Biometrics 68, 825–836
September 2012
DOI: 10.1111/j.1541-0420.2012.01744.x

Evolutionary Factor Analysis of Replicated Time Series

Giovanni Motta[1,*] and Hernando Ombao[2,]**

[1]Department of Quantitative Economics, Maastricht University P.O.Box 616, 6200 MD Maastricht, The Netherlands

[2]Department of Statistics, University of California Irvine Irvine, California 92697, U.S.A.

[*]*email:* dr.giovannimotta@gmail.com

[**]*email:* hombao@uci.edu

Summary. In this article, we develop a novel method that explains the dynamic structure of multi-channel electroencephalograms (EEGs) recorded from several trials in a motor–visual task experiment. Preliminary analyses of our data suggest two statistical challenges. First, the variance at each channel and cross-covariance between each pair of channels evolve over time. Moreover, the cross-covariance profiles display a common structure across all pairs, and these features consistently appear across all trials. In the light of these features, we develop a novel evolutionary factor model (EFM) for multi-channel EEG data that systematically integrates information across replicated trials and allows for smoothly time-varying factor loadings. The individual EEGs series share common features across trials, thus, suggesting the need to pool information across trials, which motivates the use of the EFM for replicated time series. We explain the common co-movements of EEG signals through the existence of a small number of common factors. These latent factors are primarily responsible for processing the visual–motor task which, through the loadings, drive the behavior of the signals observed at different channels. The estimation of the time-varying loadings is based on the spectral decomposition of the estimated time-varying covariance matrix.

Key words: Electroencephalography; Factor models; Local stationarity; Principal components.

1. Introduction

In this article, we develop a novel statistical method based on time-varying principal components for analyzing replicated nonstationary multichannel electroencephalograms (EEGs) which were collected for the purpose of investigating the brain dynamics during processing of a motor–visual task. The dataset consists of 62 channel electroencephalograms (EEGs) recorded from one healthy subject in a hand-guided visual–motor experiment. These EEGs were recorded over several replicated identical trials. Preliminary analyses of the data demonstrate that the multiple channels are cross-correlated and that the EEGs exhibit nonstationary behavior. In Figure 1, we plot two series observed at two different locations; the time-varying variance of one series and the time-varying covariance between the two series. The covariance matrix is time-varying, indicating that the series have time-varying moments and hence, are nonstationary.

Moreover, the entries of the covariance matrix vary over time in a very similar way. This suggests the use of dimension reduction techniques that exploit the co-movements. For this reason, we projected the 62 series on the main three evolutionary (i.e., time-varying) principal components, obtaining the so-called evolutionary common components. The similarity between the original series and the estimated common components (that are extracted from the 62×62 time-varying covariance matrix using only three factors) confirm the validity of our dimension reduction method for nonstationary time series.

The multichannel EEGs display nonstationary behavior because their variances and cross-covariances appear to change slowly over time (within a trial). To account for this feature, we shall adopt the concept of *local stationarity* introduced by Dahlhaus (1997) which provides a rigorous framework for the treatment of nonstationarity. The definition of local stationarity, based on rescaled time $u = \frac{t}{T}$, where T is the length of the time series, gives a meaningful asymptotic theory of statistical inference. In addition to nonstationarity, the EEGs appear to be highly cross-correlated (strong correlation between channels). In fact, when considered jointly as a multivariate time series, the EEGs exhibit a common behavior driven by some latent factors.

Our proposed methodology shares a common goal with that developed in Prado, West, and Krystal (2001), namely the application of nonstationary latent factor models for multivariate EEG series. In particular, as in Prado et al. (2001), we adopt time-varying parameters to model the nonstationarity. However, there are three differences in the approaches. First, the time-varying coefficients in Prado et al. (2001) are stochastic, whereas in our approach they are deterministic. We believe that this is an important feature of our approach because it is reasonable to assume that the loadings that weight the latent factors change over time (during the experiment) in a deterministic way. Second, the loadings in the model by Prado et al. (2001) have an auto-regressive prior, whereas our loadings are completely nonparametric and thus our model is more flexible. However, if the parametric model is

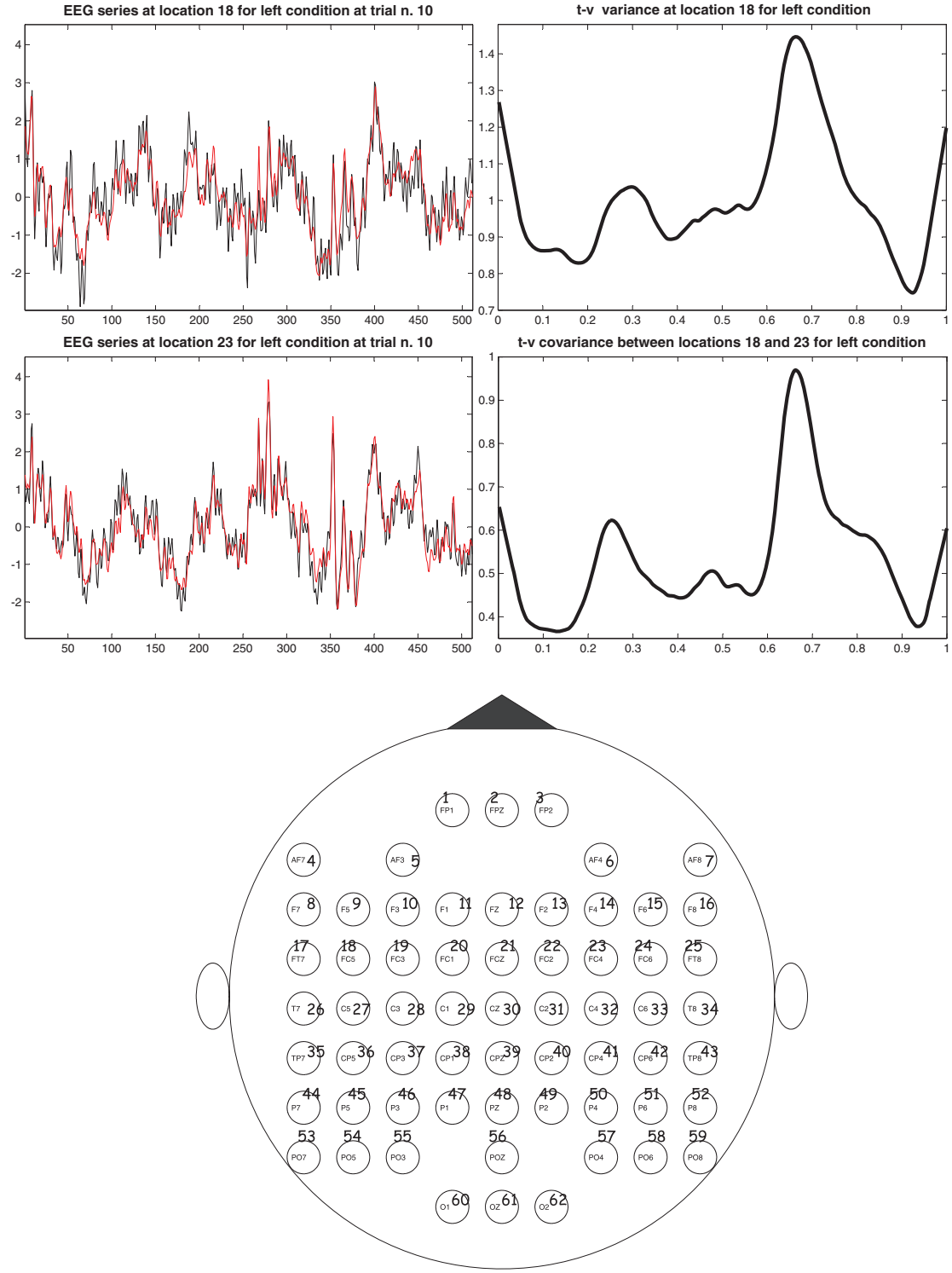

Figure 1. The subject was commanded to move the joystick many times, with a total of $R = 118$ trials. For each trial, a time series is recorded for an interval of one second. The start of the time series is 500 milliseconds before the command. The end of the time series is 500 milliseconds after the command. Here, one trial will have a total of $T = 512$ time points. On the left side, we plot EEG data $Y_i(t, r)$ versus the estimated common components $\widehat{X}_i(t, r)$, for $i = 18, 23$, $t = 1, \ldots, T$ and $r = 10$. On the right-top side, we plot the time-varying variance of $Y_{18}(t, r)$, whereas the right-bottom side is the estimated time-varying covariance between $Y_{18}(t, r)$ and $Y_{23}(t, r)$ in rescaled time $u \in (0, 1)$. We only report the results for left condition. The shapes of the time-varying covariances corresponding to right condition are very similar to those reported in this figure. This figure appears in color in the electronic version of this article.

adequate in representing the latent factors, then the approach in Prado et al. (2001) gives further significant interpretation of the dynamics of the factors. Third, the latent factors in our model are allowed to vary across several trials. Our approach can, thus, be easily extended to the situation where one needs to consider trial-specific effects by adding some random term (perhaps as a multiplier) to the time-varying loadings. One would imagine that this is also possible in the approach of Prado et al. (2001) by adding another layer or hierarchy in the Bayesian analysis.

A traditional tool for dimension reduction is the Principal Components Analysis (PCA). For nonstationary time series, Ombao and Ho (2006) employed the spectral representation of a locally stationary multichannel signal and used a localized version of PCA in the frequency domain to extract the relevant nonredundant information from massive high dimensional EEG signals. In a related approach, Ombao, von Sachs, and Guo (2005) used smooth localized complex exponentials (SLEX) waveforms, and developed a method to extract nonredundant spectral information by applying a time-varying eigenvalue-eigenvector decomposition to the time-varying SLEX spectral density matrix.

