INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

DISCUSSION PAPER

2016/13

Functional mixed effects wavelet estimation for spectra of replicated time series

Chau, J. and R. von Sachs

Functional mixed effects wavelet estimation for spectra of replicated time series

Joris Chau†

ISBA, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Rainer von Sachs

ISBA, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract. Motivated by spectral analysis of replicated brain signal time series, we propose a functional mixed effects approach to model replicate-specific spectral densities as random curves varying about a deterministic population-mean spectrum. In contrast to existing work, we do not assume the replicate-specific spectral curves to be independent, i.e. there may exist explicit correlation between different replicates in the population. By projecting the replicate-specific curves onto an orthonormal wavelet basis, estimation and prediction is carried out under an equivalent linear mixed effects model in the wavelet coefficient domain. To cope with potentially very localized features of the spectral curves, we develop estimators and predictors based on a combination of generalized least squares estimation and nonlinear wavelet thresholding, including asymptotic confidence sets for the population-mean curve. We derive L_2-risk bounds for the nonlinear wavelet estimator of the population-mean curve –a result that reflects the influence of correlation between different curves in the replicate population– and consistency of the estimators of the inter- and intra-curve correlation structure in an appropriate sparseness class of functions. To illustrate the proposed functional mixed effects model and our estimation and prediction procedures, we present several simulated time series data examples and we analyze a motivating brain signal dataset recorded during an associative learning experiment.

Keywords: Spectral analysis, replicated time series, functional mixed effects model, wavelet thresholding, between-curve correlation, nonparametric confidence sets.

1. Introduction

Spectral analysis of replicated time series has recently gained growing interest; for instance in the field of brain data analysis, where it is common to collect time series data (such as EEG, fMRI, or local field potential data) from multiple subjects, or over multiple trials in an experiment. In this context the emphasis is on estimation of population spectral characteristics of the collection of time series, rather than on spectral behavior of individual time series. Applications beyond the field of brain data analysis can be found in: biomedical experiments, geophysical and financial data analysis, or speech modeling, when the inferential focus is not on the mean responses of the replicated time series but on the stochastic variation of the replicated times series about their means. While there is an extensive literature on spectral analysis and inference of individual

†*Address for correspondence*: Joris Chau, Institute of Statistics, Biostatistics, and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.
E-mail: `j.chau@uclouvain.be`

time series, this is not necessarily the case for replicated time series, and existing approaches mostly work under simplifying assumptions such as independent or at least uncorrelated time series replications, which if not satisfied can lead to statistically inefficient estimators or even give misleading inferences.

We address the specific problem of analyzing spectra of replicated time series showing potentially very localized features, allowing for explicit correlation between the time series replicates. Examples of replicated time series data with correlation between replicates include: time series data collected from multiple subjects in an experiment with possible correlation between different subjects due to unknown covariates (age, gender, etc.), or time series data collected over multiple trials of an experiment, where the spectral characteristics of the time series replicates evolve over the course of the experiment. A particular motivating example is the spectral analysis of brain data trial-replicated time series in the context of learning experiments, such a dataset is analyzed in Section 7. As pointed out by Morris (2014) and also Fiecas and Ombao (2016) there is a strong need to generalize existing approaches in this direction, although only few modifications to these assumptions have been developed by now.

In the context of second-order spectral analysis of stationary replicated time series, Diggle and Al Wasel (1997) introduce a log-linear mixed effects model for independent time series replicates, which is generalized by Hernandez-Flores et al. (1999) and Iannaccone and Coles (2001) by considering nonparametric estimation of the fixed effects curve. Freyermuth et al. (2010) develop a tree-structed wavelet method for log-spectral estimation also in the case of independent replicated time series, whereas Krafty et al. (2011) introduce a more general mixed-effects approach based on spline smoothing of empirical log-spectra handling two-level nested designs with replicated time series for each of S independent subjects, where different time series replicates within a subject are allowed to be correlated based on known covariates. Qin et al. (2009) consider a covariate-indexed functional fixed effects model for time-varying spectra of independent replicated nonstationary time series, and Martinez et al. (2013) apply the Bayesian wavelet-based mixed effects approach developed by Morris and Carroll (2006) to model time-varying spectra of replicated nonstationary time series, allowing for potential correlation between the time series replicates induced by the experimental design. More recently, in the context of learning experiments, Fiecas and Ombao (2016) model log-spectra of replicated nonstationary time series trials by including a replicate-time effect that evolves over the course of the experiment. In a general functional data analysis context, not aimed at spectral analysis of time series in particular, nonparametric functional mixed effects models have been considered among others by Guo (2002), and Qin and Guo (2006) using smoothing-spline approaches, and by Aston et al. (2010) using functional principal components. In order to avoid the modelling of functional data by inherently smooth curves, wavelet-based approaches have been considered by Morris and Carroll (2006) and Morris et al. (2008) using Bayesian wavelet shrinkage methods, by Giacofci et al. (2013) using nonlinear wavelet thresholding, and by Antoniadis and Sapatinas (2007) focusing on inference in a wavelet-based functional mixed effects model (see Morris (2014) for a comprehensive overview).

In this work we introduce an additive two-layer functional mixed effects model in the frequency domain for a collection $\{X_s(t)\}_{s=1,\ldots,S}$ of S individual time series, each with discrete observations over time. The time series replicates are modeled to have random replicate-specific log-spectra, which consist of a fixed effects curve on the first layer, (population-average or -mean log-spectrum), additional to replicate-specific random effects curves on the second layer. We contribute in particular by allowing for explicit correlation between the random effects curves, and we model this correlation in an appropriate way to allow for its fully nonparametric estimation, where it is important to note that we dispose only of a single realization for each of the S time series replicates. As we observe only noisy versions of the replicate-specific log-spectra, through the log-periodograms of each replicate, we face a curve estimation (curve denoising) problem. Here, we borrow strength from existing wavelet methodology on denoising curves showing potentially very localized features, and we tailor nonlinear wavelet thresholding to the specific situation of a functional mixed effects model in the frequency domain. By projecting the discretely sampled replicate-specific log-periodograms onto an orthonormal wavelet basis, we obtain an equivalent finite-dimensional linear mixed effects model in the wavelet coefficient domain. This allows us to apply traditional linear mixed model estimation methods combined with nonlinear wavelet thresholding in a unified framework for both the fixed- and random effects empirical wavelet coefficients. To achieve simultaneous estimation of the fixed effects curve and the correlation structure between different random effects curves, we propose an easy-to-implement iterative generalized least squares estimation algorithm. We complete the model by proposing predictors of the individual replicate-specific log-spectra, as well as asymptotic confidence regions for the population-mean log-spectrum, which may be interesting on their own as the literature on inference in the context of nonlinear wavelet thresholding estimators is relatively limited.

On the theoretical side, we contribute by deriving consistency results for the estimators of the fixed effects curve and the variance-covariance-structure of the random effects curves, where we consider asymptotics in both T the time series length, and S the replicate sample size. In particular, we derive bounds on the L_2-risk of the nonlinear wavelet estimator of the fixed effects curve over an appropriate sparseness class of ℓ_0-constrained functions. We obtain a rate of convergence that is equal to the minimax rate of estimation in the absence of random effects, plus an additional term that accounts for the fact that the population-mean curve is estimated by an empirical (generalized least squares) mean over S time series replicates, clearly demonstrating the effect of correlation between the different random effects curves.

The structure of the paper is as follows. In Section 2 we introduce the model set-up in both the frequency and wavelet coefficient domain with an appropriate combined ℓ_0-sparseness constraint for the fixed- and random effects that allows for general inhomogeneous functional behavior over frequency. We impose some conditions on the variance-covariance structure of the random effects, since consistent estimation in a totally unstructured covariance matrix is infeasible. In Section 3 we present estimators for the different ingredients in the model, the population-mean log-spectrum, and the within- and between-replicate covariance and correlation matrices, and we also propose predictors for the replicate-specific log-spectra. As estimation of the first- and second-order structure in the model are interrelated, we present an efficient iterative generalized least squares algorithm to simultaneously estimate both ingredients. In Section 4 we provide the necessary estimation theory; we derive risk bounds on the nonlinear wavelet estimator of the population-mean log-spectrum, which are insightful as they highlight the influence of possible correlation between the random effects curves on the first-order estimation problem. Further, we give conditions for consistency of the within- and between-replicate covariance matrices and some insights in the principals of the proofs. In Section 5 we derive asymptotic confidence regions for the population-mean log-spectrum based on the nonlinear wavelet estimator. In Section 6 we give numerical results on the performance of the estimation and inference procedures for simulated time series data, and in Section 7 we analyze a motivating data example consisting of brain signal time series data recorded over the course of an associative learning experiment.

2. Methodology

2.1. Model setup

Let $\{X_s(t)\}_{t>0}$ be replicated mean-zero second-order stationary time series for replicates $s = 1, \ldots, S$, with absolutely summable covariance functions $\sum_{h=-\infty}^{\infty} |\mathrm{Cov}(X_s(t), X_s(t+h))| < \infty$. If we observe discretely sampled time series $\{X_s(t),\ t = 1, \ldots 2T\}$, the raw bias-corrected log-periodograms at frequencies $\omega_\ell = \ell/(2T) \in [0, 1)$ are given by,

$$Y_s^f(\omega_\ell) = \log \frac{1}{2T} \left| \sum_{t=1}^{2T} X_s(t) \exp(-2\pi i \omega_\ell t) \right|^2 + \gamma \tag{1}$$

where $\gamma \approx 0.577$ is a bias-correction equal to the Euler-Mascheroni constant (see Wahba (1980)). For convenience, we consider the time series length to be dyadic $2T = 2^J$ in order to avoid additional complications in the subsequent wavelet estimation. We also note that it suffices to consider the log-periodograms only over the range of frequencies $\omega_\ell \in [0, 1/2)$, i.e. indices $\ell = 0, \ldots, T-1$, since the log-spectra are $[0, 1]$-periodic and symmetric in $\omega_\ell = 1/2$.

2.1.1. *Frequency domain functional mixed model*

We model the replicate-specific log-spectra as random curves varying about a deterministic population-mean log-spectrum, which is common to all replicates, see Figure 2 for a simulated example. Similar approaches are considered in Diggle and Al Wasel (1997), Freyermuth et al. (2010), and Krafty et al. (2011) to model the (log-)spectra of stationary replicated time series. We express the raw log-periodograms in terms of the following functional mixed effects model:

$$\begin{aligned} Y_s^f(\omega_\ell) &= H_s^f(\omega_\ell) + E_s^f(\omega_\ell), \qquad s = 1, \ldots, S, \quad \ell = 0, \ldots, T-1 \\ &= h^f(\omega_\ell) + U_s^f(\omega_\ell) + E_s^f(\omega_\ell) \end{aligned} \tag{2}$$

where,

(a) $h^f \in L_2([0, 1/2])$ is a deterministic population-mean log-spectrum (a functional fixed effect).

(b) $\{U_s^f, s = 1, \ldots, S\}$ are mean-zero Gaussian random processes (functional random effects) with realizations in $L_2([0, 1/2])$ for each replicate s. The assumptions on the variance-covariance structure of the functional random effects are detailed in Section 2.2.

(c) $E_s^f(\omega_\ell) \xrightarrow{d} \log(\chi_2^2/2) + \gamma$ are asymptotically independent noise terms, with $\mathbb{E}[E_s^f(\omega_\ell)] = o_T(1)$ and $\mathrm{Var}(E_s^f(\omega_\ell)) = \sigma_e^2 + o_T(1)$, where $\sigma_e^2 := \pi^2/6$ by Wahba (1980). Note that for $\omega_0 = 0$, $E_s^f(\omega_0) \xrightarrow{d} \log(\chi_1^2) + \gamma$, but since the influence of this term is negligible for T large enough, we consider the error term $E_s^f(\omega_0)$ to have the same asymptotic distribution as the other error terms in our subsequent analysis (as in Moulin (1994), Krafty et al. (2011), and Freyermuth et al. (2010)). The noise terms $E_s^f(\omega_\ell)$ are assumed to be independent between different replicates and independent of the functional random effects U_s^f for all s, ℓ.

2.1.2. *Wavelet domain linear mixed model*

Let $\phi : \mathbb{R} \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be sufficiently smooth *father* and *mother* wavelet functions, and define the infinite set of translated and dilated functions:

$$\begin{aligned} \phi_{J_0,m}(x) &= 2^{J_0/2}\phi(2^{J_0/2}x - m) \\ \psi_{j,m}(x) &= 2^{j/2}\psi(2^j x - m), \qquad j \geq J_0, m \in \mathbb{Z} \end{aligned} \tag{3}$$

It can be shown that, under certain assumptions on ϕ and ψ (see Daubechies et al. (1992)), the system in eq.(3) generates an orthonormal basis of $L_2(\mathbb{R})$ and, for all $f \in L_2(\mathbb{R})$,

$$f(x) = \sum_{m=0}^{2^{J_0}-1} \alpha_m \phi_{J_0,k}(x) + \sum_{j=J_0}^{\infty} \sum_{m=0}^{2^j-1} \beta_{j,m}\psi_{j,m}(x)$$

with,

$$\alpha_m = \langle f, \phi_{J_0,m}\rangle = \int f(x)\phi_{J_0,m}(x)\,dx, \qquad \beta_{j,m} = \langle f, \psi_{j,m}\rangle = \int f(x)\psi_{j,m}(x)\,dx$$

where the series converges in $L_2(\mathbb{R})$. There exist several ways to transform a wavelet basis on $L_2(\mathbb{R})$ to a wavelet basis for functions with bounded support $L_2([0,1])$, see for example (Daubechies et al., 1992, Chapter 10). Hereafter we consider periodizing the wavelet basis, since the log-spectra are periodic functions in $L_2([0,1])$. Also, without loss of generality we fix the primary resolution scale at $j_0 = 0$, and for ease of notation we compress the scale and location indices (j,m) into a single scale-location index k: here ψ_1 corresponds to the wavelet scaling function at resolution scale and location $(j,m) = (0,0)$, and ψ_k with $k > 1$ corresponds to the wavelet detail function at $(j,m) = (\lfloor \log_2(m-1)\rfloor, m - (2^{\lfloor \log_2(m-1)\rfloor} + 1)$. Since the log-periodograms are sampled over a discrete grid of frequencies, instead of true wavelet coefficients we compute empirical wavelet coefficients:

$$d_k = \langle f, \psi_k\rangle_T = \frac{1}{T}\sum_{\ell=0}^{T} f(x_\ell)\psi_k(x_\ell) = \int f(x)\psi_k(x)dx + o_T(1)$$

which can be written in matrix notation as $\boldsymbol{d} = \boldsymbol{W}_{\mathcal{B}}\boldsymbol{f}$, where $\boldsymbol{f} \in \mathbb{R}^T$ is a discrete sampled function and $\boldsymbol{W}_{\mathcal{B}}$ is an orthogonal $T \times T$-matrix associated to the wavelet basis $\mathcal{B} = \{\psi_k\}_{k=1}^{\infty}$. In practice, these coefficients are computed in $O(T)$ time using the discrete wavelet transform. Note that by orthogonality of $\boldsymbol{W}_{\mathcal{B}}$ the inverse transform is just, $\boldsymbol{f} = \boldsymbol{W}'_{\mathcal{B}}\boldsymbol{d}$, where $\boldsymbol{W}'_{\mathcal{B}}$ denotes the transpose of $\boldsymbol{W}_{\mathcal{B}}$.

Projecting the discrete sampled frequency domain model in eq.(2) onto the wavelet basis $\mathcal{B}$ via its discrete wavelet transform, we observe that there is a one-to-one relation between the function mixed model in the frequency domain and a linear mixed model in the wavelet coefficient domain given by,

$$\begin{aligned} Y_{sk} &= H_{sk} + \epsilon_{sk}, \qquad s = 1,\ldots,S, \quad k = 1,\ldots,T \\ &= h_k + U_{sk} + \epsilon_{sk} \end{aligned} \tag{4}$$

where,

(a) $\boldsymbol{h} = (h_1, \ldots, h_T)' = \boldsymbol{W}_{\mathcal{B}} \boldsymbol{h}^f \in \ell^2$ with $\boldsymbol{h}^f = (h^f(\omega_0), \ldots, h^f(\omega_{T-1}))' \in \mathbb{R}^T$. This is a sequence of fixed effect wavelet coefficients shared by all replicates in the population.
(b) $\boldsymbol{U}_s = (U_{s1}, \ldots, U_{sT})' = \boldsymbol{W}_{\mathcal{B}} \boldsymbol{U}_s^f$ with $\boldsymbol{U}_s^f = (U_s^f(\omega_0), \ldots, U_s^f(\omega_{T-1}))'$ for $s = 1, \ldots, S$. The replicate-specific vector $\boldsymbol{U}_s$ is a sequence of Gaussian random effects wavelet coefficients. The assumptions on the variance-covariance structure of the vectors $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_S$ are detailed below.
(c) $\boldsymbol{\epsilon}_s = (\epsilon_{s1}, \ldots, \epsilon_{sT})' = \boldsymbol{W}_{\mathcal{B}} \boldsymbol{E}_s^f$ with $\boldsymbol{E}_s^f = (E_s^f(\omega_0), \ldots, E_s^f(\omega_{T-1}))'$ for $s = 1, \ldots, S$. The vectors $\boldsymbol{\epsilon}_s$ are sequences of asymptotically independent wavelet noise coefficients with $\mathbb{E}[\epsilon_{sk}] = o_T(1)$ and $\mathrm{Var}(\epsilon_{sk}) = \sigma_e^2/T + o_T(T^{-1})$. The noise coefficients are independent between different replicates and independent of the random effects sequences for all s, k.

2.2. Covariance matrix assumptions

Let $\boldsymbol{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_S)'$ be the stacked $S \times T$-matrix of random effects sequences $\boldsymbol{U}_s$. One of our main interests is in allowing for explicit correlation between the random effects sequences of different replicates, therefore we will not assume the covariance matrix of $\mathrm{vec}(\boldsymbol{U})$ to be diagonal as is the case in Diggle and Al Wasel (1997), Krafty et al. (2011), and Freyermuth et al. (2010). However, some structure on the covariance matrix $\mathrm{Cov}(\mathrm{vec}(\boldsymbol{U}))$ is necessary, since consistent estimation in a totally unstructured matrix would be impossible (using only ST observations). We consider structural assumptions on the variance-covariance matrix of $\mathrm{vec}(\boldsymbol{U})$ as proposed in Morris and Carroll (2006) and Morris et al. (2008) in a general functional data analysis context. Since the frequency domain model and the wavelet coefficient domain model are equivalent representations, the structural assumptions in the wavelet domain automatically transfer to assumptions on the variance-covariance structure of the functional random effects in the frequency domain. We assume that the covariance matrix $\boldsymbol{G} := \mathrm{Cov}(\mathrm{vec}(\boldsymbol{U}))$ consists out of the Kronecker product of a $T \times T$ within-replicate diagonal covariance matrix and an $S \times S$ between-replicate correlation matrix:

$$\begin{aligned}
\boldsymbol{G} &= \begin{pmatrix} \sigma_{u1}^2 & 0 & \ldots & 0 \\ 0 & \sigma_{u2}^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_{uT}^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1S} \\ \rho_{21} & 1 & \ldots & \rho_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{S1} & \rho_{S2} & \ldots & 1 \end{pmatrix} \\
&:= \boldsymbol{G}_T \otimes \boldsymbol{G}_S
\end{aligned}$$

where $\boldsymbol{G}_S$ is symmetric and positive-semidefinite. By considering a diagonal *within-replicate* covariance matrix $\boldsymbol{G}_T$, the random effects coefficients are assumed to be uncorrelated between scale-locations $k = 1, \ldots, T$. The variance components are heterogeneous across coefficients, therefore allowing for very general spatially inhomogeneous behavior of the random effects curves in the frequency domain. The unstructured *between-replicate* correlation matrix $\boldsymbol{G}_S$ allows for correlation between the random effects coefficients of different replicates at matching scale-locations k. We observe that the correlation $\rho_{ss'}$ between two different replicates remains the same across all locations. This is in order to keep the dimensions of the working covariance matrices relatively small, but also perhaps more importantly to allow for consistent estimation of $\rho_{ss'}$ as the time series sample size increases. Note that $\boldsymbol{G}$ is symmetric and positive-semidefinite, since it is the Kronecker product of two symmetric positive-semidefinite matrices. Also, there is no identification problem between the two matrices $\boldsymbol{G}_T$ and $\boldsymbol{G}_S$, since $\boldsymbol{G}_S$ is restricted to have unit diagonal.

2.3. Functional space assumptions

In order to develop the necessary estimation theory, we impose some regularity (smoothness) conditions on the realized replicate-specific sequences in the wavelet coefficient domain, or equivalently, on the realized discretely sampled curves in the frequency domain. In particular, we assume that the fixed and random effects sequences are *asymptotically sparse* elements of the ℓ_0-sequence space with respect to the wavelet basis $\mathcal{B}$, defined as:

$$\ell_{0,T}(C) = \{\boldsymbol{x} \in \mathbb{R}^T : \|\boldsymbol{x}\|_0 = C\}$$

such that $\|\boldsymbol{x}\|_0 = \#\{k : x_k \neq 0\}$.

Assumption (A1). Let $\boldsymbol{h} \in \ell_{0,T}(k_{h,T})$, with set of indices of non-zero coefficients $K_{h,T}$. We assume that $k_{h,T} = |K_{h,T}| \to \infty$ as $T \to \infty$, but $k_{h,T} = o(T)$. Let $\boldsymbol{\sigma}_u^2 = (\sigma_{u1}^2, \ldots, \sigma_{uT}^2)' \in \ell_{0,T}(k_{u,T})$ with set of indices of non-zero coefficients $K_{u,T}$, such that $\sup_k \sigma_{uk}^2 < \infty$. We assume that $K_{u,T} \subseteq K_{h,T}$ and $k_{u,T} = |K_{u,T}| \to \infty$ as $T \to \infty$, but $k_{u,T} = o(T)$.

These regularity conditions assert that the fixed and realized random effects sequences or curves increase in complexity with T (almost surely for the random effects), but at a slower rate than T. Furthermore, we make the assumption $K_{u,T} \subseteq K_{h,T}$, this is a convenient way to ensure that the population-mean curve h^f and the realized replicate-specific curves $h_s^f | u_s^f = h^f + u_s^f$ share the same smoothness properties, which is an important requirement for the coherency of interpretation in a functional mixed effects model, as discussed in Guo (2002), Qin and Guo (2006), and Antoniadis and Sapatinas (2007). We note that the ℓ_0-constraints are placed on the empirical wavelet coefficients, not on the true wavelet coefficients. This means that we require smoothness conditions only on the discretely sampled curves (and not on the continuous curves) in the frequency domain.

2.3.1. *Classical ℓ_0-Gaussian sequence model*

Before presenting the estimation procedure, we recall some useful results on nonlinear thresholding methods in a classical Gaussian sequence model under ℓ_0-sparsity constraints. Consider the ℓ_0-Gaussian sequence model:

$$y_i = \theta_i + \epsilon_n z_i, \qquad i = 1, \ldots, n \tag{5}$$

with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)' \in \ell_{0,n}(k_n)$ and $z_1, \ldots, z_n \overset{\text{iid}}{\sim} N(0, 1)$. The minimax ℓ_2-risk of estimation for $\boldsymbol{\theta}$ in an ℓ_0-sequence model is defined as:

$$R(\ell_{0,n}(k_n), \epsilon_n) := \inf_{\hat{\boldsymbol{\theta}}} \sup_{\boldsymbol{\theta} \in \ell_{0,n}(k_n)} \mathbb{E}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$$

where $\|\cdot\|$ denotes the Euclidian norm. Under the ℓ_0-sparsity constraint $k_n \to \infty$ as $n \to \infty$, but $k_n = o(n)$, the minimax estimation risk for $\boldsymbol{\theta}$ satisfies,

$$R(\ell_{0,n}(k_n), \epsilon_n) \sim 2\epsilon_n^2 k_n \log(n/k_n)$$

and it is well-known that hard (or soft) nonlinear thresholding of the coefficients asymptotically achieves the minimax risk. In particular, the hard nonlinear thresholding estimator $\hat{\theta}_i = y_i \mathbf{1}\{|y_i| \geq \lambda_n\}$ for $i = 1, \ldots, n$, with $\lambda_n = \epsilon_n \sqrt{2\log(n/k_n)}$ is an asymptotic minimax estimator, see Johnstone (2015) for a detailed proof. We note that the nonlinear thresholding estimators $\{\hat{\theta}_i\}_{i=1,\ldots,n}$ are nonadaptive in the sense that the threshold λ_n depends on the (typically unknown) smoothness space parameter k_n. Abramovich et al. (2006) show that in the context of ℓ_0-Gaussian sequence models, using False Discovery Rate (FDR) nonlinear thresholding, it is possible to asymptotically achieve the minimax risk without requiring knowledge of the smoothness space parameter k_n. The FDR thresholding estimators are easily found through the following *step-up* procedure:

(a) Given observations $y_1, \ldots, y_n$ according to eq.(5), form the order statistics:

$$|\boldsymbol{y}|_{(1)} \leq |\boldsymbol{y}|_{(2)} \leq \ldots \leq |\boldsymbol{y}|_{(n)}$$

and compare them to the series of right-tail Gaussian quantiles $t_k = \epsilon_n \widetilde{\Phi}^{-1}(q/2 \cdot k/n)$, where $q \in (0, 1)$ is an FDR tuning parameter, and $\widetilde{\Phi}(t) = 1 - \Phi(t)$ with $\Phi(\cdot)$ the standard normal cumulative distribution function (CDF).

