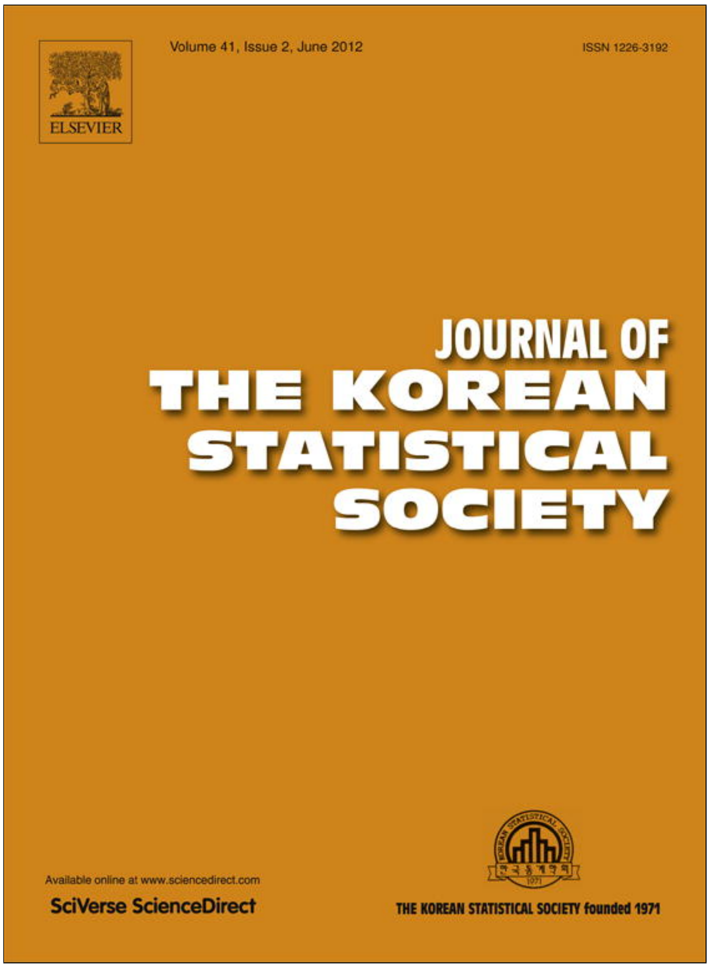
Volume 41, Issue 2, June 2012
ISSN 1226-3192
ELSEVIER
JOURNAL OF
THE KOREAN
STATISTICAL
SOCIETY
Available online at www.sciencedirect.com
SciVerse ScienceDirect
THE KOREAN STATISTICAL SOCIETY founded 1971

Journal of the Korean Statistical Society 41 (2012) 171–172

Contents lists available at SciVerse ScienceDirect

Journal of the Korean Statistical Society

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

Discussion

Discussion: Time-threshold maps: Using information from wavelet reconstructions with all threshold values simultaneously

Hernando Ombao

Department of Statistics, University of California at Irvine, United States

ARTICLE INFO

Article history:
Received 23 February 2012
Available online 9 March 2012

I would first like to point out the broad connections of this work with those in the literature and then highlight what I believe are its major contributions.

Consider the model $Y(s_i) = f(s_i) + \epsilon(s_i)$, where f is the unknown function that has to be estimated from the data $Y(s_i)$ via a reconstruction of thresholded wavelet coefficients. One motivation for developing the time-threshold map is that using only a single threshold on a wavelet coefficient could miss potentially important features in f. A different threshold might produce a different estimate that brings out a different feature f.

This "single-threshold-per-coefficient" approach is broadly related to the standard practice in statistics, where we often either simply *assume* a model to be true or *select one* best model from a class of models using some objective criterion. We then proceed with any estimation and inference by pretending that the selected model is the correct model, that is, the selected model is the true model that generated the observed data. This approach is risky, because it ignores uncertainty in the model selection step. It also ignores the fact that the class of models is merely an approximation of the underlying truth and that each member of the class could potentially capture or explain different aspects of the underlying function.

To account for model uncertainty in statistical inference, one can apply the Bayesian model averaging (BMA) approach. Here, the final estimate is some weighted average of estimates from different models and the weights are proportional to the posterior probability of each model (see Hoeting & Madigan, 1999 for a tutorial on the implementation of BMA). Madigan and Raftery (1994) note that BMA gives better predictive ability than using only the single selected model. Another related area is on curve aggregation (see Bunea, Ombao, & Auguste, 2006), where the final estimate is some weighted average of several curve estimates.