In this article, we propose a new factor model for high dimensional nonstationary time series under the setting where there are several multichannel time series obtained from replicated trials. The nonstationarity in the data is fully explained by the factor loadings which are smooth functions of time and estimated by the eigenvectors of a nonparametric estimator of the time-varying covariance matrix. Our approach, inspired by Motta, Hafner, and von Sachs (2011), has the following optimal properties. From the point of view of *modeling*, our model does not assume any particular parametric form for the factors and hence does not suffer from model misspecification. From the *estimation* viewpoint, our methodology is flexible as the underlying model is nonparametric and thus fully adaptive. We also benefit from an automatic (data-driven) bandwidth selection procedure (see Section 3.1 below). From a *practical* view point, our method systematically integrates variation across trials and could be extended to allow for testing of differences across experimental conditions and patient groups. From the *theoretical* point of view, our methodology is consistent as it gives mean-squared consistent results. It combines information from several trials in a computationally efficient manner, in the sense that we benefit from a faster rate of convergence of the estimators. The increase in the speed of the rate is given by the square root of the number of trials.

The main contributions of this article are as follows. We develop a parsimonious representation of the multivariate time series with evolutionary second-order properties (i.e., time-varying covariance). Moreover, we rigorously develop a nonparametric method for the estimation of and inference on the stochastic properties of the time-varying weights that load the underlying factors.

Throughout the article we use bold-unslanted letters for matrices, bold-slanted letters for vectors and unbold (normal) letters for scalars. Random operators are denoted with superscript when they depend on diverging sizes (namely, T, P, and R). When needed, vectors and matrices are indexed by subscripts denoting their size. We denote by $\mathrm{tr}(\cdot)$ the trace operator, by $\mathrm{rk}(\mathbf{A})$ the rank of a matrix $\mathbf{A}$, by $\mathbf{I}_n$ the identity matrix of dimension n, by $\mathbf{O}_{m,n}$ the null matrix of dimension $m \times n$ and by $\|\cdot\|$ the Frobenius (Euclidean) norm, that is, $\|\mathbf{A}\| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$. We call trials the (independent) R replications of the same experiment, and channels (or locations, equivalently) the spots where the P electrodes are located. They are indexed, respectively, as $r = 1, \ldots, R$, and $i = 1, \ldots, P$. The time series index is $t = 1, \ldots, T$, where T is the sample size.

2. Background on Evolutionary Factor Analysis

Linear factor models are both intuitively appealing and highly relevant for analyzing EEG data because they are a model-based dimension-reduction approach for characterizing the complex temporal dynamics of a high dimensional multivariate time series using only a few common factors. Thus, this approach has a strong potential for providing a statistically-principled basis for making inference on massive high dimensional brain data.

In traditional factor models, the process is treated as being temporally stationary which appears to be restrictive because, over long time periods, factor loadings are highly unlikely to remain constant. A promising approach is to model unconditional variances and covariances via nonparametric estimation, which imposes very little structure on the unconditional covariance matrix while being very easy to estimate.

Motta et al. (2011) generalized the asymptotic theory of Factor Analysis given by Bai (2003) to the locally stationary case from both the identification and the estimation points of view. Given a $P \times 1$ vector of observations $\boldsymbol{Y}$, they consider a factor model with $P \times q$ time-varying factor loadings $\boldsymbol{\Lambda}$

$$\begin{aligned} \boldsymbol{Y}_P^T(t) &= \boldsymbol{X}_P^T(t) + \boldsymbol{Z}_P(t) \\ &= \boldsymbol{\Lambda}_P\left(\frac{t}{T}\right)\boldsymbol{F}(t) + \boldsymbol{Z}_P(t), \qquad t = 1, \ldots, T, \end{aligned} \tag{1}$$

with $P \times 1$ common components $\boldsymbol{X}$, $q \times 1$ factors $\boldsymbol{F}$ and $P \times 1$ idiosyncratic components $\boldsymbol{Z}$. The common component $\boldsymbol{X}$ describes the co-movements of all the series, the idiosyncratic component $\boldsymbol{Z}$ is specific to each particular series. The basic idea is to consider the loadings as smooth functions of rescaled time, rendering the process nonstationary while the factors remain stationary. However, the assumption that loadings are smooth permits considering the process as locally stationary and enables us to estimate the model using nonparametric methods. Similarly to Ombao and Ho (2006), the technique employed to model the nonstationarity is the rescaling of time to the unit interval, but the difference is that model (1) has a factor structure. The time-varying covariance matrix $\boldsymbol{\Sigma}_P(\frac{t}{T})$ of $\boldsymbol{Y}_P^T(t)$ in (1) is estimated consistently by smoothing the cross-products of the observations. Then to estimate the loadings, the factors and the common components of model (1), Motta et al. (2011) applied a time-varying (i.e., a time-localized version of) PCA to the estimator $\widehat{\boldsymbol{\Sigma}}_P(\frac{t}{T})$ of $\boldsymbol{\Sigma}_P(\frac{t}{T})$ in time domain. One limitation of the above approaches is that they do not directly handle multivariate nonstationary time series data that are obtained repeatedly across several independent trials. In this article, we develop a novel model that builds on the ideas and framework of cutting-edge work on evolutionary factor analysis.

3. Extracting Evolutionary Principal Components from Replicated Time Series

In this section, we set up a model for high-dimensional EEG datasets observed over repeated trials. We assume that the observations $\boldsymbol{Y}(t,r)$ at time t and rth trial can be represented by a factor model of the form,

$$\begin{aligned}\boldsymbol{Y}_P(t,r) &= \boldsymbol{X}_P(t,r) + \boldsymbol{Z}_P(t,r),\\ t &= 1,\dots,T, \qquad r = 1,\dots,R,\end{aligned}$$

where T is the sample size and R is the number of trials. The processes $\boldsymbol{Y}_P(t,r)$, $\boldsymbol{X}_P(t,r)$, and $\boldsymbol{Z}_P(t,r)$ are $P\times 1$ vectors of observed time series, common components and idiosyncratic errors, respectively. We allow the common component to have a reduced rank structure with time-varying loadings. More precisely, following Motta et al. (2011), we assume that there exists a function $\boldsymbol{\Lambda}_P(u)$ defined on the unit interval $u\in(0,1)$ such that,

$$\boldsymbol{X}_P^T(t,r) = \boldsymbol{\Lambda}_P\Big(\frac{t}{T}\Big)\boldsymbol{F}(t,r), \tag{2}$$

where $\boldsymbol{\Lambda}$ is an $P\times q$ matrix of smoothly varying loadings, and where $\boldsymbol{F}$ is a zero-mean $q\times 1$ vector of orthogonal common factors with variance $\boldsymbol{\Sigma}^F$. Equation (2) implies that the common component $\boldsymbol{X}_P^T(t,r)$ is a sequence of stochastic processes whose structure depends on T, the sample size. As a consequence, $\boldsymbol{Y}_P^T(t,r)$ is also nonstationary, and the model is written as a sequence of models,

$$\begin{aligned}\boldsymbol{Y}_P^T(t,r) &= \boldsymbol{X}_P^T(t,r) + \boldsymbol{Z}_P(t,r),\\ t &= 1,\dots,T, \qquad r = 1,\dots,R.\end{aligned} \tag{3}$$

Hence, defining $\boldsymbol{\Sigma}_P^Z = \mathbb{V}\mathrm{ar}\{\boldsymbol{Z}_P(t,r)\}$ we can define, for all $u\in(0,1)$, a matrix-valued smooth function $\boldsymbol{\Sigma}_P(u)$ given by

$$\boldsymbol{\Sigma}_P(u) = \boldsymbol{\Sigma}_P^X(u) + \boldsymbol{\Sigma}_P^Z, \tag{4}$$

where
$$\boldsymbol{\Sigma}_P^X(u) := \boldsymbol{\Lambda}_P(u)\boldsymbol{\Sigma}^F\boldsymbol{\Lambda}_P'(u), \tag{5}$$

such that the covariance matrix of $\boldsymbol{Y}_P^T(t,r)$ can be defined as

$$\begin{aligned}\boldsymbol{\Sigma}_P\Big(\frac{t}{T}\Big) &:= \mathbb{V}\mathrm{ar}\left\{\boldsymbol{Y}_P^T(t)\right\}\\ &= \boldsymbol{\Sigma}_P^X\Big(\frac{t}{T}\Big) + \boldsymbol{\Sigma}_P^Z, \qquad t = 1,\dots,T,\end{aligned}$$

where $\boldsymbol{\Sigma}_P^X(\frac{t}{T}) := \boldsymbol{\Lambda}_P(\frac{t}{T})\boldsymbol{\Sigma}^F\boldsymbol{\Lambda}_P'(\frac{t}{T})$ is the time-varying covariance matrix of the common components $\boldsymbol{X}_P^T(t,r)$, $t=1,\dots,T$, $r=1,\dots,R$. The loadings are estimated by the eigenvectors of an estimator of $\boldsymbol{\Sigma}_P(u)$. To estimate the loadings consistently, we need $\boldsymbol{\Sigma}_P(u)$ to be continuously differentiable. This is guaranteed by assuming that the entries of the $P\times q$ matrix of loadings $\boldsymbol{\Lambda}_P(u)$ are continuously differentiable in u. Moreover, to identify the number q of factors, we assume that for all $u\in[0,1]$, the $P\times q$ matrix of loadings $\boldsymbol{\Lambda}_P(u)$ has rank q.

