Data-Adaptive Estimation of Time-Varying Spectral Densities

Anne van Delft and Michael Eichler

Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany; Department of Quantitative Economics, Maastricht University, Maastricht, Netherlands

ABSTRACT
This article introduces a data-adaptive nonparametric approach for the estimation of time-varying spectral densities from nonstationary time series. Time-varying spectral densities are commonly estimated by local kernel smoothing. The performance of these nonparametric estimators, however, depends crucially on the smoothing bandwidths that need to be specified in both time and frequency direction. As an alternative and extension to traditional bandwidth selection methods, we propose an iterative algorithm for constructing localized smoothing kernels data-adaptively. The main idea, inspired by the concept of propagation-separation, is to determine for a point in the time-frequency plane the largest local vicinity over which smoothing is justified by the data. By shaping the smoothing kernels nonparametrically, our method not only avoids the problem of bandwidth selection in the strict sense but also becomes more flexible. It not only adapts to changing curvature in smoothly varying spectra but also adjusts for structural breaks in the time-varying spectrum. Supplementary materials, including the R package tvspecAdapt containing an implementation of the routine, are available online.

1. Introduction
Spectral analysis of time series data has been of interest for many years and has a varied history owing to applications in a wide range of disciplines such as geophysics, astronomy, sound analysis, analysis of medical data, or yet of economical data. There exists a rather extensive literature on spectral analysis of weakly stationary processes and statistical techniques are well developed (Cramér 1942; Bartlett 1950; Grenander and Rosenblatt 1957; Cooley and Tukey 1965; Brillinger 1981). However, in many applications the time series at hand show some nonstationary behavior in the sense that the oscillations described by the spectrum change over time. For instance, such behavior can be observed for the brain activity in various brain regions during associating learning experiments (Fiecas and Ombao 2016). In such cases, the assumption of weak stationarity often still seems plausible over shorter time periods, that is, the series can locally be well approximated by a stationary time series and its local oscillations can be described by the spectrum of the approximating stationary series. Empirically, this idea is reflected, for example, in the use of the spectrogram or the segmented periodogram for time-frequency data analysis (e.g., Sandsten 2016).

Theoretically, several definitions of time-varying spectra have been proposed in the literature (e.g., Priestley 1965; Subba Rao 1970; Martin 1985; Hallin 1986). Most definitions depend on the length $T$ of the time series and hence are problematic for statistical inference. For a unifying approach, Dahlhaus (1996b) developed the concept of locally stationary processes, where the nonstationary time series is embedded in a sequence of series that share the same dynamics over an increasing number of observations. This framework not only yields a unique definition of a time-varying spectrum but also allows meaningful asymptotic approximations to the sampling distribution of localized estimators and test statistics.

Like in the stationary case, time-varying spectral densities are commonly estimated by kernel smoothing of some raw spectral estimator such as the segmented periodogram or the preperiodogram (Neumann and von Sachs 1997). An important problem that is common to all kernel smoothing estimators is the selection of proper bandwidths since the accuracy of the resulting estimators has been found to be quite sensitive to the choice of bandwidths (e.g., Eichler, Motta, and von Sachs 2011). Theoretical optimal bandwidths (Dahlhaus 1996b) depend on the unknown underlying spectral density, and no guidelines are available on how to set them in practice (Dahlhaus 2009). To our knowledge, data-adaptive schemes suitable in the context of time-frequency analysis have not yet been considered. However, under specific parametric assumptions there are methods available based on the segmented periodogram. For example, Sergides and Paparoditis (2009) and Preuß, Vetter, and Dette (2011) used an integrated version of the segmented periodogram to test for semiparametric hypotheses, which avoids selection of bandwidths in frequency direction. The major drawback of using the segmented periodogram as an underlying estimator is that a fixed bandwidth in time direction, namely, the length of segments, must be set. This leads to seriously biased estimates in case of structural breaks in the spectrum.

In this article, we propose an alternative approach that circumvents the problem of classical bandwidth selection by data-adaptive shaping of the smoothing kernel. The procedure is inspired by the propagation-separation approach, an adaptive implementation of which was developed by Neumann and von Sachs (2009).
weighing scheme where the iteratively updated weights gain from previously aggregate information. This method was first introduced by Polzehl and Spokoiny (2006) in the context of local likelihood models and is closely related to previous works in this direction (e.g., Lepskii 1991; Polzehl and Spokoiny 2000). The general principle is to determine, based on some measure of homogeneity, for each design point the maximal local neighborhood that can be used for smoothing. Starting with a small initial neighborhood, the smoothing region is iteratively extended to include new data points for which the hypothesis of homogeneity can still be maintained (propagation) while data points are excluded whenever it is violated. This approach has demonstrated to be useful in a variety of problems such as image denoising and classification (e.g., Belomestny and Spokoiny 2007; Tabelow et al. 2008).

In our adaptation of the propagation-separation approach, at each iteration of the algorithm the neighborhood used for estimating the time-varying spectral density \( f(u, \lambda) \) at rescaled time \( u \in [0, 1] \) and frequency \( \lambda \in [-\pi, \pi] \) is described by weights \( W_{u, \lambda}(\cdot, \cdot) \) that define the shape of the kernel. These weights are derived from the spectral estimates constructed at the previous step. The effective neighborhood for local smoothing is then given by the points for which the corresponding weights are nonzero. As a discrepancy measure for the deviation from homogeneity, we use the squared relative difference between the corresponding spectral estimates. Compared to classical kernel estimates with either global or local bandwidth, the advantage of our approach is that the smoothing kernel itself is adjusted in terms of shape and effective bandwidth for each point separately. This flexibility in shaping the smoothing kernel completely data-adaptively is of particular importance when structural breaks are present. In that case, our estimator can gain precision from smoothing in one direction without getting severely biased by smoothing across the break. The algorithm has been additionally robustified to be less affected by so-called cross-terms, which pose a serious problem in high-resolution time-frequency analysis. We emphasize that our work is related to aforementioned existing literature on data-adaptive schemes but that these are not directly applicable in the setting of time-frequency analysis.

Our article is organized as follows. In Section 2, we provide the background on locally stationary processes and estimation of time-varying spectral densities and briefly describe the notion of cross-terms. In Section 3, we present the algorithm and explain the importance of the various steps. In Section 4, we illustrate the properties of the proposed estimator by three examples and examine its performance in a simulation study. Finally, the approach is illustrated by an application to local field potential (LFP) recordings in Section 5. The routine and further technical details are provided as supplementary material.