(b) Note the last crossing time:

$$k^* := \max\{k : |\boldsymbol{y}|_{(k)} \geq t_k\}$$

(c) Estimate $\boldsymbol{\theta}$ by the hard nonlinear thresholding estimator:

$$\hat{\theta}_i = y_i \mathbf{1}\{|y_i| \geq t_{k^*}\}, \qquad i = 1, \ldots, n \tag{6}$$

If the FDR tuning parameter is chosen as $q = q_n \in [\gamma/\log(n), 1/2]$ for some $\gamma > 0$, then Abramovich et al. (2006) show that under the same regularity condition as before, $k_n \to \infty$ as $n \to \infty$, but $k_n = o(n)$, the FDR thresholding estimator $\hat{\boldsymbol{\theta}}$ satisfies,

$$\sup_{\boldsymbol{\theta} \in \ell_{0,n}(k_n)} \mathbb{E}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = (1 + o(1)) R(\ell_{0,n}(k_n), \epsilon_n)$$

which implies that $\hat{\boldsymbol{\theta}}$ is an asymptotic minimax estimator.

3. Estimation procedure

3.1. Population-mean log-spectrum h^f

We estimate the population-mean log-spectrum $h^f(\omega_\ell)$ at frequencies $\omega_\ell \in [0, 1/2)$ by the projection estimator,

$$\hat{h}^f(\omega_\ell) = \sum_{k=1}^{T} \hat{h}_k(\boldsymbol{Y}_k)\psi_k(\omega_\ell), \qquad \ell = 0, \ldots, T-1$$

the inverse discrete wavelet transform with respect to the basis $\mathcal{B} = \{\psi_k\}_{k=1}^{\infty}$ of the estimated fixed effects sequence of coefficients $\hat{\boldsymbol{h}}(\boldsymbol{Y}) := \{\hat{h}_k(\boldsymbol{Y}_k)\}_{k=1}^{T}$, where $\boldsymbol{Y}_k = (Y_{1k}, \ldots, Y_{Sk})'$. The sequence $\hat{\boldsymbol{h}}(\boldsymbol{Y})$ is based on thresholded generalized least squares estimators,

$$\hat{h}_k(\boldsymbol{Y}_k) \quad = \quad (\boldsymbol{w}_k' \boldsymbol{Y}_k)\mathbf{1}\{k \in \widehat{K}_h(\boldsymbol{Y})\}, \qquad k = 1, \ldots, T$$

where,

(a) $\boldsymbol{w}_k = (w_{1k}, \ldots, w_{Sk})'$ are generalized least squares weights depending on the between-replicate correlation structure through $w_{sk} = \left(\sum_{i=1}^S \boldsymbol{V}_{k[i,s]}^{-1}\right)\left(\sum_{i,j=1}^S \boldsymbol{V}_{k[i,j]}^{-1}\right)^{-1}$. Here, $\boldsymbol{V}_k$ denotes the asymptotic covariance matrix of $\boldsymbol{Y}_k$ given by $\boldsymbol{V}_k = \sigma_{uk}^2 \boldsymbol{G}_S + \frac{\sigma_e^2}{T}\mathrm{I}_S$, with I_S the $S \times S$-identity matrix.

(b) $\widehat{K}_h(\boldsymbol{Y}) = \{k : |\frac{1}{S}\sum_{s=1}^S Y_{sk}| \geq \lambda_{h,T}\}$ is the estimated set of indices of non-zero coefficients with universal threshold $\lambda_{h,T} = \sqrt{\sigma_e^2/(ST)}\sqrt{2\log(T/k_{h,T})}$. The motivation for this thresholded set comes from the observation that the replicate-specific sequences – conditional on the random effects– are *independent* between replicates and, individually for each replicate, follow an $\ell_{0,T}(k_{h,T})$-sequence model with the same set of non-zero coefficients $K_{h,T}$ for each replicate and noise variance approximately σ_e^2/T. In the unconditional case, it follows that the distributional behavior of the sequences at scale-locations of zero coefficients ($k \notin K_{h,T}$) remains the same, which allows for the same control on the number of false positives as in the conditional case. Moreover, under some regularity conditions, the empirical wavelet noise coefficients are asymptotically normal ($T \to \infty$), which asymptotically justifies the threshold choice $\lambda_{h,T}$ based on a Gaussian sequence model (see Section 4.1 for a more detailed discussion).

3.2. Random-effects covariance matrices $\boldsymbol{G}_T$ and $\boldsymbol{G}_S$

3.2.1. *Estimation of the vector of variance components $\boldsymbol{\sigma}_u^2$*

The within-replicate random effects covariance matrix $\boldsymbol{G}_T$ is assumed to be diagonal, with vector of variance components $\boldsymbol{\sigma}_u^2 = (\sigma_{u1}^2, \ldots, \sigma_{uT}^2)' \in \ell_{0,T}(k_{u,T})$ on the diagonal. This vector is estimated by the thresholded sample-variances,

$$\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k) = \left\{\frac{1}{S}\sum_{s=1}^S \left(Y_{sk} - \hat{h}_k(\boldsymbol{Y}_k)\right)^2 - \frac{\sigma_e^2}{T}\right\}_+ \mathbf{1}\left\{k \in \widehat{K}_u(\boldsymbol{Y})\right\}$$

with estimated set of indices of non-zero variance components $\widehat{K}_u(\boldsymbol{Y}) = \{k : |T_k(\boldsymbol{Y}_k)| \geq \lambda_{u,T}\}$ such that,

$$\begin{aligned} T_k(\boldsymbol{Y}_k) &= \log\left(\frac{1}{S}\sum_{s=1}^S \left(Y_{sk} - \hat{h}_k(\boldsymbol{Y}_k)\right)^2\right) - \log\left(\frac{2\sigma_e^2}{ST}\right) - \psi^{(0)}\left(\frac{S}{2}\right) \qquad (7) \\ \lambda_{u,T} &= \sqrt{\psi^{(1)}(S/2)}\sqrt{2\log(T)} \end{aligned}$$

where $\psi^{(0)}(\cdot)$ and $\psi^{(1)}(\cdot)$ denote the digamma and trigamma function. The motivation for this thresholded set, which has similar structure as the thresholded set $\widehat{K}_h(\boldsymbol{Y})$, comes from the observation that –for uncorrelated replicates– the vector $\{T_k(\boldsymbol{Y}_k)\}_{k=1}^T$ is variance stabilizing and behaves approximately as an $\ell_{0,T}(k_{u,T})$-Gaussian sequence model with noise variance $\psi^{(1)}(S/2)$. This justifies the threshold choice $\lambda_{u,T}$ based on a Gaussian sequence model. For a general between-replicate correlation matrix $\boldsymbol{G}_S$, it follows that the distributional behavior of the zero variance components with indices $k \notin K_{u,T}$ remains the same, and this allows for the same control on the number of false positives as in the uncorrelated case. We note that the threshold

$\lambda_{u,T}$ is slightly more conservative than the asymptotic minimax universal threshold as in Section 2.3.1. The reason for this is that in the context of a Gaussian sequence model, under the more conservative threshold, *both* the number of false positives and the number of false negatives in the estimated set of indices of non-zero variance components tend to zero almost surely (Corollary 1). Under the asymptotic minimax threshold, this only holds true for the number of false negatives.

3.2.2. *Estimation of the between-replicate correlation coefficients ρ_{ij}*

The between-replicate correlation matrix $\boldsymbol{G}_S$ is estimated elementwise by considering the following sample-correlation based estimators,

$$\hat{\rho}_{ij}(\boldsymbol{Y}) = \frac{1}{\hat{k}_u} \sum_{k \in \widehat{K}_u} \frac{(Y_{ik} - \hat{h}_k(\boldsymbol{Y}_k))(Y_{jk} - \hat{h}_k(\boldsymbol{Y}_k))}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta}, \qquad i, j = 1, \ldots, S, \quad i \neq j \tag{8}$$

where $\delta > 0$ is a small constant that ensures that the denominator is bounded away from zero, and $\widehat{K}_u = \widehat{K}_u(\boldsymbol{Y})$ is the estimated set of indices of non-zero variance components with cardinality $\hat{k}_u = |\widehat{K}_u(\boldsymbol{Y})|$. The intuition behind this estimator comes from the fact that $\mathrm{Cov}(Y_{ik}, Y_{jk}) = \sigma^2_{uk}\rho_{ij}$ for each $k \in K_{u,T}$, whereas $\mathrm{Cov}(Y_{ik}, Y_{jk}) = 0$ for each $k \notin K_{u,T}$ due to $\sigma^2_{uk} = 0$.

3.2.3. *Finite-sample correction of $\widehat{\boldsymbol{G}}_S$*

The estimated matrix $\widehat{\boldsymbol{G}}_S(\boldsymbol{Y})$ with off-diagonal elements $\hat{\rho}_{ij}(\boldsymbol{Y})$ is only an approximate correlation matrix. It is symmetric and has unit diagonal, but is not guaranteed to be positive-semidefinite. By Higham (2002), we compute the correlation matrix $\widetilde{\boldsymbol{G}}_S(\boldsymbol{Y})$ with minimum distance in Frobenius-norm to the originally estimated matrix $\widehat{\boldsymbol{G}}_S(\boldsymbol{Y})$,

$$\widetilde{\boldsymbol{G}}_S(\boldsymbol{Y}) = \arg\min_{\boldsymbol{X}=\boldsymbol{X}^T} \{ \|\widehat{\boldsymbol{G}}_S(\boldsymbol{Y}) - \boldsymbol{X}\|_F : \boldsymbol{X} \text{ is a correlation matrix} \}$$

However, replacing the matrix $\widehat{\boldsymbol{G}}_S(\boldsymbol{Y})$ by the new matrix $\widetilde{\boldsymbol{G}}_S(\boldsymbol{Y})$, the estimated variance components $\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k)$ are no longer properly scaled (i.e. $\hat{\sigma}^2_{uk}\widehat{\boldsymbol{G}}_S \neq \hat{\sigma}^2_{uk}\widetilde{\boldsymbol{G}}_S$). Instead, we consider the rescaled estimators $\tilde{\sigma}^2_{uk}(\boldsymbol{Y})$ satisfying $\|\hat{\sigma}^2_{uk}\widehat{\boldsymbol{G}}_S\|_F = \|\tilde{\sigma}^2_{uk}\widetilde{\boldsymbol{G}}_S\|_F$, which are easily obtained through,

$$\tilde{\sigma}^2_{uk}(\boldsymbol{Y}) = \hat{\sigma}^2_{uk}(\boldsymbol{Y}_k)\frac{\|\widehat{\boldsymbol{G}}_S(\boldsymbol{Y})\|_F}{\|\widetilde{\boldsymbol{G}}_S(\boldsymbol{Y})\|_F}$$

Note that the rescaling does not affect the estimated zero variance components corresponding to $k \notin \widehat{K}_u(\boldsymbol{Y})$, i.e. if $\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) = 0$ then also $\tilde{\sigma}^2_{uk}(\boldsymbol{Y}) = 0$.

3.3. *Iterative estimation scheme*

In order to estimate the population-mean log-spectrum $h^f(\omega_\ell)$, we consider a generalized least squares estimator with weights depending on the between-replicate correlation structure. On the other hand, estimation of the random effects covariance and correlation matrices $\boldsymbol{G}_T$ and $\boldsymbol{G}_S$ depends on the population-mean sequence $\boldsymbol{h}$, since the sample-variances and sample-correlations need to be centered about their respective means. This is typically the case in linear mixed mdoel estimation where one allows for between-replicate correlation, and in this context. one easy-to-implement approach is to consider an iterative-generalized least squares scheme (see e.g. Jiang (2007)). First, we compute the thresholded ordinary least squares estimator of $\boldsymbol{h}$ by equally weighting each of the observations across replicates. This does not require any information on the between-replicate correlation structure. Next, we iterate between estimation of $\boldsymbol{G}_T$ and $\boldsymbol{G}_S$ given the estimate of $\boldsymbol{h}$, and estimation of $\boldsymbol{h}$ given estimates of $\boldsymbol{G}_T$ and $\boldsymbol{G}_S$, and we continue iterating until some convergence criterion is satisfied.

Remark. Under a similar random effects variance-covariance structure, considering only the linear part of the estimators (without the thresholding), Jiang et al. (2007) show that an iterative-generalized least squares scheme converges exponentially with probability tending to one as the number of replicates S increases.

3.4. Replicate-specific log-spectra H_s^f

The replicate-specific log-spectra $H_s^f(\omega_\ell)$ at frequencies $\omega_\ell \in [0, 1/2)$ are predicted by the projection estimators,

$$\widehat{H}_s^f(\omega_\ell) = \sum_{k=1}^{T} \widehat{H}_{sk}(\boldsymbol{Y})\psi_k(\omega_\ell), \qquad s = 1, \ldots, S, \quad \ell = 0, \ldots, T-1$$

the inverse discrete wavelet transform with respect to the basis $\mathcal{B}$ of the predicted replicate-specific sequence of coefficients $\widehat{\boldsymbol{H}}_s(\boldsymbol{Y}) := \{\widehat{H}_{sk}(\boldsymbol{Y})\}_{k=1}^T$. For ease of notation we drop the dependence of the estimators on $\boldsymbol{Y}$, which should be clear from the context. Prediction in the wavelet coefficient domain reduces to prediction in a linear mixed model, and we can find estimated predictors of the random effects sequences $\boldsymbol{U}_k = (U_{1k}, \ldots, U_{Sk})'$ through,

$$\widehat{\boldsymbol{U}}_k(\boldsymbol{Y}) = \widehat{\boldsymbol{G}}_k \widehat{\boldsymbol{V}}_k^{-1}(\boldsymbol{Y}_k - \hat{h}_k \mathbf{1}_S)$$

Here, $\widehat{\boldsymbol{G}}_k = \tilde{\sigma}_{uk}^2 \widetilde{\boldsymbol{G}}_S$ and $\widehat{\boldsymbol{V}}_k = \tilde{\sigma}_{uk}^2 \widetilde{\boldsymbol{G}}_S + \frac{\sigma_e^2}{T} \mathrm{I}_S$ are plug-in estimators, and $\mathbf{1}_S$ denotes an S-dimensional vector of ones. Note that if $\widehat{\boldsymbol{G}}_k$ and $\widehat{\boldsymbol{V}}_k^{-1}$ are replaced by the true matrices $\boldsymbol{G}_k$ and $\boldsymbol{V}_k^{-1}$, and $\hat{\boldsymbol{h}}$ is replaced by the best linear unbiased estimator, then the $\widehat{\boldsymbol{U}}_k$ are the best linear unbiased predictors of the random effects sequences $\boldsymbol{U}_k$ in a linear mixed model, see Searle et al. (1992). In general, due to the nonlinear thresholding of coefficients, $\hat{\boldsymbol{h}}$ ceases to be an unbiased estimator of $\boldsymbol{h}$. By combining the estimator for the fixed effects sequence $\boldsymbol{h}$ and the predictors for the random effects sequences $\boldsymbol{U}_k$, the replicate-specific sequences of coefficients $\boldsymbol{H}_s = \{h_k + U_{sk}\}_{k=1}^T$ are predicted through,

$$\widehat{\boldsymbol{H}}_s(\boldsymbol{Y}) = \left\{\hat{h}_k(\boldsymbol{Y}_k) + \widehat{U}_{sk}(\boldsymbol{Y})\right\}_{k=1}^T$$

4. Estimation theoretical results

4.1. Risk bounds on the estimator of h^f

In this section we derive finite-sample upper bounds on the ℓ_2-risk of the estimated fixed effects sequence of coefficients $\hat{\boldsymbol{h}}(\boldsymbol{Y}) = \{\hat{h}_k(\boldsymbol{Y}_k)\}_{k=1}^T$ with respect to $\boldsymbol{h}$. By Parseval's relation, the

ℓ_2-risk in the wavelet coefficient domain is approximately equal to the L_2-risk of the projected estimators in the frequency domain,

$$\mathbb{E}\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2 = \frac{1}{T}\mathbb{E}\|\hat{\boldsymbol{h}}^f - \boldsymbol{h}^f\|^2 \to \mathbb{E}\|\hat{h}^f - h^f\|_{L_2}^2, \qquad \text{as } T \to \infty$$

Thus the same expression also gives an upper bound for the L_2-risk of the estimated population-mean log-spectrum $\hat{h}^f$ as the error in approximating the integral by a sum becomes negligible. The proof is based on the observation that, under some regularity conditions, the non-Gaussian linear mixed model in the wavelet coefficient domain (eq.(4)) is asymptotically equivalent to a Gaussian linear mixed model ($T \to \infty$). This allows us to calculate the ℓ_2-risk of the sequence $\hat{\boldsymbol{h}}(\boldsymbol{Y})$ first under an accompanying Gaussian sequence model, and relate this to the ℓ_2-risk under the non-Gaussian sequence model. The asymptotic equivalence between the two models is based on a uniform asymptotic normality result for the empirical wavelet noise coefficients. Neumann (1996) establishes this uniform asymptotic normality for the empirical wavelet noise coefficients of periodogram ordinates for a general non-Gaussian process $X(t)$ in the context of a single long time series. Since the technical considerations in Neumann (1996) are not the main focus of this paper, here we derive uniform asymptotic normality of the empirical wavelet noise coefficients of log-periodogram ordinates only for weakly dependent Gaussian processes $X_s(t)$ as in (Brillinger, 1981, Chapter 5) by combining limit theorems from Rudzkis et al. (1978) and cumulant bounds derived by Taniguchi (1979). Asymptotic normality of empirical wavelet noise coefficients of the log-periodogram has already been suggested without proof by Gao (1997), Moulin (1994), and Freyermuth et al. (2010) under the approximate additive noise model with $\epsilon_k \sim (0, \pi^2/6)$. In order for a certain summation effect to work it will only be possible to derive asymptotic normality for empirical wavelet coefficients bounded away from the finest wavelet scale. For this reason, we make the additional assumption that, for increasing T, the set of non-zero fixed effects coefficients is bounded away from the finest wavelet scale. This means that the finest wavelet scale contains virtually only noise and no signal as T increases, which is a common assumption in the wavelet literature. In an ordinary signal plus noise model, this allows for instance for estimation of the noise variance via the empirical coefficients located at the finest wavelet scale (see Vidakovic (1999)).

Assumption (A2). Define the set,

$$J_{T,\alpha} := \{1\} \cup \{k \geq 2 \,|\, 2^{\lfloor \log_2(k-1) \rfloor} \leq CT^{1-\alpha}\} \tag{9}$$

for some constant $C > 0$. We assume that there exist some $T^* > 0$ and $0 < \alpha^* < 1$, such that for $T \geq T^*$, $K_{h,T} \subseteq J_{T,\alpha^*}$.

Assumption (A3). The replicated time series $\{X_s(t)\}_{t>0}$ are Gaussian processes such that $\sum_{h=-\infty}^{\infty} |h||\mathrm{Cov}(X_s(t), X_s(t+h))| < \infty$ for all $s = 1, \ldots, S$.

Theorem 1. Under (A1)-(A3), let α such that $0 < \alpha \leq \alpha^*$. We consider the estimators,

$$\hat{h}_k(\boldsymbol{Y}_k) = \left(\boldsymbol{w}_k' \boldsymbol{Y}_k\right) \mathbf{1}\{|\bar{\boldsymbol{Y}}_k| \geq \lambda_{h,T}, k \in J_{T,\alpha}\}, \qquad k = 1, \ldots, T$$

where $\bar{\boldsymbol{Y}}_k = \frac{1}{S}\sum_{s=1}^{S} Y_{sk}$ and $\lambda_{h,T} = \sqrt{\sigma_e^2/(ST)}\sqrt{2\log(T/k_{h,T})}$. For T sufficiently large, the

ℓ_2-risk of $\hat{\boldsymbol{h}}(\boldsymbol{Y})$ satisfies,

$$\sup_{\boldsymbol{h} \in \ell_{0,T}(k_{h,T})} \mathbb{E}\|\hat{\boldsymbol{h}}(\boldsymbol{Y}) - \boldsymbol{h}\|^2 \lesssim \frac{k_{h,T}}{ST} \log\left(\frac{T}{k_{h,T}}\right) + k_{u,T}\left(\sup_k \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k - \frac{\sigma_e^2}{TS}\right) \tag{10}$$

where $\lesssim$ denotes the inequality $\leq$ up to a multiplicative positive constant.

The first term on the right-hand side in eq.(10) is equivalent to the minimax rate of estimation in an $\ell_{0,T}(k_{h,T})$-Gaussian sequence model with noise variance of order $(ST)^{-1}$. The second term arises from introducing the random effects and is an upper bound on the integrated error made in estimating $\boldsymbol{h}$ by taking a weighted sample average over a finite number of replicates. We observe that $\boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k = \mathrm{Var}(\boldsymbol{w}_k'\boldsymbol{Y}_k)$ for $\boldsymbol{Y}_k \sim (h_k\boldsymbol{1}_S, \boldsymbol{V}_k)$, and this term is minimized by the generalized least squares weights $\boldsymbol{w}_k$ according to Section 3.1. We note that for sub-optimal ordinary least squares weights $\boldsymbol{w}_k = (\frac{1}{S}, \ldots, \frac{1}{S})'$ the second term reduces to:

$$k_{u,T}\left(\sup_k \boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k - \frac{\sigma_e^2}{ST}\right) \quad = \quad k_{u,T} \sup_k \frac{\sigma_{uk}^2 \boldsymbol{1}_S' \boldsymbol{G}_S \boldsymbol{1}_S}{S^2}$$

where $\boldsymbol{1}_S'\boldsymbol{G}_S\boldsymbol{1}_S$ is just the entrywise sum over all elements in $\boldsymbol{G}_S$. This implies that, if $\frac{k_{u,T}}{S^2}\boldsymbol{1}_S'\boldsymbol{G}_S\boldsymbol{1}_S \to 0$ as $S, T \to \infty$, the thresholded ordinary least squares estimator remains a consistent estimator of $\boldsymbol{h}$. In the case of uncorrelated replicates this term is for instance of the order $k_{u,T}/S$. Note that, although being nonnegative, the term $\boldsymbol{1}_S'\boldsymbol{G}_S\boldsymbol{1}_S$ is not a matrix norm, moreover it decreases as replicates become more negatively correlated. This might seem surprising, but can be illustrated by the following simple bi-replicate example: suppose one observes two random replicate-curves that are highly negatively correlated, with high probability the true population mean-curve lies in between the two replicate-curves and the error term due the random effects should therefore be smaller than in the independent curve situation; if the curves are perfectly negatively correlated, the true population mean-curves lies exactly in between the two replicate-curves, and the error term due to random effects should disappear completely.

4.1.1. *Key components in the proof of* Theorem 1

We outline the proof of Theorem 1 through several key lemmas. The first lemma gives the uniform asymptotic normality result for empirical wavelet coefficients of the log-periodogram ordinates and relates the ℓ_2-risk of $\hat{\boldsymbol{h}}_\lambda(\boldsymbol{Y})$ under the sequence model in the wavelet coefficient domain (eq.(4)) to the ℓ_2-risk of $\hat{\boldsymbol{h}}_\lambda(\boldsymbol{\xi})$ under an accompanying Gaussian sequence model. Here, $\lambda \geq 0$ is an arbitrary nonlinear threshold.