The covariance matrix $\boldsymbol{\Sigma}_P^Z$ of the idiosyncratic components is a sequence of covariance matrices with uniformly bounded eigenvalues, that is, the $\sup_P \mathrm{v}_{1P}^Z$ is uniformly bounded over $P\in\mathbb{N}$, where v_{1P}^Z denotes the largest eigenvalue of $\boldsymbol{\Sigma}_P^Z$.

Moreover, the factor process $\boldsymbol{F}(t,r)$ and the idiosyncratic errors $\boldsymbol{Z}_P(t,r)$ are orthogonal at all leads, lags, and trials, that is, $\mathbb{E}\big\{\boldsymbol{Z}_P(t,r)\boldsymbol{F}(t-k,s)'\big\} = \boldsymbol{O}_{P,q}$ for all $r,s=1,\dots,R$, all $P\in\mathbb{N}$ and all $t,k\in\mathbb{Z}$. A more detailed list of assumptions on our model is given in the Web Appendix. Under these assumptions, we say that the observations in (3) follow an evolutionary factor model (EFM).

Remark 1 (Evolutionary common components). The common components process $\boldsymbol{X}_P^T$ is a triangular array of P-dimensional random vectors whose structure does not only depend on t, but also on T. In the locally stationary setting, letting T tend to infinity does not mean extending the data to the future anymore. In the rescaled time framework, letting T tend to infinity means that we have in the sample $\boldsymbol{X}_P^T(1,r),\dots,\boldsymbol{X}_P^T(T,r)$ more and more 'observations' for each value of $\boldsymbol{\Lambda}(\frac{t}{T})$. For local estimation at time t this implies that increasing T is equivalent to increasing the sampling rate in a local neighborhood of $\frac{t}{T}$ on the domain $[0,1]$.

Remark 2 (Stochastic versus deterministic). In our model, the stochastic processes $\boldsymbol{Y}_P(t,r)$, $\boldsymbol{F}(t,r)$, $\boldsymbol{X}_P(t,r)$ and $\boldsymbol{Z}_P(t,r)$ depend on trial r, $r=1,\dots,R$, whereas the deterministic functions $\boldsymbol{\Lambda}_P(u)$, $\boldsymbol{\Sigma}^F$, $\boldsymbol{\Sigma}_P(u)$, $\boldsymbol{\Sigma}_P^X(u)$ and $\boldsymbol{\Sigma}_P^Z$ do not.

Remark 3 (Stationary Factors). The assumption that the process $\boldsymbol{F}(t,r)$ is stationary process is not a serious constraint. For example, suppose that $\boldsymbol{F}(t,r)$ is a nonstationary factor process with time-varying representation $\boldsymbol{F}(t,r) = \boldsymbol{\Xi}\big(\frac{t}{T}\big)\boldsymbol{\eta}(t,r)$, where $\boldsymbol{\eta}(t,r)$ is a stationary orthonormal white noise process and $\boldsymbol{\Xi}(u)$ is a smooth function of time. Then the common component $\boldsymbol{X}(t,r)$ can be rewritten as $\boldsymbol{X}(t,r) = \boldsymbol{\Lambda}(\frac{t}{T})\,\boldsymbol{\Xi}(\frac{t}{T})\,\boldsymbol{\eta}(t,r) = \widetilde{\boldsymbol{\Lambda}}(\frac{t}{T})\,\boldsymbol{\eta}(t)$. Thus the process can be represented as in (2) with the factor process $\widetilde{\boldsymbol{F}}(t,r) = \boldsymbol{\eta}(t,r)$ satisfying the above assumptions.

Remark 4 (Loadings and covariances). It is well known that EEG signals exhibit a high variability over trials; and that averaging over a large number of trials allows one to remove the noise and thus to recover the underlying signal. EEG recordings obtained following stimulus presentation have a low amplitude in comparison with the background activity. Consequently, they are barely visible in single-trials. The usual way of improving the ratio between the signal power and the background EEG power, that is, the signal-to-noise ratio (SNR), is by averaging the response of several trials (see Chiappa 1997).

In this respect, the fact that the loadings do not depend on r is a crucial assumption. This implies that there is a dynamic behavior within a trial but that this dynamic behavior does not change from one trial to the next. This assumption is crucial for estimation (see later), as $\boldsymbol{\Lambda}$ is estimated from the average over trials of the estimated covariance matrices.

3.1 *Estimation Steps*

In this section, we describe the estimation steps. Then, in Section 3.2, we derive the asymptotic properties of our estimators.

(1) At each time point, we estimate the covariance matrix by averaging over trials the cross-products of the observations. That is, the $P\times P$ pre-estimator $\widetilde{\boldsymbol{\Sigma}}_P^T(t)$ of the covariance matrix $\boldsymbol{\Sigma}_P$ at time t is the average (over

trails) of the $P \times P$ matrices $\{\boldsymbol{Y}_P^T(t,r)\boldsymbol{Y}_P^T(t,r)'\}$

$$\widetilde{\boldsymbol{\Sigma}}_P^T(t) := \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{Y}_P^T(t,r)\boldsymbol{Y}_P^T(t,r)', \qquad t = 1,\ldots,T.$$

Then, we define the estimator of $\boldsymbol{\Sigma}_P(u)$ in (4) at rescaled time u as the smoothed version (over rescaled time) of the estimates $\widetilde{\boldsymbol{\Sigma}}_P^T(s)$ for those values $\frac{s}{T}$ around u:

$$\widehat{\boldsymbol{\Sigma}}_P^{TR}(u;h) := \frac{1}{T}\sum_{s=1}^{T} \widetilde{\boldsymbol{\Sigma}}_P^T(s) K_h\left(u - \frac{s}{T}\right), \qquad u \in (0,1), \tag{6}$$

where $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$ is the rescaled version of a second order kernel, and $h \equiv h_{TR}$ is the sequence of smoothing bandwidths that tends to zero as $T, R \to \infty$.

(2) Before estimating the $P \times q$ matrix of loadings, we need to determine the number $\widehat{q}$ of factors, that is, the number of columns of the estimated matrix of loadings. The estimation of q is based on the $P \times P$ diagonal matrix $\widehat{\mathbf{V}}_P^{TR}(u)$ containing the P eigenvalues of $\widehat{\boldsymbol{\Sigma}}_P^{TR}(u)$. This will be detailed in section 4, where we apply our methodology to the data.

(3) Extract the $P \times q$ matrix of eigenvectors $\widehat{\boldsymbol{\Lambda}}_P^{TR}(u)$ corresponding to the largest eigenvalues of the $P \times P$ matrix $\frac{1}{P}\widehat{\boldsymbol{\Sigma}}_P^{TR}(u)$ collected in the $q \times q$ diagonal matrix $\widehat{\mathbf{V}}_q^{PTR}(u)$:

$$\frac{1}{P}\widehat{\boldsymbol{\Sigma}}_P^{TR}(u)\widehat{\boldsymbol{\Lambda}}_P^{TR}(u) = \widehat{\boldsymbol{\Lambda}}_P^{TR}(u)\widehat{\mathbf{V}}_q^{PTR}(u), \qquad u \in (0,1). \tag{7}$$

The matrix $\widehat{\mathbf{V}}_q^{PTR}(u)$ is the upper-left $q \times q$ corner of the $P \times P$ matrix $\widehat{\mathbf{V}}_P^{TR}(u)$. Although $\widehat{\mathbf{V}}_q^{PTR}(u)$ is of finite size $(q \times q)$, its asymptotic behavior depends on P.

(4) Define the $q \times 1$ principal components at time t for the r-th trial, as the projection of the data $\boldsymbol{Y}_P^T(t,r)$ at time t and trial r on the orthonormal eigenvectors $\widehat{\boldsymbol{\Lambda}}_P^{TR}(\frac{t}{T})$,

$$\widehat{\boldsymbol{F}}^{PTR}(t,r) = \frac{1}{P}\widehat{\boldsymbol{\Lambda}}_P^{TR}\left(\frac{t}{T}\right)' \boldsymbol{Y}_P^T(t,r). \tag{8}$$

(5) Define the estimated common components as the the product estimator

$$\widehat{\boldsymbol{X}}_P^{TR}(t,r) = \widehat{\boldsymbol{\Lambda}}_P^{TR}\left(\frac{t}{T}\right)\widehat{\boldsymbol{F}}^{PTR}(t,r).$$

Remark 5 (Bandwidth and Kernel selection). The estimator $\widehat{\boldsymbol{\Sigma}}_P(u;h)$ in (6) depends on the bandwidth sequence h. In our application to EEG data, the bandwidth $\widehat{h}$ was selected adaptively from the data using the local *plug-in algorithm* by Gasser, Kneip, and Köhler (1991) and Brockmann, Gasser, and Herrmann (1993). The technical computation of the bandwidth-selection procedure is described in Herrmann (1997). The basic idea of plug-in estimation is to obtain a large-sample approximation to the mean integrated squared error (MISE) of the estimator of the entries $\widehat{\sigma}_{i,j}(u;h)$ of $\widehat{\boldsymbol{\Sigma}}_P(u;h)$; then to minimize the resulting analytical expression with respect to h to obtain the asymptotic optimal bandwidth $\widehat{h}$; and finally to replace the unknown terms in $\widehat{h}$ by their estimators. The smoother we used is the Epanechnikov kernel which is of the form $K(x) = \frac{3}{4}(1-x^2)\ \mathbb{I}(|x| \le 1)$. Among all nonnegative kernels with compact support, this kernel minimizes the asymptotic MISE of $\widehat{\boldsymbol{\Sigma}}_P(u;\widehat{h})$.

