The Annals of Applied Statistics
2011, Vol. 5, No. 1, 88–90
DOI: 10.1214/10-AOAS398F
Main article DOI: 10.1214/10-AOAS398

DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE?

BY PETER CRAIGMILE[1] AND BALA RAJARATNAM[2]

The Ohio State University and Stanford University

Professors McShane and Wyner have written a thought-provoking paper that intends to challenge some of the conventional wisdom in the paleoclimate literature. Rather than commenting on the merits of the entire methodology we focus on one topic. Namely, we discuss theoretical and practical aspects of the use of the least absolute shrinkage and selection operator [Tibshirani (1996)], more popularly known as the "Lasso," in the context of paleoclimate reconstruction.

It is important to acknowledge at first sight that the Lasso seems like a natural candidate in the paleoclimate context, since one is immediately faced with a larger number of proxies, compared to the number of data points [e.g., in McShane and Wyner (2010) (hereafter MW), Section 3.2, the response variable is of length 149 whereas there are 1138 predictors]. It is clear that standard regression-based variable selection techniques will not work. The sheer number of predictors does indeed warrant a need for regularization. Many techniques are available for such problems, including popular methods such as ridge regression and principal component regression.

As pointed out by MW the "Lasso tends to choose sparse $\widehat{\beta}^{\text{Lasso}}$ thus serving as a variable selection methodology and alleviating the $p \gg n$ problem." This point is very well taken. The model selection capability of the Lasso has made it very relevant in this era of high throughput data and rapidly changing information technology. Consequently the Lasso has been useful in biomedical and genomic applications where genes are often in the tens of thousands, compared to much fewer subjects. Biomedical scientists often wish to isolate a few, but important genes that are related to disease conditions. The Lasso "zeroes out" smaller coefficients and thus can be used for model selection.

In a more abstract setting, consider a statistical model such as a regression model which has a low signal-to-noise ratio where the coefficient vector is not

Received September 2010.

[1]Supported in part by NSF Grants DMS-06-04963 and DMS-09-06864.

[2]Supported in part by NSF Grants DMS-09-06392, AGS-1003823 and Grants SU-WI-EVP10, SUFSC08-SUFSC10-SMSCVISG0906.

Key words and phrases. Paleoclimate reconstruction, LASSO, multiproxy data, time series dependence.