Lemma 1. Under (A1) and (A3), uniformly in $k \in J_{T,\alpha}$, for arbitrary $0 < \alpha < 1$,

$$P\left((Y_{sk} - h_k)/\sigma_{k,T} \geq x_s\right) = (1 + o_T(1))(1 - \Phi(x_s)) \qquad \text{for all } s = 1, \ldots, S$$

with $-\infty < x_s \leq \Delta_T \sim T^\nu$ for some $\nu > 0$, where $\sigma_{k,T}^2 = \sigma_{uk}^2 + \sigma_e^2/T$. Furthermore,

$$\sum_{k \in J_{T,\alpha}} \mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^2] = (1 + o_T(1)) \sum_{k \in J_{T,\alpha}} \mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] + O(S^{-\mu}T^{-\mu+1})$$

for arbitrary $0 < \mu < \infty$, and with $\boldsymbol{\xi}_k \sim N(h_k\boldsymbol{1}_S, \boldsymbol{V}_k)$ Gaussian random vectors.

This lemma implies that it suffices to derive an ℓ_2-risk upper bound of $\hat{\boldsymbol{h}}(\boldsymbol{Y})$ under the accompanying Gaussian sequence model. In the lemma below, we derive exact expressions of the mean-squared errors of $\hat{h}_k(\boldsymbol{\xi}_k)$ with respect to h_k under the Gaussian model, which is equivalent to exact normality of the empirical wavelet noise coefficients.

Lemma 2. Suppose that $\epsilon_{1k}, \ldots, \epsilon_{Sk} \overset{\text{iid}}{\sim} N(0, \sigma_e^2/T)$, such that $\boldsymbol{\xi}_k \sim N(h_k \mathbf{1}_S, \boldsymbol{V}_k)$ for each $k = 1, \ldots, T$. An exact expression of the mean squared error of $\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k)$ with threshold $\lambda \geq 0$ is given by,

$$\begin{aligned}\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] \;=\;& \left(2\boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k - h_k^2\right) + \left(h_k^2 - \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k\right) \left(\Phi\left(\frac{\sqrt{S}(\lambda - h_k)}{\tilde{\sigma}_{k,T}}\right)\right. \\ & \left. + \Phi\left(\frac{\sqrt{S}(\lambda + h_k)}{\tilde{\sigma}_{k,T}}\right)\right) + \frac{(\boldsymbol{w}_k' \boldsymbol{V}_k \mathbf{1}_S)^2}{S\tilde{\sigma}_{k,T}^2}\left(\frac{\sqrt{S}(\lambda - h_k)}{\tilde{\sigma}_{k,T}}\right)\phi\left(\frac{\sqrt{S}(\lambda - h_k)}{\tilde{\sigma}_{k,T}}\right) \\ & + \frac{(\boldsymbol{w}_k' \boldsymbol{V}_k \mathbf{1}_S)^2}{S\tilde{\sigma}_{k,T}^2}\left(\frac{\sqrt{S}(\lambda + h_k)}{\tilde{\sigma}_{k,T}}\right)\phi\left(\frac{\sqrt{S}(\lambda + h_k)}{\tilde{\sigma}_{k,T}}\right)\end{aligned}$$

where $\tilde{\sigma}_{k,T}^2 = \frac{\sigma_{uk}^2}{S} \mathbf{1}_S' \boldsymbol{G}_S \mathbf{1}_S + \sigma_e^2/T$, and $\phi(\cdot)$ denotes the standard normal probability density function. Furthermore, for each k

$$\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] \leq \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k + \lambda^2$$

These upper bounds are sharp when either $\lambda/\tilde{\sigma}_{k,T} \to \infty$, in which case $\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] \uparrow \lambda^2$, or when $\lambda/\tilde{\sigma}_{k,T} \to 0$, in which case $\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] \uparrow \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k$.

The derivations follow from straightforward calculus and the proofs are therefore omitted. The expression for the mean squared error generalize an expression for the mean squared error of a nonlinear hard threshold estimator in a classical Gaussian sequence model (eq.(5)) found in Bruce and Gao (1996). Note that we find back the expression in Bruce and Gao (1996) when $S = 1$ and $\tilde{\sigma}_{k,T}^2 = 1$. The main result in Theorem 1 on the ℓ_2-risk of $\hat{\boldsymbol{h}}(\boldsymbol{Y})$ now follows from combining Lemma 1, the upper bounds in Lemma 2, and plugging in the threshold $\lambda_{h,T}$. The complete proof is deferred to the Appendix.

4.2. Consistent estimation of $\boldsymbol{G}_T$ and $\boldsymbol{G}_S$

In this section, we derive some asymptotic results for the estimators $\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k)$ of the variance components in the within-replicate covariance matrix $\boldsymbol{G}_T$, and the estimators $\hat{\rho}_{ij}(\boldsymbol{Y})$ of the correlation coefficients in the between-replicate correlation matrix $\boldsymbol{G}_S$. Theorem 2 below gives the main results on consistent estimation of the set of non-zero variance components $K_{u,T}$, uniform consistency of the estimated variance components, and marginal consistency of the estimated between-replicate correlation coefficients for an appropriately chosen threshold $\lambda_{u,T}$. The derived results crucially rely on the condition $\|\boldsymbol{G}_S\|_F = o(S)$, which controls the level of correlation between different replicates. Essentially it requires that the effective number of uncorrelated replicates increases with the total number of replicates S. To illustrate, for uncorrelated replicates $\|\boldsymbol{G}_S\|_F/S = 1/\sqrt{S}$, whereas for perfectly correlated replicates $\|\boldsymbol{G}_S\|_F/S = 1$.

In order to simplify the proofs, as in Krafty et al. (2011), Gao (1997), Moulin (1994), and Freyermuth et al. (2010), we work under the approximate model where the empirical wavelet noise coefficients are mean zero with variance σ_e^2/T, which is the asymptotic version of the model as $T \to \infty$. Note that we do not assume normality of the empirical wavelet noise coefficients, nor independence between different scale-locations k within a single replicate.

Theorem 2. Suppose that $\epsilon_{sk} \sim (0, \sigma_e^2/T)$ with $\mathbb{E}[\epsilon_{sk}^4] = O(T^{-2})$ for each $s = 1, \ldots, S$ and $k = 1, \ldots, T$, and that there exist uniform consistent estimators $\sup_k |\hat{h}_k - h_k| = o_p^{S,T}(1)$, with $|\hat{h}_k - h_k| = o_p^{S,T}(T^{-1/2})$ for $k \notin K_{u,T}$. If $\|\boldsymbol{G}_S\|_F = o(S)$, and $C \le \lambda_{u,T} = o(\log(T))$ for some constant $C > 0$, then

$$P(\widehat{K}_u(\boldsymbol{Y}) = K_{u,T}) \quad \to \quad 1, \qquad \text{as } S, T \to \infty \tag{11}$$

$$\sup_k |\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k) - \sigma_{uk}^2| \quad \xrightarrow{P} \quad 0, \qquad \text{as } S, T \to \infty \tag{12}$$

If $0 < \delta \le \inf_{k \in K_{u,T}} \sigma_{uk}^2$ for $\hat{\rho}_{ij}(\boldsymbol{Y})$ in eq.(8), then for each i, j with $i \neq j$,

$$\hat{\rho}_{ij}(\boldsymbol{Y}) \quad \xrightarrow{P} \quad \rho_{ij}, \qquad \text{as } S, T \to \infty$$

The conditions $\mathbb{E}[\epsilon_{sk}^4] = O(T^{-2})$ and $|\hat{h}_k - h_k| = o_p^{S,T}(T^{-1/2})$ for $k \notin K_{u,T}$ are needed in order to control the number of false positives in the set of estimated non-zero variance components in eq.(11). Under (A1)-(A3), by the cumulant bounds derived in the proof of Theorem 1 (see eq.(30)), it follows that $\mathbb{E}[\epsilon_{sk}^4] = O(T^{-2})$ for each s, k. Furthermore, if $\log(T)/S \to 0$ as $S, T \to \infty$, it can be verified that $|\hat{h}_k - h_k| = o_p^{S,T}(T^{-1/2})$ for $k \notin K_{u,T}$ for the nonlinear estimators $\hat{h}_k(\boldsymbol{Y}_k)$ in Theorem 1 with ordinary least squares weights $\boldsymbol{w}_k = (\frac{1}{S}, \ldots, \frac{1}{S})'$.
The following corollary gives the theoretical justification for the statistics $T_k(\boldsymbol{Y}_k)$, which are based on the accompanying Gaussian sequence model where the empirical wavelet noise coefficients are exactly normally distributed. Under the Gaussian sequence model, it is possible to adopt the threshold $\lambda_{u,T} = \sqrt{\psi^{(1)}(S/2)}\sqrt{2\log(T - k_{u,T})}$, which converges to zero if $\log(T)/S \to 0$ as $S, T \to \infty$, while preserving the consistency result that both the number of false positives and false negatives in the estimated set of non-zero variance components is zero with probability tending to one.

Corollary 1. Suppose that $\epsilon_{sk} \overset{\text{iid}}{\sim} N(0, \sigma_e^2/T)$ for each $s = 1, \ldots, S$ and $k = 1, \ldots, T$, such that $\boldsymbol{\xi}_k \sim N(h_k \boldsymbol{1}_S, \boldsymbol{V}_k)$, and consider the statistics $T_k(\boldsymbol{\xi}_k)$ as in eq.(7) with $\hat{h}_k(\boldsymbol{\xi}_k)$ replaced by the true coefficients h_k. For uncorrelated replicates ($\boldsymbol{G}_S = \mathrm{I}_S$), the vector $\{T_k(\boldsymbol{\xi}_k)\}_{k=1}^T$ converges to an $\ell_{0,T}(k_{u,T})$-Gaussian sequence model as $S \to \infty$,

$$\begin{cases} \dfrac{1}{\sqrt{\psi^{(1)}(S/2)}} \left(T_k(\boldsymbol{\xi}_k) - \log\left(\frac{\sigma_{uk}^2 T + \sigma_e^2}{\sigma_e^2} \right) \right) \quad \xrightarrow{d} \quad N(0,1), \quad \text{if } k \in K_{u,T} \\ \dfrac{1}{\sqrt{\psi^{(1)}(S/2)}} T_k(\boldsymbol{\xi}_k) \quad \xrightarrow{d} \quad N(0,1), \quad \text{if } k \notin K_{u,T} \end{cases} \tag{13}$$

For correlated replicates, with general correlation matrix $\boldsymbol{G}_S$, it remains true that for $S \to \infty$,

$$\frac{1}{\sqrt{\psi^{(1)}(S/2)}} T_k(\boldsymbol{\xi}_k) \quad \xrightarrow{d} \quad N(0,1), \quad \text{if } k \notin K_{u,T}$$

If $\|\boldsymbol{G}_S\|_F = o(S)$ and $\sqrt{\psi^{(1)}(S/2)}\sqrt{2\log(T-k_{u,T})} \leq \lambda_{u,T} = o(\log(T))$, then as in eq.(11)

$$P(\widehat{K}_u(\boldsymbol{\xi}) = K_{u,T}) \to 1 \qquad \text{as } S, T \to \infty.$$

5. Confidence regions

In this section, we develop asymptotic confidence regions for the discretely sampled population mean log-spectrum, where it is important to take into account the possible correlation between different replicate-specific curves to avoid the use of erroneous confidence sets. In the wavelet coefficient domain, the estimated sequence of fixed effect coefficients is a nonlinear biased estimator of $\boldsymbol{h}$, and for this reason it is generally difficult to derive asymptotic confidence bounds directly from the asymptotic distribution of $\hat{\boldsymbol{h}}(\boldsymbol{Y})$, even under Gaussian model assumptions. As proposed in Robins and van der Vaart (2006) and Genovese and Wasserman (2005) among others, instead we consider an estimator of the squared norm $\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2$ (conditional on $\hat{\boldsymbol{h}}$), and we derive the asymptotic distribution of this estimator instead of the asymptotic distribution of the original estimator $\hat{\boldsymbol{h}}$. Asymptotic ℓ_2-confidence regions for $\boldsymbol{h}$ can then be constructed by restricting the norm of $\boldsymbol{h}$ with respect to the estimated sequence $\hat{\boldsymbol{h}}$. Moreover, by the norm equivalence between the functional domain and wavelet coefficient domain, we can easily transfer the confidence regions for $\boldsymbol{h}$ in the wavelet domain to confidence regions for $\boldsymbol{h}^f$ in the frequency domain. For convenience, we work under the approximate model assumption that the wavelet noise coefficients are exactly normally distributed, with mean zero and variance σ_e^2/T. The derived confidence regions are therefore approximate in the sense that they are based on asymptotic distributional behavior of the estimator of the pivot quantity ($S \to \infty$), but also on the fact that the empirical wavelet noise coefficients are only asymptotically normally distributed ($T \to \infty$) under appropriate model conditions as discussed in Section 4.1.

The method is based on the assumption that we can split the sample $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T)' \in \mathbb{R}^{ST}$, with $\boldsymbol{\xi}_k \sim N(h_k \mathbf{1}_S, \boldsymbol{V}_k)$, into two sets of independent observations $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)} \in \mathbb{R}^{ST}$. Suppose that the covariance matrices $\boldsymbol{V}_k$ are known, one simple approach to split the sample into two independent samples at the cost of making the variance twice as large is to consider,

$$\boldsymbol{\xi}_k^{(1)} := \boldsymbol{\xi}_k - \boldsymbol{X}_k, \qquad \boldsymbol{\xi}_k^{(2)} := \boldsymbol{\xi}_k + \boldsymbol{X}_k \qquad \text{for all } k = 1, \ldots, T$$

where the vectors $\boldsymbol{X}_k \sim N(\mathbf{0}, \boldsymbol{V}_k)$ are independent of $\boldsymbol{\xi}_k$. It is easily verified that $\boldsymbol{\xi}_k^{(1)}, \boldsymbol{\xi}_k^{(2)} \sim N(h_k \mathbf{1}_S, 2\boldsymbol{V}_k)$ and $\boldsymbol{\xi}_k^{(1)}, \boldsymbol{\xi}_k^{(2)}$ are independent random vectors for each $k = 1, \ldots, T$. We estimate the sequence $\boldsymbol{h}$ using only the observations in $\boldsymbol{\xi}^{(1)}$, and construct the confidence regions from the additional independent set of observations $\boldsymbol{\xi}^{(2)}$ conditional on $\hat{\boldsymbol{h}}(\boldsymbol{\xi}^{(1)})$. The nature of the estimator $\hat{\boldsymbol{h}}$ is irrelevant for the construction of the confidence region, however, since the radius of the confidence region is proportional to $\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|$, better estimators $\hat{\boldsymbol{h}}$ (in terms of ℓ_2-risk) will lead to smaller confidence regions. We note that the same procedure can be carried out the other way around, that is, compute $\hat{\boldsymbol{h}}$ based only on the observations $\boldsymbol{\xi}^{(2)}$, and then construct confidence regions based on $\boldsymbol{\xi}^{(1)}$ given $\hat{\boldsymbol{h}}(\boldsymbol{\xi}^{(2)})$. In the remainder, we consider constructing confidence regions for $\boldsymbol{h}$ only through the first procedure.

Suppose that we have split the sample into two independent parts $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}$ and have computed $\hat{\boldsymbol{h}} := \hat{\boldsymbol{h}}(\boldsymbol{\xi}^{(1)})$. The next step is to find an estimator of the pivot quantity $\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2$, conditional on

$\hat{\boldsymbol{h}}$, using only the set of observations $\boldsymbol{\xi}^{(2)}$. We consider the unbiased estimator,

$$\widehat{R}(\boldsymbol{\xi}^{(2)}, \hat{\boldsymbol{h}}) = \sum_{k=1}^{T} \left[\sum_{s=1}^{S} \left[w_{sk} (\xi_{sk}^{(2)} - \hat{h}_k)^2 \right] - 2\sigma_{k,T}^2 \right]$$

where $\boldsymbol{w}_k = (w_{1k}, \ldots, w_{Sk})'$ is a vector of (generalized least squares) weights such that $\sum_s w_{sk} = 1$. Straightforward calculus shows that, conditional on $\hat{\boldsymbol{h}}$, $\widehat{R}(\boldsymbol{\xi}^{(2)}, \hat{\boldsymbol{h}})$ is an unbiased estimator of $\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2$ and,

$$\tau^2(\boldsymbol{h}, \hat{\boldsymbol{h}}) := \mathrm{Var}\left(\widehat{R}(\boldsymbol{\xi}^{(2)}, \hat{\boldsymbol{h}})\right) = \sum_{k=1}^{T} \left[8\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F^2 + 8(h_k - \hat{h}_k)^2 \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k \right]$$

where $\mathrm{diag}(\boldsymbol{w}_k)$ is a diagonal matrix with elements $\boldsymbol{w}_k = (w_{1k}, \ldots, w_{Sk})'$ on the diagonal.

Theorem 3. Suppose that $\|\boldsymbol{\Gamma}_k\|_1 / \|\boldsymbol{\Gamma}_k\|_F \to 0$ as $S \to \infty$, where $\boldsymbol{\Gamma}_k := \boldsymbol{V}_k^{1/2} \mathrm{diag}(\boldsymbol{w}_k) \boldsymbol{V}_k^{1/2}$ with $\boldsymbol{V}_k^{1/2}$ a symmetric matrix square root of $\boldsymbol{V}_k$. For a given confidence level $1 - \alpha$, consider the confidence set

$$\widehat{C}_\alpha(\boldsymbol{\xi}) = \left\{ \boldsymbol{h} \in \ell_2 \ : \ \|\boldsymbol{h} - \hat{\boldsymbol{h}}\| \leq \sqrt{z_\alpha \tau(\boldsymbol{h}, \hat{\boldsymbol{h}}) + \widehat{R}(\boldsymbol{\xi}^{(2)}, \hat{\boldsymbol{h}})} \right\}$$

with standard normal quantile z_α, and $\hat{\boldsymbol{h}} = \hat{\boldsymbol{h}}(\boldsymbol{\xi}^{(1)})$ such that $\boldsymbol{\xi}_k^{(1)}, \boldsymbol{\xi}_k^{(2)} \sim N(h_k \mathbf{1}_S, 2\boldsymbol{V}_k)$ with $\boldsymbol{\xi}_k^{(1)}, \boldsymbol{\xi}_k^{(2)}$ independent for all $k = 1, \ldots, T$. Then,

$$\liminf_{S \to \infty} \inf_{\boldsymbol{h} \in \ell_2} P(\boldsymbol{h} \in \widehat{C}_\alpha(\boldsymbol{\xi})) \geq 1 - \alpha$$

The validity of the asymptotic confidence regions relies on the condition $\|\boldsymbol{\Gamma}_k\|_1 / \|\boldsymbol{\Gamma}_k\|_F \to 0$, which requires the maximum absolute row sum (or column sum by symmetry) of $\boldsymbol{\Gamma}_k$ to be dominated by its Frobenius norm for increasing S. This condition implies that the number of relevant principal components of the matrix $\boldsymbol{\Gamma}_k$ is increasing, or in other words, the vector of eigenvalues of $\boldsymbol{\Gamma}_k$ should not be dominated by one or a few large values as S increases. Although in a somewhat different spirit than the condition $\|\boldsymbol{G}_S\|_F = o(S)$, this condition also implies that the effective number of independent replicates should increase with the total number of replicates S. Note that considering a vector of equal weights $\boldsymbol{w}_k = (\frac{1}{S}, \ldots, \frac{1}{S})'$, this condition can be restated in terms of the between-replicate correlation matrix as $\|\boldsymbol{G}_S\|_1 / \|\boldsymbol{G}_S\|_F \to 0$ for $S \to \infty$. This ratio has an optimal rate $1/\sqrt{S}$, when $\boldsymbol{G}_S$ is equal to the identity matrix, and it can be verified that this condition implies $\|\boldsymbol{G}_S\|_F = o(S)$, (the other direction does not hold). Furthermore, since $\frac{1}{\sqrt{T}} \|\boldsymbol{h}^f - \hat{\boldsymbol{h}}^f\| = \|\boldsymbol{h} - \hat{\boldsymbol{h}}\|$, if we consider the scaled confidence regions in the frequency domain given by,

$$\widehat{C}_\alpha^f(\boldsymbol{\xi}) = \left\{ \boldsymbol{h}^f \in \ell_2 : \|\boldsymbol{h}^f - \hat{\boldsymbol{h}}^f\| \leq \sqrt{T} \sqrt{z_\alpha \tau(\boldsymbol{h}, \hat{\boldsymbol{h}}) + \widehat{R}(\boldsymbol{\xi}^{(2)}, \hat{\boldsymbol{h}})} \right\}$$

By Theorem 3, the asymptotic coverage probability also satisfies,

$$\liminf_{S \to \infty} \inf_{\boldsymbol{h}^f \in \ell_2} P(\boldsymbol{h}^f \in \widehat{C}_\alpha^f(\boldsymbol{\xi})) \geq 1 - \alpha$$

Remark. Note that the confidence regions are constructed under the assumption that the covariance matrices $\boldsymbol{V}_k$ are known. In practice, these covariance matrices are unknown, and we therefore replace them by plug-in estimators $\widehat{\boldsymbol{V}}_k$. The generalized least squares weights $\boldsymbol{w}_k$, which typically also depend on the covariance matrices $\boldsymbol{V}_k$, can be replaced for instance by the sub-optimal ordinary least squares weights $\boldsymbol{w}_k = (\frac{1}{S}, \ldots, \frac{1}{S})'$. This does not change the asymptotic normality result of the estimator $\widehat{R}(\boldsymbol{\xi}^{(2)}, \hat{\boldsymbol{h}})$ in the proof of Theorem 3, but it comes at the cost of increasing its variance, thereby increasing the radius of the confidence regions.

6. Simulated data examples

In this section, we assess the finite-sample performance of the developed estimators by some simulated data examples. In Algorithm 1 below, we describe a procedure to simulate replicated time series $\{X_s(t), s = 1, \ldots, S\}$ with random log-spectra H_s^f by means of discrete Cramér representations. In short, given a wavelet basis $\mathcal{B}$, population-mean transfer function g^f (with population-mean log-spectrum $h^f(\omega) = \log(\text{Mod}(g^f(\omega))^2)$), within-replicate covariance matrix $\boldsymbol{G}_T$, and between-replicate correlation matrix $\boldsymbol{G}_S$, we generate replicate-specific random transfer functions $G_s^f(\omega)$ which are inserted into discrete Cramér representations to generate the replicated time series. In Algorithm 1, ξ_ℓ^s denotes a complex-valued normal random variable with independent real and imaginary parts, such that $\xi_\ell^s = \xi_{-\ell}^{s*}$, where * denotes the complex conjugate.

Algorithm 1

$\boldsymbol{h}^f \leftarrow \log(\text{Mod}(\boldsymbol{g}^f)^2)$
$\boldsymbol{h} \leftarrow \text{DWT}_{\mathcal{B}}(\boldsymbol{h}^f)$, the discrete wavelet transform w.r.t. the basis $\mathcal{B}$.