3.2 *Asymptotic Estimation Theory*

In this section, we study the asymptotic properties of the estimators defined in Section 3.1. In particular, we focus on the estimators $\widehat{\boldsymbol{\Sigma}}_P^{TR}(u)$, $\widehat{\mathbf{V}}_q^{PTR}(u)$ and $\widehat{\boldsymbol{\Lambda}}_P^{TR}\left(\frac{t}{T}\right)$. The properties of these three estimators are highlighted in Sections 3.2.1, 3.2.2, and 3.2.3, respectively, while the formal results and their proofs are given in the Appendix and the Web Appendix, respectively. The properties of the estimators $\widehat{\boldsymbol{F}}^{PTR}(t,r)$ and $\widehat{\boldsymbol{X}}_P^{TR}(t,r)$ are highlighted in Section 3.2.4 and detailed in the Web Appendix.

3.2.1 *Evolutionary covariance.* The first result is about consistency of our nonparametric estimator of the time-varying covariance $\boldsymbol{\Sigma}_P(u)$. In our Proposition 1 (see Appendix), we show that the estimator $\widehat{\boldsymbol{\Sigma}}_P^{TR}(u;h)$ converges to $\boldsymbol{\Sigma}_P(u)$ as T and R tend to infinity. We emphasize that, for the consistency of the estimator $\widehat{\boldsymbol{\Sigma}}_P^{TR}(u;h)$ we only need that $T \to \infty$, whereas the divergence of the number of trials R is only assumed to improve even further the quality of the estimates.

Remark 6 (Increased rate of convergence with multiple trials). The result in Proposition 1 shows that the possibility of observing a collection of time series generated from the same experiment (multiple trials) increases the speed of convergence of $\widehat{\boldsymbol{\Sigma}}$ to $\boldsymbol{\Sigma}$ by $\sqrt{R}$. As a consequence, we do not need to smooth too much within a trial because we have many trials and thus we essentially smooth by averaging over trials. By not smoothing too much, we control bias and preserve high resolution in rescaled time. In the Web Appendix we show that if $\widetilde{h}_T$ is the optimal bandwidth for $\widetilde{\boldsymbol{\Sigma}}$, then the optimal bandwidth for $\widehat{\boldsymbol{\Sigma}}$ is $\widehat{h}_{TR} = R^{-\frac{1}{5}}\widetilde{h}_T$. Hence the multi-trial situation helps us to get estimates which are better localized in time.

The estimation of the covariance matrix is the first step toward estimating factor models, and it is from this estimator that all the other estimators are derived. The consistency of the estimators of the number of factors and the loadings depends on, respectively, the consistency of the eigenvalues and the eigenvectors of $\widehat{\boldsymbol{\Sigma}}_P^{TR}(u)$. The eigenvalues are used to estimate the number of factors, whereas the eigenvectors are estimates of the evolutionary loadings. The orthonormal eigenvectors are such that,

$$P^{-1}\widehat{\boldsymbol{\Lambda}}_P^{TR}(u)'\widehat{\boldsymbol{\Lambda}}_P^{TR}(u) = \mathbf{I}_q \quad \text{for all } u \in (0,1); \tag{9}$$

then by (7) we can rewrite the $q \times q$ matrix of estimated eigenvalues as,

$$\frac{1}{P}\widehat{\mathbf{V}}_q^{PTR}(u) = P^{-1}\widehat{\boldsymbol{\Lambda}}_P^{TR}(u)'\left\{\frac{1}{P}\widehat{\boldsymbol{\Sigma}}_P^{TR}(u)\right\}\widehat{\boldsymbol{\Lambda}}_P^{TR}(u). \tag{10}$$

In section 3.2.2 below, we illustrate two important properties of the matrix $\widehat{\mathbf{V}}_q^{PTR}(u)$ in (10). First, the size of this matrix is an estimate of the number of factors. Second, its rescaled version $\frac{1}{P}\widehat{\mathbf{V}}_q^{PTR}(u)$ converges to a well defined matrix as T and P tend to infinity.

Remark 7 (Double Asymptotics). The asymptotic results presented in Sections 3.2.2–3.2.4 below (and detailed in the Appendix and the Web Appendix) hold for $P, T \to \infty$. We use the concept of double asymptotics, where both cross-section size and sample size go to infinity simultaneously. We apply the techniques in Motta et al. (2011) to our context where we need to systematically integrate common information across all replicated trials. We emphasize that for the consistency of the estimators we only need that $P, T \to \infty$, whereas the divergence of the number of trials R is only assumed to improve even further the quality of the estimates. We also emphasize that P, T or R are allowed to grow to infinity without any restriction. For ease of presentation, from now on we will skip the dependency of the estimators on the bandwidth sequence h_{TR}.

3.2.2 *Evolutionary eigenvalues.* Let $\mathrm{v}_{1P}(u) \geq \mathrm{v}_{2P}(u) \cdots \geq \mathrm{v}_{PP}(u)$ be the set of P eigenvalues of $\boldsymbol{\Sigma}_P(u)$, the evolutionary matrix of the observations. In Proposition 2 (see Appendix) we show that under the assumptions of our model, only the largest q eigenvalues $\mathrm{v}_{1P}(u), \mathrm{v}_{2P}(u), \ldots, \mathrm{v}_{qP}(u)$ diverge as P increases while the remaining $(P-q)$ stay bounded. This is an important property that we utilized to estimate the number of factors, because the eigenvalues of $\widehat{\boldsymbol{\Sigma}}_P(u)$ are estimates of the eigenvalues of $\boldsymbol{\Sigma}_P(u)$. Analogously, let $\widehat{\mathrm{v}}_{1P}(u) \geq \widehat{\mathrm{v}}_{2P}(u) \cdots \geq \widehat{\mathrm{v}}_{PP}(u)$ be the set of P eigenvalues in decreasing order of $\widehat{\boldsymbol{\Sigma}}_P(u)$. Estimating q is equivalent to fix the size of the matrix $\widehat{\mathbf{V}}_q^{PTR}(u) = \mathrm{diag}\{\widehat{\mathrm{v}}_{1P}(u), \widehat{\mathrm{v}}_{2P}(u), \ldots, \widehat{\mathrm{v}}_{qP}(u)\}$. In Section 4.2.2, we consider a sequence of estimated evolutionary covariances and look at the number q of eigenvalues that diverge as the size of the estimated covariance increase.

In Proposition 3 (see Appendix) we show that the matrix $\frac{1}{P}\widehat{\mathbf{V}}_q^{PTR}(u)$ converges to $\mathbf{V}(u)$, a well-defined q-dimensional diagonal matrix. To give an idea of the proof of Proposition 3, we can decompose the overall error $\|\frac{1}{P}\widehat{\mathbf{V}}_q^{PTR}(u) - \mathbf{V}(u)\|$ as the following:

(i) *estimation error:* $\|\frac{1}{P}\widehat{\mathbf{V}}_q^{PTR}(u) - \frac{1}{P}\mathbf{V}_q^P(u)\|$;
(ii) *approximation error:* $\|\frac{1}{P}\mathbf{V}_q^P(u) - \mathbf{V}(u)\|$.

The estimation error in (i) depends on the difference between the largest q eigenvalues of $\frac{1}{P}\widehat{\boldsymbol{\Sigma}}_P(u)$ and the largest q eigenvalues of $\frac{1}{P}\boldsymbol{\Sigma}_P(u)$, collected in the matrix $\frac{1}{P}\mathbf{V}_q^P(u)$. The fact that it tends to zero follows from the consistency of the estimator $\widehat{\boldsymbol{\Sigma}}_P^{TR}(u)$. In contrast to the stochastic estimation error, the approximation error in (ii) is purely deterministic; it only depends on P and neither on T nor on R, and comes from the approximate factor structure of our model, namely the assumption of uniformly bounded eigenvalues.

The next result is about the asymptotic behavior of the time-varying eigenvectors.

3.2.3 *Evolutionary eigenvectors.* Analogously to the stationary case, the loadings can only be estimated up to a transformation because only the product $\boldsymbol{\Sigma}_P^X(u) = \boldsymbol{\Lambda}_P(u)\boldsymbol{\Sigma}^F\boldsymbol{\Lambda}_P(u)'$ is identifiable.

In our Theorem 1 (see Appendix), we show that the $P \times q$ matrix of estimated loadings $\widehat{\boldsymbol{\Lambda}}_P^{TR}(u)$ converges to a linear transformation $\boldsymbol{\Lambda}_P(u)\mathbf{H}(u)$ of the true loading matrix $\boldsymbol{\Lambda}_P(u)$, where $\mathbf{H}(u)$ is a $q \times q$ invertible matrix.