2. Locally Stationary Processes

Let \( X_t \) be a nonstationary process that has been observed at times \( t = 1, \ldots, T \). For frequency-domain based analysis of the process, we follow the approach by Dahlhaus (1996b) and view the process \( X_t \) as part of a sequence of processes \( \{X_{t, T}, t = 1, \ldots, T\} \), \( t \in \mathbb{N} \), where \( X_{t, T} \) has the representation

\[
X_{t, T} = \sum_{j \in \mathbb{Z}} a_{t, T}(j) e^{i\lambda j}
\]

for some weakly stationary white-noise process \( \{e_t\} \) with \( \mathbb{E}(e_t) = 0 \) and \( \mathbb{E}(e_t^2) = 1 \). The processes \( X_{t, T} \) for different \( T \) are related by (approximately) sharing the same dynamics locally. More precisely, we assume that there exist functions \( a(u, \lambda) \) on \([0, 1] \times [-\pi, \pi] \) such that

(i) \( \sup_{j} |a_{t, T}(j)| \leq K \ell(j)^{-1} \) for all \( j \in \mathbb{Z} \),

(ii) \( |a(u, j)| \leq K \ell(j)^{-1} \) for all \( j \in \mathbb{Z} \),

(iii) \( \sup_{j} \sum_{t=1}^{T} \left| a_{t, T}(j) - a(j - \lambda) \right| \leq K \),

(iv) \( V(a(e, j)) \leq K \ell(j)^{-1} \) for all \( j \in \mathbb{Z} \),

where

\[
V(g) = \sup \left\{ \sum_{k=1}^{m} \left| g(x_k) - g(x_{k-1}) \right| : 0 \leq x_0 < \cdots < x_m \leq 1, \ m \in \mathbb{N} \right\}
\]

is the total variation of a function \( g \) on \([0, 1] \) and

\[
\ell(j) = \max \{ 1, |j| \log^{+} |j| \}
\]

for some constant \( \zeta > 0 \). A sequence of processes \( \{X_{t, T}\} \) satisfying the above assumptions (i) to (iv) is called locally stationary (e.g., Dahlhaus and Polonik 2009; Dahlhaus 2009).

The above representation implies that, locally about a point \( u \ T \) for some \( u \in [0, 1] \), the processes \( \{X_{t, T}\} \) can be approximated by the weakly stationary process

\[
X_{t}^{(u)} = \sum_{j \in \mathbb{Z}} a(u, j) e^{i\lambda j}, \quad t \in \mathbb{Z},
\]

and thus the oscillating behavior of the processes \( \{X_{t, T}\} \) can be described by the spectral density

\[
f(u, \lambda) = \frac{1}{2\pi} A(u, \lambda) A(u, \lambda)^*, \quad (1)
\]

where

\[
A(u, \lambda) = \sum_{j \in \mathbb{Z}} a(u, j) e^{-i\lambda j}.
\]

The function \( f(u, \lambda) \) in (1) is called the time-varying spectral density of the locally stationary process \( \{X_{t, T}\} \) at frequency \( \lambda \in [-\pi, \pi] \) and rescaled time \( u \in [0, 1] \).

In this article, we are more generally interested in the estimation of the time-varying spectral densities of processes \( \{X_{t, T}\} \) that possibly exhibit structural breaks in time. Therefore, we impose the following condition on the process \( \{X_{t, T}\} \).

Assumption 2.1. \( \{X_{t, T}, t = 1, \ldots, T, T \in \mathbb{N} \} \) is a piecewise locally stationary process with time-varying spectral density \( f(u, \lambda) \) that is twice differentiable in \( \lambda \) and piecewise twice differentiable in \( u \) with bounded derivatives in both directions.

Nonparametric approaches for the estimation of the time-varying spectral density are based on the fact that

\[
\gamma(u, k) = \int_{-\pi}^{\pi} f(u, \lambda) e^{ik\lambda} d\lambda,
\]

defines a localized auto-covariance function. Using estimators \( \hat{\gamma}_{T}(u, k) \) for \( \gamma(u, k) \), we obtain

\[
\hat{f}_{T}(u, \lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{\gamma}_{T}(u, k) e^{-ik\lambda}.
\]
as an estimator for the time-varying spectral density at rescaled time \( t = \frac{u}{T} \) and frequency \( \lambda \). For instance, taking auto-covariance estimators on segments of length \( 2m \),

\[
\hat{y}_T(\frac{u}{T}, k) = \frac{1}{2m} \sum_{j=-m}^{m} X_jX_{j+k}
\]

for \( k \geq 0 \) and \( \hat{y}_T(\frac{u}{T}, k) = \hat{y}_T(\frac{u}{T}, -k) \) for \( k < 0 \), we obtain the segmented periodogram (e.g., Dahlhaus 1996b; Preuß, Vetter, and Dette 2011), which treats the series over a segment of length \( 2m \) about the time point \( u T \) as stationary. While this approach inherits the good properties of the periodogram in the stationary case, the restriction to lags at most 2 implies that the preperiodogram of the signal \( \hat{y}_T(u, k) \) for \( k = 0 \) and \( k \geq 2 \) is the so-called cross-term or interference term with respect to \( k = 1 \).

Here, the first two terms on the right-hand side are the prepe-

roriodogram of the signal \( \hat{X} + \hat{Y} \) is given by

\[
J_T^{(X+Y)}(\frac{u}{T}, \lambda) = J_T^{(X)}(\frac{u}{T}, \lambda) + J_T^{(Y)}(\frac{u}{T}, \lambda) + 2 \Re J_T^{(X,Y)}(\frac{u}{T}, \lambda).
\]

The preperiodogram exhibits a much better time–frequency concentration than the segmented periodogram and has many other useful properties such as interpretability of the time and frequency marginals, instantaneous frequency and group delay, and weak finite support (see Sandsten 2016). These properties however come at a cost in the form of nondisappearing cross-terms or interference terms (e.g., Sandsten 2016). To understand the nature of these terms, consider a signal with two components \( X \) and \( Y \). The quadratic superposition principle implies that the preperiodogram of the signal \( X + Y \) is given by

\[
J_T^{(X+Y)}(\frac{u}{T}, \lambda) = \sum_{i=1}^{N(u/r)} \sum_{j=1}^{N(\lambda/r)} W_{r,i}^{(k)}(s, j) J_T(u, \lambda_i), \quad k = 0, \ldots, k_{\text{max}},
\]

where \( \lambda_i = \frac{\pi}{T} \) for \( j = 1 \), \ldots, \( T \) denote the Fourier frequencies and \( C = \sum_{i,j} K_i((\lambda - \lambda_j)/b_{i,T}) K_j((u - s/T)/b_{i,T}) \) is the normalization constant. The properties of (3) have been investigated in the setting of empirical spectral processes in Dahlhaus (2009) and Dahlhaus and Polonik (2009). In particular, it has been shown that, under suitable conditions on the rates of the smoothing bandwidths, the estimator is asymptotically normal (Theorem 3.2, Example 4.1 of Dahlhaus 2009). In the following, we refer to (3) as the nonadaptive estimator.