$\boldsymbol{U} \leftarrow \boldsymbol{0}_{S\times T}$, an $(S \times T)$-matrix of zeroes.
For $k = 1, \ldots, T$,
  if $\sigma_{uk}^2 > 0$ **then** put $\boldsymbol{U}_{[,k]} \sim N(\boldsymbol{0}, \sigma_{uk}^2 \boldsymbol{G}_S)$

For $s = 1, \ldots, S$,
  $U_s^f \leftarrow \text{I-DWT}_{\mathcal{B}}(\boldsymbol{U}_{[s,]})$, the inverse discrete wavelet transform w.r.t. the basis $\mathcal{B}$.
  $G_s^f \leftarrow \boldsymbol{g}^f \sqrt{\exp(U_s^f)}$
  $X_s(t) \leftarrow \frac{1}{\sqrt{2T}} \sum_{\ell=-(T-1)}^{T} G_s^f(\omega_\ell) \exp(i2\pi\omega_\ell t)\xi_\ell^s$

6.1. Population-mean log-spectrum h^f

We consider data generated under a single population-mean log-spectrum h^f coming from an ARMA(2, 2) process with parameters $\boldsymbol{\phi} = (-0.2, -0.9)$ and $\boldsymbol{\theta} = (0, 1)$. The log-spectrum of this ARMA(2, 2) process is particularly difficult to estimate due to some sharp local features. The transfer function of an ARMA(p, q) process with AR parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$, MA parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$, and white noise variance σ_w^2 is given by,

$$g^f(\omega) = \sigma_w \frac{\theta(e^{-i2\pi\omega})}{\phi(e^{-i2\pi\omega})}, \qquad \omega \in [0, 1)$$

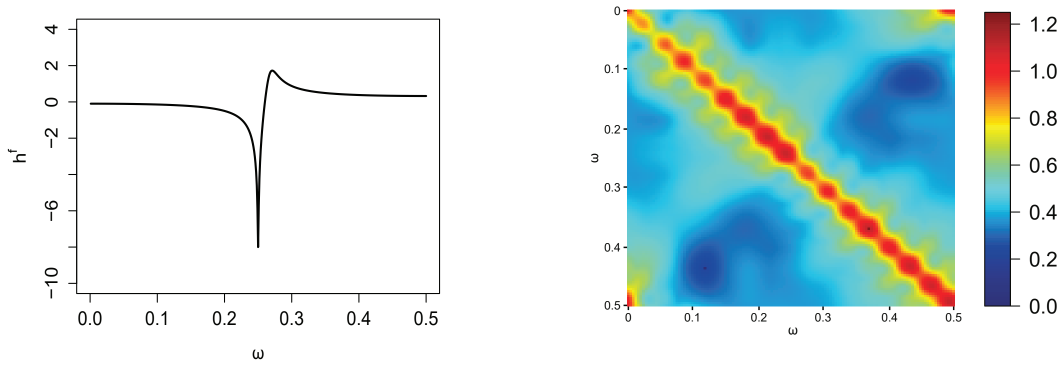

Figure 1. Population-mean log-spectrum (left), and within-replicate covariance matrix $\boldsymbol{G}_T^f$ (right).

where $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$. In Figure 1 (left), the considered population-mean log-spectrum $\boldsymbol{h}^f = (h^f(\omega_0), \ldots, h^f(\omega_{T-1}))'$ with $\sigma_w^2 = 1$ and $T = 512$ is shown for $\omega_\ell \in [0, 1/2)$. In fact, the presented curve is a relatively sparse ℓ_0-approximation of $\boldsymbol{h}^f$ under a Daubechies extremal-phase wavelet basis $\mathcal{B}$ with $N = 6$ vanishing moments, where we have thresholded all coefficients with $|h_k| < T^{-1}$. Here, we have used the `WaveThresh` package in `R`, see (Nason, 2010, Chapter 2) for more details.

6.2. Random effects covariance matrices

For the $T \times T$-dimensional covariance matrix $\boldsymbol{G}_T$, we consider the diagonal matrix $\mathrm{diag}(\boldsymbol{\sigma}_u^2)$ with set of indices of non-zero variance components given by,

$$K_{u,T} = \{k \in K_{h,T} \; : \; \lfloor \log_2(k-1) \rfloor < J\}$$

which are simply all the indices contained in the wavelet scales $\{j \; : \; 0 \leq j < J\}$ with the additional constraint that $K_{u,T} \subseteq K_{h,T}$. For the magnitudes of the variance components, we consider σ_{uk}^2 decaying with a factor 2 per increasing wavelet scale, i.e. for some constant $C > 0$, let $\sigma_{u1}^2 = C$ and define for $k > 1$,

$$\sigma_{uk}^2 = \begin{cases} C \cdot 2^{-\lfloor \log_2(k-1) \rfloor - 2} & \text{if } k \in K_{u,T} \\ 0 & \text{if } k \notin K_{u,T} \end{cases}$$

In Figure 1 (right), the within-replicate covariance matrix in the frequency domain $\boldsymbol{G}_T^f := \boldsymbol{W}'_{\mathcal{B}} \boldsymbol{G}_T \boldsymbol{W}_{\mathcal{B}}$ is shown using again a Daubechies extremal-phase wavelet basis with $N = 6$ vanishing moments and parameters $C = 0.5$, $J = 4$, which are also the values used in the subsequent simulation studies. Under this specific model, Figure 2 shows generated random log-spectra for several independent replicates and corresponding simulated replicate-specific time series.

For the between-replicate $S \times S$-dimensional correlation matrix $\boldsymbol{G}_S$ we consider two different scenarios,

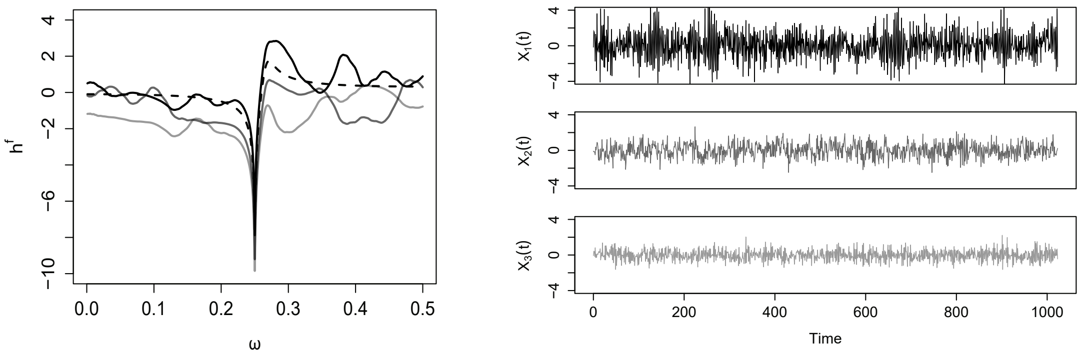

Figure 2. Replicate-specific log-spectra (left), and simulated replicated time series (right).

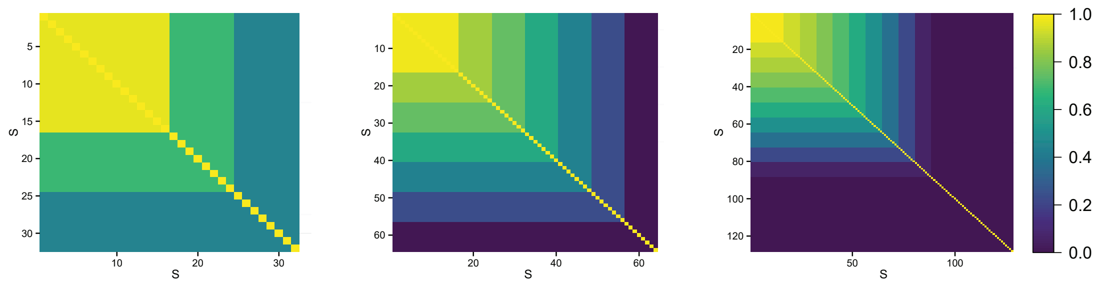

Figure 3. Contour correlation matrices for $S = \{32, 64, 128\}$.

(a) A symmetric block-diagonal matrix containing one single $S/2 \times S/2$-dimensional block of highly correlated replicates with $\rho_{ij} = 0.9$ for $1 \leq i, j \leq S/2$. The constructed correlation matrix satisfies $\|\boldsymbol{G}_S\|_F = o(S)$ and is positive semi-definite.

(b) A symmetric contour-matrix that consists of layers of block matrices with decaying levels of correlation, see Figure 3. The layers are chosen such that again $\|\boldsymbol{G}_S\|_F = o(S)$ and $\boldsymbol{G}_S$ is positive semi-definite.

6.3. Simulation study

We assess the performance of the proposed estimation procedure and compare this to the performance of several related alternatives. First, we consider a naive ordinary least squares approach (OLS). In this naive approach we individually smooth the replicate-specific log-periodograms using FDR thresholding with tuning parameter $q = 0.001$ and estimate $\boldsymbol{h}^f$ by averaging the smoothed curves over replicates, thereby not taking into account the between-replicate dependence structure. Second, we consider the non-adaptive iterative-generalized least squares approach, where the smoothness space parameter $k_{h,T}$ is assumed to be known and the set $K_{h,T}$ is estimated using the universal threshold $\lambda_{h,T}$ according to Section 3.1. Third, we consider an adaptive iterative-generalized least squares approach, in which $k_{h,T}$ is not known. In order to estimate the set $K_{h,T}$ we use FDR thresholding with tuning parameters $q = \{0.1, 0.001\}$. Finally, in order to assess the increase in estimation error due to the iteration scheme, we also compute oracle estimators $\hat{\boldsymbol{h}}^f$ and $\widehat{\boldsymbol{G}}_S(\hat{\boldsymbol{h}})$, where we assume the true generalized least squares-weights (depending on $\boldsymbol{G}_S$) to be known in estimating $\boldsymbol{h}^f$, so that no iteration of the estimators is required. Note that in this final scenario we do not assume knowledge of $k_{h,T}$ nor $k_{u,T}$, and the set $K_{h,T}$ is estimated as in the third scenario with FDR tuning parameter $q = 0.001$. The performance of the estimator $\hat{\boldsymbol{h}}^f = (\hat{h}^f(\omega_0), \ldots, \hat{h}^f(\omega_{T-1}))'$ is assessed through the squared error averaged over Fourier frequencies. Similarly, the performance of the estimators $\widehat{\boldsymbol{G}}_T$ and $\widehat{\boldsymbol{G}}_S$ is evaluated through the squared error averaged over matrix elements. In Table 1 we show average squared errors with in parenthesis corresponding standard errors ($\hat{\sigma}/\sqrt{M}$) for $M = 1000$ replicated simulation experiments.

6.3.1. *Simulation results*

It should not come as a surprise that knowledge of the true generalized least squares-weights significantly improves the estimation error. This is seen by comparing the estimation error for $\boldsymbol{h}^f$ of the naive ordinary least squares approach, which does not take into account the between-replicate correlation structure, with that of the oracle estimator. The iterative-generalized least squares scheme then inflates the estimation error for $\boldsymbol{h}^f$ relative to the oracle estimator, but still outperforms the ordinary least squares approach under all of the considered scenarios. Another important observation is that the performance of the adaptive estimators –regardless of the choice of the FDR tuning parameter– is similar to the performance of the nonadaptive estimators in essentially all of the considered scenarios. The adaptive estimators even slightly outperform the nonadaptive estimators in some cases, which may be caused by the fact that the asymptotic minimax threshold $\lambda_{h,T}$ is somewhat conservative in a finite-sample situation. It is therefore quite safe to transfer the derived rates for the risk of $\hat{\boldsymbol{h}}^f$ in the nonadaptive estimation case (with $k_{h,T}$ known) also to the adaptive estimation case (with $k_{h,T}$ unknown).
In Figure 4 and 5, we give some visual representations of the estimates under a block-diagonal between-replicate correlation structure ($T = 512, S = 64$). The left-column of Figure 4 shows the estimated quantities of interest for a single simulation experiment using the adaptive approach with FDR tuning parameter $q = 0.001$. The right-column shows the average estimates over $M = 1000$ replications of the simulation experiment. Figure 5 shows true and predicted random effects curves for a single simulation experiment under the same scenario as in Figure 4. It is observed that the estimated between-replicate correlation matrices $\widehat{\boldsymbol{G}}_S$ are somewhat biased downwards. This is an artifact caused by the iterative estimation scheme, and is understood by the following argument: in estimating $\boldsymbol{G}_S$, the replicate-specific curves are centered around $\hat{\boldsymbol{h}}^f$ (instead of the true $\boldsymbol{h}^f$), however $\hat{\boldsymbol{h}}^f$ is itself a sample average over the replicate-specific curves, therefore the replicate-specific curves will typically appear to be more negatively correlated around $\hat{\boldsymbol{h}}^f$ than they are in reality around the true curve $\boldsymbol{h}^f$. Note that, under the conditions in Section 4, this artifact disappears asymptotically due to consistency of the covariance matrix estimators.

Table 1. Average squared errors (and standard errors $\hat{\sigma}/M$) of the estimates of $\boldsymbol{h}^f$ ($\times 10^{-1}$), $\boldsymbol{G}_T$ ($\times 10^{-4}$), and $\boldsymbol{G}_S$ ($\times 10^{-1}$).

		$T = 512$			$T = 1024$		
S	Approach	ASE($\hat{\boldsymbol{h}}^f$)	ASE($\widehat{\boldsymbol{G}}_T$)	ASE($\widehat{\boldsymbol{G}}_S$)	ASE($\hat{\boldsymbol{h}}^f$)	ASE($\widehat{\boldsymbol{G}}_T$)	ASE($\widehat{\boldsymbol{G}}_S$)
Block-diagonal correlation matrix $\boldsymbol{G}_S$ (Case (a))							
32	OLS	2.74 (0.05)	-	-	2.53 (0.05)	-	-
	Non-adaptive	2.15 (0.04)	0.65 (0.02)	1.29 (0.01)	2.12 (0.04)	0.31 (0.01)	1.24 (0.01)
	Adapt. ($q = 0.1$)	2.10 (0.04)	0.66 (0.02)	1.30 (0.01)	2.08 (0.04)	0.31 (0.01)	1.24 (0.01)
	Adapt. ($q = 0.001$)	2.15 (0.05)	0.64 (0.02)	1.31 (0.01)	2.12 (0.04)	0.31 (0.01)	1.25 (0.01)
	Oracle ($q = 0.001$)	1.18 (0.02)	0.42 (0.01)	0.78 (0.01)	1.05 (0.02)	0.20 (0.01)	0.69 (0.01)
64	OLS	2.61 (0.05)	-	-	2.42 (0.05)	-	-
	Non-adaptive	1.70 (0.03)	0.52 (0.02)	1.38 (0.01)	1.76 (0.04)	0.27 (0.01)	1.36 (0.01)
	Adapt. ($q = 0.1$)	1.69 (0.04)	0.52 (0.02)	1.39 (0.01)	1.73 (0.04)	0.28 (0.01)	1.35 (0.01)
	Adapt. ($q = 0.001$)	1.75 (0.04)	0.51 (0.02)	1.41 (0.01)	1.74 (0.04)	0.27 (0.01)	1.36 (0.01)
	Oracle ($q = 0.001$)	0.86 (0.01)	0.24 (0.01)	0.80 (0.01)	0.74 (0.01)	0.12 (0.005)	0.72 (0.01)
128	OLS	2.56 (0.05)	-	-	2.38 (0.05)	-	-
	Non-adaptive	1.57 (0.03)	0.41 (0.01)	1.56 (0.01)	1.61 (0.03)	0.21 (0.01)	1.55 (0.01)
	Adapt. ($q = 0.1$)	1.56 (0.03)	0.41 (0.01)	1.57 (0.01)	1.61 (0.04)	0.21 (0.01)	1.56 (0.01)
	Adapt. ($q = 0.001$)	1.58 (0.03)	0.40 (0.01)	1.58 (0.01)	1.62 (0.04)	0.21 (0.01)	1.57 (0.01)
	Oracle ($q = 0.001$)	0.69 (0.01)	0.23 (0.01)	0.96 (0.01)	0.58 (0.01)	0.09 (0.01)	0.90 (0.01)
Contour correlation matrix $\boldsymbol{G}_S$ (Case (b))							
32	OLS	7.21 (0.16)	-	-	7.24 (0.16)	-	-
	Non-adaptive	6.64 (0.15)	1.95 (0.03)	4.34 (0.03)	6.45 (0.15)	1.04 (0.02)	4.01 (0.03)
	Adapt. ($q = 0.1$)	6.60 (0.15)	1.93 (0.03)	4.35 (0.03)	6.41 (0.15)	1.04 (0.02)	4.03 (0.03)
	Adapt. ($q = 0.001$)	6.56 (0.15)	1.92 (0.03)	4.31 (0.03)	6.35 (0.15)	1.03 (0.02)	4.05 (0.03)
	Oracle ($q = 0.001$)	5.95 (0.13)	1.26 (0.03)	3.40 (0.02)	5.85 (0.13)	0.76 (0.02)	2.98 (0.01)
64	OLS	4.61 (0.10)	-	-	4.60 (0.10)	-	-
	Non-adaptive	2.66 (0.06)	1.07 (0.06)	1.95 (0.02)	2.65 (0.06)	0.54 (0.02)	1.93 (0.02)
	Adapt. ($q = 0.1$)	2.64 (0.06)	1.05 (0.05)	1.95 (0.02)	2.63 (0.06)	0.54 (0.02)	1.94 (0.02)
	Adapt. ($q = 0.001$)	2.67 (0.06)	1.05 (0.05)	1.98 (0.02)	2.66 (0.06)	0.53 (0.02)	1.95 (0.02)
	Oracle ($q = 0.001$)	1.48 (0.03)	0.53 (0.02)	0.99 (0.01)	1.35 (0.03)	0.24 (0.01)	0.85 (0.01)
128	OLS	2.70 (0.05)	-	-	2.64 (0.05)	-	-
	Non-adaptive	1.72 (0.03)	0.84 (0.07)	1.34 (0.01)	1.65 (0.03)	0.37 (0.03)	1.30 (0.01)
	Adapt. ($q = 0.1$)	1.71 (0.03)	0.84 (0.07)	1.35 (0.01)	1.64 (0.03)	0.37 (0.03)	1.30 (0.01)
	Adapt. ($q = 0.001$)	1.72 (0.03)	0.85 (0.07)	1.36 (0.01)	1.67 (0.04)	0.36 (0.03)	1.31 (0.01)
	Oracle ($q = 0.001$)	0.75 (0.01)	0.28 (0.01)	0.91 (0.01)	0.64 (0.01)	0.11 (0.005)	0.85 (0.01)

Table 2. Empirical coverage ($\times 10^2$) of confidence regions for $\boldsymbol{h}^f$.

	$\alpha = 0.05$				$\alpha = 0.1$			
	$T = 512$		$T = 1024$		$T = 512$		$T = 1024$	
	$S = 64$	$S = 128$	$S = 64$	$S = 128$	$S = 64$	$S = 128$	$S = 64$	$S = 128$
Block-diagonal correlation matrix $\boldsymbol{G}_S$ (Case (a))								
Asymptotic (Scen. 1)	95.1	93.5	95.6	93.7	88.7	86.0	88.5	87.2
Asymptotic (Scen. 2)	96.7	96.5	98.3	98.2	86.1	84.4	89.8	89.7
Bootstrap	99.1	99.8	96.7	97.5	98.2	99.2	92.7	95.1
Contour correlation matrix $\boldsymbol{G}_S$ (Case (b))								
Asymptotic (Scen. 1)	99.5	96.4	99.6	96.7	96.9	90.3	97.2	91.1
Asymptotic (Scen. 2)	97.2	97.2	98.4	98.5	86.4	85.6	90.5	90.1
Bootstrap	97.3	99.5	95.5	96.2	94.4	99.0	89.6	93.5

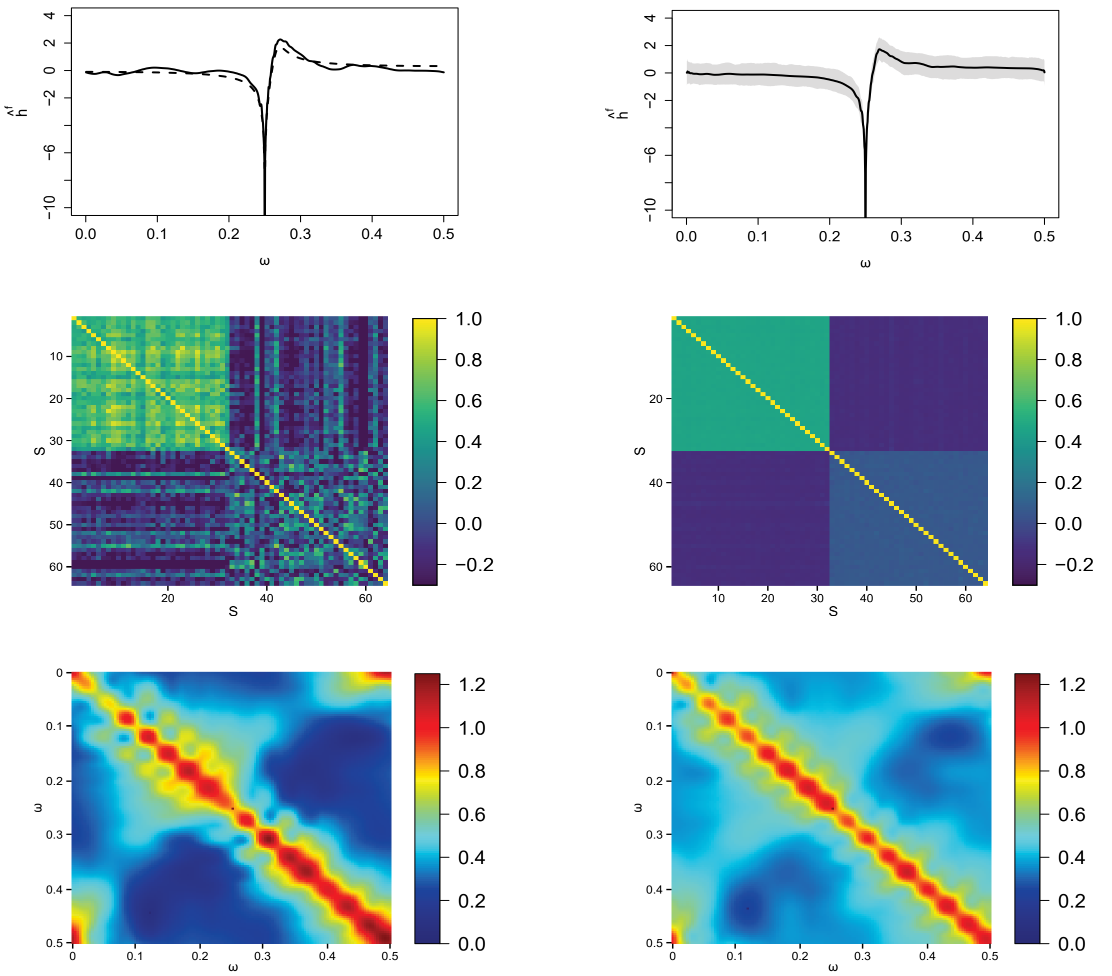

Figure 4. Estimates of $\boldsymbol{h}^f$, $\boldsymbol{G}_S$, and $\boldsymbol{G}_T^f$ for a single simulation experiment (left-column). Average estimates over 1000 simulation experiments, with $(0.01, 0.99)$-empirical pointwise quantiles of $\hat{\boldsymbol{h}}^f$ (right-column).

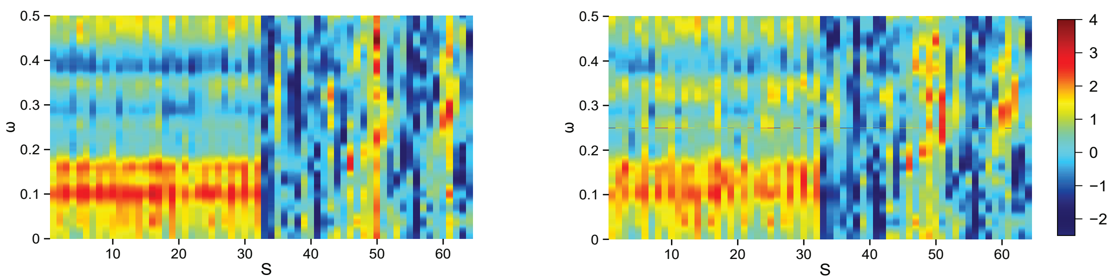

Figure 5. True random effects curves $U_s^f(\omega)$ (left), and predicted random effects curves $\widehat{U}_s^f(\omega)$ (right) for a single simulation experiment.

6.4. Confidence region coverage

We also assess the validity of the constructed confidence regions for $\boldsymbol{h}^f$ by computing their empirical coverage under some of the simulated models considered before. Since we are interested in the (negative) impact on the empirical coverage caused by replacing the true covariance matrices $\boldsymbol{V}_k$ by estimates $\widehat{\boldsymbol{V}}_k$ we consider two different scenarios. In the first scenario, we assume the true matrices $\boldsymbol{V}_k$ to be known, both in performing the sample splitting procedure and in constructing the confidence regions. In the second scenario, we consider the matrices $\boldsymbol{V}_k$ to be unknown, therefore replacing them by plug-in estimators $\widehat{\boldsymbol{V}}_k$. For the weight vectors $\boldsymbol{w}_k$, we consider ordinary least squares weights, equally weighting each replicate. As a benchmark procedure, we also compute the empirical coverage of parametric bootstrap confidence regions for $\boldsymbol{h}^f$ with $B = 1000$ bootstrap samples as proposed in Krafty et al. (2011) and Fiecas and Ombao (2016). The bootstrap confidence regions are constructed using the true covariance matrices $\boldsymbol{V}_k$, and in this sense are oracle confidence regions, that should be compared to the asymptotic confidence regions under the first scenario. In Table 2, the empirical coverage for the different approaches is shown for $M = 5000$ replicated simulation experiments ($M = 1000$ for the bootstrap confidence regions), the simulation results are shown only for the adaptive approach (i.e. $k_{h,T}$ unknown) using FDR tuning parameter $q = 0.001$, the results for the other approaches considered in Table 1 are similar.

