COMSTA: 6349 Model 3Gsc pp. 1–11 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Q1 Cross-validated wavelet block thresholding for non-Gaussian errors

Q2 K. McGinnity [a], R. Varbanov [b,*], E. Chicken [b]

[a] *Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, Virginia 22311, USA*
[b] *Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306, USA*

A R T I C L E I N F O

Article history:
Received 15 January 2016
Received in revised form 22 September 2016
Accepted 25 September 2016
Available online xxxx

Keywords:
Wavelets
Thresholding
Nonparametric function estimation

A B S T R A C T

Wavelet thresholding generally assumes independent, identically distributed normal errors when estimating functions in a nonparametric regression setting. VisuShrink and SureShrink are just two of the many common thresholding methods based on this assumption. When the errors are not normally distributed, however, few methods have been proposed. A distribution-free method for thresholding wavelet coefficients in nonparametric regression is described, which unlike some other non-normal error thresholding methods, does not assume the form of the non-normal distribution is known. Improvements are made to an existing even–odd cross-validation method by employing block thresholding and level dependence. The efficiency of the proposed method on a variety of non-normal errors, including comparisons to existing wavelet threshold estimators, is shown on simulated data.

1. Introduction

Wavelet thresholding has been a staple of statistical functional estimation for years. Donoho and Johnstone (1994, 1995, 1998) introduced methods for thresholding the wavelet coefficients derived from the wavelet transformation of the observed data in nonparametric regression:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where the ε_i are independent and identically distributed (i.i.d.) Gaussian errors with mean zero and constant variance σ^2 and the sample points $x_i = i/n$ are equally spaced over an interval. By thresholding observed wavelet coefficients representing noise, a smooth estimate of the underlying can be obtained. The assumptions on the errors have been loosened in only a handful of papers on wavelet thresholding.

Neumann and von Sachs (1995) discuss wavelet thresholding methods in non-Gaussian and non-i.i.d. situations. The main idea of their paper is that, in many situations, asymptotic normality can be proven and traditional thresholding methods can be used. Given independent observations, they demonstrate a way to show equivalence to the Gaussian case via strong approximations. They also derive asymptotic normality in the case of weak dependence.

Antoniadis and Fryzlewicz (2006) propose a scale-dependent wavelet thresholding procedure for Gaussian noise, and then extend it to include non-Gaussian noise. However, the paper assumes not only that the non-normal errors are i.i.d.

* Corresponding author.
E-mail address: r.varbanov@stat.fsu.edu (R. Varbanov).

http://dx.doi.org/10.1016/j.csda.2016.09.010
0167-9473/

Please cite this article in press as: McGinnity, K., et al., Cross-validated wavelet block thresholding for non-Gaussian errors. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.09.010

with mean zero, but also that they follow a known specified distribution. They determine a suitable threshold for each resolution level by mimicking the arguments of Donoho and Johnstone in the Gaussian case.

Pensky and Sapatinas (2007) investigate the performance of Bayes factor estimators in wavelet regression models with i.i.d. non-Gaussian errors. They choose a general distribution η_j for the errors and assume they possess symmetric PDFs on the reals that are unimodal, positive, and finite at zero. One advantage of their method is that knowledge of the true distribution of the errors is not needed in order to obtain an optimal estimator of f. However, their estimators are only preferable for irregular functions with high peaks, and produce sub-optimal results when compared with other methods under certain prior distributions.

Nason (1996) introduces an even–odd cross-validation method for choosing the threshold parameter in wavelet shrinkage. His statistic compares an interpolated wavelet estimator from the even reconstructed data to the odd noisy data and vice versa over various threshold values, then applies a sample size correction.

In this paper, we propose a completely nonparametric method to threshold wavelet coefficients that enhances Nason's cross-validation method by incorporating level-dependent block thresholding. Block thresholding thresholds wavelet coefficients in groups, rather than individually, with the goal of increasing precision by utilizing information about neighboring coefficients (Cai, 1999). Nason's method uses term-by-term thresholding, so it is reasonable to ask if incorporating blocking will have an analogous effect here.

Nason also makes use of a global threshold, the same threshold value for all considered coefficients. This is similar to VisuShrink of Donoho and Johnstone (1994). However, level-dependent thresholding has also been shown to have advantages over universal thresholds. For example, SureShrink (Donoho and Johnstone, 1995), a level-dependent thresholding method, has been shown to have lower mean squared error (MSE) than VisuShrink. Each of these modifications, blocking and level dependence, improves performance with distribution-based thresholds and are thus natural considerations for attempting to improve cross-validation thresholding.

Our method does not put any assumptions on the errors except that they are i.i.d. and centered at zero. Unlike Neumann and von Sachs (1995), we do not discuss asymptotic normality, but instead develop a method specifically meant to handle non-Gaussian errors. Nor do we require that the distribution of the errors be known, as do Antoniadis and Fryzlewicz (2006). Unlike that of Pensky and Sapatinas (2007), no proper choice of prior is required for our method.

This paper is divided as follows. Section 2 provides a brief background on wavelets, wavelet notation, and wavelet thresholding methods before the details of the proposed estimator are described in Section 3. Section 4 contains a simulation comparison of the proposed estimator to the Nason estimator, VisuShrink, and other current methods which may assume normal errors. A discussion of the results and methods is given in the final section.

2. Background

2.1. Wavelets

Wavelets are an orthogonal series representation of functions in the space of square-integrable functions $L_2(\mathbb{R})$. Ogden (1997) and Vidakovic (1999) offer good introductions to wavelet methods and their properties. Let ϕ and ψ represent the father and mother wavelet functions, respectively. There are many choices for these two functions, see Daubechies (1992). Here, ϕ and ψ are chosen to be compactly supported and to generate an orthonormal basis. Let

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)$$

and

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$$

be the translations and dilations of ϕ and ψ, respectively. For any fixed integer j_0,

$$\{\phi_{j_0 k}, \psi_{jk} | j \geq j_0, k \text{ an integer}\}$$

is an orthonormal basis for $L_2(\mathbb{R})$. Let

$$\xi_{jk} = \langle f, \phi_{jk}\rangle$$

and

$$\theta_{jk} = \langle f, \psi_{jk}\rangle$$

be the usual inner product of a function $f \in L_2(\mathbb{R})$ with the wavelet basis functions. Then f can be expressed as an infinite series:

$$f(x) = \sum_k \xi_{j_0 k}\phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty}\sum_k \theta_{jk}\psi_{jk}(x). \tag{2}$$

The function f is not known and must be estimated. This is done using the discrete wavelet transform (DWT) of Mallat (1999). If f is sampled as a vector of dyadic length $n = 2^J$ for some positive integer J, then the DWT will provide a total of n

Please cite this article in press as: McGinnity, K., et al., Cross-validated wavelet block thresholding for non-Gaussian errors. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.09.010