3. Propagation–Separation Approach in the Time–Frequency Plane

A well-known problem in the application of kernel estimators is the selection of appropriate bandwidths as the quality of the resulting estimates depends critically on the chosen bandwidth (e.g., Eichler, Motta, and von Sachs 2011). For time-varying spectral estimators, various asymptotic results are available (see, e.g., Dahlhaus 2009; Dahlhaus and Polonik 2009), but determining the corresponding optimal smoothing bandwidths in practice is still an open problem. As an alternative to standard bandwidth selection methods, we propose an iterative algorithm to determine at each point in the time–frequency plane the shape of the smoothing kernel data-adaptively. Our method is based on the propagation–separation approach by Polzehl and Spokoiny (2006). Starting with small bandwidths for the initial estimates, the bandwidths are increased in each iteration to allow smoothing over larger regions of homogeneity (propagation) while detection of differences between estimates of the preceding iteration leads to penalization and reduction of kernel weights to stop further smoothing (separation).

To illustrate the main idea, consider the white-noise signal \( X_t, t = 1, \ldots, 500 \), in Figure 1(a) with time-varying variance (Figure 1(b)) and a structural break at time \( t_{\text{break}} = 280 \). As the variance is constant for time points \( 1 \leq t \leq t_{\text{break}} \), the smoothing bandwidth in time direction can be chosen as large as possible, but due to the structural break smoothing should not extend across time \( t_{\text{break}} \). The corresponding (nonnormalized) smoothing kernels in time direction about time points \( u_0 = 0.2 \) and \( u_0 = 0.4 \) (in rescaled time) are depicted as the upper two curves in Figure 1(c). For times \( t_{\text{break}} < t \leq 500 \), the variance changes smoothly over time and the bandwidths of the smoothing kernels needs to be adapted accordingly; the two lower curves in Figure 1(c) show the smoothing kernels about the time points \( u_0 = 0.6 \) and \( u_0 = 0.9 \). Finally, we note that since the spectral density is constant over frequencies the smoothing kernels should be unconstrained with maximal bandwidth in frequency direction.

3.1. The Algorithm

For data-adaptive kernel estimation of the time-varying spectral density, we consider a sequence of weighted averages

\[
\hat{f}_T(u, \lambda) = \frac{1}{N(u/r)} \sum_{s,j} W_{r,i}^{(k)}(s, j) J_T(u, \lambda_i), \quad k = 0, \ldots, k_{\text{max}},
\]

where \( u = \frac{u}{T} \) and \( \lambda_i = \frac{\pi}{T} \) for \( r = 1, \ldots, T \).

\[
N^{(k)}(r, i) = \sum_{s,j} W_{r,i}^{(k)}(s, j)
\]
The local smoothing kernel at the point \((u, \lambda_i)\) is asymptotically additive relaxation toward the most recent estimate. This relaxation arises from the usual bias at peaks and troughs (e.g., Dahlhaus 1989) where the amount of shift is determined by the integrated penalty factor, which indicates saturation in the growth of the effective smoothing region.

Below we first present the complete algorithm in detail. In the next subsection, we then provide further explanations and additional comments.

Initialization \((k = 0)\).

For initial bandwidths \(b_{i, T}^{(0)}\) and \(b_{i, T}^{(0)}\) compute initial kernel estimates

\[
\hat{f}^{(0)}(u, \lambda_i) = \frac{1}{N^{(0)}(r, i)} \sum_{(s, j) \in B^{(0)}(r, i)} K_i(u - u_s, \lambda_i) f_t(u_s, \lambda_j),
\]

and local averages

\[
\tilde{f}^{(0)}(u, \lambda_i) = \frac{1}{sB^{(0)}(r, i)} \sum_{(s, j) \in B^{(0)}(r, i)} \hat{f}^{(0)}(u_s, \lambda_j),
\]

where \(N^{(0)}(r, i) = \sum_{s, j} K_i(\lambda_j - \lambda_i)/b_{i, T}^{(0)} K_i(u_s - u_t)/b_{i, T}^{(0)}\) and \(B^{(0)}(r, i) = \{ (s, j) : |u_s - u_t| < b_{i, T}^{(0)}, |\lambda_j - \lambda_i| < b_{i, T}^{(0)}\}\) for \(k = 0, \ldots, k_{\text{max}}\).

Parameters: Smoothing kernels \(K_i\) and \(K_t\) (default \(K_i(x) = K_t(x) = 6(\frac{1}{4} - x^2) 1_{[-1/2, 1/2]}(x)\))

Iteration \(k - 1 \rightarrow k\).

(a) Increase kernel bandwidths: \(b_{i, T}^{(k)} = 4b_{i, T}^{(k - 1)}\) and \(b_{i, T}^{(k)} = \alpha_i b_{i, T}^{(k - 1)}\).

Parameters: Growth rates \(\alpha_i\) and \(\alpha\) (default values \(\alpha_i = \alpha = 1.2\)).

(1) Penalty step.

(b) From discrepancies \(\Delta(\hat{f}^{(k - 1)}(u, \lambda_i), \hat{f}^{(k - 1)}(u, \lambda_j))\) (defined in (5)) between previous estimates \(\hat{f}^{(k - 1)}(u, \lambda_i)\) and \(\hat{f}^{(k - 1)}(u, \lambda_j)\) compute the penalties for the deviation from homogeneity

\[
\tilde{P}_{r, t}^{(k)}(s, j) = K_i \left( \frac{\hat{f}^{(k - 1)}(u, \lambda_i) - \hat{f}^{(k - 1)}(u, \lambda_j)}{\delta^{(k)}(s, j)} \right),
\]

where \(K_i\) is a decreasing nonnegative penalty kernel with bounded support \([0, c_p]\); constrain final penalty \(P_{c, t}^{(k)}(s, j)\) to be radially nonincreasing:

\[
P_{c, t}^{(k)}(s, j) = \min \left( \tilde{P}_{r, t}^{(k)}(s, j), (1 - \alpha) P_{c, t}^{(k)}(s - 1, j) + \alpha P_{c, t}^{(k)}(s - 1, j - 1) \right)
\]
for points \((s, j)\) such that \((s, j) = \beta (s - 1, j - \alpha)\) for some \(\beta > 1\) and \(\alpha \in [0, 1]\) and

\[
P_{r, t}(s, j) = \min \left( \hat{P}_{r, t}(s, j), (1 - \alpha) P_{r, t}^{(k)}(s - 1, j) + \alpha P_{r, t}^{(k)}(s - 1, j - 1) \right)
\]

for points \((s, j)\) such that \((s, j) = (\beta \alpha, \alpha - \alpha)\) for some \(\beta > 1\) and \(\alpha \in [0, 1]\).