7. Analysis of brain signals: replicated LFP time series

To conclude, we analyze brain signal data recorded during an associative learning experiment. The dataset consists of local field potential (LFP) time series traces, measuring electrical activity in the brain over the course of the experiment, see Gorrostieta et al. (2012) and Fiecas and Ombao (2016) for a more detailed description. During the experiment, a non-human male macaque learns the association between one set of objects (four pictures) and another set of objects (doors located in four different quadrants of the visual field) by means of trial-and-error. In each trial, the macaque was first presented with a picture and then required to select one of the four doors. Each time the macaque made the correct association it was given a reward in the form of a small quantity of juice. Over the course of the associative learning experiment, the electrical activity in the brain of the macaque was measured using local field potentials. Local field potentials measure electrical activity in the brain directly via chronically implanted probes, in contrast to other commonly-used non-invasive recording techniques, such as electroencephalograms (EEGs) and functional magnetic resonance imaging (fMRI). In this analysis, we consider univariate local field potential time series data recorded in the nucleus accumbens (NAc) region, which is a region in the brain that has been demonstrated to be highly implicated in cognitive processes involving memory and reward. After preprocessing of the LFP time series data, there remains a total of 590 (univariate) time series traces of length 2048 sampled at 1000 Hz, thus roughly corresponding to 2 seconds of data. The goal of the analysis is to study trial-population spectral characteristics, and in particular the evolving spectral properties of the time series over the course of the experiment. In Figure 6, we show the recorded LFP time series for three trials (start, middle, and end of the experiment) and their corresponding raw periodograms to highlight the dominating oscillatory frequencies present in the LFP time series. The raw periodograms in Figure 6 show clear common frequency behavior across the three different trials, as seen from the peaks which remain located at the same frequencies across trials. The left image of Figure

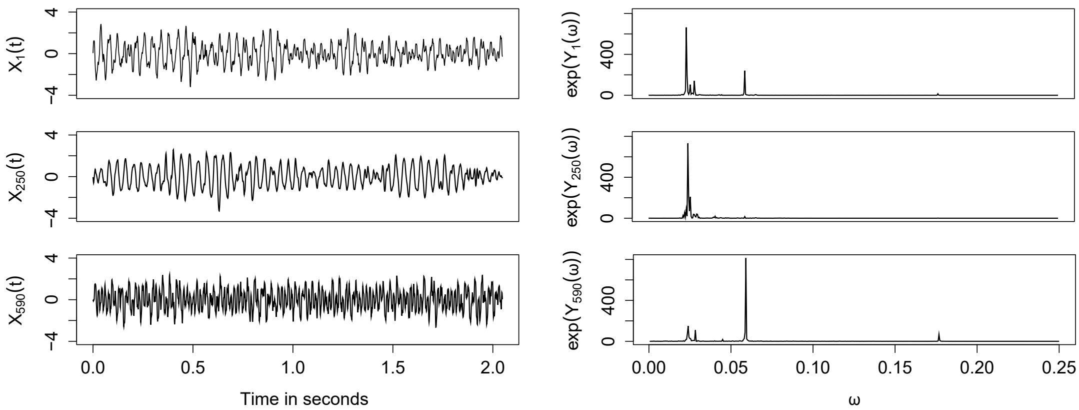

Figure 6. LFP time series (left) and corresponding raw periodograms (right) for trials 1, 250, and 590.

8 shows initial smoothed log-periodograms across all the trials in the experiment, using FDR thresholding with tuning parameter $q = 0.001$, without taking into account the between-trial dependence structure. Note that instead of the individually smoothed log-periodograms, we show blockwise-average log-periodograms over 10 adjacent individual trials in order to improve visibility of the image. In general, the log-periodograms in Figure 8 (left) display very common frequency behavior across trials, however towards the end of the experiment the overall power in the middle- and high-frequency range ($\omega > 0.1$) seems to increase (except for the frequency-band around $\omega \approx 0.14$), and we also note that power in the very low-frequency range (ω close to zero) fades out after approximately half of the trials. Simply averaging the spectral estimates across all trials will not take into account, nor give us any information, about the dependence structure of the underlying brain dynamics over the course of the experiment. Therefore, we need a model that allows for explicit correlation between trial-replicates in the population. In the data analysis we consider frequency content up to 256 Hz ($\omega = 0.25$), since higher frequency behavior is typically attributed to noise and not physiological behavior in the brain, and we do not want the estimated between-trial correlation matrix to be dominated by this very high-frequency content ($\omega > 0.25$). In the top-row of Figure 7, the estimates of the population-mean log-spectrum $\boldsymbol{h}^f = (h^f(\omega_0), \ldots, h^f(\omega_{511}))'$ and population-mean spectrum $\exp(\boldsymbol{h}^f)$ are shown, with in grey several regions of interest for the neuroscientist: the α-band (8-16 Hz), the β-band (16-32 Hz), γ-band (32-100 Hz). The dotted lines in Figure 7 (top-left) correspond to the magnitude of one estimated standard deviation, around the population-mean curve, of the random effects curves in the frequency domain. The bottom row of Figure 7 shows the estimate for the between-trial correlation matrix $\boldsymbol{G}_S$ for blocks of 10 adjacent individual trials, and the estimate for the within-trial covariance matrix $\boldsymbol{G}_T^f$ in the frequency domain. Note that the estimated correlation matrix clearly demonstrates the correlation structure between the trials that we observed for the initial smoothed log-periodograms in Figure 8 (left). Trials at the beginning of the experiment are highly correlated and trials at the end of the experiment are highly correlated, and the correlation between trials decays as the lag between trials increases. This suggests that the trial-specific log-spectra evolve over the course of the learning experiment, as already discussed

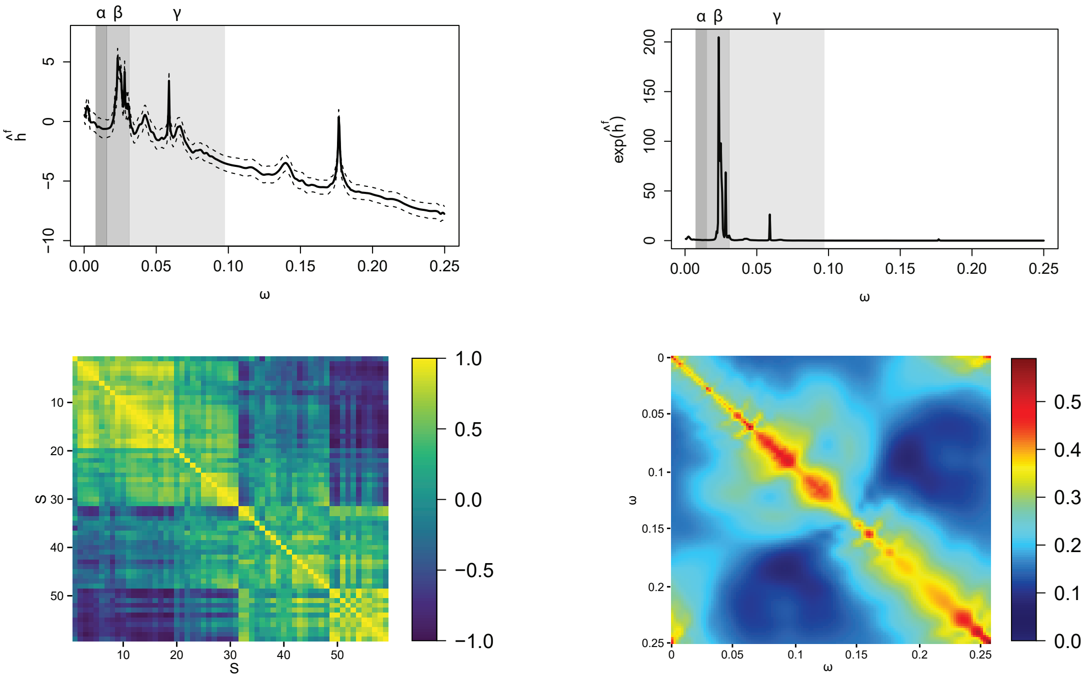

Figure 7. Estimates of $\boldsymbol{h}^f$ and $\exp(\boldsymbol{h}^f)$ (top-row), and estimates of $\boldsymbol{G}_S$ and $\boldsymbol{G}_T^f$ (bottom-row).

in Fiecas and Ombao (2016), which should be taken into account in the data analysis in order to improve estimation of the population-mean curve and especially prediction of the replicate-specific curves, but also to avoid misleading results in any subsequent inference procedures. The right image of Figure 8 shows the predicted trial-specific log-spectra (again for blocks of 10 adjacent trials). Comparing the two images in Figure 8, we observe that the predicted trial-specific log-spectra perform better in suppressing the overall noise than the individual smoothed log-periodograms, since the functional mixed-effects model pools information across different trials. On the other hand, the predicted curves are still able to capture many of the relevant features that are present in the left image for the individual smoothed log-periodograms. These features would not be as pronounced if we do not take into account the explicit correlation between trials over the course of the experiment.

8. Conclusion

In the context of spectral analysis for replicated time series, where the focus is on population spectral characteristics rather than the behavior of individual time series, we propose to model replicate-specific log-spectra as random curves based on a nonparametric functional mixed effects approach. We address the specific problem of analyzing spectra that are characterized by localized peaks or troughs by successfully using projection estimators, and in particular nonlinear wavelet thresholding. Here we benefit both from a convenient linear mixed effects structure in the wavelet coefficient domain, and the possibility to constrain the complexity of this nonpara-

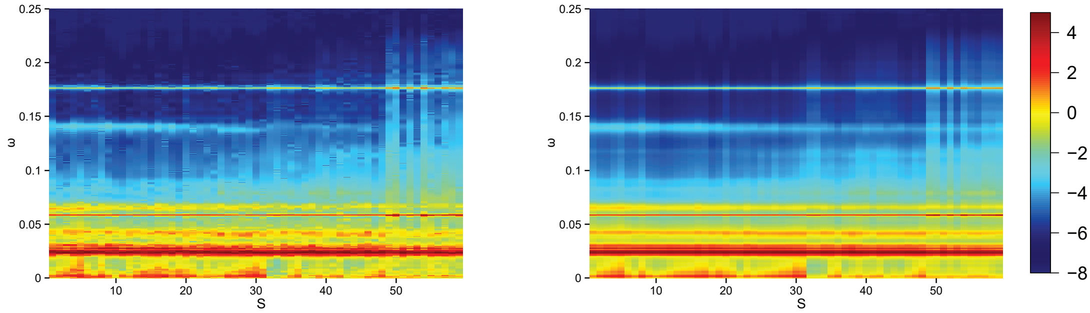

Figure 8. Individually smoothed log-periodograms (left), and predicted trial-specific log-spectra $\hat{\boldsymbol{h}}_s^f$ (right).

metric estimation problem by natural ℓ_0-sparsity constraints. The paradigm of ℓ_0-driven sparsity leads to simple constraints ensuring that the replicate-specific curves and the population-mean curve share the same level of complexity (as in Guo (2002)), but also allows to come up with practical near-optimal threshold choices for both fixed- and random effects estimation. The good performance of this smoothing device is also confirmed by our empirical results.

As an additional important ingredient, we introduce a generic correlation model for the population of random effects curves, where both intra- and inter-subject correlation are modeled in a convenient nonparametric way in the wavelet coefficient domain. The importance of including correlated functions in the general setting of functional response regression has already been underlined by Morris (2014). The author clearly expresses that the overwhelming majority of existing work cannot model correlation between curves and is hence only suitable for independently sampled functions. This is not realistic for many functional data problems, and can lead to estimators that are statistically inefficient, or even give misleading inferences. As an example of replicated time series data with explicit correlation between different replicates in the population, we analyze empirical brain signal data over the course of an associative learning experiment. There is a clear indication that the spectral behavior of trial-replicated time series evolves over the course of the experiment, and we are able to reproduce a meaningful correlation structure over time series replicates that demonstrates this evolutionary behavior over the course of the learning experiment. In order to give some notion of statistical inference, we construct nonparametric confidence sets for the population-mean log-spectrum by adapting the approach in Robins and van der Vaart (2006) to include the random effects curve uncertainty into inference on the fixed effects curve. Here again, it is important to take into account potential correlation between replicate-specific curves to avoid the use of erroneous confidence sets. We note that our fully nonparametric approach, although developed in the framework of spectral analysis for time series data, can equally well be applied in a general functional data analysis context in the presence of correlated random curves, where we benefit from the twofold adaptation properties of wavelets towards sparse and localized structures.

To conclude, we discuss two important directions in which to generalize the proposed model. First, replicated time series spectral analysis of brain data eventually calls for a *multivariate* treatment in order to reveal dependence structures between different regions in the brain through cross-spectral analysis of different components of the multivariate time series (e.g. multivariate

EEG time series data from different regions in the brain). To the best of our knowledge, no quantitative analysis that embeds replicate-specific spectral matrices into a multivariate functional mixed effects model exists so far, not even for the case of independent replicates. We are currently generalizing our functional mixed effects approach developed for replicated univariate time series to this more challenging setting. Second, there is considerable evidence (see e.g. Ombao et al. (2005)) that for long EEG recordings, the second-order stationary assumption for the time series is too strong. It is preferable to weaken this assumption and to consider for instance a variance-covariance structure that slowly changes over time. In the context of an individual time series, time-varying spectral analysis is a challenging task since it can lead to estimators with extremely high variance (see e.g. Neumann and von Sachs (1997)). We expect that the methodology presented here will become very efficient in this context, since it allows for pooling of information across the different time series replicates.

9. Appendix: Proofs

9.1. Proof of Lemma 1.

Proof. Let us write $\boldsymbol{Y}_k = \boldsymbol{h}_k^u + \boldsymbol{\epsilon}_k$, where $\boldsymbol{h}_k^u = \boldsymbol{h}_k + \boldsymbol{U}_k$, with $\boldsymbol{h}_k = h_k \mathbf{1}_S$ and $\boldsymbol{\epsilon_k} = (\epsilon_{1k}, \ldots, \epsilon_{Sk})'$ independent noise terms. Then, conditional on the random effects coefficients, we will show that, uniformly in $k \in J_{T,\alpha}$ with $\alpha > 0$,

$$\frac{P\left(\frac{\sqrt{T}}{\sigma_e}(Y_{sk} - h_{sk}^u) \geq x_s \mid U_{sk} = u_{sk}\right)}{1 - \Phi(x_s)} \overset{T\to\infty}{\to} 1 \qquad \text{for all } s = 1, \ldots, S$$

where $-\infty < x_s \leq T^\nu$ for some $\nu > 0$. In order to prove this asymptotic normality result, we derive upper bounds on the n-th order cumulants ($n \geq 2$) of $\epsilon_{sk} = \langle \boldsymbol{E}_s^f, \psi_k \rangle_T$, the empirical wavelet coefficients of the log-periodogram error terms $\boldsymbol{E}_s^f$ with respect to the wavelet basis function ψ_k. Since $\mathrm{Var}(\epsilon_{sk}) = O(T^{-1})$, the n-th order cumulant of ϵ_{sk} (for all s, k) can be written in terms of joint cumulants as,

$$\begin{aligned}
\mathrm{cum}_n(\epsilon_{sk}/\sqrt{\mathrm{Var}(\epsilon_{sk})}) \quad &\lesssim \quad \mathrm{cum}\left(\frac{1}{\sqrt{T}}\sum_{\ell_1=1}^{T} E_s^f(\omega_{\ell_1})\psi_k(\omega_{\ell_1}), \ldots, \frac{1}{\sqrt{T}}\sum_{\ell_n=1}^{T} E_s^f(\omega_{\ell_n})\psi_k(\omega_{\ell_n})\right) \\
&= \quad T^{-n/2}\sum_{\ell_1}\cdots\sum_{\ell_n}\psi_k(\omega_{\ell_1})\cdots\psi(\omega_{\ell_n})\mathrm{cum}(E_s^f(\omega_{\ell_1}), \ldots, E_s^f(\omega_{\ell_n})) \\
&= \quad T^{-n/2}\sum_{\ell_1}\psi_k(\omega_{\ell_1})^n \mathrm{cum}(E_s^f(\omega_{\ell_1}), \ldots, E_s^f(\omega_{\ell_1})) \\
&\quad\quad +T^{-n/2}\sum_{j=1}^{n}\sum_{\ell_1\neq\ell_2}\psi_k(\omega_{\ell_1})^j\psi_k(\omega_{\ell_2})^{n-j}\mathrm{cum}(E_s^f(\omega_{\ell_1}), \ldots, E_s^f(\omega_{\ell_1}), \\
&\quad\quad\quad E_s^f(\omega_{\ell_2}), \ldots, E_s^f(\omega_{\ell_2})) \\
&\quad\quad\quad \vdots \\
&\quad\quad +T^{-n/2}\sum_{\ell_1\neq\ldots\neq\ell_n}\psi_k(\omega_{\ell_1})\ldots\psi_k(\omega_{\ell_n})\mathrm{cum}(E_s^f(\omega_{\ell_1}), \ldots, E_s^f(\omega_{\ell_n})) \quad (14)
\end{aligned}$$

By Lemma 2 in Taniguchi (1979), for a stationary Gaussian process $X_s(t)$ with $\sum_{t=-\infty}^{\infty}|t||\gamma_s(t)| < \infty$ and $n \geq 2$:

$$\mathrm{cum}(E_s^f(\omega_1), \ldots, E_s^f(\omega_{n_1}), E_s^f(\omega_{n_1+1}), \ldots, E_s^f(\omega_{n_1+n_2}), \ldots, E_s^f(\omega_n)) \quad = \quad \begin{cases} O(T^{-m}) & \text{if } m \geq 2 \\ O(1) & \text{if } m = 1 \end{cases}$$

where, $n_1 + \ldots + n_m = n$, and

$$0 < \omega_1 = \ldots, = \omega_{n_1} < \omega_{n_1+1} = \ldots = \omega_{n_1+n_2} < \omega_{n-n_m+1} = \ldots = \omega_n$$

Furthermore, since wavelet basis functions at wavelet scale j are of the order $2^{j/2}$, for all $k \in J_{T,\alpha}$, $|\psi_k(\omega_\ell)| \leq CT^{(1-\alpha)/2}$ uniformly in ω_ℓ for some $C > 0$. From eq.(14) we obtain,

$$\begin{aligned}
\mathrm{cum}_n(\epsilon_{sk}/\sqrt{\mathrm{Var}(\epsilon_{sk})}) &\lesssim T^{-n/2+1}C^{n-2}(T^{(1-\alpha)/2})^{n-2}\int \psi_k(\omega)^2 d\omega + T^{-n/2}C^n B_n(T^{(1-\alpha)/2})^n \\
&\leq T^{-n/2+1}C^{n-2}(T^{(1-\alpha)/2})^{n-2} + T^{-n/2}C^n(n!)(T^{(1-\alpha)/2})^n \\
&\lesssim C^n(n!)(T^{-\alpha/2})^{n-2} \qquad (15)
\end{aligned}$$

where by orthonomality of the wavelet basis functions $\int \psi_k(\omega)^2 d\omega = 1$, and where B_n denotes the n-th Bell number satisfying $B_n \leq n!$ for all $n \in \mathbb{N}$.

By the same arguments as in Neumann (1996), due to the cumulant bounds in eq.(15) and Lemma 1 in Rudzkis et al. (1978),

$$P\left(\frac{\epsilon_{sk} - \mathbb{E}[\epsilon_{sk}]}{\sqrt{\mathrm{Var}(\epsilon_{sk})}} \geq x_s\right) = (1 + o_T(1))(1 - \Phi(x_s)) \qquad \text{for all } s = 1, \ldots, S \qquad (16)$$

for $-\infty < x_s \leq \Delta_T \sim T^{\alpha/6}$.

Then under (A3), $\mathbb{E}[E_s^f(\omega_\ell)] = O(T^{-1})$ uniformly in ω_ℓ by Taniguchi (1979), thus for $k \in J_{T,\alpha}$,

$$\mathbb{E}[\epsilon_{sk}] = \frac{1}{T}\sum_{\ell=1}^{T} \mathbb{E}[E_s^f(\omega_\ell)]\psi_k(\omega_\ell) \leq \sup_\ell |\psi_k(\omega_\ell)| O(T^{-1}) = O(T^{-(1+\alpha)/2}) \qquad (17)$$

and since $\mathrm{Var}(\epsilon_{sk}) = O(T^{-1})$, the standardized bias satisfies $b := \mathbb{E}[\epsilon_{sk}]/\sqrt{\mathrm{Var}(\epsilon_{sk})} = O(T^{-\alpha/2})$. Rewriting eq.(16) gives,

$$\frac{P\left(\frac{\sqrt{T}}{\sigma_e}\epsilon_{sk} \geq x_s\right)}{1 - \Phi(x_s)} = (1 + o_T(1))\frac{1 - \Phi(x_s + b))}{1 - \Phi(x_s)} \qquad (18)$$

Let w.l.o.g. $b \geq 0$ and fix some $c > 1$, then for $x_s \leq c$

$$\left|\frac{1 - \Phi(x_s + b)}{1 - \Phi(x_s)} - 1\right| = \frac{|\Phi(x_s + b) - \Phi(x_s)|}{1 - \Phi(x_s)} \to 0 \qquad \text{as } T \to \infty$$

On the other hand, for $c < x_s \leq \Delta_T$ by a formula for Mill's ratio (see Neumann (1996)),

$$\left|\frac{1 - \Phi(x_s + b)}{1 - \Phi(x_s)} - 1\right| \leq \frac{b\phi(x_s)}{1 - \Phi(x_s)} \leq \frac{bx_s}{1 - 1/x_s^2} \to 0 \qquad \text{as } T \to \infty$$

Since $\boldsymbol{Y}_k - \boldsymbol{h}_k^u = \boldsymbol{\epsilon}_k$, we conclude from eq.(18) that conditional on the random effects coefficients,

$$\frac{P\left(\frac{\sqrt{T}}{\sigma_e}(Y_{sk} - h_{sk}^u) \geq x_s \mid U_{sk} = u_{sk}\right)}{1 - \Phi(x_s)} \overset{T\to\infty}{\to} 1 \qquad \text{for all } s = 1, \ldots, S$$

for $-\infty < x_s \leq \Delta_T$ and uniformly in $k \in J_{T,\alpha}$. Moreover, for any given S,

$$\frac{\prod_{s=1}^S P\left(\frac{\sqrt{T}}{\sigma_e}(Y_{sk} - h_{sk}^u) \geq x_s \mid U_{sk} = u_{sk}\right)}{\prod_{s=1}^S(1 - \Phi(x_s))} \overset{T\to\infty}{\to} 1, \qquad \text{for } x_s \leq \Delta_T$$

and since conditional on the random effects coefficients the terms $Y_{sk} - h^u_{sk}$ are all independent across replicates, this is equivalent to:

$$P\left(\frac{\sqrt{T}}{\sigma_e}\mathrm{I}_S \cdot (\boldsymbol{Y}_k - \boldsymbol{h}^u_k) \geq \boldsymbol{x} \,|\, \boldsymbol{U}_k = \boldsymbol{u}_k\right) = (1 + o_T(1))P(\boldsymbol{Z}_k \geq \boldsymbol{x}), \qquad \text{for } \boldsymbol{x} \leq \boldsymbol{\Delta}_T$$

where $\boldsymbol{x} = (x_1, \ldots, x_S)' \in \mathbb{R}^S$, $\boldsymbol{\Delta}_T \sim (T^\nu, \ldots, T^\nu)' \in \mathbb{R}^S$ with $\nu > 0$, and $\boldsymbol{Z}_k \in \mathbb{R}^S$ a vector of independent standard normal random variables. Let us write $\widetilde{\boldsymbol{Y}}_k = \boldsymbol{V}_k^{-1/2}(\boldsymbol{Y}_k - \boldsymbol{h}_k)$, where $\boldsymbol{V}_k^{-1/2}$ is the square root matrix of $\boldsymbol{V}_k^{-1}$, (recall that $\boldsymbol{V}_k = \sigma^2_{uk}\boldsymbol{G}_S + \frac{\sigma_e^2}{T}\mathrm{I}_S$). The random effects coefficients are assumed to be jointly multivariate normal, therefore the unconditional version follows as well,

$$P(\widetilde{\boldsymbol{Y}}_k \geq \boldsymbol{x}) = (1 + o_T(1))P(\boldsymbol{Z}_k \geq \boldsymbol{x}), \qquad \text{for } \boldsymbol{x} \leq \boldsymbol{\Delta}_T \tag{19}$$

In the following part, we relate the mean squared error of $\hat{h}_{k,\lambda}(\boldsymbol{Y}_k)$ w.r.t. h_k to the mean squared error of $\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k)$ w.r.t. h_k, and show that they are asymptotically equivalent as $T \to \infty$. We split up,

$$\begin{aligned}
\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^2] &= \mathbb{E}\left[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^2 \mathbf{1}_{\{\boldsymbol{h}_k - \boldsymbol{V}_k^{1/2}\boldsymbol{\Delta}_T \leq \boldsymbol{Y}_k \leq \boldsymbol{h}_k + \boldsymbol{V}_k^{1/2}\boldsymbol{\Delta}_T\}}\right] \\
&\quad + \mathbb{E}\left[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^2 \mathbf{1}_{\{|\widetilde{\boldsymbol{Y}}_k| > \boldsymbol{\Delta}_T\}}\right] \\
&:= R_1 + R_2
\end{aligned} \tag{20}$$

According to eq.(19) above, there exist $C_T^{(\ell)}$ and $C_T^{(u)}$ tending to 1 as $T \to \infty$ (uniformly in $k \in J_{T,\alpha}$), such that

$$C_T^{(\ell)} P(\boldsymbol{Z}_k \geq \boldsymbol{x}) \leq P(\widetilde{\boldsymbol{Y}}_k \geq \boldsymbol{x}) \leq C_T^{(u)} P(\boldsymbol{Z}_k \geq \boldsymbol{x}) \qquad \text{for } \boldsymbol{x} \leq \boldsymbol{\Delta}_T$$

which is equivalent to,

$$C_T^{(\ell)} P(\boldsymbol{\xi}_k \geq \boldsymbol{x}) \leq P(\boldsymbol{Y}_k \geq \boldsymbol{x}) \leq C_T^{(u)} P(\boldsymbol{\xi}_k \geq \boldsymbol{x}) \qquad \text{for } \boldsymbol{x} \leq \boldsymbol{h}_k + \boldsymbol{V}_k^{1/2}\boldsymbol{\Delta}_T \tag{21}$$

with $\boldsymbol{\xi}_k = \boldsymbol{h}_k + \boldsymbol{V}_k^{1/2}\boldsymbol{Z}_k \sim N(\boldsymbol{h}_k, \boldsymbol{V}_k)$.

