ANALYSIS OF FUNCTIONAL CORRELATIONS

by

Scott D. Rothenberger

B.S. Physics, University of Pittsburgh, 2007

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts & Sciences

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Scott D. Rothenberger

It was defended on

November 20, 2014

and approved by

Dr. Robert T. Krafty

Dr. Satish Iyengar

Dr. Yu Cheng

Dr. Sungkyu Jung

Dr. Martica Hall

Dissertation Advisors: Dr. Robert T. Krafty,

Dr. Satish Iyengar

ANALYSIS OF FUNCTIONAL CORRELATIONS

Scott D. Rothenberger, PhD

University of Pittsburgh, 2014

Technological advances have led to an increase in the collection of high-dimensional, nearly continuously sampled signals. Evolutionary correlations between such signals are salient to many studies, as they provide important information about associations between different dynamic processes and can be used to understand how these processes relate to larger complex mechanisms. Despite the large number of methods for analyzing functional data that have been explored in the past twenty-five years, there is a dearth of methods for analyzing functional correlations. This dissertation introduces new methods for addressing three questions pertaining to functional correlations. First, we address the problem of estimating a single functional correlation by developing a smoothing spline estimator and accompanying bootstrap procedure for forming confidence intervals. Next, we consider the problem of testing the equivalence of two functional correlations from independent samples by developing a novel adaptive Neyman testing procedure. Lastly, we address the problem of testing the equivalence of two functional correlations from dependent samples by extending the adaptive Neyman test to this more complicated setting, and by embedding the problem in a state-space framework to formulate a practical Kalman filter-based algorithm for its implementation. These methods are motivated by questions in sleep medicine and chronobiology and are used to analyze the dynamic coupling between delta EEG power and high frequency heart rate variability during sleep.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Researchers are often interested in examining how the relationship between variables measured on the same sample of subjects evolves over time. Specifically, an investigator may want to model the population correlation coefficient between two variables as a smooth curve and make inferential claims about the time-varying nature of the population correlation. However, to our knowledge, no procedures have been developed thus far for the analysis of functional correlations. The establishment of a methodology for effective estimation, reliable point-wise inference and formal comparisons of temporal correlation functions would be a novel contribution to the field of functional data analysis.

Formulating a method for functional correlation estimation and inference is not a straightforward endeavor, as a direct application of existing techniques for functional data may lead to substandard results. Bayesian confidence intervals, the most common tool for smoothing spline inference, might be constructed for point-wise measures of uncertainty of a functional correlation; however, such intervals have no overall significant level and often suffer from serious coverage problems. In addition, the assumption of independent errors does not hold for functional correlations; correlations between two variables measured on the same subjects at different times will be dependent, making it difficult to accurately assess standard errors. Furthermore, formal point-wise hypothesis testing procedures for functional data with satisfactory empirical properties have thus far been elusive.

The situation becomes even more complex when testing for the overall equivalence of two correlation functions. The simplest case arises when the functions come from two independent groups of subjects; observed correlations will be correlated over time within each group, but the two correlation curves will be independent. When comparing two different correlation functions from the same group of subjects, though, correlations will be dependent

both within and between observed correlation curves. Even in classical statistics, the usual methods for testing equality of two simple population correlation coefficients do not apply when the sample correlation coefficients have been calculated based on data from the same individuals. Several tests have been developed in an attempt to address this issue. But, as demonstrated by Dunn and Clark [13, 14], their performances vary greatly based on a variety of factors. With no a priori knowledge of the dependence between the two sample correlation coefficients used in the test, one must settle for a test procedure with low power to preserve the significance level. Comparing correlated correlations in the non-functional setting can be complicated, but procedures exist. When comparing correlated functional correlations, a new methodology must be developed, as none exists.

The goals of the dissertation are:

1. To establish an effective method for analyzing a functional correlation
2. To develop a formal test of equivalence of two functional correlations from independent samples
3. To develop a formal test of equivalence of two functional correlations from correlated samples.

To meet the first goal, we propose analyzing the functional correlation between two variables measured on the same subjects through a smoothing spline model on the Fisher's transformation scale. Several approaches for obtaining confidence intervals are explored. Our findings led us to advocate a novel bootstrap procedure based on the large sample Gaussian distribution of Fisher transformed correlations. This bootstrap-based procedure allows one to investigate how the correlation between two variables evolves over time using nonparametric measures of point-wise uncertainty to account for autocorrelated errors.