3.2.4 *Estimated factors and common components.* Finally, in the Web Appendix we also prove the consistency of the estimated factors and common components. In particular, we show that the vector $\widehat{\boldsymbol{F}}^{PTR}(t,r)$ of estimated factors converges to $\mathbf{H}(\frac{t}{T})^{-1}\boldsymbol{F}(t,r)$, where $\mathbf{H}(\frac{t}{T})^{-1}$ is the inverse of $\mathbf{H}(\frac{t}{T})$. Moreover, the estimated common component $\widehat{X}_i^{TR}(t) := \widehat{\boldsymbol{\lambda}}_i^{TR}(\frac{t}{T})'\widehat{\boldsymbol{F}}^{TR}(t,r)$ converges in probability to $X_i^T(t,r)$, the common component of the ith series at time t. Note that, unlike the estimators of $\boldsymbol{\Lambda}_P(\frac{t}{T})$ or $\boldsymbol{F}(t,r)$, the estimated common component is identified because the indeterminacy of $\boldsymbol{\Lambda}_P(\frac{t}{T})$ and $\boldsymbol{F}(t,r)$ due to the $q \times q$ transformation matrix $\mathbf{H}(\frac{t}{T})$ cancels out in the product between $\widehat{\boldsymbol{\lambda}}_i^{TR}(\frac{t}{T})'$ (that converges to $\boldsymbol{\lambda}_i(\frac{t}{T})'\mathbf{H}\big(\frac{t}{T}\big)$) and $\widehat{\boldsymbol{F}}^{TR}(t,r)$ (that converges to $\mathbf{H}(\frac{t}{T})^{-1}\boldsymbol{F}(t,r)$).

Remark 8 (Asymptotic framework with bounded number of channel locations). Our results are based on assumptions that reflect empirical observations on the EEG data, which are well represented by an approximate factor model where: (i) a large P is needed to recover the common structure (cross-section direction), whereas (ii) a large R is useful to improve the quality of the estimates (time direction) and for inference. Principal components analysis is a dimension reduction technique that finds "a natural justification" when (i) the data are thought as being generated according to an approximate factor model, and (ii) P is large. A framework with bounded P would be justified if the data would show strong evidence for a reduced rank structure of the covariance of the observations. That is, if the $P \times P$ covariance matrix of the data $\boldsymbol{\Sigma}_P$ has q nonzero eigenvalues and $P-q$ eigenvalues that are *exactly zero*, rather than bounded. This is rarely observed in empirical applications. For example, in our application we find that $\widehat{\boldsymbol{\Sigma}}_P(r)$, the 62×62 sample covariance of the observations, has rank 62 for all $r = 1, \ldots, 118$. The case in which $\boldsymbol{\Sigma}_P$ (the covariance matrix of $\boldsymbol{Y}$) has rank q can be well represented by $\boldsymbol{Y} \equiv \boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{F}$ or, put differently, $\boldsymbol{Z} \equiv \mathbf{0}$. In this case the approximation error is identically zero for any P, and the q eigenvectors corresponding to the q largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_P$ consistently estimate (up to transformation) $\boldsymbol{\Lambda}$ for any P.

4. Analysis of the Multichannel EEG Data

4.1 *Data Description and Preprocessing*

We analyzed the multichannel EEG data, collected in a visual–motor experiment, using our novel EFM. The EEG signals were recorded over a montage of $P = 62$ channels (see Figure 1) at the sampling rate of 512 Hz. A band-pass filter of 0.02–100 Hz was applied before analysis. The subject was required to make quick displacements of a hand-held joystick from a central position either to the right or to the left from center as instructed by a visual cue. The visual cue was randomly selected for each trial. There were a total of $R = 118$ trials for each of the two conditions. Each trial had length of $T = 512$ time points recorded over a period of 1000 milliseconds—covering $(-500, 500)$ milliseconds where the reference point of 0 is the time of presentation of the visual cue. For each trial, both linear and quadratic trends were removed and the EEGs were further filtered using a 4th order low-pass Butterworth filter with stopband at 60 Hz. From the observed time series, we estimated the time-varying covariances in Figure 1 (right-bottom side) according to the

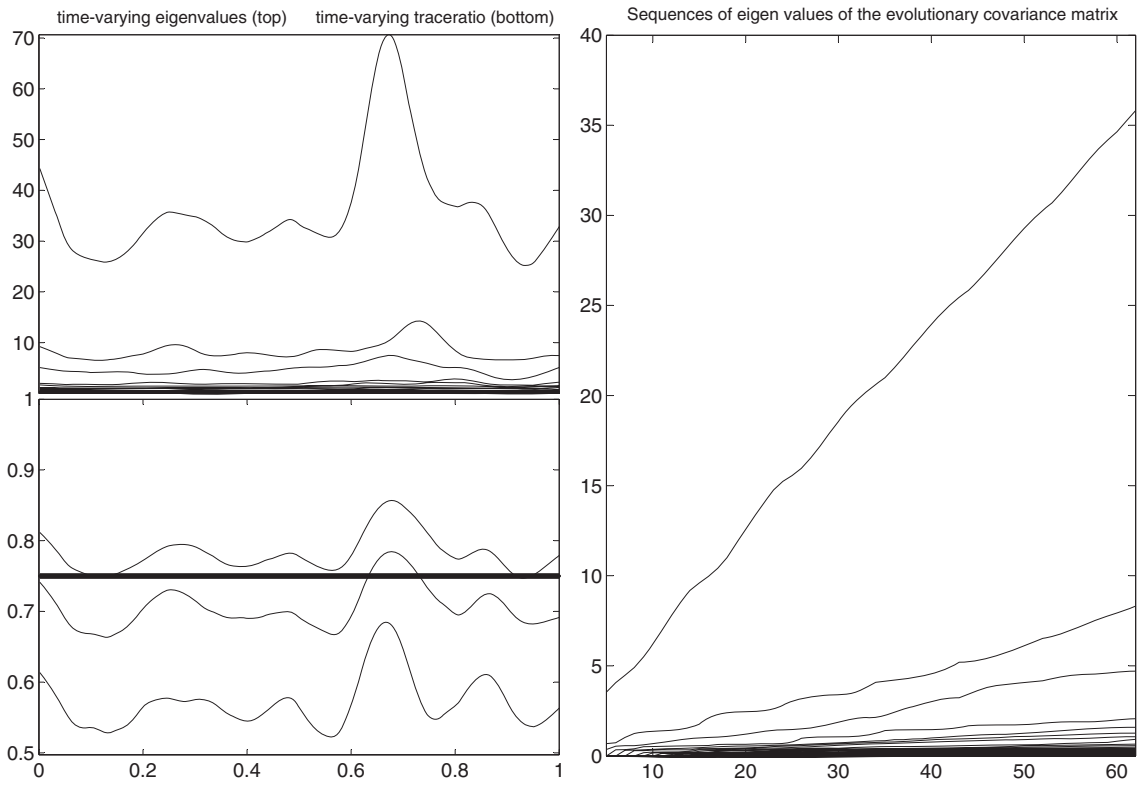

Figure 2. Left-top: 62 time-varying eigenvalues contained in the $P \times P$ diagonal matrix $\widehat{\mathbf{V}}_P(u)$, $u \in (0,1)$. Left-bottom: time-varying trace ratio $\widehat{\rho}_j(u)$, $u \in (0,1)$, $j = 1 \ldots, \hat{q}$, defined in (11). The lowest curve is $\widehat{\rho}_1(u)$, the curve in the middle is $\widehat{\rho}_2(u)$, and the highest curve is $\widehat{\rho}_3(u)$. Because $\widehat{\rho}_3(u) \geq 0.75$ for all u, we select $\widehat{q} = 3$ factors. Right: sequences of eigenvalues $\overline{\mathbf{V}}_p$, average over time of $\widehat{\mathbf{V}}_p(u)$, obtained from the estimated time-varying covariances $\widehat{\mathbf{\Sigma}}_p(u)$; $p = 5, \ldots, P = 62$. There are $\widehat{q} = 3$ diverging eigenvalues. We only report the eigenvalues for left condition. The shapes of the eigenvalues corresponding to right condition are very similar to those reported in this figure.

methodology described in Section 3.1. To estimate the loadings and the common components, we need to select the number of factors.

4.2 *Determining the Number of Factors*

To estimate the number of factors, we adopted two criteria, both based on the estimated evolutionary eigenvalues $\widehat{\mathbf{V}}_P(u)$ of the evolutionary covariance $\widehat{\mathbf{\Sigma}}_P(u)$. The first criterion is based on the explained variance, whereas the second one defines the estimated number of factors as the number of diverging eigenvalues of the covariance matrix.

4.2.1 *Time-varying trace ratio.* The plots in Figure 2 suggest that there are a total of three primary underlying brain factors involved in visual–motor processing. To have a relative measure of explained variance we define the time-varying trace ratio as,

$$\widehat{\rho}_j(u) := \frac{\sum_{k=1}^{j} \widehat{\mathrm{v}}_k(u)}{\sum_{i=1}^{P} \widehat{\mathrm{v}}_i(u)}, \qquad j = 1, \ldots, \hat{q} = 3, \tag{11}$$

which is plotted on the left-bottom side of Figure 2. The plots suggest that three factors explain more than 75% of the global variability. These three factors appear to be consistent with recording the relevant neural processes for these visual–motor actions in Bédard and Sanes (2009). These electrode sites include the frontal leads (FC) to measure activity related to premotor processing, the central leads (C) to measure activity related to motor performance, and the parietal (P) and occipital (O) leads to measure activity related to visual–motor transformations.

4.2.2 *Diverging eigenvalues of the evolutionary covariance.* To estimate the number of factors q, we apply the result of Proposition 2 in the Appendix (see also Section 3.2.2 above).

