The Annals of Applied Statistics
2008, Vol. 2, No. 2, 474–477
DOI: 10.1214/08-AOAS137B
Main article DOI: 10.1214/07-AOAS137

DISCUSSION OF: TREELETS—AN ADAPTIVE MULTI-SCALE BASIS FOR SPARSE UNORDERED DATA

BY PETER J. BICKEL[1] AND YA'ACOV RITOV[2]

University of California and The Hebrew University of Jerusalem

We divide our comments on this very interesting paper into two parts following its own structure:

1. The use of treelets in connection with the correlation matrix of $\mathbf{X} = (X_1, \dots, X_p)^\mathsf{T}$ for which we have n i.i.d. copies, or as the authors refer to it, "unsupervised learning."
2. The use of treelets as a step in best fitting the linear regression of X_1 on $(X_2, \dots, X_p)^\mathsf{T}$.

1. Unsupervised learning. The authors' emphasis is on the method as a useful way of representing data analogous to a wavelet representation where $\mathbf{X} = \mathbf{X}(t)$ with t genuinely identified with a point on the line and observation at p time points, but where the time points have been permuted.

As such, this can be viewed as a clustering method which, from their examples, gives very reasonable answers. However, to make more general theoretical statements and to permit comparison to other methods, they necessarily introduce the model

$$\mathbf{X} = \sum_{j=i}^{K} U_j v_j + \sigma Z_j, \tag{1}$$

where $\mathbf{U} = (U_1, \dots, U_K)^\mathsf{T}$ is an unobservable vector, the v_j are fixed unknown vectors, and $\mathbf{Z} \sim N_p(0, J_p)$, where J_p is the identity, N_p is the p dimensional Gaussian distribution, and $\mathbf{U}, \mathbf{Z}$ are independent.

At this point, we are a bit troubled by the authors' analysis. We believe a key point, that is only stressed implicitly by the authors, is that the population tree structure, as defined, is only a function of the population covariance matrix. This is clear at Step 1, and follows since the Jacobi transformations depend only on the covariance and variances of the coordinates involved. This raises a problematic issue. If $\mathbf{U}$, and hence $\mathbf{X}$, has a Gaussian distribution, then the structure as postulated

Received November 2007; revised November 2007.

[1] Supported in part by NSF Grant DMS-06-05236.

[2] Supported in part by an ISF grant.