Parameters: Cut-off \(c_p\) for penalties (default \(c_p = \chi^2_{1, 0.9}\); penalty kernel \(K_p\) (default \(K_p(x) = 1 - (x/c_p)^2\) for \(x \leq c_p\) and zero elsewhere)

(c) Compute intermediate estimator

\[
\tilde{f}^{(k)}(u_r, \lambda_i) = \frac{1}{N^{(k)}(r, i)} \sum_{(s, j) \in B^{(k)}(r, i)} \tilde{W}_{r, t}(s, j) f_T(u_s, \lambda_j)
\]

(6)

from adapted kernel weights

\[
\tilde{W}_{r, t}(s, j) = K_r\left( \frac{s-r}{h_r}\right) K_r\left( \frac{u_s-u_r}{h_r}\right) p_{r, t}(s, j),
\]

(7)

where \(\tilde{N}^{(k)}(r, i) = \sum_{s, j} \tilde{W}_{r, t}(s, j)\) and \(\tilde{M}^{(k)}(r, i) = \sum_{s, j} \tilde{W}_{r, t}(s, j)^2\).

(d) Compute total measure of penalization

\[
\pi^{(k)}(r, i) = \frac{1}{2} \left( \max (\pi_{r+}^{(k)}(r, i), \pi_{r-}^{(k)}(r, i)) + \max (\pi_{r+}^{(k)}(r, i), \pi_{r-}^{(k)}(r, i)) \right)
\]

from directional penalization measures in time directions

\[
\pi_{r+}^{(k)}(r, i) = \frac{1}{C_{r+}} \sum_{(s, j) \in B^{(k)}(r, i)} p_{r, t}(s, j) \sqrt{(s-r)^2 - (j-i)^2},
\]

\[
\pi_{r-}^{(k)}(r, i) = \frac{1}{C_{r-}} \sum_{(s, j) \in B^{(k)}(r, i)} p_{r, t}(s, j) \sqrt{(s-r)^2 + (j-i)^2}
\]

with \(C_{r+} = \sum_{(s, j) \in B^{(k)}(r, i)} \sqrt{(s-r)^2 - (j-i)^2}\) and similarly \(\pi_{r+}^{(k)}(r, i)\) and \(\pi_{r-}^{(k)}(r, i)\) in frequency directions. Additionally compute the overall penalization at \((u_r, \lambda_i)\) as

\[
\pi^{(k)}_{all}(u_r, \lambda_i) = \frac{1}{\#B^{(k)}(r, i)} \sum_{(s, j) \in B^{(k)}(r, i)} p_{r, t}(s, j).
\]

(e) Compute the signal-to-noise ratio

\[
s^{(k)}(u_r, \lambda_i) = \frac{\tilde{f}^{(k)}_{hom}(u_r, \lambda_i)}{\sigma^{(k)}_{hom}(u_r, \lambda_i)}
\]

(8)

from local mean and variance of \(\tilde{f}^{(k)}\)

\[
\tilde{f}^{(k)}_{hom}(u_r, \lambda_i) = \frac{1}{\#B_{hom}(r, i)} \sum_{(s, j) \in B_{hom}(r, i)} \tilde{f}^{(k)}(u_s, \lambda_j),
\]

\[
\sigma^{(k)}_{hom}(u_r, \lambda_i)^2 = \frac{1}{\#B_{hom}(r, i)} \sum_{(s, j) \in B_{hom}(r, i)} \left( \tilde{f}^{(k)}(u_s, \lambda_j) - \tilde{f}^{(k)}_{hom}(u_r, \lambda_i) \right)^2
\]

over local region of homogeneity

\[
B_{hom}(r, i) = \{(s, j) : |u_s - u_r| < b^{(0)}_{L, T}, |\lambda_j - \lambda_i| < b^{(0)}_{h_{L, T}}, p_{r, t}(s, j) > \frac{1}{2}\}.
\]

(2) Relaxation I: Cancellation of negative cross-terms

(f) If \(\tilde{f}^{(k)}(u_r, \lambda_i) < \tilde{f}^{(k-1)}(u_r, \lambda_i) < 0\) set

\[
\tilde{f}^{(k)}(u_r, \lambda_i) = \tilde{f}^{(k-1)}(u_r, \lambda_i);
\]

similarly set \(\tilde{N}^{(k)}(r, i), \tilde{M}^{(k)}(r, i), \tilde{f}^{(k)}_{hom}(u_r, \lambda_i), \) and \(s^{(k)}(u_r, \lambda_i)\) to their previous values \(\tilde{N}^{(k-1)}(r, i), \tilde{M}^{(k-1)}(r, i), \tilde{f}^{(k-1)}_{hom}(u_r, \lambda_i), \) and \(s^{(k-1)}(u_r, \lambda_i)\) respectively.

(3) Relaxation II: Shift toward smoothed version of intermediate estimates

(g) Compute local averages of \(\tilde{f}^{(k)}(u, \lambda), \) the kernel normalization, and the total penalization measure

\[
\tilde{f}^{(k)}_{ave}(u_r, \lambda_i) = \frac{1}{N^{(k)}(r, i)} \sum_{(s, j) \in B^{(k)}(r, i)} K_t\left( \frac{u_s-u_r}{c_p}\right) \tilde{f}^{(k)}(u_s, \lambda_j),
\]

(9)

\[
\tilde{N}^{(k)}_{ave}(u_r, \lambda_i) = \frac{1}{N^{(k)}(r, i)} \sum_{(s, j) \in B^{(k)}(r, i)} K_t\left( \frac{u_s-u_r}{c_p}\right) \tilde{N}^{(k)}(s, j),
\]

(10)

\[
\tilde{f}^{(k)}_{ave}(r, i) = \frac{1}{N^{(k)}(r, i)} \sum_{(s, j) \in B^{(k)}(r, i)} K_t\left( \frac{u_s-u_r}{c_p}\right) \pi^{(k)}(s, j).
\]

(11)

(h) From the signal-to-noise ratio \(s^{(k)}(u_r, \lambda_i)\) defined in (e) compute the relaxation parameter

\[
\theta^{(k)}(r, i) = \min \left( \epsilon_s, s^{(k)}(u_r, \lambda_i), 1 \right),
\]

for all points \((u_r, \lambda_i)\) for which the stability condition \(\tilde{N}^{(k)}(r, i) \geq \alpha_1 \tilde{N}^{(k)}_{ave}(r, i)\) holds and set \(\theta^{(k)}(r, i) = 1\) if the stability condition is violated; in that case, we additionally set \(\tilde{N}^{(k)}(r, i) = N^{(k-1)}(r, i)\) and \(\tilde{M}^{(k)}(r, i) = M^{(k-1)}(r, i)\).