In the argument below we use that for $g : \mathbb{R}^S \to \mathbb{R}$ measurable, if $\boldsymbol{a} \in \mathbb{R}^S$ is such that $P(g(\boldsymbol{X}) \geq g(\boldsymbol{a})) = 1$, then

$$\mathbb{E}[g(\boldsymbol{X})] = \int g(\boldsymbol{x})\, dP(\boldsymbol{X} \leq \boldsymbol{x}) = g(\boldsymbol{a}) + \int_{\boldsymbol{a}}^{(\infty,\ldots,\infty)} P(\boldsymbol{X} \geq \boldsymbol{x})\, dg(\boldsymbol{x}) \tag{22}$$

Let $g(\boldsymbol{x}) = (\hat{h}_{k,\lambda}(\boldsymbol{x}) - h_k)^2 \mathbf{1}_{\{\boldsymbol{x} \in \boldsymbol{B}_k\}}$, where we recall that $\hat{h}_{k,\lambda}(\cdot)$ is the (deterministic) thresholding rule that defines our estimator, and let $\boldsymbol{a} = \inf_{\boldsymbol{x} \in \boldsymbol{B}_k}(\hat{h}_{k,\lambda}(\boldsymbol{x}) - h_k)^2$, with $\boldsymbol{B}_k = \{\boldsymbol{x} : \boldsymbol{h}_k - \boldsymbol{V}_k^{1/2}\boldsymbol{\Delta}_T \leq \boldsymbol{x} \leq \boldsymbol{h}_k + \boldsymbol{V}_k^{1/2}\boldsymbol{\Delta}_T\}$. Note that $\boldsymbol{a}$ exists and is attained for $\boldsymbol{x} \in \boldsymbol{B}_k$, since $\boldsymbol{B}_k$ is closed and bounded and therefore compact.

Using eq.(21) and eq.(22), we can upper bound R_1 by,

$$\begin{aligned}
R_1 &= \int g(\boldsymbol{x})\, dP(\boldsymbol{Y}_k \leq \boldsymbol{x}) \\
&= g(\boldsymbol{a}) + \int_{\boldsymbol{a}}^{(\infty,\ldots,\infty)} P(\boldsymbol{Y}_k \geq \boldsymbol{x})\, d\big[(\hat{h}_{k,\lambda}(\boldsymbol{x}) - h_k)^2 \mathbf{1}_{\{\boldsymbol{x} \in \boldsymbol{B}_k\}}\big] \\
&\leq g(\boldsymbol{a}) + C_T^{(u)} \int_{\boldsymbol{a}}^{(\infty,\ldots,\infty)} P(\boldsymbol{\xi}_k \geq \boldsymbol{x})\, d\big[(\hat{h}_{k,\lambda}(\boldsymbol{x}) - h_k)^2 \mathbf{1}_{\{\boldsymbol{x} \in \boldsymbol{B}_k\}}\big] \\
&\leq (C_T^{(u)} \vee 1)\left\{g(\boldsymbol{a}) + \int_{\boldsymbol{a}}^{(\infty,\ldots,\infty)} P(\boldsymbol{\xi}_k \geq \boldsymbol{x})\, d\big[(\hat{h}_{k,\lambda}(\boldsymbol{x}) - h_k)^2 \mathbf{1}_{\{\boldsymbol{x} \in \boldsymbol{B}_k\}}\big]\right\} \\
&= (C_T^{(u)} \vee 1)\mathbb{E}\left[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2 \mathbf{1}_{\{\boldsymbol{\xi}_k \in \boldsymbol{B}_k\}}\right]
\end{aligned} \tag{23}$$

Completely analogous, we can lower bound R_1 by,

$$R_1 \quad \geq \quad (C_T^{(\ell)} \wedge 1)\mathbb{E}\left[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2 \mathbf{1}_{\{\boldsymbol{\xi}_k \in \boldsymbol{B}_k\}}\right] \tag{24}$$

Combining eq.(23) and eq.(24), and using that $C_T^{(\ell)}, C_T^{(u)} \to 1$ as $T \to \infty$, we conclude that

$$R_1 = (1 + o_T(1))\mathbb{E}\left[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2 \mathbf{1}_{\{\boldsymbol{\xi}_k \in \boldsymbol{B}_k\}}\right] \tag{25}$$

uniformly in $k \in J_{T,\alpha}$. For the other term R_2 in eq.(20), with $\boldsymbol{\Delta}_T = (\Delta_{T,1}, \ldots, \Delta_{T,S})' = (T^\nu, \ldots, T^\nu)$ for some $\nu > 0$, using an upper bound for multivariate Gaussian tail probabilities we find that,

$$\begin{aligned} P(|\widetilde{\boldsymbol{Y}}_k| \geq \boldsymbol{\Delta_T}) &\leq C_T^{(u)} P(\boldsymbol{Z}_k \geq \boldsymbol{\Delta}_T) \\ &\leq C_T^{(u)} \left(\prod_{s=1}^{S} \Delta_{T,s}\right)^{-1} (2\pi)^{-S/2} \exp\left(-\frac{1}{2}\boldsymbol{\Delta}_T' \mathrm{I}_S \boldsymbol{\Delta}_T\right) \\ &\lesssim C_T^{(u)} T^{-S\nu} \exp\left(-\frac{S}{2}T^{2\nu}\right) = O((ST)^{-\mu}) \end{aligned} \tag{26}$$

for arbitrary $0 < \mu < \infty$, since the exponential rate above decays faster than any arbitrary polynomial rate.

Furthermore, it can be verified that

$$\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^4] \leq \mathbb{E}[(|\boldsymbol{w}_k'\boldsymbol{Y}_k - h_k| + |h_k|)^4] \leq \max_s \mathbb{E}[(|Y_{sk} - h_k| + |h_k|)^4] \tag{27}$$

By Parseval's relation,

$$\sup_k |h_k| = \|\boldsymbol{h}\|_\infty \leq \|\boldsymbol{h}\|_2 = \frac{1}{\sqrt{T}}\|\boldsymbol{h}^f\|_2 = \|h^f\|_{L_2} + o_T(1) = O(1) \tag{28}$$

where the last equality is due to the fact that $h^f \in L_2([0, 1/2])$ (here $\|\cdot\|_2$ denotes the Euclidian norm). By Jensen's inequality,

$$\mathbb{E}[|Y_{sk} - h_k|^n] \leq \sqrt{\mathbb{E}[(Y_{sk} - h_k)^{2n}]} = \sqrt{\mathbb{E}[(U_{sk} + \epsilon_{sk})^{2n}]} \qquad \text{for } 1 \leq n \leq 4 \tag{29}$$

By eq.(17) and the cumulant bounds in eq.(15) for all $s = 1, \ldots, S$,

$$\begin{aligned} \mathrm{cum}_1(\epsilon_{sk}) &= O(T^{-1/2}T^{-\alpha/2}) = O(T^{-1/2}) \\ \mathrm{cum}_n(\epsilon_{sk}) &= O(T^{-n/2}(T^{-\alpha/2})^{n-2}) = O(T^{-n/2}) \qquad \text{for } n \geq 2 \end{aligned}$$

Therefore,

$$\mathbb{E}[\epsilon_{sk}^n] = O\left(\sum_{m=1}^{n} \prod_{i_1,\ldots,i_m \,:\, i_1+\ldots+i_m=n, i_j \geq 1} |\mathrm{cum}_{i_j}(\epsilon_{sk})|\right) = O(T^{-n/2}) \tag{30}$$

Also, since the random effects coefficients are assumed to be Gaussian $\mathbb{E}[U_{sk}^n] = O(1)$ for $1 \leq n \leq 8$, we obtain from eq.(27-30) that

$$\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^4] = O(1) \tag{31}$$

Combining the Cauchy-Schwarz inequality and eq.(26) and eq.(31) we find that,

$$R_2 \leq \sqrt{P(|\widetilde{\boldsymbol{Y}}_k| \geq \Delta_T)}\sqrt{\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^4]} = O((ST)^{-\mu})$$

for arbitrary $0 < \mu < \infty$, (abusing notation for μ). Combining eq.(20), eq.(25), and eq.(32) yields,

$$\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^2] \quad = \quad (1 + o_T(1))\mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] + O((ST)^{-\mu}) \tag{32}$$

Since the above equation holds uniformly over $k \in J_{T,\alpha}$, with $|J_{T,\alpha}| = O(T^{1-\alpha})$ for $\alpha > 0$, it follows that,

$$\begin{aligned}\sum_{k\in J_{T,\alpha}} \mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{Y}_k) - h_k)^2] &= (1+o_T(1)) \sum_{k\in J_{T,\alpha}} \mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] + \sum_{k\in J_{T,\alpha}} O((ST)^{-\mu}) \\ &= (1+o_T(1)) \sum_{k\in J_{T,\alpha}} \mathbb{E}[(\hat{h}_{k,\lambda}(\boldsymbol{\xi}_k) - h_k)^2] + O(S^{-\mu}T^{-\mu+1})\end{aligned}$$

for arbitrary $0 < \mu < \infty$, which concludes the proof.

9.2. Proof of Theorem 1

Proof. We split up, using Lemma 1:

$$\begin{aligned}\mathbb{E}\|\hat{\boldsymbol{h}}(\boldsymbol{Y}) - \boldsymbol{h}\|^2 &= \sum_{k\in J_{T,\alpha}} \mathbb{E}[(\hat{h}_k(\boldsymbol{Y}_k) - h_k)^2] + \sum_{k\notin J_{T,\alpha}} h_k^2 \\ &= (1+o_T(1)) \sum_{k\in J_{T,\alpha}} \mathbb{E}[(\hat{h}_k(\boldsymbol{\xi}_k) - h_k)^2] + O(S^{-\mu}T^{1-\mu}) + \sum_{k\notin J_{T,\alpha}} h_k^2 \qquad (33)\end{aligned}$$

for arbitrary $0 < \mu < \infty$. For T sufficiently large ($T \geq T^*$) since $0 < \alpha \leq \alpha^*$ with T^*, α^* as in (A2), the last term on the right-hand side disappears by (A2). It remains to show that the remaining part behaves according to the claimed rates.

By (A1), $\boldsymbol{h} \in \ell_{0,T}(k_{h,T})$ and $\boldsymbol{\sigma}_u^2 \in \ell_{0,T}(k_{u,T})$ with $K_{u,T} \subseteq K_{h,T}$. We decompose the first sum on the right-hand side above into three different regions $\{k \notin K_{h,T}\}$, $\{k \in K_{h,T} \setminus K_{u,T}\}$, and $\{k \in K_{h,T} \cap K_{u,T}\}$, and upper bound by Lemma 2,

$$\begin{aligned}\sum_{k\in J_{T,\alpha}} \mathbb{E}[\hat{h}_k(\boldsymbol{\xi}_k) - h_k)^2] \leq\ & (T - k_{h,T}) \left(\frac{2\sigma_e^2}{TS} \left(1 - \Phi\left(\frac{\sqrt{TS}\lambda_{h,T}}{\sigma_e} \right) \right) + \frac{2\sigma_e \lambda_{h,T}}{\sqrt{TS}} \phi\left(\frac{\sqrt{TS}\lambda_{h,T}}{\sigma_e} \right) \right) \\ & + (k_{h,T} - k_{u,T}) \left(\frac{\sigma_e^2}{TS} + \lambda_{h,T}^2 \right) + k_{u,T} \sup_k \left[\boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k + \lambda_{h,T}^2 \right] \qquad (34)\end{aligned}$$

We observe that $\lambda_{h,T} = \frac{\sigma_e}{\sqrt{TS}} \sqrt{2\log(T/k_{h,T})} > \frac{\sigma_e}{\sqrt{TS}}$ for T large, since $k_{h,T}/T \to 0$ as $T \to \infty$. Therefore, for T sufficiently large,

$$\begin{aligned}\frac{2\sigma_e^2}{TS} \left(1 - \Phi\left(\frac{\sqrt{TS}\lambda_{h,T}}{\sigma_e} \right) \right) &= \frac{2\sigma_e^2}{TS} \int_{\sqrt{TS}\lambda_{h,T}/\sigma_e}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2 \right) dz \\ &\leq \frac{2\sigma_e^2}{TS} \frac{\sigma_e}{\sqrt{TS}\lambda_{h,T}} \int_{\sqrt{TS}\lambda_{h,T}/\sigma_e}^{\infty} \frac{z}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2 \right) dz \\ &\leq \frac{2\sigma_e \lambda_{h,T}}{\sqrt{TS}} \phi\left(\frac{\sqrt{TS}\lambda_{h,T}}{\sigma_e} \right) \\ &= \frac{2\sigma_e^2}{TS} \sqrt{2\log\left(\frac{T}{k_{h,T}} \right)} \phi(0) \frac{k_{h,T}}{T} \qquad (35)\end{aligned}$$

where in the second step we use $\frac{z}{\sqrt{TS}\lambda_{h,T}/\sigma_e} \geq 1$ and in the third step $\frac{\sigma_e^2}{TS} \leq \lambda_{h,T}^2$ and $\int_\lambda^\infty z\phi(z)dz = \phi(\lambda)$. By eq.(35) and plugging in $\lambda_{h,T}$, the right-hand side in eq.(34) is upper bounded by:

$$\begin{aligned}
\sum_{k\in J_{T,\alpha}} \mathbb{E}[\hat{h}_k(\boldsymbol{\xi}_k) - h_k)^2] \quad &\leq \quad (T - k_{h,T})\left(\frac{4\sigma_e^2}{TS}\sqrt{2\log\left(\frac{T}{k_{h,T}}\right)}\phi(0)\frac{k_{h,T}}{T}\right) \\
&\quad +(k_{h,T} - k_{u,T})\left(\frac{\sigma_e^2}{TS} + \frac{2\sigma_e^2}{TS}\log\left(\frac{T}{k_{h,T}}\right)\right) \\
&\quad +k_{u,T}\sup_k\left[\boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k + \frac{2\sigma_e^2}{TS}\log\left(\frac{T}{k_{h,T}}\right)\right] \\
&\leq \quad \frac{k_{h,T}\sigma_e^2}{TS}\left(1 + 4\phi(0)\sqrt{2\log\left(\frac{T}{k_{h,T}}\right)} + 2\log\left(\frac{T}{k_{h,T}}\right)\right) + k_{u,T}\left(\sup_k \boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k - \frac{\sigma_e^2}{TS}\right) \\
&\lesssim \quad \frac{k_{h,T}}{TS}\log\left(\frac{T}{k_{h,T}}\right) + k_{u,T}\left(\sup_k \boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k - \frac{\sigma_e^2}{TS}\right) \qquad (36)
\end{aligned}$$

By plugging eq.(36) into eq.(33), we obtain for T sufficiently large

$$\mathbb{E}\|\hat{\boldsymbol{h}}(\boldsymbol{Y}) - \boldsymbol{h}\|^2 \quad \lesssim \quad \frac{k_{h,T}}{TS}\log\left(\frac{T}{k_{h,T}}\right) + k_{u,T}\left(\sup_k \boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k - \frac{\sigma_e^2}{TS}\right) + S^{-\mu}T^{1-\mu}$$

Since $0 < \mu < \infty$ can be chosen arbitrarily large, for $\mu > 2$ the term $\frac{k_{h,T}}{TS}\log\left(\frac{T}{k_{h,T}}\right)$ dominates $S^{-\mu}T^{1-\mu}$, thus concluding the proof.

9.3. *Proof of Theorem 2*

9.3.1. *Proof of the convergence in eq.*(11)

Proof. First we show that the inclusion $K_{u,T} \subseteq \widehat{K}_u(\boldsymbol{Y})$ holds with probability tending to 1 as $S, T \to \infty$. Since $\widehat{K}_u(\boldsymbol{Y}) = \{k : |T_k(\boldsymbol{Y}_k)| \geq \lambda_{u,T}\}$, it suffices to show that

$$P\left(\inf_{k\in K_{u,T}} |T_k(\boldsymbol{Y}_k)| \geq \lambda_{u,T}\right) \to 1, \qquad \text{as } S, T \to \infty \qquad (37)$$

Writing $\sigma_{k,T}^2 = \sigma_{uk}^2 + \sigma_e^2/T$ and $\hat{\sigma}_k^2(\boldsymbol{Y}_k) = \frac{1}{S}\sum_{s=1}^S (Y_{sk} - \hat{h}_k)^2$, since $\inf_{k\in K_{u,T}} \sigma_{k,T}^2 \geq \delta > 0$ we can lower bound,

$$\begin{aligned}
\inf_{k\in K_{u,T}} |T_k(\boldsymbol{Y}_k)| \quad &= \quad \inf_{k\in K_{u,T}}\left|\log(\hat{\sigma}_k^2(\boldsymbol{Y}_k)) + \log\left(\frac{T}{\sigma_e^2}\right) - \left(\log\left(\frac{2}{S}\right) + \psi^{(0)}\left(\frac{S}{2}\right)\right)\right| \\
&\geq \quad \inf_{k\in K_{u,T}}\left|\log(\sigma_{k,T}^2) + \log\left(\frac{T}{\sigma_e^2}\right) - \left(\log\left(\frac{2}{S}\right) + \psi^{(0)}\left(\frac{S}{2}\right)\right)\right| \\
&\quad - \sup_{k\in K_{u,T}}\left|\log(\hat{\sigma}_k^2(\boldsymbol{Y}_k) - \log(\sigma_{k,T}^2)\right|
\end{aligned}$$

It can be verified that $\psi^{(0)}(x) = \log(x) + O(x^{-1})$, therefore $|\log(2/S) + \psi^{(0)}(S/2)| = o_S(1)$, and the first term on the right-hand side is seen to grow at the rate $\sim \log(T)$ for S, T increasing. For the second term, we show that $\sup_{k\in K_{u,T}} |\log(\hat{\sigma}_k^2(\boldsymbol{Y}_k) - \log(\sigma_{k,T}^2)| = o_p^{S,T}(1)$. First, decompose

$$\begin{aligned}\hat{\sigma}_k^2(\boldsymbol{Y}_k) &= \frac{1}{S}\sum_{s=1}^{S}((Y_{sk}-h_k)+(h_k-\hat{h}_k))^2 \\ &= \underbrace{\frac{1}{S}\sum_{s=1}^{S}(Y_{sk}-h_k)^2}_{\text{(i)}} + \underbrace{\frac{2}{S}\sum_{s=1}^{S}(Y_{sk}-h_k)(h_k-\hat{h}_k)}_{\text{(ii)}} + \underbrace{\frac{1}{S}\sum_{s=1}^{S}(h_k-\hat{h}_k)^2}_{\text{(iii)}} \qquad (38)\end{aligned}$$

For $k \in K_{u,T}$, term (ii) and term (iii) uniformly converge to zero in probability as $S,T \to \infty$, which follows from the fact that $\sup_k |h_k - \hat{h}_k| = o_p^{S,T}(1)$ and $\frac{2}{S}\sum_{s=1}^{S}(Y_{sk}-h_k) = O_p(1)$ since $\mathrm{Var}(Y_{sk}) < \infty$.

For term (i), by linearity of the expectation and recalling that $Y_{sk} = h_k + U_{sk} + \epsilon_{sk}$,

$$\mathbb{E}\left[\frac{1}{S}\sum_{s=1}^{S}(Y_{sk}-h_k)^2\right] = \frac{1}{S}\sum_{s=1}^{S}\mathrm{Var}(U_{sk}+\epsilon_{sk}) = \sigma_{uk}^2 + \sigma_e^2/T = \sigma_{k,T}^2$$

where we use that by assumption $\epsilon_{sk} \sim (0, \sigma_e^2/T)$. Also,

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{S}\sum_{s=1}^{S}(Y_{sk}-h_k)^2\right) &= \frac{1}{S^2}\sum_{s=1}^{S}\mathrm{Var}((U_{sk}+\epsilon_{sk})^2) + \frac{1}{S^2}\sum_{s\neq s'}\mathrm{Cov}((U_{sk}+\epsilon_{sk})^2, (U_{s'k}+\epsilon_{s'k})^2) \\ &= \frac{1}{S^2}\sum_{s=1}^{S}\mathrm{Var}((U_{sk}+\epsilon_{sk})^2) + \frac{1}{S^2}\sum_{s\neq s'}\mathrm{Cov}(U_{sk}^2, U_{s'k}^2)\end{aligned}$$

where in the second step we use that $\epsilon_{sk} \perp U_{s'k}$ for all s, s', $\epsilon_{sk} \perp \epsilon_{s'k}$ for $s \neq s'$, and the fact that $\mathbb{E}[\epsilon_{sk}] = 0$ for all s, k. We observe that $\mathrm{Var}((U_{sk}+\epsilon_{sk})^2) = O(1)$ uniformly over $k = 1, \ldots, T$, since $U_{sk} \sim N(0, \sigma_{uk}^2)$ with $\sup_k \sigma_{uk}^2 < \infty$, and $\epsilon_{sk} \sim (0, \sigma_e^2/T)$ with $\mathbb{E}[\epsilon_{sk}^4] = O(T^{-2})$ is independent of U_{sk} for all s, k. Moreover, by Gaussianity of the random effects coefficients, it can be verified that

$$\mathrm{Cov}(U_{sk}^2, U_{s'k}^2) = 2\rho_{ss'}^2\sigma_{uk}^4$$

Therefore, using that $\sup_k \sigma_{uk}^2 < \infty$,

$$\mathrm{Var}\left(\frac{1}{S}\sum_{s=1}^{S}(Y_{sk}-h_k)^2\right) = O(S^{-1}) + 2\sigma_{uk}^4\frac{\|\boldsymbol{G}_S\|_F^2}{S^2} = O(S^{-1}) + O\left(\frac{\|\boldsymbol{G}_S\|_F^2}{S^2}\right)$$

which converges to zero uniformly over $k \in \{1, \ldots, T\}$, since $\|\boldsymbol{G}_S\|_F/S \to 0$ by assumption. By Chebychev's inequality $\frac{1}{S}\sum_{s=1}^{S}(Y_{sk}-h_k)^2 \xrightarrow{P} \sigma_{k,T}^2$ uniformly over $k \in K_{u,T}$ as $S \to \infty$, and by Slutsky's lemma it follows from eq.(38) that $\hat{\sigma}_k^2(\boldsymbol{Y}_k) \xrightarrow{P} \sigma_{k,T}^2$ uniformly over $k \in K_{u,T}$ as $S,T \to \infty$.