We computed, for a grid of time points, the evolutionary covariance matrix estimator $\widehat{\mathbf{\Sigma}}_P(u)$ in (6). Then we computed the eigenvalues of the upper-left $p \times p$ sub-matrices $\widehat{\mathbf{\Sigma}}_p(u)$, $p = 1, \ldots, P$. We report on the right side of Figure 2 the plot of the average over rescaled-time of the eigenvalues $\widehat{\mathrm{v}}_{jp}(u)$ of $\widehat{\mathbf{\Sigma}}_p(u)$, $j = 1, \ldots, p$, $p = 1, \ldots, P$. In the plot we indicated on the horizontal axis the number of cross-sectional units p, which obviously is maximum when the whole sample $P = 62$ is considered. There are $\widehat{q} = 3$ diverging eigenvalues, whereas all the others stay bounded as P increases.

4.3 *Spatio-temporal Factor Loadings*

Because we know the exact location of the 62 electrodes, we can draw conclusions about the locations of the corresponding underlying sixty-two (time-varying) loadings. It is in this sense that the loadings are (temporally and) spatially varying. Some typical time-varying loadings of the factors

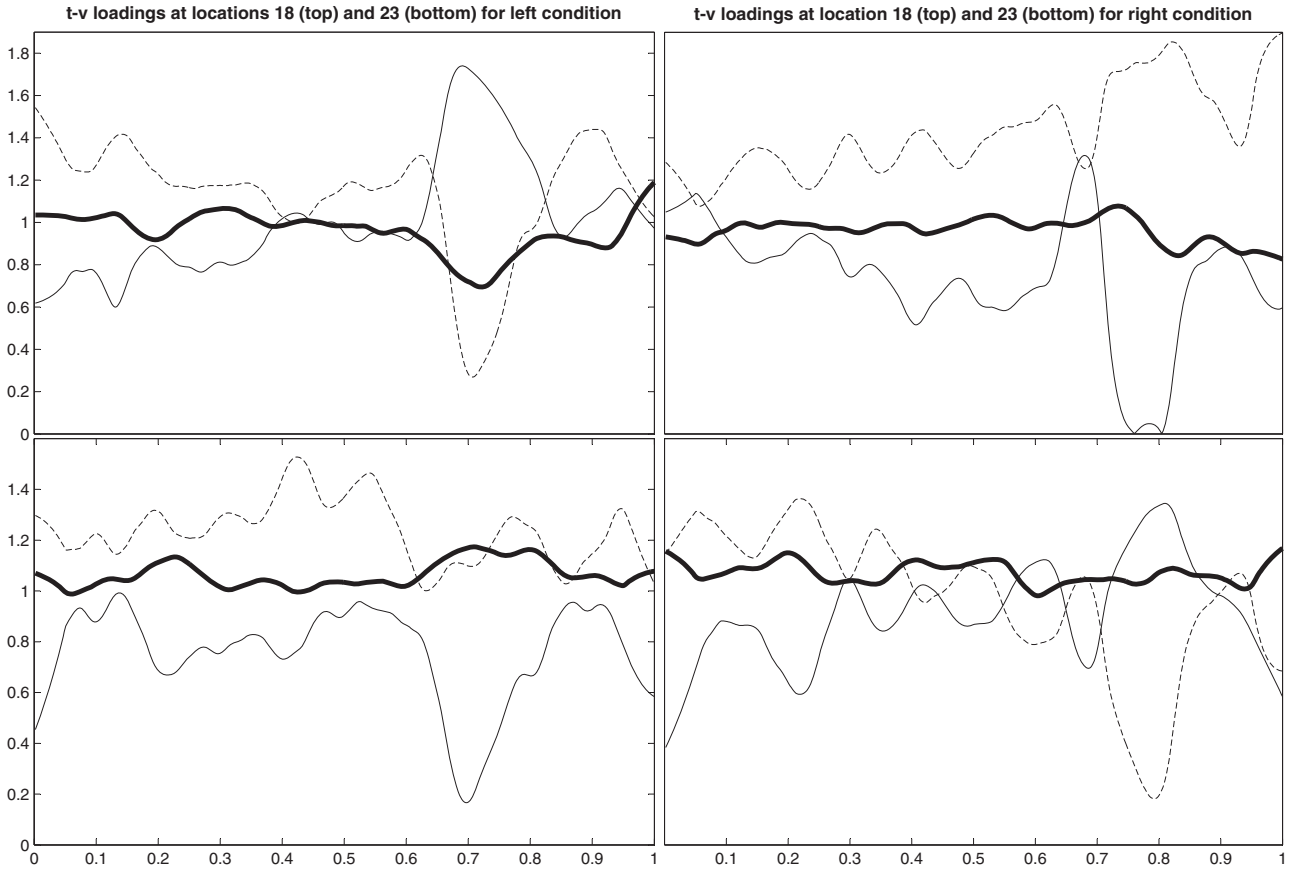

Figure 3. Estimated time-varying loadings over $\widehat{q} = 3$ factors. The entries are the functions $|\widehat{\lambda}_{ij}(u)|$ for $i = 18, 23$, $j = 1, 2, 3$, $u \in (0, 1)$. Bold lines: $j = 1$; solid lines: $j = 2$; dashed lines: $j = 3$. Left: left move; right: right move.

associated with the left and right movement are displayed in Figure 3. Then, the absolute value of the weights were collapsed into four distinct time periods: $(-500, 0)$ milliseconds; $(0, 50)$ milliseconds; $(50, 250)$ milliseconds and $(250, 500)$ milliseconds. In Figure 4, we plot the resulting spatially and temporally varying loadings. As the loadings corresponding to the first factor are constant over time, we focus on Factors 2 and 3. Moreover, we only plot the two central time periods, that is, $(0, 50)$ milliseconds and $(50, 250)$ milliseconds. The first period corresponds to prestimulus—this was a 500 millisecond interval before the visual stimulus was presented; the second corresponds to the very short period immediately following stimulus presentation; the third and fourth correspond to the post-stimulus response with the fourth period, typically capturing the moment when the participant actually responded by moving the joystick. Very often, event-related potentials display interesting patterns at about 100 and 200 milliseconds poststimulus. Studies have reported the $N100$ and $P200$ components (or the $N1 - P2$ complex) that appear following audio, somato-sensory, and visual stimuli (see Warnke, Remschmidt, and Hennighausen 1994). Thus, the third time period corresponds to processing the instruction presented. Hand-movement, in response to the stimulus, takes place around $300 - 350$ milliseconds post-stimulus presentation.

The weights corresponding to factor 1 appear to be uniformly distributed over the scalp topography. This suggests that factor 1 corresponds to the process that engages the entire brain cortex. Moreover, the weights visually appear to be almost constant over time suggesting that factor 1 can be interpreted to correspond to "baseline" brain activity that is persistent in most cognitive tasks. These spatial distributions on the scalp are consistent across the leftward and rightward movement—suggesting that they represent a stable, physiological process. However, our formal test detected some statistically significant differences between the leftward and rightward movement but these differences appear to be mild relative to those captured by factor 2 and factor 3. While still an issue of debate within the neuroscience community, this factor can be interpreted as a "default" brain connectivity. Here, we note that our model was able to objectively extract this feature from the data. This could perhaps serve as an addition to a growing list of empirical evidence of the default network.

In contrast, the weights for factor 2 appear to be concentrated in the temporal and fronto-central regions and provide connections between the left and right brain hemispheres. This very interesting result suggests contra-lateral connectivity or co-activation associated with this visuo–motor task. Our model was able to capture this activation in the

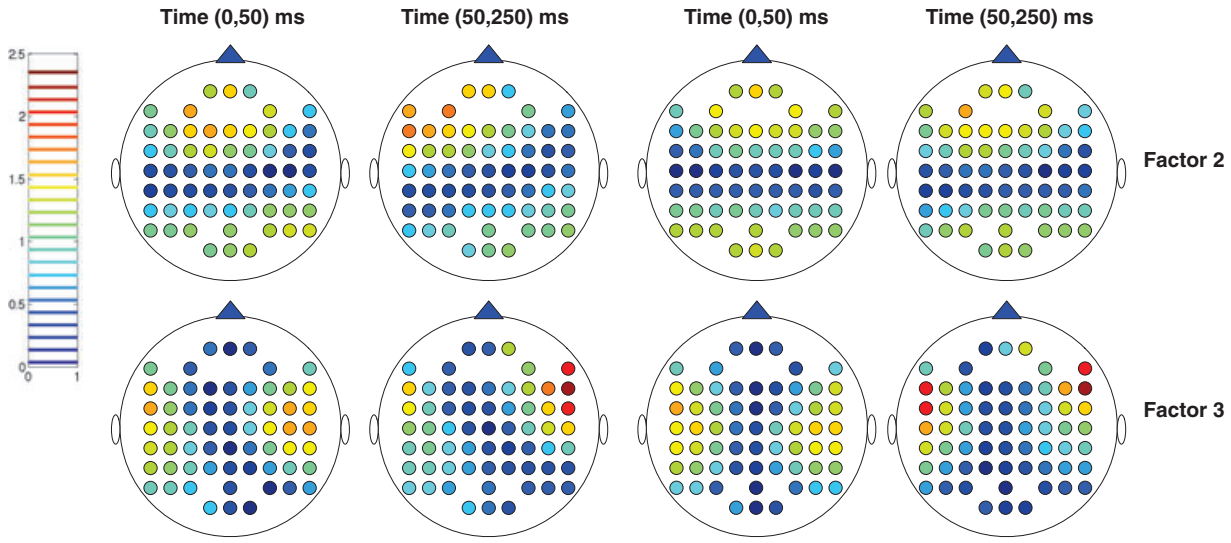

Figure 4. Spatio-temporal factor-loadings. The first row shows the factor loadings that weight the second factor, while the second row shows those weighting the third factor. The first two columns correspond to left condition, whereas the last two to right condition. This figure appears in color in the electronic version of this article.

cortical regions that are primarily involved in visual and motor processing. The weights for factor 3 are concentrated on regions that were not implicated by factor 2 and capture the anterior–posterior pathways—in contrast to the contra-lateral inter-hemispheric pathway captured by factor 2. These results are also verified by data from other neuroimaging modality which we now describe. Several empirical data suggest functional changes in the contra-lateral primary motor cortex, supplementary motor area (SMA), premotor cortex of both hemispheres. However, to the best of our knowledge, this is the first analysis that has objectively demonstrated and confirmed these results using factor analysis of time series data. This is consistent with the analysis of fMRI experiments from Roland et al. (1980) that show that SMA is a higher order supra-motor center involved in the generation and programming of complex movements. We also report convergent findings from electrical stimulation and lesion studies in animals suggesting that voluntary motor control is hierarchically organized within the cortex (see Goldberg 1985, Wise 1985).