Parameters: Threshold \(\epsilon_s\) for the signal-to-noise ratio (default \(\epsilon_s = 2\)

(i) Obtain new estimate \(\tilde{f}^{(k+1)}\) as weighted average of \(\tilde{f}^{(k)}\) and smoothed version \(\tilde{f}^{(k)}_{ave}\)

\[
\tilde{f}^{(k+1)}_{rel}(u_r, \lambda_i) = \left( 1 - \theta^{(k)}(r, i) \right) \tilde{f}^{(k)}_{ave}(u_r, \lambda_i) + \theta^{(k)}(r, i) \tilde{f}^{(k)}_{rel}(u_r, \lambda_i);
\]

similarly shift the local mean and the total penalization measure toward their smoothed versions

\[
\tilde{f}^{(k+1)}_{rel}(r, i) = \left( 1 - \theta^{(k)}(r, i) \right) \tilde{f}^{(k)}_{ave}(r, i) + \theta^{(k)}(r, i) \tilde{f}^{(k+1)}_{rel}(r, i);
\]

\[
\tilde{f}^{(k+1)}_{ave}(r, i) = \left( 1 - \theta^{(k)}(r, i) \right) \tilde{f}^{(k)}_{ave}(r, i) + \theta^{(k)}(r, i) \tilde{f}^{(k+1)}_{ave}(r, i).
\]

(4) Relaxation III: Shift toward estimate from previous iteration

(j) Compute final estimate \(\tilde{f}^{(k)}\) as weighted average of \(\tilde{f}^{(k)}_{rel}\) and \(\tilde{f}^{(k-1)}\)

\[
\tilde{f}^{(k)}(u_r, \lambda_i) = \left( 1 - K_p (\pi^{(k)}_{rel}(r, i) \epsilon_p) \right) \tilde{f}^{(k+1)}_{rel}(u_r, \lambda_i) + K_p (\pi^{(k)}_{rel}(r, i) \epsilon_p) \tilde{f}^{(k-1)}_{ave}(u_r, \lambda_i),
\]
Similarly $\hat{f}^{(k)}(u_r, \lambda_i)$, $N^{(k)}(r, i)$, and $M^{(k)}(r, i)$ are obtained from $\hat{f}^{(k)}_{rel}(u_r, \lambda_i)$, $\hat{N}^{(k)}(r, i)$, and $\hat{M}^{(k)}(r, i)$, respectively, and their versions from the previous iteration.

**Loop.** Repeat Steps (1) to (4) until the average overall penalty over all design points,

$$\bar{\pi}^{(k)}_{all} = \frac{1}{T} \sum_{r,i=1}^{T} \pi^{(k)}_{all}(u_r, \lambda_i), \quad (12)$$

where $\pi^{(k)}_{all}(u_r, \lambda_i)$ is the overall penalty defined in (d), indicates that the majority of all smoothing kernels are shrinking, that is, until

$$\bar{\pi}^{(k)}_{all} < \frac{1}{\alpha_1} \alpha_1, \quad (13)$$

or until the maximal number $k_{max}$ of iterations with $b^{(k_{max})}_{all} = b^{(k_{max})}_{all} \approx 1$ is reached. In case the algorithm terminates by (13), estimates are returned for

$$k_{final} = \max \{ k | \bar{\pi}^{(k)}_{all} \geq \left( \frac{1}{\alpha_1} + \frac{3}{\alpha_1} \alpha_1 \right)^{-1} \}.$$

### 3.2. Further Details

We now provide further details on the initialization parameters and on the various steps.

#### 3.2.1. Parameters for Initialization

**Initial bandwidth parameters $b^{(0)}_{l,T}$ and $b^{(0)}_{l,T}$.** The choice of initial bandwidth faces the usual trade-off of being able to retrieve details in the signal while not having too strong distortions due to noise. The algorithm offers default values set to $b^{(0)}_{l,T} = b^{(0)}_{l,T} = \sqrt{\log T^{1.9}/2\pi T}$, which are conservative compared to the CLT condition underlying the distribution of the penalty statistic (Assumption 1.1(ii) of the online supplement). Additionally the algorithm has the option to automatically improve the initial bandwidths if increasing them slightly could reduce the percentage of negative initial estimates.

**Choice for kernel functions $K_t$, $K_f$, and $K_p$:** The default smoothing kernels $K_t$ and $K_f$ of the algorithm have been shown to yield the smallest mean squared error (Dahlhaus 1996a) in case of local homogeneity. In combination, we have found the concave penalty kernel $K_p$ most appropriate. The cut-off value for the penalty kernel is based on the asymptotic distribution of the penalty statistic under local homogeneity.

#### 3.2.2. Penalty Step

**Steps (b) and (e):** For the relative squared error in the discrepancy, we use a local average of the estimates over a small region to improve stability of the penalization step. To prevent bias from high curvature or structural breaks, the local average is taken only over points that are judged as belonging to the same homogenous region by the discrepancy.

**Step (d):** The penalization measures in step (d) give an indication about the growth of the effective smoothing region, which is required for controlling relaxation toward previous estimates as well as for stopping of the algorithm. To allow effective smoothing regions to grow further even in the close neighborhood of structural breaks that lead to strong penalization on one side, penalization is measured along the time and frequency directions separately. For this, penalties are weighted by the component of their normalized vectors in the direction of interest.

#### 3.2.3. Relaxation I and II

The first two relaxation steps control for the effect of cross-terms and reduce their presence iteratively. Cross-terms often lead to highly oscillating positive and negative spikes in the preperiodogram that make recovering of the spectral signal very difficult. Empirically, areas that are dominated by cross-terms can be identified by low signal-to-noise ratio. Relaxation II stabilizes the areas affected by cross-terms by applying a local smoothing based on the local signal-to-noise ratio.

**Step (f):** Negative cross-terms can lead to negative estimates that over iterations can destabilize the estimation. We therefore limit the effect by replacing negative estimates by their least observed versions prior to Relaxation II.