For the uniform convergence in probability of $\log(\hat{\sigma}_k^2(\boldsymbol{Y}_k))$ over $k \in K_{u,T}$, we show that for any $\epsilon, \gamma > 0$, there exist S_0, T_0 sufficiently large (depending on only ϵ, γ) such that

$$P(|\log(\hat{\sigma}_k^2(\boldsymbol{Y}_k)) - \log(\sigma_{k,T}^2)| > \epsilon) < \gamma \qquad (39)$$

for $S > S_0, T > T_0$ and all $k \in K_{u,T}$. For arbitrary $\epsilon > 0$, by the law of total probability,

$$P(|\log(\hat{\sigma}_k^2) - \log(\sigma_{k,T}^2)| > \epsilon) \;=\; P(|\log(\hat{\sigma}_k^2) - \log(\sigma_{k,T}^2)| > \epsilon : \hat{\sigma}_k^2 \geq \delta/2)P(\hat{\sigma}_k^2 \geq \delta/2) + P(\hat{\sigma}_k^2 < \delta/2)$$

Since $\log(x)$ has bounded derivative on $x \in [\delta/2, \infty)$ for δ bounded away from zero, it is uniformly continuous on the domain $[\delta/2, \infty)$. Thus, there exists $\epsilon_1 > 0$ such that $|x - x_0| \leq \epsilon_1$ implies $|\log(x) - \log(x_0)| \leq \epsilon$ for all $x, x_0 \in [\delta/2, \infty)$. Using that $\inf_{k \in K_{u,T}} \sigma_k^2 \geq \delta$, for this choice of ϵ_1 we get,

$$\begin{aligned} P(|\log(\hat{\sigma}_k^2) - \log(\sigma_{k,T}^2)| > \epsilon : \hat{\sigma}_k^2 \geq \delta/2) P(\hat{\sigma}_k^2 \geq \delta/2) &\leq P(|\hat{\sigma}_k^2 - \sigma_{k,T}^2| > \epsilon_1 : \hat{\sigma}_k^2 \geq \delta/2) P(\hat{\sigma}_k^2 \geq \delta/2) \\ &\leq P(|\hat{\sigma}_k^2 - \sigma_{k,T}^2| > \epsilon_1) \end{aligned}$$

By the uniform convergence in probability of $\hat{\sigma}_k^2(\boldsymbol{Y}_k)$, there exist $S_0^{(1)}, T_0^{(1)}$ such that

$$P(|\hat{\sigma}_k^2 - \sigma_{k,T}^2| > \epsilon_1) < \gamma/2$$

for $S > S_0^{(1)}, T > T_0^{(1)}$ and all $k \in K_{u,T}$. Similarly, using again that $\inf_{k \in K_{u,T}} \sigma_{k,T}^2 \geq \delta$, there exist $S_0^{(2)}, T_0^{(2)}$ such that

$$P(\hat{\sigma}_k^2 < \delta/2) < P(|\hat{\sigma}_k^2 - \sigma_{k,T}^2| > \delta/2) < \gamma/2$$

for $S > S_0^{(2)}, T > T_0^{(2)}$ and all $k \in K_{u,T}$. The uniform convergence in probability in eq.(39) now follows from the above arguments with $S_0 = S_0^{(1)} \vee S_0^{(2)}$ and $T_0 = T_0^{(1)} \vee T_0^{(2)}$.

We conclude that the left-hand side inside the probability in eq.(37) grows at the rate $\sim \log(T)$ in probability for increasing S, T. On the other hand, $\lambda_{u,T} = o(\log(T))$ by assumption. Combining these two results implies that $K_{u,T} \subseteq \widehat{K}_u(\boldsymbol{Y})$ with probability tending to 1 as $S, T \to \infty$.

Next, we show that the other inclusion $\widehat{K}_u(\boldsymbol{Y}) \subseteq K_{u,T}$ also holds with probability tending to 1 as $S, T \to \infty$, which is equivalent to showing

$$P\left(\sup_{k \notin K_{u,T}} |T_k(\boldsymbol{Y}_k)| \geq \lambda_{u,T}\right) \to 0, \qquad \text{as } S, T \to \infty$$

For $k \notin K_{u,T}$, eq.(38) term (ii) and term (iii) are $o_p^{S,T}(T^{-1})$, which is obtained by combining $\sup_{k \notin K_{u,T}} |h_k - \hat{h}_k| = o_p^{S,T}(T^{-1/2})$ and $(Y_{sk} - h_k) = O_p(T^{-1/2})$ since $\mathrm{Var}(Y_{sk}) = \sigma_e^2/T$. Term (i) in eq.(38) satisfies uniformly over $k \notin K_{u,T}$,

$$\frac{1}{S}\sum_{s=1}^{S}(Y_{sk} - h_k)^2 = \mathbb{E}[(Y_{sk} - h_k)^2] + O_p\left(\sqrt{\mathrm{Var}\left(\frac{1}{S}\sum_{s=1}^{S}(Y_{sk} - h_k)^2\right)}\right) = \frac{\sigma_e^2}{T} + o_p^{S,T}(T^{-1}) \tag{40}$$

since $\mathrm{Var}((Y_{sk} - h_k)^2) = \mathrm{Var}(\epsilon_{sk}^2) = O(T^{-2})$ for $k \notin K_{u,T}$. Combining the three terms it follows that,

$$\sup_{k \notin K_{u,T}} \left| \frac{T}{\sigma_e^2} \hat{\sigma}_k^2(\boldsymbol{Y}_k) - 1 \right| \xrightarrow{P} 0, \qquad \text{as } S, T \to \infty \tag{41}$$

By continuity of the logarithm, for arbitrary $\epsilon > 0$ and fixing $x_0 = 1$, there exists $\epsilon_1 = \epsilon_1(x_0)$ independent of k, such that $|x - x_0| \leq \epsilon_1$ implies $|\log(x) - \log(x_0)| \leq \epsilon$. For this choice of ϵ_1,

$$P(|\log(T\hat{\sigma}_k^2(\boldsymbol{Y}_k)/\sigma_e^2) - \log(1)| > \epsilon) \leq P(|T\hat{\sigma}_k^2(\boldsymbol{Y}_k)/\sigma_e^2 - 1| > \epsilon_1)$$

Since ϵ_1 only depends on $x_0 = 1$, and $T\hat{\sigma}_k^2(\boldsymbol{Y}_k)/\sigma_e^2 \xrightarrow{P} 1$ uniformly over $k \notin K_{u,T}$, we obtain

$$\sup_{k \notin K_{u,T}} |\log(\hat{\sigma}_k^2(\boldsymbol{Y}_k)) + \log(T/\sigma_e^2)| \xrightarrow{P} 0, \qquad \text{as } S, T \to \infty$$

Therefore, by the triangle inequality,

$$\sup_{k \notin K_{u,T}} |T_k(\boldsymbol{Y}_k)| \le \sup_{k \notin K_{u,T}} |\log(\hat{\sigma}_k^2(\boldsymbol{Y}_k)) + \log(T/\sigma_e^2)| + |\log(2/S) + \psi^{(0)}(S/2)| \xrightarrow{P} 0, \quad \text{as } S,T \to \infty$$

since $|\log(2/S) + \psi^{(0)}(S/2)| = o_S(1)$. On the other hand, $\lambda_{u,T} \geq C$ for some constant $C > 0$ by assumption. Combining these two results implies that $\widehat{K}_u(\boldsymbol{Y}) \subseteq K_{u,T}$ with probability tending to 1 as $S, T \to \infty$. To conclude, since both $P(K_{u,T} \subseteq \widehat{K}_u(\boldsymbol{Y})) \to 1$ and $P(\widehat{K}_u(\boldsymbol{Y}) \subseteq K_{u,T}) \to 1$ as $S, T \to \infty$, this also implies $P(\widehat{K}_u(\boldsymbol{Y}) = K_{u,T}) \to 1$ as $S, T \to \infty$.

9.3.2. *Uniform consistency of the estimators of σ_{uk}^2*

Proof. From the proof in Section 9.3.1 we already know that $\sup_k |\hat{\sigma}_k^2(\boldsymbol{Y}_k) - \sigma_{k,T}^2| \xrightarrow{P} 0$ as $S, T \to \infty$ for $k \in \{1, \ldots, T\}$. It remains to show that $\sup_k |\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k) - \sigma_{uk}^2| = o_p^{S,T}(1)$ with $k \in \{1, \ldots, T\}$. For the linear part of the estimator, consider the function $g(x) = \{x - \frac{\sigma_e^2}{T}\}_+$ which is uniformly continuous on $x \in [0, \infty)$ since $|g(x) - g(x_0)| \leq |x - x_0|$ for all $x, x_0 \in [0, \infty)$. Therefore, by a similar argument as in Section 9.3.1,

$$\sup_k \left| \left\{ \hat{\sigma}_k^2(\boldsymbol{Y}_k) - \frac{\sigma_e^2}{T} \right\}_+ - \sigma_{uk}^2 \right| \xrightarrow{P} 0, \qquad \text{as } S, T \to \infty \tag{42}$$

In order to show the uniform convergence in probability of the nonlinear estimators $\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k)$ over $k \in \{1, \ldots, T\}$, it suffices to show that for any $\epsilon, \gamma > 0$, there exist S_0, T_0 sufficiently large (depending only on ϵ, γ) such that

$$P(|\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k) - \sigma_{uk}^2| > \epsilon) < \gamma$$

for $S > S_0, T > T_0$ and all $k \in \{1, \ldots, T\}$.

First, consider the case $k \in K_{u,T}$, by the law of total probability for arbitrary $\epsilon > 0$,

$$\begin{aligned} P(|\hat{\sigma}_{uk}^2 - \sigma_{uk}^2| > \epsilon \mid k \in K_{u,T}) &= P(|\hat{\sigma}_{uk}^2 - \sigma_{uk}^2| > \epsilon \mid k \in \widehat{K}_u) P(k \in \widehat{K}_u \mid k \in K_{u,T}) \\ &\quad + P(|\hat{\sigma}_{uk}^2 - \sigma_{uk}^2| > \epsilon \mid k \notin \widehat{K}_u) P(k \notin \widehat{K}_u \mid k \in K_{u,T}) \\ &\leq P(|\{\hat{\sigma}_k^2 - \sigma_e^2/T\}_+ - \sigma_{uk}^2| > \epsilon) + P(k \notin \widehat{K}_u \mid k \in K_{u,T}) \end{aligned}$$

For any $\gamma > 0$, there exist $S_0^{(1)}, T_0^{(1)}$ such that for $S > S_0^{(1)}, T > T_0^{(1)}$ we have $P(|\{\hat{\sigma}_k^2 - \sigma_e^2/T\}_+ - \sigma_{uk}^2| > \epsilon) < \gamma/2$ for all $k \in K_{u,T}$ due to eq.(42), and $P(k \notin \widehat{K}_u \mid k \in K_{u,T}) < \gamma/2$ for all $k \in K_{u,T}$, since $P(K_{u,T} \subseteq \widehat{K}_u(\boldsymbol{Y})) \to 1$. Thus for any $\epsilon, \gamma > 0$,

$$P(|\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k) - \sigma_{uk}^2| > \epsilon) < \gamma$$

for $S > S_0^{(1)}, T > T_0^{(1)}$ and all $k \in K_{u,T}$.

Second, consider the case $k \notin K_{u,T}$, again by the law of total probability for arbitrary $\epsilon > 0$,

$$\begin{aligned} P(|\hat{\sigma}_{uk}^2 - \sigma_{uk}^2| > \epsilon \mid k \notin K_{u,T}) &= P(|\hat{\sigma}_{uk}^2 - \sigma_{uk}^2| > \epsilon \mid k \in \widehat{K}_u) P(k \in \widehat{K}_u \mid k \notin K_{u,T}) \\ &\quad + P(|\hat{\sigma}_{uk}^2 - \sigma_{uk}^2| > \epsilon \mid k \notin \widehat{K}_u) P(k \notin \widehat{K}_u \mid k \notin K_{u,T}) \\ &\leq P(|\{\hat{\sigma}_k^2 - \sigma_e^2/T\}_+ - \sigma_{uk}^2| > \epsilon) + P(\sigma_{uk}^2 > \epsilon \mid k \notin K_{u,T}) \end{aligned}$$

We note that $P(\sigma_{uk}^2 > \epsilon \mid k \notin K_{u,T}) = 0$ for all $\epsilon > 0$ since $\sigma_{uk}^2 = 0$ for $k \notin K_{u,T}$, and again by eq.(38) for any $\gamma > 0$ there exist $S_0^{(2)}, T_0^{(2)}$ such that $P(|\{\hat{\sigma}_k^2 - \sigma_e^2/T\}_+ - \sigma_{uk}^2| > \epsilon) < \gamma$ for all $k \notin K_{u,T}$. Thus for any $\epsilon, \gamma > 0$,

$$P(|\hat{\sigma}_{uk}^2(\boldsymbol{Y}_k) - \sigma_{uk}^2| > \epsilon) < \gamma$$

for $S > S_0^{(2)}$, $T > T_0^{(2)}$ and all $k \notin K_{u,T}$. Let $T_0 = T_0^{(1)} \vee T_0^{(2)}$ and $S_0 = S_0^{(1)} \vee S_0^{(2)}$, then the uniform convergence in probability with $k \in \{1, \ldots, T\}$ in accordance with eq.(12) follows from the above arguments.

9.3.3. *Consistency of the estimators of ρ_{ij}*

Proof. We show consistency of the estimators $\hat{\rho}_{ij}(\boldsymbol{Y})$ marginally for each i, j with $i \neq j$. Writing $\widehat{K}_u = \widehat{K}_u(\boldsymbol{Y})$ and $\hat{k}_u = |\widehat{K}_u(\boldsymbol{Y})|$, we decompose:

$$
\begin{aligned}
\hat{\rho}_{ij}(\boldsymbol{Y}) &= \frac{1}{\hat{k}_u} \sum_{k \in \widehat{K}_u} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta} \\
&= \frac{k_{u,T}}{\hat{k}_u} \Bigg(\underbrace{\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta}}_{\text{(i)}} + \underbrace{\frac{1}{k_{u,T}} \sum_{k \in \widehat{K}_u \setminus K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta}}_{\text{(ii)}} \\
&\qquad \underbrace{- \frac{1}{k_{u,T}} \sum_{k \in K_{u,T} \setminus \widehat{K}_u} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta}}_{\text{(iii)}} \Bigg)
\end{aligned}
\tag{43}
$$

In part (i) below it is shown that term (i) converges to ρ_{ij} in probability as $S, T \to \infty$. In part (ii) it is shown that the terms (ii) and (iii) converge to zero in probability as $S, T \to \infty$. Then, combining these results, and observing that $k_{u,T}/\hat{k}_u \xrightarrow{P} 1$ as $S, T \to \infty$ since $P(\widehat{K}_u = K_{u,T}) \to 1$, an application of Slutsky's lemma yields the claimed result.

Part (i) From the proof in Section 9.3.2, we know that $\sup_{k \in K_{u,T}} |\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) - \sigma^2_{uk}| = o_p^{S,T}(1)$. Therefore, term (i) in eq.(43) can be rewritten as,

$$
\begin{aligned}
\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta} &= \frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\sigma^2_{uk}} \cdot \frac{\sigma^2_{uk}}{(\sigma^2_{uk} + o_p^{S,T}(1)) \vee \delta} \\
&= (1 + o_p^{S,T}(1)) \cdot \left(\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\sigma^2_{uk}} \right)
\end{aligned}
\tag{44}
$$

where we also use that $\inf_{k \in K_{u,T}} \sigma^2_{uk} \geq \delta$. We show that $\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\sigma^2_{uk}} \xrightarrow{P} \rho_{ij}$ as $S, T \to \infty$. Writing out further,

$$
\begin{aligned}
\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\sigma^2_{uk}} &= \underbrace{\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - h_k)(Y_{jk} - h_k)}{\sigma^2_{uk}}}_{\text{(i)}} + \underbrace{\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{jk} - h_k)(\hat{h}_k - h_k)}{\sigma^2_{uk}}}_{\text{(ii)}} \\
&\quad + \underbrace{\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - h_k)(\hat{h}_k - h_k)}{\sigma^2_{uk}}}_{\text{(iii)}} + \underbrace{\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(\hat{h}_k - h_k)^2}{\sigma^2_{uk}}}_{\text{(iv)}}
\end{aligned}
\tag{45}
$$

Term (ii), (iii), and (iv) converge to zero in probability as $S, T \to \infty$. This is observed from combining $\sup_k (\hat{h}_k - h_k) = o_p^{S,T}(1)$, $\inf_{k \in K_{u,T}} \sigma^2_{uk} > 0$, and respectively $\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} (Y_{jk} - h_k) = O_p(1)$

for term (ii) and $\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} (Y_{ik} - h_k) = O_p(1)$ for term (iii).

It remains to show that term (i) in eq.(45) converges to ρ_{ij} in probability. By linearity of the expectation,

$$\mathbb{E}\left[\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - h_k)(Y_{jk} - h_k)}{\sigma^2_{uk}}\right] = \frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{\sigma^2_{uk}\rho_{ij}}{\sigma^2_{uk}} = \rho_{ij}$$

Thus, by Chebychev's inequality:

$$P\left(\left|\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - h_k)(Y_{jk} - h_k)}{\sigma^2_{uk}} - \rho_{ij}\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\mathrm{Var}\left(\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - h_k)(Y_{jk} - h_k)}{\sigma^2_{uk}}\right) \tag{46}$$

Since $E[Y_{ik}^4] = O(1)$ for all $i = 1, \ldots, S$, by Cauchy-Schwarz's inequality $\sup_{k \in K_{u,T}} \mathrm{Var}((Y_{ik} - h_k)(Y_{jk} - h_k))/\sigma^4_{uk} = O(1)$ for all $i \neq j$. Recalling that $Y_{ik} = h_k + U_{ik} + \epsilon_{ik}$, it can be verified that,

$$\mathrm{Cov}\left(\frac{(Y_{ik} - h_k)(Y_{jk} - h_k)}{\sigma^2_{uk}}, \frac{(Y_{ik'} - h'_k)(Y_{jk'} - h'_k)}{\sigma^2_{uk'}}\right) = \frac{1}{\sigma^2_{uk}\sigma^2_{uk'}}\mathrm{Cov}(\epsilon_{ik}, \epsilon_{ik'})\mathrm{Cov}(\epsilon_{jk}, \epsilon_{jk'})$$

where we use that, $\mathrm{Cov}(U_{ik}, U_{jk'}) = 0$ for all i, j and $k \neq k'$, $U_{ik} \perp \epsilon_{jk'}$ for all i, j and k, k', and $\epsilon_{ik} \perp \epsilon_{jk'}$ for all $i \neq j$ and k, k'. We argue that ϵ_{ik} and $\epsilon_{ik'}$ are asymptotically uncorrelated for all i and $k \neq k'$. In the frequency domain, the noise terms at different frequencies $E_i^f(\omega_\ell)$ and $E_i^f(\omega_{\ell'})$ for $\ell \neq \ell'$ are asymptotically independent (see Brillinger (1981)), therefore

$$\mathrm{Cov}(\boldsymbol{E}_i^f) = \sigma_e^2 \mathrm{I}_T + \boldsymbol{\Delta}_i^f, \qquad \text{s.t. } \|\boldsymbol{\Delta}_i^f\|_F = o^T(1)$$

where $\boldsymbol{E}_i^f = (E_i^f(\omega_0), \ldots, E_i^f(\omega_{T-1}))'$. Projecting to the coefficient domain with orthogonal discrete wavelet transform-matrix $\boldsymbol{W}$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$, this yields

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \mathrm{Cov}(\boldsymbol{W}\boldsymbol{E}_i^f) &= \sigma_e^2 \boldsymbol{W}\mathrm{I}_T\boldsymbol{W}' + \boldsymbol{W}\boldsymbol{\Delta}_i^f\boldsymbol{W}' \\ &= \frac{\sigma_e^2}{T}\mathrm{I}_T + \boldsymbol{W}\boldsymbol{\Delta}_i^f\boldsymbol{W}'\end{aligned}$$

and by the rotational invariance of the Frobenius-norm,

$$\|\boldsymbol{W}\boldsymbol{\Delta}_i^f\boldsymbol{W}'\|_F = \|\boldsymbol{\Delta}_i^f\boldsymbol{W}'\boldsymbol{W}\|_F = \frac{1}{T}\|\boldsymbol{\Delta}_i^f\|_F = o^T(T^{-1})$$

from which we conclude that ϵ_{ik} and $\epsilon_{ik'}$ are asymptotically uncorrelated for all $k \neq k$. Combining the arguments above, it follows that,

$$\mathrm{Var}\left(\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - h_k)(Y_{jk} - h_k)}{\sigma^2_{uk}}\right) \lesssim \frac{1}{k_{u,T}} + \frac{(k_{u,T}^2 - k_{u,T})}{k_{u,T}^2} \cdot o^T(T^{-1}) \to 0, \quad \text{as } T \to \infty$$

using that $k_{u,T} \to \infty$ as $T \to \infty$. From eq.(45), eq.(46) and Slutsky's lemma, we conclude that,

$$\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\sigma^2_{uk}} \xrightarrow{P} \rho_{ij} \qquad \text{as } S, T \to \infty$$

Returning to eq.(44),

$$\frac{1}{k_{u,T}} \sum_{k \in K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta} = (1 + o_p^{S,T}(1)) \cdot (\rho_{ij} + o_p^{S,T}(1)) \xrightarrow{P} \rho_{ij} \qquad \text{as } S, T \to \infty$$

Part (ii) First we show that term (ii) in eq.(43) converges to zero in probability. Since $\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta \geq \delta$,

$$P\left(\left|\frac{1}{k_{u,T}} \sum_{k \in \widehat{K}_u \setminus K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\hat{\sigma}^2_{uk}(\boldsymbol{Y}_k) \vee \delta}\right| > \epsilon\right) \leq P\left(\left|\frac{1}{k_{u,T}} \sum_{k \in \widehat{K}_u \setminus K_{u,T}} \frac{(Y_{ik} - \hat{h}_k)(Y_{jk} - \hat{h}_k)}{\delta}\right| > \epsilon\right)$$

Therefore it suffices to show that the probability on the right-hand side converges to zero for all $\epsilon > 0$ as $S, T \to \infty$. For ease of notation write $\zeta_k := \frac{(Y_{ik}-\hat{h}_k)(Y_{jk}-\hat{h}_k)}{\delta}$. By the law of total probability, for all $\epsilon > 0$,

$$\begin{aligned}
P\left(\left|\frac{1}{k_{u,T}} \sum_{k \in \widehat{K}_u \setminus K_{u,T}} \zeta_k\right| > \epsilon\right) &= P\left(\left|\frac{1}{k_{u,T}} \sum_{k \in \widehat{K}_u \setminus K_{u,T}} \zeta_k\right| > \epsilon \ : \ \#\{\widehat{K}_u \setminus K_{u,T}\} = 0\right) P\left(\#\{\widehat{K}_u \setminus K_{u,T}\} = 0\right) \\
&\quad + P\left(\left|\frac{1}{k_{u,T}} \sum_{k \in \widehat{K}_u \setminus K_{u,T}} \zeta_k\right| > \epsilon \ : \ \#\{\widehat{K}_u \setminus K_{u,T}\} > 0\right) P\left(\#\{\widehat{K}_u \setminus K_{u,T}\} > 0\right) \\
&\leq 0 \cdot P\left(\#\{\widehat{K}_u \setminus K_{u,T} = 0\right) + 1 \cdot P\left(\#\{\widehat{K}_u \setminus K_{u,T}\} > 0\right)
\end{aligned}$$

Since $P(\widehat{K}_u = K_{u,T}) \to 1$ as $S, T \to \infty$, also $P(\#\{\widehat{K}_u \setminus K_{u,T}\} = 0) \to 1$, thus the right-hand side above converges to zero as $S, T \to \infty$.

Completely analogous, using that $P(\#\{K_{u,T} \setminus \widehat{K}_u\} = 0) \to 1$ as $S, T \to \infty$, we find that term (iii) in eq.(43) converges to zero in probability as $S, T \to \infty$.

9.4. Proof of Lemma 3.

Proof. The proof consists of two parts. In the first part the convergence in distribution in eq.(13) is shown. In the second part we show that $\widehat{K}_u(\boldsymbol{\xi}) = K_{u,T}$ with probability tending to 1 as $S, T \to \infty$.