To formally compare functional correlations, adaptive Neyman hypothesis tests of equivalence of two correlation functions, motivated by methods of Fan and Lin (1998) [20], are developed. These tests are developed for both the setting where the samples from two groups are independent and when these samples are dependent. The tests adaptively overcome the curse of dimensionality, avoid the bias incurred from smoothing-based approaches, and are well-suited for the analysis of functional correlations where inherently only one sam-

ple correlation curve is provided from a group of signals. Connections between these tests and state-space models are established to allow the Kalman filter to be used to formulate algorithms for practical implementation.

The two correlation functions need not be independent; they may be from two distinct groups of subjects, or from the same group of subjects measured under two different conditions or during two different time periods.

The remainder of this dissertation is organized as follows. Chapter 2 discusses the motivating study for the proposed methods, Rothenberger et al. (2014) [58], which will be revisited as the main example in each subsequent chapter. Briefly, the time-varying correlation between delta electroencephalographic (EEG) power and high frequency heart rate variability (HF-HRV) during sleep is examined for a cohort of 197 midlife women enrolled in the SWAN Sleep Study. To our knowledge, this is the first study to explore the dynamic sleep-HRV relationship in women, and the first study to model correlations between slow-wave sleep and nocturnal HRV as continuous functions of time.

Chapter 3 describes the analysis of a single functional correlation. Our novel hypothesis test of equivalence of independent correlation functions is presented in Chapter 4, and empirical significance and power calculations are evaluated using simulations and compared to that of Bayesian confidence bands for hypothesis testing. The procedure developed for testing independent functions is extended in Chapter 5 to provide a new methodology for comparing dependent correlation functions using a novel adaptive Neyman test for dependent samples. The empirical significance and power of the test are evaluated by simulation. Lastly, conclusions are drawn in Chapter 6.

2.0 MOTIVATING STUDY: DELTA EEG POWER AND HF-HRV

The formulation of the proposed methods for analyzing functional correlations was motivated by our study of delta EEG power and HF-HRV during non-rapid eye movement (NREM) sleep in midlife women [58]. The novelty of our study with respect to the basic understanding of nocturnal physiology and the role of sleep in relation to health and functioning is introduced in this section of the dissertation. The primary goals of this motivating application are presented, and the utility of treating correlations as continuous functions of time when examining the interaction between physiological systems is elucidated.

2.1 INTRODUCTION

The autonomic nervous system (ANS) is responsible for the unconscious regulation of internal organs and glands. The sympathetic branch of the ANS functions in physiological actions which require a quick reaction ("fight or flight" response), while the parasympathetic branch functions in activities which do not require an immediate response ("rest and digest"). Spectral analysis of HRV, which is the elapsed time between consecutive heartbeats, is used to quantify distinct components of cardiac autonomic tone during sleep; in particular, the high frequency band of the HRV power spectrum (HF-HRV; 0.15-4 Hz) is a measure of cardiac parasympathetic activity. Spectral analysis of electroencephalograms (EEG) is used to quantify sleep depth; deep sleep is characterized by an abundance of EEG power in the delta band (0.5-4 Hz) during non-rapid eye movement (NREM) sleep.

Mounting evidence suggests that sleep is an important determinant of health and functioning, including cardiometabolic disease risk [8, 9, 22, 42, 60, 63]. Altered autonomic tone,

as measured by decreased heart rate variability (HRV), may represent one pathway through which sleep affects health and functioning [8, 42, 22]. Alterations in HRV have been observed in sleep apnea and insomnia, which are the two most common sleep disorders seen in primary care settings [47, 73]. Heart rate variability is decreased during both sleep and wakefulness in patients with sleep apnea compared to good sleeper controls across the lifespan [31, 38, 45]. Heart rate variability also appears to normalize in conjunction with successful continuous positive airway pressure (CPAP) treatment [23, 35]. Although the evidence is less conclusive in insomnia, some studies have observed decreased HRV during sleep in patients with insomnia compared to good sleeper controls [5, 32, 33].

Sleep and HRV are both regulated, in part, by autonomic nervous system activity. Non-rapid eye movement (NREM) sleep is characterized by relatively greater parasympathetic tone, indicated by greater high frequency HRV (HF-HRV), while rapid eye movement (REM) sleep and wakefulness show increased sympathetic nervous system activity [4, 51, 64]. Gradations in HRV are seen within NREM sleep, with lower levels of HF-HRV seen during stage 1 sleep and higher levels seen during stage 3 and 4 "slow-wave" sleep [4, 66]. Studies that have evaluated cardiac autonomic tone in relation to sleep have often used a "discrete epoch" approach in which spectral analysis of HRV is measured during five- to ten-minute epochs corresponding to specific stages of sleep (e.g., stage N3 sleep, rapid eye movement (REM) sleep). More nuanced methodological approaches, including those that utilize two minute arousal-free discrete epochs, have shown that fluctuations in HRV are attributable to the changing distribution of sleep stages [67, 68].