**Step (g):** We use a signal-to-noise ratio to assess the reliability of the estimates. The signal-to-noise ratio is obtained by the ratio of a local average and an estimate of the local standard deviation. The local region over which these measures are computed is chosen such that it only includes points within a small neighborhood, as given by the starting bandwidths, that have been identified by the discrepancy measure to potentially belong to the same homogenous region.

**Steps (h) to (j):** Full weight is provided to a local average if the signal-to-noise ratio is less than $e_1$ or a large cross-term is detected. Large cross-terms can be identified as points of which the sum of weights is extremely low compared to the local average sum of weights of the same step. By imposing local smoothing in these regions, we allow to distinguish signal from noise iteratively enabling the detection of breaks as well as smooth patterns.

#### 3.2.4. Relaxation III and Early Stopping

**Step (k):** The purpose of the last relaxation step is to stabilize the estimate if the total penalization measure indicates that no further smoothing over larger regions is required. In each step, the bandwidths of the neighborhood over which can be smoothed are increased. Meaning that, unless the smoothing region covers a homogenous area, the penalty kernel becomes less pronounced as the distance kernels flatten. This causes a bias that we control for by bending the estimator to its previous estimator according to the total penalization measure.

**Early stopping:** In principle, the algorithm continues until the entire plane is searched for each point, that is, until $b^{(k_{max})}_{all} = b^{(k_{max})}_{all} \approx 1$. The algorithm stops earlier if (12) is satisfied, that is, if smoothing kernels are shrinking for the majority of design points. Further smoothing will then no longer lead to improved estimates and the returned estimates are specified by the largest $k$ for which $\bar{\pi}^{(k)}_{all}$ is bounded away from the stopping threshold $1/\alpha_1 \alpha_1$.

### 4. Simulations

In this section, we first illustrate our data-adaptive estimation method by application to simulated data from three processes.
that cover three types of possible scenarios: structural break with otherwise constant spectrum, smooth time-varying spectrum, and time-varying spectrum that also exhibits a structural break. In all examples, the default parameters are taken to demonstrate that our method can be expected to work well for a wide range of processes without requiring process-specific choice of parameters. Furthermore, to investigate the performance of our data-adaptive spectral estimator quantitatively, we compare it in a simulation study to a kernel spectral estimator with global bandwidth chosen such that the mean square error is minimized. As the optimal choice of the bandwidth depends on the unknown spectral density, this estimator can be seen as an oracle estimator. Throughout this section, we therefore refer to this estimator as the oracle estimator.

### 4.1. Structural Break White Noise

The first process that we consider is a white-noise process with an upward shift in the variance of the process at some time $t_0$. More specifically, data $X_1, \ldots, X_T$ were simulated from the following model:

$$X_t = \begin{cases} 
\varepsilon_t & \text{if } t \leq t_0 \\
\sqrt{10}\varepsilon_t & \text{if } t > t_0 
\end{cases}$$

with $\varepsilon_t \sim \mathcal{N}(0, 1)$ and $t_0 = 0.5625 T$ where $T = 1024$. The corresponding time-varying spectral density is given by

$$f_X(u, \lambda) = \frac{1}{2\pi} 1_{|u| \leq t_0/T} + \frac{10}{2\pi} 1_{|u| > t_0/T}$$

and is depicted in Figure 2(a). Figure 2(b) shows the corresponding preperiodogram. Compared with the ordinary periodogram, it exhibits much more variation which completely blurs the piecewise constant form of the density. The algorithm stopped after $k_{\text{final}} = k_{\text{max}} = 14$ iterations, that is, the search region was extended until covering the entire plane. The resulting data-adaptive estimate is given in Figure 2(d). It clearly shows two levels for the spectral density. Figure 2(e) and 2(f) depicts, respectively, the final penalty and the adaptive kernel for the point $(u, \lambda) = (\frac{1}{2}, 0)$ in the middle of the plane. Once the break is detected, the penalty kernel forces the weights down to zero. This results in an asymmetric smoothing kernel that is “cut off” and thus succeeds in separating the areas on both sides of the break (compare with the example in Figure 1). To illustrate the effect of penalization, Figure 2(c) shows a nonadaptive kernel estimate as given in (3) computed with the same bandwidths as used in the final step of our iterative method. Not surprisingly, the presence of the break is completely smoothed out.

To examine the performance of our estimator, we compared it with the oracle estimator over 100 repetitions of process $X_t$. Figure 3 depicts the mean curves and pointwise boxplots at frequency $\lambda = 0$ for different time points $u$. It can be observed that close the break the adaptive estimator varies much more than the oracle estimator. This can be explained by the fact that the adaptive estimator clearly detects the break but detect its location slightly before or slightly after the true break point in the different repetitions. The oracle estimator does not detect it but simply smooths it out. Away from the break, our estimator is much more precise and is very close to the true spectrum. The good performance of our estimator compared to any nonadaptive kernel estimator with global bandwidths can also be seen from the root integrated mean square errors and the integrated mean absolute errors listed in Table 1.

---

**Figure 2.** White noise with structural break: (a) True spectrum, (b) preperiodogram, (c) nonadaptive estimate, (d) adaptive estimate, (e) penalty kernel, and (f) smoothing kernel (scaled by a factor $10^5$) of the adaptive estimate at $(u, \lambda) = (\frac{1}{2}, 0)$.
4.2. Locally Stationary Series

As a second process, we consider a time-varying moving average of order 1 given by

$$Y_t = \cos \left(2 \pi \frac{t}{T}\right) \varepsilon_t - \left(\frac{1}{2}\right)^2 \varepsilon_{t-1}$$

with $\varepsilon_t \sim N(0, 1)$ and $T = 1024$. The time-varying spectral density of this process, given by

$$f(u, \lambda) = \frac{1}{2\pi} \left( \cos(2\pi u)^2 - 2u^2 \cos(2\pi u) \cos(\lambda) + u^4 \right).$$

is depicted in Figure 4(a). The spectrum exhibits a peak in the middle of the time-frequency plane with smooth hill-sides in frequency direction and steeper ones in time direction. The preperiodogram is given in Figure 4(b) and indicates low signal frequency direction and steeper ones in time direction. The true spectral density lies almost always within the range of observed estimates, which is not the case for the oracle estimator. Summarizing we find that although the performance is very similar, the adaptive estimator seems to capture the curvature slightly better than the oracle estimator.

The error measures in Table 1 indicate that both the adaptive and the oracle estimator give comparable results. Figure 5 provides more details about the behavior of the two estimators. As can be observed, the oracle estimator shows similar bias at peaks and troughs despite the much smaller variation in the lower regions. In contrast, the bias of the adaptive estimator adjusts to the level of variation and thus is less biased in the troughs.