Part (i) For uncorrelated replicates ($\boldsymbol{G}_S = \mathrm{I}_S$), we can write $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T \overset{\text{iid}}{\sim} N(\boldsymbol{0}, (\sigma^2_{uk} + \sigma^2_e/T)\mathrm{I}_S)$, and thus

$$\log\left\{\frac{1}{S}\sum_{s=1}^{S}(\xi_{sk} - h_k)^2\right\} \overset{d}{=} \log(\sigma^2_{uk} + \sigma^2_e/T) + \log(A^2_S/S), \qquad \text{for } k = 1, \ldots, T$$

for some $A^2_S \sim \chi^2_S$. The term $\log(A^2_S/S)$ tends to a normal distribution as $S \to \infty$, and by Pav (2015),

$$\begin{aligned}
\mathbb{E}\left[\log(A^2_S/S)\right] &= \log(2/S) + \psi^{(0)}(S/2) \\
\mathrm{Var}(\log(A^2_S/S)) &= \psi^{(1)}(S/2)
\end{aligned}$$

where $\psi^{(0)}(\cdot)$ and $\psi^{(1)}(\cdot)$ denote the digamma and trigamma function. Combining the above results with the definition of $T_k(\boldsymbol{\xi}_k)$ implies the weak convergence in eq.(13). For correlated replicates, the noise coefficients ϵ_{sk} remain independent across replicates, and since $\sigma^2_{uk} = 0$ for $k \notin K_{u,T}$, it remains true that,

$$\frac{1}{\sqrt{\psi^{(1)}(S/2)}} T_k(\boldsymbol{\xi}_k) \xrightarrow{d} N(0, 1) \quad \text{if } k \notin K_{u,T} \tag{47}$$

where the convergence is uniformly over $k \notin K_{u,T}$, since $\sigma^2_{uk} = 0$ implies that $T_k(\boldsymbol{\xi}_k)$ is independent of k.

Part (ii) By the same argument as in Section 9.3.1, it follows that $K_{u,T} \subseteq \widehat{K}_u(\boldsymbol{\xi})$ with probability tending to 1 as $S, T \to \infty$, using that $\lambda_{u,T} = o(\log(T))$ as in Theorem 2.

In order to show $\widehat{K}_u(\boldsymbol{\xi}) \subseteq K_{u,T}$ with probability tending to 1 as $S, T \to \infty$, we use a standard argument (see Johnstone (2015) Section 8.3, 8.10). By the uniform weak convergence in eq.(47) for $k \notin K_{u,T}$,

$$P\left(\frac{T_k(\boldsymbol{\xi}_k)}{\sqrt{\psi^{(1)}(S/2)}} \leq x\right) = (1 + R_S)P(Z \leq x), \qquad Z \sim N(0,1)$$

where $R_S = o_S(1)$ is uniform over $k \notin K_{u,T}$. By independence across indices k and using Gaussian tail probabilities, we upper bound

$$\begin{aligned}
P\left(\sup_{k \notin K_{u,T}} \frac{|T_k(\boldsymbol{\xi}_k)|}{\sqrt{\psi^{(1)}(S/2)}} \geq \sqrt{2\log(T - k_{u,T})}\right) &\leq 1 - \left(1 - 2P\left(\frac{T_k(\boldsymbol{\xi}_k)}{\sqrt{\psi^{(1)}(S/2)}} \geq \sqrt{2\log(T - k_{u,T})}\right)\right)^{T-k_{u,T}} \\
&= 1 - \left(1 - 2(1 + R_S)\widetilde{\Phi}\left(\sqrt{2\log(T - k_{u,T})}\right)\right)^{T-k_{u,T}} \\
&\leq 2(T - k_{u,T})(1 + R_S)\frac{\phi(\sqrt{2\log(T - k_{u,T})})}{\sqrt{2\log(T - k_{u,T})}} \\
&= \frac{2(1 + R_S)}{\sqrt{\pi\log(T - k_{u,T})}} \to 0, \qquad \text{as } S, T \to \infty
\end{aligned}$$

from which we conclude that also $\widehat{K}_u(\boldsymbol{\xi}) \subseteq K_{u,T}$ with probability tending to 1 as $S, T \to \infty$.

9.5. Proof of Theorem 3

Proof. For ease of notation we write $\boldsymbol{\xi}_k := \boldsymbol{\xi}_k^{(2)} \sim N(\boldsymbol{h}_k, 2\boldsymbol{V}_k)$ with $\boldsymbol{h}_k = h_k \mathbf{1}_S$. In the first part of this proof we show that, conditional on $\hat{\boldsymbol{h}} = \hat{\boldsymbol{h}}(\boldsymbol{\xi}^{(1)})$, the estimators $\widehat{R}(\boldsymbol{\xi}, \hat{\boldsymbol{h}})$ are asymptotically normal, i.e.

$$\frac{\widehat{R}(\boldsymbol{\xi}, \hat{\boldsymbol{h}}) - \|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})} \xrightarrow{d} N(0,1), \qquad \text{as } S \to \infty \tag{48}$$

Asymptotic confidence regions can then be constructed (unconditional on $\hat{\boldsymbol{h}}$) based on Gaussian quantiles.

Conditional on $\hat{\boldsymbol{h}}$, we decompose

$$\begin{aligned}
\frac{\widehat{R}(\boldsymbol{\xi}, \hat{\boldsymbol{h}}) - \|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})} &= \frac{1}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})}\left(\sum_{k=1}^{T}\left[\sum_{s=1}^{S}\left[w_{sk}(\xi_{sk} - \hat{h}_k)^2\right] - 2\sigma_{k,T}^2 - (\hat{h}_k - h_k)^2\right]\right) \qquad (49) \\
&= \underbrace{\sum_{k=1}^{T}\frac{1}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})}\left[\sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k)^2 - 2\sigma_{k,T}^2\right]}_{\text{(i)}} + \underbrace{\sum_{k=1}^{T}\frac{2(h_k - \hat{h}_k)}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})}\sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k)}_{\text{(ii)}}
\end{aligned}$$

where we use that $\sum_s w_{sk} = 1$ for each $k \in \{1, \ldots, T\}$.

First, we derive the asymptotic distribution of term (i). Write $\boldsymbol{Z}_k = \frac{1}{\sqrt{2}}\boldsymbol{V}_k^{-1/2}(\boldsymbol{\xi}_k - \boldsymbol{h}_k)$, such that

$\boldsymbol{Z}_k \sim N(\boldsymbol{0}, \mathrm{I}_S)$. Here $\boldsymbol{V}_k^{-1/2}$ is a symmetric matrix square root of $\boldsymbol{V}_k^{-1}$. We can rewrite,

$$\begin{aligned}
\sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k)^2 &= (\boldsymbol{\xi}_k - \boldsymbol{h}_k)'\mathrm{diag}(\boldsymbol{w}_k)(\boldsymbol{\xi}_k - \boldsymbol{h}_k) \\
&= (\boldsymbol{\xi}_k - \boldsymbol{h}_k)'\boldsymbol{V}_k^{-1/2}\boldsymbol{V}_k^{1/2}\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k^{1/2}\boldsymbol{V}_k^{-1/2}(\boldsymbol{\xi}_k - \boldsymbol{h}_k) \\
&= \boldsymbol{Z}_k'\boldsymbol{\Gamma}_k\boldsymbol{Z}_k \\
&= \boldsymbol{Z}_k'\boldsymbol{P}_k'\boldsymbol{\Lambda}_k\boldsymbol{P}_k\boldsymbol{Z}_k
\end{aligned}$$

with $\boldsymbol{P}_k'\boldsymbol{\Lambda}_k\boldsymbol{P}_k$ the eigendecomposition of $\boldsymbol{\Gamma}_k = 2\boldsymbol{V}_k^{1/2}\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k^{1/2}$, such that $\boldsymbol{P}_k\boldsymbol{Z}_k \sim N(\boldsymbol{0}, \boldsymbol{P}_k'\boldsymbol{P}_k) \stackrel{d}{=} N(\boldsymbol{0}, \mathrm{I}_S)$. It follows that,

$$\sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k)^2 = \boldsymbol{Z}_k'\boldsymbol{P}_k\boldsymbol{\Lambda}_k\boldsymbol{P}_k\boldsymbol{Z}_k \stackrel{d}{=} \sum_{s=1}^{S} \lambda_{sk}A_{sk}^2$$

with $\boldsymbol{\lambda}_k = (\lambda_{1k}, \ldots, \lambda_{sk})'$ the eigenvalues of $\boldsymbol{\Gamma}_k$ and $A_{1k}^2, \ldots, A_{Sk}^2 \stackrel{\text{iid}}{\sim} \chi_1^2$. Furthermore,

$$\sum_{s=1}^{S} \lambda_{sk} = \mathrm{tr}(\boldsymbol{\Lambda}_k) = \mathrm{tr}(\boldsymbol{\Gamma}_{\boldsymbol{k}}) = 2\mathrm{tr}(\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k) = 2\sigma_{k,T}^2$$

since $\sum_s w_{sk} = 1$, and

$$\|\boldsymbol{\lambda}_k\| = \sqrt{\mathrm{tr}(\boldsymbol{\Lambda}_k^2)} = \sqrt{\mathrm{tr}\left(\boldsymbol{\Gamma}_k'\boldsymbol{\Gamma}_k\right)} = 2\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F$$

Term (i) in eq.(49) can now be rewritten as,

$$\begin{aligned}
\sum_{k=1}^{T} \frac{1}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})}\left[\sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k)^2 - 2\sigma_{k,T}^2\right] &\stackrel{d}{=} \sum_{k=1}^{T} \frac{\sqrt{8}\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F \cdot \sum_{s=1}^{S} \frac{\lambda_{sk}}{\|\boldsymbol{\lambda}_k\|}(A_{sk}^2 - 1)/\sqrt{2}}{\sqrt{\sum_{k=1}^{T} 8\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F^2 + 8(h_k - \hat{h}_k)^2\boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k}} \\
&:= C^{(1)}\sum_{k=1}^{T}\left[B_k^{(1)} \cdot \sum_{s=1}^{S} \frac{\lambda_{sk}}{\|\boldsymbol{\lambda}_k\|}(A_{sk}^2 - 1)/\sqrt{2}\right] \qquad (50)
\end{aligned}$$

with,

$$B_k^{(1)} := \frac{\sqrt{8}\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F}{\sqrt{\sum_{k=1}^{T} 8\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F^2}}, \qquad C^{(1)} := \frac{\sqrt{\sum_{k=1}^{T} 8\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F^2}}{\sqrt{\sum_{k=1}^{T} 8\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F^2 + 8(h_k - \hat{h}_k)^2\boldsymbol{w}_k'\boldsymbol{V}_k\boldsymbol{w}_k}}$$

such that $\|\boldsymbol{B}^{(1)}\| = 1$, and $C^{(1)} \in [0, 1]$. We show that $\sum_{s=1}^{S} \frac{\lambda_{sk}}{\|\boldsymbol{\lambda}_k\|}(A_{sk}^2 - 1)/\sqrt{2} \xrightarrow{d} N(0, 1)$ independently for all $k = 1, \ldots, T$ by the Lindeberg-Feller central limit theorem. First note that,

$$\mathbb{E}\left[\sum_{s=1}^{S} \frac{\lambda_{sk}}{\|\boldsymbol{\lambda}_k\|}(A_{sk}^2 - 1)/\sqrt{2}\right] = 0, \qquad \mathrm{Var}\left(\sum_{s=1}^{S} \frac{\lambda_{sk}}{\|\boldsymbol{\lambda}_k\|}(A_{sk}^2 - 1)/\sqrt{2}\right) = \frac{\sum_{s=1}^{S} \lambda_{sk}^2}{\|\boldsymbol{\lambda}_k\|^2} = 1$$

Writing $X_{sk} := (A_{sk}^2 - 1)/\sqrt{2}$, the Lindeberg conditions are satisfied if:

$$\lim_{S\to\infty} \sum_{s=1}^{S} \frac{\lambda_{sk}^2}{\|\boldsymbol{\lambda}_k\|^2}\mathbb{E}\left[X_{sk}^2\mathbf{1}\left\{\frac{|\lambda_{sk}|}{\|\boldsymbol{\lambda}_k\|}|X_{sk}| > \epsilon\right\}\right] = 0 \qquad \text{for all } \epsilon > 0$$

Since $\sum_{s=1}^{S} \frac{\lambda_{sk}^2}{\|\boldsymbol{\lambda}_k\|^2} = 1$, it suffices to show that $\sup_s \mathbb{E}\left[X_{sk}^2 \mathbf{1}\left\{\frac{|\lambda_{sk}|}{\|\boldsymbol{\lambda}_k\|}|X_{sk}| > \epsilon\right\}\right] \to 0$ as $S \to \infty$, which holds if $\sup_s |\lambda_{sk}|/\|\boldsymbol{\lambda}_k\| \to 0$ as $S \to \infty$. This is seen by combining Cauchy-Schwarz's inequality and the fact that $\mathbb{E}[X_{sk}^4] < \infty$ (see also van der Vaart (2000) Ex. 2.28). Note that by the triangle inequality and the Gershgorin circle theorem,

$$\sup_s |\lambda_{sk}| \leq \sup_s \{|\lambda_{sk} - \boldsymbol{\Gamma}_{k[s,s]}| + |\boldsymbol{\Gamma}_{k[s,s]}|\} \leq \sup_s \sum_{i=1}^{S} |\boldsymbol{\Gamma}_{k[s,i]}| = \|\boldsymbol{\Gamma}_k\|_1$$

Therefore, since by assumption $\|\boldsymbol{\Gamma}_k\|_1/\|\boldsymbol{\Gamma}_k\|_F \to 0$,

$$\frac{\sup_s |\lambda_{sk}|}{\|\boldsymbol{\lambda}_k\|} \leq \frac{\|\boldsymbol{\Gamma}_k\|_1}{\|\boldsymbol{\Gamma}_k\|_F} \to 0 \qquad \text{as } S \to \infty$$

By an application of the Lindeberg-Feller central limit theorem, we conclude that $\sum_{s=1}^{S} \frac{\lambda_{sk}}{\|\boldsymbol{\lambda}_k\|}(A_{sk}^2 - 1)/\sqrt{2} \xrightarrow{d} N(0,1)$ for all $k = 1, \ldots, T$.

Remark. Under the stronger assumption $\kappa(\boldsymbol{\Gamma}_k) < \infty$, where $\kappa(\cdot)$ denotes the condition number (the maximum eigenvalue divided by the minimum eigenvalue), all the eigenvalues λ_{sk} for $s = 1, \ldots, S$ are of the same order, and it follows that,

$$\frac{\sup_s |\lambda_{sk}|}{\|\boldsymbol{\lambda}_k\|} \lesssim \frac{1}{\sqrt{S}} \to 0 \quad \text{as } S \to \infty$$

which is also sufficient for the Lindeberg conditions to hold.

Next, we derive the distribution of term (ii) in eq.(49). We note that for each k, $\sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k)$ is a mean-zero Gaussian random variable, with variance

$$\mathrm{Var}\left(\sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k)\right) = \boldsymbol{w}_k' \mathrm{Cov}(\boldsymbol{\xi}_k - \boldsymbol{h}_k)\boldsymbol{w}_k = 2\boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k$$

Therefore,

$$\begin{aligned}
\sum_{k=1}^{T} \frac{2(h_k - \hat{h}_k)}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})} \sum_{s=1}^{S} w_{sk}(\xi_{sk} - h_k) \quad &\stackrel{d}{=} \quad \sum_{k=1}^{T} \frac{\sqrt{8}(h_k - \hat{h}_k)\sqrt{\boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k} \cdot Z_k}{\sqrt{\sum_{k=1}^{T} 8\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F^2 + 8(h_k - \hat{h}_k)^2 \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k}} \\
&:= \quad C^{(2)} \sum_{k=1}^{T} B_k^{(2)} Z_k \qquad (51)
\end{aligned}$$

where $Z_1, \ldots, Z_T \stackrel{\text{iid}}{\sim} N(0,1)$ and,

$$B_k^{(2)} := \frac{\sqrt{8}(h_k - \hat{h}_k)\sqrt{\boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k}}{\sqrt{\sum_{k=1}^{T} 8(h_k - \hat{h}_k)^2 \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k}}, \qquad C^{(2)} := \frac{\sqrt{\sum_{k=1}^{T} 8(h_k - \hat{h}_k)^2 \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k}}{\sqrt{\sum_{k=1}^{T} 8\|\mathrm{diag}(\boldsymbol{w}_k)\boldsymbol{V}_k\|_F^2 + 8(h_k - \hat{h}_k)^2 \boldsymbol{w}_k' \boldsymbol{V}_k \boldsymbol{w}_k}}$$

such that $\|\boldsymbol{B}^{(2)}\| = 1$, and $C^{(2)} \in [0,1]$.

Combining eq.(49), eq.(50) and eq.(51) we conclude that, conditional on $\hat{\boldsymbol{h}}$,

$$\begin{aligned}
\frac{\widehat{R}(\boldsymbol{\xi}, \hat{\boldsymbol{h}}) - \|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})} \quad &\stackrel{d}{=} \quad C^{(1)} \sum_{k=1}^{T} \left[B_k^{(1)} \cdot \sum_{s=1}^{S} \frac{\lambda_{sk}}{\|\boldsymbol{\lambda}_k\|}(A_{sk}^2 - 1)/\sqrt{2}\right] + C^{(2)} \sum_{k=1}^{T} B_k^{(2)} Z_k \\
&\xrightarrow{d} \quad C^{(1)} Z_1 + C^{(2)} Z_2 \sim N(0,1) \quad \text{as } S \to \infty \qquad (52)
\end{aligned}$$

where we use that $\|\boldsymbol{B}^{(1)}\| = \|\boldsymbol{B}^{(2)}\| = 1$ and also $\|\boldsymbol{C}\| = 1$ with $\boldsymbol{C} = (C^{(1)}, C^{(2)})'$, combined with the fact that a standard normal random vector is invariant under rotation by a vector of norm 1.

By the asymptotic normality result in eq.(52), for a given confidence level $1 - \alpha$

$$\liminf_{S\to\infty} \inf_{\boldsymbol{h}\in\ell_2} P\left(\frac{\widehat{R}(\boldsymbol{\xi}, \hat{\boldsymbol{h}}) - \|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2}{\tau(\boldsymbol{h}, \hat{\boldsymbol{h}})} \geq -z_\alpha \,\Big|\, \hat{\boldsymbol{h}} \right) \geq 1 - \alpha$$

with z_α a standard normal quantile. By Fatou's lemma, the asymptotic unconditional coverage probability is also at least $1 - \alpha$ (see Robins and van der Vaart (2006) Section 2). Therefore,

$$\liminf_{S\to\infty} \inf_{\boldsymbol{h}\in\ell_2} P(\boldsymbol{h} \in \widehat{C}_\alpha(\boldsymbol{\xi})) \geq 1 - \alpha$$

where,

$$\widehat{C}_\alpha(\boldsymbol{\xi}) = \left\{ \boldsymbol{h} \in \ell_2 : \|\boldsymbol{h} - \hat{\boldsymbol{h}}\| \leq \sqrt{z_\alpha \tau(\boldsymbol{h}, \hat{\boldsymbol{h}}) + \widehat{R}(\boldsymbol{\xi}, \hat{\boldsymbol{h}})} \right\}$$

which concludes the proof.

Acknowledgments

We thank the UC Irvine Space-Time Modeling Group, and in particular Hernando Ombao for useful discussions regarding this work, and Dr. Emad Eskandar (Massachussetts General Hospital) for the local field potential data that was used to illustrate the methodology. This research is supported by IAP research network P7/06 of the Belgian government (Belgian Science Policy), and by the contract 'Projet d'Actions de Recherche Concertées' (ARC) No. 12/17-045 of the 'Communauté française de Belgique' granted by the Académie universitaire Louvain. The first author gratefully acknowledges funding from the Belgian Fund for Scientific Research (FRIA/FRS-FNRS).

References

Abramovich, F., Y. Benjamini, D. Donoho, and I. Johnstone (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist. 34*(2), 584–653.

Antoniadis, A. and T. Sapatinas (2007). Estimation and inference in functional mixed-effects models. *Computational Statistics & Data Analysis 51*(10), 4793–4813.

Aston, J., J. Chiou, and J. Evans (2010). Linguistic pitch analysis using functional principal component mixed effect models. *J. R. Statist. Soc. C 59*(2), 297–317.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.

Bruce, A. and H. Gao (1996). Understanding waveshrink: Variance and bias estimation. *Biometrika 83*(4), 727–745.

Daubechies, I. et al. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

Diggle, P. and I. Al Wasel (1997). Spectral analysis of replicated biomedical time series. *J. R. Statist. Soc. C 46*(1), 31–71.

Fiecas, M. and H. Ombao (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *J. Am. Statist. Ass.*. (Accepted).

Freyermuth, J., H. Ombao, and R. von Sachs (2010). Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *J. Am. Statist. Ass. 105*(490), 634–646.

Gao, H. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis 18*(3), 231–251.

Genovese, C. and L. Wasserman (2005). Confidence sets for nonparametric wavelet regression. *Ann. Statist.*, 698–729.

Giacofci, M., S. Lambert-Lacroix, G. Marot, and F. Picard (2013). Wavelet-based clustering for mixed-effects functional models in high dimension. *Biometrics 69*(1), 31–40.

Gorrostieta, C., H. Ombao, R. Prado, S. Patel, and E. Eskandar (2012). Exploring dependence between brain signals in a monkey during learning. *Journal of Time Series Analysis 33*(5), 771–778.

Guo, W. (2002). Functional mixed effects models. *Biometrics 58*(1), 121–128.

Hernandez-Flores, C., J. Artiles-Romero, and P. Saavedra-Santana (1999). Estimation of the population spectrum with replicated time series. *Computational Statistics & Data Analysis 30*(3), 271–280.

Higham, N. (2002). Computing the nearest correlation matrix; a problem from finance. *IMA Journal of Numerical Analysis 22*(3), 329–343.

Iannaccone, R. and S. Coles (2001). Semiparametric models and inference for biomedical time series with extra-variation. *Biostatistics 2*(3), 261–276.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer.

Jiang, J., Y. Luan, Y. Wang, et al. (2007). Iterative estimating equations: Linear convergence and asymptotic properties. *Ann. Statist. 35*(5), 2233–2260.

Johnstone, I. (2015). Gaussian estimation: Sequence and multiresolution models. (Unpublished manuscript).

Krafty, R., M. Hall, and W. Guo (2011). Functional mixed effects spectral analysis. *Biometrika 98*(3), 583–598.

Martinez, J., K. Bohn, R. Carroll, and J. Morris (2013). A study of mexican free-tailed bat chirp syllables: Bayesian functional mixed models for nonstationary acoustic time series. *J. Am. Statist. Ass. 108*(502), 514–526.

Morris, J. (2014). Functional regression. *arXiv preprint arXiv:1406.4068*.

Morris, J., P. Brown, R. Herrick, K. Baggerly, and K. Coombes (2008). Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. *Biometrics 64*(2), 479–489.

Morris, J. and R. Carroll (2006). Wavelet-based functional mixed models. *J. R. Statist. Soc. B 68*(2), 179–199.

Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Processing 42*(11), 3126–3136.

Nason, G. (2010). *Wavelet Methods in Statistics with R*. New York: Springer.

Neumann, M. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-gaussian time series. *Journal of Time Series Analysis 17*(6), 601–633.

Neumann, M. and R. von Sachs (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist. 25*(1), 38–76.

Ombao, H., R. von Sachs, and W. Guo (2005). SLEX analysis of multivariate nonstationary time series. *J. Am. Statist. Ass. 100*(470), 519–531.

Pav, S. (2015). Moments of the log non-central chi-square distribution. *arXiv preprint arXiv:1503.06266*.

Qin, L. and W. Guo (2006). Functional mixed-effects model for periodic data. *Biostatistics 7*(2), 225–234.

Qin, L., W. Guo, and B. Litt (2009). A time-frequency functional model for locally stationary time series data. *Journal of Computational and Graphical Statistics 18*(3), 675–693.

Robins, J. and A. van der Vaart (2006). Adaptive nonparametric confidence sets. *Ann. Statist. 34*(1), 229–253.

Rudzkis, R., L. Saulis, and V. Statulevičius (1978). A general lemma on probabilities of large deviations. *Lithuanian Mathematical Journal 18*(2), 226–238.

Searle, S., G. Casella, and C. E. McCulloch (1992). *Variance Components*. New Jersey: John Wiley & Sons.

Taniguchi, M. (1979). On estimation of parameters of gaussian stationary processes. *Journal of Applied Probability 16*(3), 575–591.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge U.K.: Cambridge university press.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley & Sons.

Wahba, G. (1980). Automatic smoothing of the log periodogram. *J. Am. Statist. Ass. 75*(369), 122–132.