We also formally tested for the differences between the loadings for the rightward versus leftward movements using the Wilcoxon rank sum test. In the same spirit as in Figure 4, the weights were collapsed into four distinct time periods (the same as before). More precisely, let $\widehat{\lambda}_{ij}^{\text{left}}(t,r)$ and let $\widehat{\lambda}_{ij}^{\text{right}}(t,r)$ be the loadings corresponding to location i, factor j, period t, trial r, and directions left and right, respectively (with $i=1,\ldots,P=62$, $j=1,2,3$, $t=1,2,3,4$, $r=1,\ldots,R=118$). Then, for each i, j and t we tested the two-sided hypothesis that the (absolute value of the) $R\times 1$ vectors $\widehat{\boldsymbol{\lambda}}_{ij}^{\text{left}}(t)=\{\widehat{\lambda}_{ij}^{\text{left}}(t,1),\ldots,\widehat{\lambda}_{ij}^{\text{left}}(t,R)\}'$ and $\widehat{\boldsymbol{\lambda}}_{ij}^{\text{right}}(t)=\{\widehat{\lambda}_{ij}^{\text{right}}(t,1),\ldots,\widehat{\lambda}_{ij}^{\text{right}}(t,R)\}'$ come from distributions with equal medians. We applied the test individually with a confidence level $\alpha=0.05$, as well as simultaneously (a multiple comparison) using the Bonferroni's correction, with a confidence level α/P. The two sets of data are assumed to come from continuous distributions that are identical except possibly for a location shift, but are otherwise arbitrary.

Table 1

Number of locations where the differences between left loadings and right loadings are statistically significant, according to the Wilcoxon rank sum test, at the confidence level α (in brackets are the number of locations where the differences between left loadings and right loadings are statistically significant at the level α/P)

	Factor 1	Factor 2	Factor 3
[−500,0]	12 (**0**)	2 (**0**)	13 (**0**)
[0,50]	0 (**0**)	2 (**1**)	2 (**0**)
[50,250]	26 (**15**)	23 (**4**)	10 (**1**)
[250,500]	0 (**0**)	21 (**8**)	21 (**0**)

The results are reported in Tables 1 and 2. The number of locations i where the differences between $|\widehat{\boldsymbol{\lambda}}_{ij}^{\text{left}}(t)|$ and $|\widehat{\boldsymbol{\lambda}}_{ij}^{\text{right}}(t)|$ are statistically significant are reported in Table 1. In Table 2 we report, only for those locations i where the test rejects the null of equal distribution between left and right condition, three quantiles of the distributions of the $R\times 1$ vectors $\boldsymbol{\delta}_{ij}(t)$, where $\boldsymbol{\delta}_{ij}(t)=\{|\widehat{\lambda}_{ij}^{\text{right}}(t,1)|-|\widehat{\lambda}_{ij}^{\text{left}}(t,1)|,\ldots,|\widehat{\lambda}_{ij}^{\text{right}}(t,R)|-|\widehat{\lambda}_{ij}^{\text{left}}(t,R)|\}'$.

Table 1 indicates that the most significant differences between the leftward and rightward movement appear *after* the stimulus. For this reason, we show in Table 2 only the results for the intervals 50–250 and 250–500 milliseconds.

In the interval 50–250 milliseconds poststimulus presentation, factor 2 loadings at the left anterior channels were larger for the rightward condition. Interestingly, on the same anterior channels, the loadings for the leftward condition was actually larger. For the contra-lateral pathway, it appears that

Table 2

Indexes of the locations where the right loadings are significantly different the from left loadings, see Table 1. For these indexes, we report three quantiles of the distributions of the differences between right loadings and left loadings (in bold are the locations where the differences between left-loadings and right-loadings are statistically significant at the level α/P).

	Factor 1				Factor 2				Factor 3			
Period	i	$Q_{0.1}$	$Q_{0.5}$	$Q_{0.9}$	i	$Q_{0.1}$	$Q_{0.5}$	$Q_{0.9}$	i	$Q_{0.1}$	$Q_{0.5}$	$Q_{0.9}$
[50,250]	1	−0.51	s0.11	0.72	**3**	−0.79	0.36	1.31	6	−1.06	−0.24	0.93
	3	−0.60	−0.07	0.28	4	−1.44	−0.34	0.99	12	−1.23	−0.16	0.66
	4	−0.39	0.17	0.65	**6**	−0.84	0.36	1.29	**13**	−1.12	−0.20	0.62
	5	−0.53	0.19	0.75	7	−1.21	0.45	1.62	14	−1.19	−0.20	1.01
	6	−0.68	−0.08	0.39	8	−1.35	−0.19	1.12	22	−1.03	−0.17	0.51
	7	−0.75	−0.24	0.25	9	−1.22	−0.24	0.89	24	−1.02	−0.11	0.74
	8	−0.35	0.32	0.77	10	−1.05	−0.23	0.72	25	−1.19	−0.19	0.95
	9	−0.42	0.25	0.74	13	−0.75	0.11	1.12	41	−0.84	0.19	1.14
	10	−0.45	0.18	0.72	14	−0.84	0.23	1.12	61	−0.93	−0.15	0.76
	14	−0.69	−0.10	0.45	15	−1.02	0.29	1.16	62	−1.12	−0.18	0.68
	15	−0.80	−0.20	0.29	18	−1.24	−0.19	0.87				
	16	−0.97	−0.26	0.20	19	−0.90	−0.23	0.70				
	17	−0.36	0.24	0.92	23	−0.72	0.16	1.02				
	18	−0.23	0.15	0.73	25	−1.27	−0.11	0.84				
	19	−0.37	0.11	0.55	27	−1.00	−0.15	0.81				
	23	−0.51	−0.08	0.43	29	−0.81	−0.13	0.53				
	24	−0.61	−0.18	0.24	33	−0.98	−0.21	0.61				
	25	−0.69	−0.21	0.29	**34**	−1.28	−0.29	0.58				
	26	−0.38	0.12	0.61	42	−1.02	−0.18	0.65				
	27	−0.31	0.08	0.55	**43**	−1.15	−0.30	0.71				
	28	−0.30	0.06	0.40	44	−0.55	0.02	1.04				
	33	−0.45	−0.08	0.34	45	−0.70	0.12	0.88				
	34	−0.50	−0.14	0.39	48	−0.59	0.16	0.89				
	35	−0.40	0.12	0.41								
	43	−0.51	−0.13	0.44								
	44	−0.44	0.05	0.47								
[250,500]					1	−1.05	−0.17	0.60	6	−1.24	−0.16	0.85
					3	−0.63	0.28	1.29	9	−1.08	0.12	1.45
					4	−1.57	−0.43	0.59	10	−0.90	0.23	1.10
					5	−1.14	−0.22	0.80	14	−1.13	−0.28	0.81
					6	−0.67	0.32	1.38	15	−1.41	−0.21	0.84
					7	−1.05	0.37	1.45	16	−1.61	−0.22	1.06
					8	−1.94	−0.39	0.70	17	−1.29	0.30	1.79
					9	−1.51	−0.44	0.69	18	−1.06	0.39	1.33
					10	−1.11	−0.18	0.62	19	−0.89	0.16	0.99
					12	−0.80	0.17	0.98	23	−1.01	−0.09	0.70
					13	−0.56	0.21	1.07	24	−1.07	−0.27	0.84
					14	−0.71	0.43	1.28	25	−1.18	−0.22	1.10
					15	−0.86	0.24	1.30	27	−0.92	0.28	1.11
					17	−1.41	−0.14	0.84	32	−0.88	0.16	1.25
					18	−1.24	−0.21	0.68	38	−0.84	−0.17	0.46
					22	−0.61	0.16	1.09	39	−1.12	−0.16	0.79
					23	−0.64	0.27	1.07	41	−0.78	0.26	1.23
					25	−1.23	−0.23	0.92	42	−0.76	0.10	1.16
					31	−0.53	0.18	0.61	47	−0.94	−0.14	0.62
					34	−1.21	−0.16	0.83	50	−0.61	0.13	0.85
					43	−1.08	−0.22	0.70	57	−0.75	0.10	0.94

the factor 2 loadings are dominantly larger for the leftward condition. The most interesting differences were captured by factor 2 in the interval 250–500 milliseconds poststimulus presentation. The largest differences were observed on the left temporal–parietal channels where the loadings for the leftward motion were larger than the rightward motion. These results are reported only for a single subject. However, the model is promising and can be extended to analyzing multivariate time series data from multiple subjects. In terms of the temporal dynamics of the brain process or the evolution of the weights over time, we note that factors 1 and 2 for both conditions are more-or-less stationary but that the nonstationarity is mostly captured by factor 3. We observe that in both conditions, the weights are concentrated mostly

on the fronto-parietal regions starting from 50 milliseconds poststimulus presentation.