In particular, we note that the true spectral density lies almost always within the range of observed estimates, which is not the case for the oracle estimator. We emphasize that the chosen bandwidth of the oracle estimator is computed from the true spectral density and therefore that any estimator with global bandwidths will show larger errors.

4.3. Structural Break in Locally Stationary Series

We now combine the above two processes. More precisely, we consider a white-noise process that at $t_0 = 0.4 T$ turns into a moving average process with similar dynamics as in the previous example but shifted in time. Thus, we have

$$Z_t = \begin{cases} 
\varepsilon_t & \text{if } t \leq t_0 \\
\cos \left(2\pi \left(\frac{t}{T} - \frac{1}{2}\right)\right) \varepsilon_t - \left(\frac{1}{2}\right)^2 \varepsilon_{t-1} & \text{if } t > t_0
\end{cases}$$

with $\varepsilon_t \sim N(0, 1)$. The corresponding time-varying spectral density is given by

$$f(u, \lambda) = \frac{\sigma^2}{2\pi} 1_{[u \leq \lambda/T]} + g(u - \frac{1}{2}, \lambda) 1_{[u > \lambda/T]},$$

where $g$ is the nonadaptive kernel estimate $\hat{f}_{\text{na}}$ with optimal bandwidth $\hat{b}_{\text{na}}$.

Table 1. Mean (standard deviation) of the root integrated mean square error (RIMSE) and integrated mean absolute error (IMAE) for the adaptive estimator $\hat{f}_{\text{na}}$ and the oracle kernel estimator with optimal bandwidth $\hat{f}_{\text{opt}}$, obtained over 100 repetitions of the three processes $X_t$, $Y_t$, and $Z_t$.

| Process | $\text{RIMSE}$ | $\text{MAE}$ |
|---------|----------------|--------------|
| $X_t$   | $0.182$ (0.044) | $0.270$ (0.032) | $0.092$ (0.036) | $0.174$ (0.027) |
| $Y_t$   | $0.034$ (0.008) | $0.033$ (0.007) | $0.021$ (0.003) | $0.022$ (0.004) |
| $Z_t$   | $0.028$ (0.004) | $0.031$ (0.003) | $0.020$ (0.003) | $0.024$ (0.003) |
where \( g(u, \lambda) \) is as in (15). These type of spectra can, for example, occur when a signal is constant for a while and then receives a stimulus. This time-varying spectrum is interesting as the peak and the break are both close in distance as well as in level and hence are difficult to resolve. The adaptive estimate in Figure 6(d) was obtained after \( k_{\text{final}} = 8 \) iterations. Both curvature and the break are clearly discernible, and the estimate closely resembles the true spectrum shown in Figure 6(a).

Compared with the results for process \( X_t \), the flat part of the white-noise component is slightly rougher. The corresponding nonadaptive estimated spectrum \( \hat{f}_{\text{na}} \) in Figure 6(c)) evidently suffers from over-smoothing. Figure 6(e) and 6(f) depicts the penalty and smoothing kernel for the point \( (u, \lambda) = \left( \frac{1}{2}, 0 \right) \) which lies in the trough between the break and the peak. Strong penalization in time direction and close to none in frequency direction is visible. This is in accordance with the different slopes in the two directions.

Compared to the previous case, the simulation study reveals a much better performance of the adaptive estimator than the oracle estimator as shown by the error measures in Table 1. This is also visible in Figure 7. Similar as for process \( Y_t \), the adaptive estimator shows less bias in the lower regions and similar
results elsewhere. Slightly more variation is observed for the adaptive estimator close to the break, but otherwise it exhibits less variation than the oracle estimator. This is in line with the findings for process $X_t$ and can be explained similarly. Overall, we conclude that the adaptive estimate again captures the features of the true spectrum better than the oracle estimator.

5. Application to Local Field Potentials

As an application of our method, we consider local field potential (LFP) recordings of the nucleus accumbens of a male macaque monkey during an associative learning experiment.
These type of data as well as other types of brain data are known to exhibit nonstationary behavior that results in localized signal in the time-varying spectrum, often located in a small frequency band. Existing methods to resolve such signals cannot provide information upon the nature of nonstationarity. Given the ability of our method to resolve localized signal and to find both break structures and smooth patterns, we can determine what type of nonstationary behavior characterizes such data.

During each trial of this experiment, the monkey was shown four pictures and then had to select one of four doors. If the monkey made the correct association between picture and door he would receive a reward. In total, the learning experiment, which was conducted at the Neurosurgery Department at the Massachusetts General Hospital, consisted of 675 trials. Figure 8 depicts the LFP recordings of one of the trials which is of length $T = 2048$. The figure indicates changing dynamics over time. A first inspection of the preperiodogram (Figure 9(a) and 9(b)) indicates most of the neuronal activity is centered in a small frequency band close to zero. To resolve the narrow peak(s), a small bandwidth is required in frequency direction. In particular, it is well-known from the stationary setting, the width of the main lobe from the kernel function should be no larger and preferably half the size than the bandwidth of the narrowest peak in the spectrum. Moreover, there is also clear evidence of nonstationary behavior in time direction. Given the length of data, even the relatively conservative default starting bandwidths would out-smooth this behavior. Based on the width of the signal we therefore fix the starting bandwidths to the very low values $b^{(0)}_{1,T} = b^{(0)}_{1,T} = 0.025$. The rest of the parameters are set to the default values.

The estimated spectrum is given in Figure 9(c) which was obtained at $k_{\text{final}} = 6$. Signal is estimated in a small frequency band around zero as can be seen more clearly in Figure 9(d). Figure 9(e) and 9(f) depicts the estimated spectrum locally around the point $(u, \lambda) = (\frac{1}{2}, 0)$ together with the penalty kernel obtained for the point $(u, \lambda) = (\frac{1}{2}, 0)$. We observe some definite changes in both width, location, and magnitude of the peak over time when looking at the various graphs. Although steep changes occur, they seem smooth in nature. A large spike is visible around the end of the data stretch, possibly indicating the response of the monkey’s brain to receiving the award. It is worth noting that despite the low starting bandwidths and the large percentage of cross-terms, our method overall appears to capture local structures well.

6. Conclusion

In this article, we introduced a data-driven approach to estimate spectral densities of nonstationary time series, a long time open problem in the analysis of time-varying spectral analysis. We propose to iteratively determine the optimal shape of the smoothing kernel by determining the maximal neighborhood over which smoothing is justified by the data. This flexibility in shaping the smoothing kernel completely data-adaptively is of particular importance when structural breaks are present. A major problem in high-resolution time-frequency analysis are so-called cross-terms. Our algorithm is specifically designed to control for these cross-terms, which leads to estimates that benefit from the good time-frequency concentration of the
preperiodogram while not suffering from distortions due to cross-terms. A simulation study indicates that our method captures the underlying dynamics of various spectra very accurately and better than any kernel estimator with global bandwidth.

