tests. An application of the algorithm on identifying turning points and trends of data from the Dow Jones stock exchange index is presented. A Fortran implementation of our algorithm provides shorter computation times in practice than the complexity indicates in theory. Further, the single monotonicity problem has found many applications in statistical data analyses within various contexts. More generally, piecewise monotonicity is a property that occurs in a wide range of underlying functions and some important applications of it may be found in detrending data for identifying periodicities (eg. business cycles), or in estimating turning points of a function that is known only by some measurements of its values.

Keywords: Approximation; Data fitting; Divided difference; Dynamic programming; Isotonic; Least squares; Peak; Piecewise monotonic; Residual test; Smoothing; Trend test; Turning point

☆ Supported in part by a Herakleitus Grant of University of Athens.

* Corresponding author.
E-mail address: demetri@econ.uoa.gr (I.C. Demetriou).

doi:10.1016/j.csda.2004.05.029

1. Introduction

A smooth function $f(x)$ is measured at the abscissae $x_1 < x_2 < \cdots < x_n$ and the measurements (data) $\{\varphi_i \cong f(x_i) : i = 1, 2, \ldots, n\}$ contain large uncorrelated errors. The errors tend to cause more sign changes in the sequence of the first differences $\varphi_{i+1} - \varphi_i,\ i = 1, 2, \ldots, n-1$, than one would expect in the sequence of the first differences of $f(x)$. In order to reduce the number of peaks and troughs of the data, Demetriou and Powell (1991) proposed and studied a data smoothing method by imposing a *prescribed* number, say $k-1$, of sign changes on the first differences of the smoothed values $\{y_i : i=1, 2, \ldots, n\}$. Ideally, $k-1$ is the number of sign changes in the first derivative of $f(x)$.

Therefore the constraints on $\{y_i : i=1, 2, \ldots, n\}$ are that there exist integers $\{t_i : i=1, 2, \ldots, k-1\}$ satisfying the conditions

$$1 = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_k = n, \tag{1.1}$$

such that the inequalities

$$\begin{aligned} &y_{t_j} \leqslant y_{t_j+1} \leqslant \cdots \leqslant y_{t_{j+1}}, \quad j \text{ even}, \\ &y_{t_j} \geqslant y_{t_j+1} \geqslant \cdots \geqslant y_{t_{j+1}}, \quad j \text{ odd}, \end{aligned} \tag{1.2}$$

hold. We denote by $M(k, n)$ the set of all numbers $\{y_i : i = 1, 2, \ldots, n\}$ that satisfy these constraints. Specifically, the method minimizes the sum of squares

$$\Phi(y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} (y_i - \varphi_i)^2, \tag{1.3}$$

over $M(k, n)$. We call the smoothed values *optimal piecewise monotonic approximation* to the data or *best fit* and we also call optimal the associated values $\{t_j : j = 1, 2, \ldots, k-1\}$. There are about $\mathrm{O}(n^k)$ combinations of positions of the integer variables $\{t_j : j = 1, 2, \ldots, k-1\}$ for solving this problem, but Demetriou and Powell have developed a dynamic programming method that generates the required fit in only $\mathrm{O}(kn^2)$ computer operations. However, it is sometimes inefficient that the user should choose in advance a suitable value of k.

We propose an extension to this method that identifies automatically an adequate value for k at the expense of very little additional work. The extension combines dynamic programming with a test that attempts to distinguish between genuine trends and data errors along the computational process. The new algorithm begins from the fit to the data that minimizes (1.3) subject to the monotonicity constraints $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n$, and iteratively adds monotonic sections to the fit. In each iteration, the best approximation subject to the given monotonic sections is calculated and then the residuals are tested for trends. If trends are found then the next iteration obtains the best approximation with one more monotonic section. Otherwise the current approximation seems to meet the trends and the iterations terminate. Details of the test and the new algorithm are specified in Section 2. In order to check the robustness of this algorithm, in Section 3 we include numerical results from a simulation and compare our test against four well-known residual tests. Section 4 includes an application of the algorithm on identifying turning points and trends of data from the