5. Conclusions

The contribution of our article is threefold: modeling, estimation, and application. First, in terms of *modeling*, we developed a novel statistical model for nonstationary multivariate time series, that systematically integrates common information across (i) channels within a trial, and (ii) several trials from an experimental design.

Second, we developed *estimation* theory for our multitrial EFM. In particular, we showed that the estimated loading matrix converges to a linear transformation of the true loading matrix. To prove the consistency of the estimators, we only needed that the time series length per trial tends to infinity, whereas increasing the number of trials R was demonstrated to further improve the quality of the estimates.

Finally, to investigate the brain network involved in a visual–motor task using EEGs, we provided an *application* of our proposed model. There are several potential applications of our model. It can be used to test for differences in brain connectivity by directly comparing the spatio-temporal loadings. It can be further generalized in a clinical setting (with several patient groups) to determine differences in brain networks between groups.

We note that in neuroscience investigations, there are several imaging modalities utilized to study brain processes—such as functional magnetic resonance brain images (fMRI), magnetoencephalograms, and EEGs. Each of these modalities has its own advantages and limitations. On the one hand, fMRI is known to offer excellent spatial resolution (in the order of 1 mm^3). However, has poor temporal solution because they are recorded every 1000–2500 milliseconds. Thus, fMRI does not provide adequate information on the brain's temporal properties. On the other hand, EEGs have excellent temporal resolution (with one observation per 2–4 milliseconds). However, EEGs are not highly localized in space (offering information on the cm^3 scale rather than the mm^3 scale). EEGs continue to be used in many studies because they catch transient properties of brain signals. While this current work is an important starting point for studying brain processes, future work will entail inclusion of all available spatial information (obtained from anatomical constraints and functional activation results from fMRI) directly in our EFM.

6. Summary

In this article, we developed a novel statistical model for nonstationary multivariate time series recorded over replicated trials. Our model systematically integrates common information across (i) channels within a trial and (ii) several trials from an experimental design. Estimation theory, based on time-varying PCA, is rigorously developed for our multi-trial EFM. We prove consistency of the proposed nonparametric estimators of the (deterministic) time-varying covariance matrix and time-varying loadings. We also show the consistency of the estimators of the (stochastic) factors and common components. To investigate the underlying brain network implicated in this particular visual–motor task, we utilized our model to analyze a multi-channel EEG data recorded over replicated trials.

7. Supplementary Materials

Web Appendices referenced in Section 3 and in the Appendix, are available with this article at the Biometrics website on Wiley Online Library.

Acknowledgements

This research was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme (G.M.) and by the National Science Foundation (H.O.).

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71,** 135–171.

Brockmann, M., Gasser, T., and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association* **88,** 1302–1309.

Bédard, P., and Sanes, J. N. (2009). Gaze and hand position effects on finger-movement-related human brain activation. *Journal of Neurophysiology* **101,** 834–842.

Chiappa, K. H. (1997). *Evoked Potentials in Clinical Medicine.* Philadelphia, Lippincott-Raven.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25,** 1–37.

Gasser, T., Kneip, A., and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association* **86,** 643–652.

Goldberg, G. (1985). Supplementary motor area structure and function: Review and hypothesis. *Behavioral Brain Science* **8,** 567–588.

Herrmann, E. (1997). Local bandwidth choice in kernel regression estimation. *Journal of Computational and Graphical Statistics* **6,** 35–54.

Motta, G., Hafner, C., and von Sachs, R. (2011). Locally stationary factor models: Identification and nonparametric estimation. *Econometric Theory* **27,** 1279–1319.

Ombao, H. and Ho, R. M. (2006). Time-dependent frequency domain principal components analysis of multichannel non-stationary signals. *Computational Statistics and Data Analysis* **50**, 2339–2360.

Ombao, H., von Sachs, R., and Guo, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association* **100,** 519–531.

Prado, R., West, M., and Krystal, A. (2001). Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag-lead structure. *Applied Statistics* **50,** 95–109.

Roland, P., Larsen, B., Lassen, N. A., and Skinhoj, E. (1980). Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology* **43,** 118–136.

Warnke, A., Remschmidt, H., and Hennighausen, K. (1994). Verbal information processing in dyslexia—Data from a follow-up experiment of neuro-psychological aspects and eeg. *Acta Paedopsychiatrica* **56,** 203–208.

Wise, S. (1985). The primate primary motor cortex: past, present and preparatory. *Annual Review of Neuroscience* **8,** 1–19.

Received December 2010. *Revised December* 2011.
Accepted December 2011.

Appendix

In this appendix, we present the consistency results of the estimated evolutionary covariance (Proposition 1), and show

the properties of the evolutionary eigenvalues (Propositions 2 and 3) and eigenvectors (Theorem 1). The list of assumptions on our model and the proofs are provided in the Web Appendix.

The first result is about consistency of our nonparametric estimator of the time-varying covariance $\mathbf{\Sigma}_P(u)$. The consistency of the estimators of the factors and the loadings depends on the consistency of the eigenvalues (see Proposition 3 and the eigenvectors (see Theorem 1) of $\widehat{\mathbf{\Sigma}}_P^{TR}(u)$. For clarity of presentation, in what follows we set $h \equiv h_{TR}$.

PROPOSITION 1: *Under Assumptions 1–2,*

$$\sup_{u\in(0,1)} \alpha(R,T)\,\frac{1}{P}\big\|\widehat{\mathbf{\Sigma}}_P^{TR}\big(u;h\big)-\mathbf{\Sigma}_P(u)\big\| = O_p(1),$$

where $\alpha(R,T)=\min(\sqrt{RTh},\, h^{-1}\sqrt{R})$, *and where* $\widehat{\mathbf{\Sigma}}_P^{TR}(u;h)$ *and* $\mathbf{\Sigma}_P(u)$ *are defined, respectively, in (6) and (4). The Proof is given in the Web Appendix A.*

The following proposition shows that under our assumptions, only q eigenvalues of the evolutionary covariance matrix of the observations diverge as P increases while the remaining $P-q$ stay bounded. The Proof is given in the Web Appendix B.

PROPOSITION 2: *Under Assumptions 1–2, the first q eigenvalues of* $\mathbf{\Sigma}_P(u)$ *diverge, as* $P\to\infty$, *uniformly over* $u\in[0,1]$,

$$\lim_{P\to\infty}\inf_{u\in[0,1]} \mathrm{v}_{Pj}(u)=\infty \qquad \text{for all } j=1,\dots,q,$$

whereas the remaining eigenvalues are uniformly bounded by $\overline{\mathrm{v}}$, *that is,*

$$\limsup_{P\to\infty}\sup_{u\in[0,1]} \mathrm{v}_{Pj}(u)\le\overline{\mathrm{v}} \quad \text{for all } j>q.$$

PROPOSITION 3: *Under Assumptions 1–2,*

$$\sup_{(0,1)} \beta\big(R,T,P\big)\big\|\tfrac{1}{P}\widehat{\mathbf{V}}_q^{PTR}(u;h)-\mathbf{V}(u)\big\| = O_p(1),$$

where $\beta(R,T,P)=\min(\sqrt{RTh},\, h^{-1}\sqrt{R},\, P,\, \ell_P)$, $\widehat{\mathbf{V}}_q^{PTR}(u)$ *is defined in (10),* $\mathbf{V}(u)$ *is the q-dimensional diagonal matrix containing the eigenvalues of* $\mathbf{\Sigma}^{\Lambda}(u)\mathbf{\Sigma}^{F}$, *and where*

$$\ell_P := \sup_{u\in(0,1)} \Big\|\frac{\mathbf{\Lambda}_P'(u)\mathbf{\Lambda}_P(u)}{P}-\mathbf{\Sigma}^{\Lambda}(u)\Big\|. \tag{A.1}$$

The Proof of this Proposition is similar to the Proof of Proposition 2 by Motta et al. (2011) and thus omitted. The next result shows that the scaled norm of the distance between the estimated loading matrix and a linear transformation of the true loading matrix converges to zero in probability. The Proof of Theorem 1 is given in the Web Appendix C.

THEOREM 1: *Under Assumptions 1–2,*

$$\sup_{u\in(0,1)} \gamma(R,T,P)\,\frac{1}{\sqrt{P}}\big\|\widehat{\mathbf{\Lambda}}_P^{TR}(u;h)-\mathbf{\Lambda}_P(u)\mathbf{H}(u)\big\| = O_p(1) \tag{A.2}$$

where $\gamma(R,T,P)=\min(\sqrt{RTh},\, h^{-1}\sqrt{R}\,,\sqrt{P},\ell_P^{-1})$, $\widehat{\mathbf{\Lambda}}_P^{TR}(u;h)$ *is defined in (7),* $\mathbf{H}(u):=\{\mathbf{\Sigma}^F\}^{\frac{1}{2}}\mathbf{\Upsilon}(u)\{\mathbf{V}(u)\}^{-\frac{1}{2}}$, *and where* $\mathbf{\Upsilon}(u)$ *is the* $q\times q$ *matrix containing the orthonormal eigenvectors of the* $q\times q$ *matrix* $\{\mathbf{\Sigma}^F\}^{\frac{1}{2}}\mathbf{\Sigma}^{\Lambda}(u)\{\mathbf{\Sigma}^F\}^{\frac{1}{2}}$.