One limitation in the application of the algorithm is its complexity as in each step for the shaping of the smoothing kernel every point in the time-frequency plane must be compared with all points in a growing local neighborhood. Although our current implementation makes use of multiprocessor computing, the estimation could become computationally infeasible for very long time series and an extension to high-dimensional time series is—computationally speaking—not straightforward. This could possibly be improved by the use of graphical processing units. This is left for future work.

Supplementary Materials

Supplement: Supplementary document with further details on the distributional properties of the adaptive kernel estimator.

Rpackage: Package tvspecAdapt containing an implementation of the algorithm as described in this article. The algorithm is implemented in C++ and R (R Core Team 2014) using the Rcpp-package (Eddelbeutel and François 2011) and allows for Openmp support. The package (tvspecAdapt.tar.gz) together with documentation can be found in the root folder (RoutineDataAdapt.zip).

Acknowledgments

The authors sincerely thank the Editor, Associate Editor, and two referees for their constructive comments that helped to produce an improved revision of the original article.

Funding

This work has been supported in part by Maastricht University, the contract "Projet d’Actions de Recherche Concertées" No. 12/17-045 of the “Communauté française de Belgique,” and by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project A1, A7) of the German Research Foundation (DFG).

References

Bartlett, M. (1950), "Periodogram Analysis and Continuous Spectra," Biometrika, 37, 1–16. [244]
Belomestny, D., and Spokoiny, V. (2007), "Spatial Aggregation of Local Likelihood Estimates with Applications to Classification," Annals of Statistics, 35, 2287–2311. [245]
Brillinger, D. (1981), Time Series: Data Analysis and Theory, New York: McGraw Hill. [244]
Cooley, J., and Tukey, J. (1965), "An Algorithm for the Machine Calculation of Complex Fourier Series," Mathematics of Computation, 19, 297–301. [244]
Cramér, H. (1942), "On Harmonic Analysis in Certain Functional Spaces," Arkiv för Matematik, Astronomi och Fysik, 28B, 1–7. [244]
Dahlhaus, R. (1990), "Nonparametric High Resolution Spectral Estimation," Probability Theory and Related Fields, 85, 147–180. [247]
——— (1996a), "Asymptotic Statistical Inference for Nonstationary Processes with Evolutionary Spectra," in Athens Conference on Applied Probability and Time Series Analysis (Vol. II), Springer-Verlag, pp. 145–159. [249]
——— (1996b), "On the Kullback-Leibler Information Divergence of Locally Stationary Processes," Stochastic Process and their Applications, 62, 139–168. [244,245,246]
——— (2009), "Local Inference for Locally Stationary Time Series based on the Empirical Spectral Measure," Journal of Econometrics, 151, 101–112. [244,245,246]
Dahlhaus, R., and Polonik, W. (2009), "Empirical Spectral Processes for Locally Stationary Time Series," Bernoulli, 15, 1–39. [245,246]
Eddelbeutel, D., and François, R. (2011), "Rcpp: Seamless R and C++ Integration," Journal of Statistical Software, 40, 1–18. [255]
Eichler, M., Motta, G., and von Sachs, R. (2011), "Fitting Dynamic Factor Models to Non-Stationary Time Series," Journal of Econometrics, 163, 51–70. [244,246]
Fiecas, M., and Omboa, H. (2016), "Modeling the Evolution of Dynamic Brain Processes during an Associative Learning Experiment," Journal of the American Statistical Association, 111, 1440–1453. [244]
Grenander, U., and Rosenblatt, M. (1957), Statistical Analysis of Time Series, New York: Wiley. [244]
Hallin, M. (1986), "Nonstationary q-Dependent Processes and Time-Varying Moving Average Models: Invertibility Properties and the Forecasting Problem," Advances in Applied Probability, 32, 170–210. [244]
Heong, J., and Williams, W. J. (1992), "Alias-Free Generalized Discrete-Time Frequency Distribution," IEEE Transactions on Signal Processing, 40, 2757–2765. [246]
Lepski, O. V. (1991), "A Problem of Adaptive Estimation in Gaussian White Noise," Theory of Probability & Its Applications, 35, 454–466. [245]
Martin, W., and Flandrin, P. (1985), "Wigner-Ville Spectral Analysis of Nonstationary Processes," IEEE Transactions on Signal Processing, 33, 1461–1470. [244]
Neumann, M. H., and von Sachs, R. (1997), "Wavelet Thresholding in Anisotropic Function Classes and Application to Adaptive Estimation of Evolutionary Spectra," The Annals of Statistics, 25, 38–76. [244,246]
Polzehl, J., and Spokoiny, V. (2000), "Adaptive Weights Smoothing with Applications to Image Restoration," Journal of the Royal Statistical Society, Series B, 62, 335–354. [245]
——— (2006), "Propagation-Separation Approach for Local Likelihood Estimation," Probability Theory and Related Fields, 135, 335–362. [245,246]
Preuß, P., Vetter, M., and Dette, H. (2011), "Testing Semiparametric Hypotheses in Locally Stationary Processes," Scandinavian Journal of Statistics, 40, 417–437. [244,246]
Priestley, M. B. (1965), "Evolutionary Spectra and Non-Stationary Processes," Journal of the Royal Statistical Society, Series B, 27, 204–237. [244]
R Core Team (2014), R: A language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing, available at http://www.R-project.org/ [255]
Sandsten, M. (2016), Time-Frequency Analysis of Time-Varying Signals and Non-Stationary Processes (Lecture Notes), Lund, Sweden: Lund University, available at http://www.maths.lth.se/maststat/kurser/masm26/2016/TIMEFREKompendie.pdf [244,246]
Sergides, M., and Paparoditis, E. (2009), "Frequency Domain Tests of Semiparametric Hypotheses for Locally Stationary Processes," Scandinavian Journal of Statistics, 36, 800–821. [244]
Subba Rao, T. (1970), "The Fitting of Non-Stationary Time Series Models with Time-Dependent Parameters," Journal of the Royal Statistical Society, Series B, 32, 312–322. [244]
Tabelow, K., Polzehl, J., Spokoiny, V., and Voss, H. (2008), "Diffusion Tensor Imaging: Structural Adaptive Smoothing," NeuroImage, 39, 1763–1773. [245]