On a very specific problem of estimating the spectral matrix and coherence, Fiecas and Ombao (2011) develop an estimator which is a linear combination of some parametric estimator and a non-parametric estimator. Here, the primary estimand is a two-dimensional matrix defined on the frequency interval $(-\pi, \pi)$. The primary motivation of such an approach was that parametric estimates of the spectrum derived from an auto-regressive (AR) model with a sufficiently high order are well localized in frequency ω and hence can capture peaks that would have otherwise been flattened by a smoothing procedure. A non-parametric estimator, on the other hand, is robust to model mis-specification. Let $\widehat{\mathbf{f}}_1(\omega)$ be the parametric estimator derived from some objective criterion such as the Akaike information criterion or Bayesian information criterion, and let $\widehat{\mathbf{f}}_2(\omega)$ be a non-parametric estimator such as a smoothed periodogram matrix. The resulting estimator is $\widehat{\mathbf{f}}(\omega) = W(\omega)\widehat{\mathbf{f}}_1(\omega) + [1 - W(\omega)]\widehat{\mathbf{f}}_2(\omega)$, where the frequency-specific weight for the parametric estimator is proportional to the mean-squared error (Hilbert–Schmidt norm) of the other; that is, $W(\omega) = \mathbb{E}\|\widehat{\mathbf{f}}_1(\omega) - \mathbf{f}(\omega)\|_{\mathrm{HS}}$.

One major contribution of the paper by Fryzlewicz is that it rigorously develops the analogy of "combining estimates" under the framework of wavelet estimation. In the case of wavelet estimation, a particular threshold corresponds to an optimal estimate arising from a model that assumes a particular smoothness of the underlying function f. Fryzlewicz

E-mail address: hombao@uci.edu.

1226-3192/$ – see front matter
doi:10.1016/j.jkss.2012.02.005

introduces an interesting quantity $\Delta \mathbf{x}_\lambda = \lim_{h \to 0^+} \mathbf{x}_{\lambda+h} - \mathbf{x}_\lambda$, which indicates precise changes in the estimates as the threshold increases. This quantity could serve to identify "stable" subintervals of the domain of f from those subintervals that are sensitive to changes in the threshold.

I believe that the approach in Fryzlewicz can also be applied to obtain a "more complete" multi-resolution estimate of f. One can produce level-dependent (or frequency-band-specific) estimates of f obtained from the varying thresholds. Consequently, one can visually observe long-range patterns and highly localized patterns in f. This would be highly informative to the user who might be interested in differentiating between global patterns from transient features.

This paper opens up some new lines of research on wavelet estimation, and one area where I hope to see this work explored is in developing a principled way of combining estimates arising from different thresholds. Here, the weights may depend on the mean-squared error of the estimators and may vary across different intervals of the domain of f. For example, there may be a subinterval of the domain in which a particular set of thresholds performs better than the other sets of thresholds. I will attempt to give a more technical statement of this problem. Analogous to class of models in BMA or curve aggregation, define $\Lambda = \{\Lambda_1, \Lambda_2, \ldots\}$ to be a set of all possible vectors of coefficients under consideration. Next, define $\widehat{f_\ell}(s)$ to be the optimal wavelet estimator derived under the threshold set Λ_ℓ. Define the final aggregated estimate to be

$$\widehat{f}(s) = \sum_\ell W_\ell(s) \widehat{f_\ell}(s),$$

where the weights are allowed to vary across the domain s. The weights $W_\ell(s)$ could be inversely proportional to the mean-squared error of the corresponding estimator $\widehat{f_\ell}(s)$. The theoretical form of the weight might be derived analytically but, in practice, it could be estimated via an application of the bootstrap procedure. This approach allows the user to identify the threshold set that gives a good (versus poor) fit at certain subintervals of the domain.

References

Bunea, F., Ombao, H., & Auguste, A. (2006). Minimax adaptive spectral estimation from an ensemble of signals. *IEEE Transactions on Signal Processing*, *54*, 2865–2874.

Fiecas, M., & Ombao, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Annals of Applied Statistics*, *5*, 1102–1125.

Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, *14*, 382–417.

Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, *89*, 1535–1546.