These studies have demonstrated that sleep and HRV are correlated in a broad sense; yet converging evidence suggests that sleep and HRV are dynamically coupled over shorter time intervals [24, 51, 50] and this relationship may be altered in people with sleep disturbances such as obstructive sleep apnea (OSA) and insomnia [34, 33]. Taken as a whole, these studies suggest that the relationship between sleep and HRV varies across time as well as among individuals with disturbed sleep. That this relationship may be altered in association with disturbed sleep suggests that the dynamics of the EEG-HRV relationship warrant further investigation; such alterations might reflect variations in an underlying physiological process critical to the restorative properties of sleep.

Some studies have used analytical approaches that measure the strength of the linear association between two time series in the frequency domain, suggesting that the time delay between changes in HRV and changes in the EEG that is reliably observed in good sleepers disappears in individuals with sleep apnea or insomnia [34, 33]. While aggregation of data (e.g., discrete epochs, whole night averages) may reveal significant associations between sleep and HF-HRV, this approach may obfuscate more complex EEG-HRV relationships observed within and across NREM periods. These complex relationships may be especially important among individuals with primary sleep disorders such as sleep apnea or insomnia [33]. Thus, when evaluating cardiac autonomic activity as a mechanism through which sleep and sleep disturbances affect health and functioning, the analytical approach by which physiological data are examined in relation to one another across the night is an important methodological consideration.

In order to address this methodological consideration, we were interested in understanding if delta EEG power and HF-HRV fluctuate in relation to one another on a moment-to-moment basis, both within and across NREM sleep periods. Specifically, we were interested in modeling correlations between EEG delta power and HF-HRV during NREM sleep as smooth, continuous functions of time, primarily because the dynamics of this relationship might reflect an underlying physiological process critical to the restorative properties of delta EEG power and cardiac parasympathetic activity during sleep. At a more basic level, understanding the time-varying nature of the EEG-HRV relationship will enable researchers to more accurately assess HRV during sleep. A greater understanding of the dynamics of the EEG-HRV relationship provides a more complete picture of the basic physiology of sleep which, despite originating in the brain, is inextricably linked to peripheral physiology [43, 62]. We chose to focus on NREM delta EEG power as it is a stable and reliable quantitative measure of visually-scored slow-wave sleep, which has been linked with HF-HRV in previous studies [4, 66, 67]. Conceptually, delta EEG power and parasympathetic nervous system activity may promote physiological restoration, a putative function on NREM sleep. Although delta power can be detected during REM sleep, its expression during NREM sleep is most closely tied to its role as a marker of sleep homeostasis and sleep depth [21].

To evaluate time-varying associations between HRV and the sleep EEG, we utilized

overnight data from a sample of midlife women studied at four sites around the country: Chicago, IL; Detroit, MI; Oakland, CA; and Pittsburgh, PA. Our study seeks to address the following three aims that correspond to the methodological aims outlined in Chapter 1:

1. To examine how the correlation between delta EEG power and HF-HRV varies as a function of time during NREM sleep in midlife women
2. To examine whether temporal EEG-HRV relationships in midlife women differ as a function of sleep disordered breathing and insomnia
3. To examine if the time-varying correlation significantly changes across different NREM periods.

2.2 PARTICIPANTS AND DATA

A total of 368 women participated in the multi-site Study of Women's Health Across the Nation (SWAN) Sleep Study [28, 59]. Each study site recruited Caucasian participants and members of one racial/ethnic minority group (African American or Chinese). Eligibility for the SWAN Sleep Study was based primarily on factors known to affect sleep. Specific exclusions were regular overnight shiftwork; current menopausal hormone replacement therapy use; current chemotherapy, radiation, or oral corticosteroid use; and regular consumption of more than 4 alcoholic drinks per day.

A subset ($n = 197$) of the SWAN Sleep Study cohort was used for the current analyses. Of these participants, 19 exhibited symptoms of insomnia without sleep disordered breathing (SDB), 26 exhibited symptoms SDB without insomnia, 6 exhibited both symptoms of insomnia and SDB, and 146 did not exhibit symptoms of insomnia or SDB. Participants were not included in the present analyses if quantitative EEG or HRV data were not available due to technical problems with the polysomnography (PSG) recording ($n = 56$); if they were taking medications that affect heart rate variability (e.g., beta blockers, angiotensin-converting-enzyme (ACE) inhibitors) ($n = 57$); if they were missing covariate data ($n = 19$); or were missing too much HRV or EEG data to reliably interpolate HRV and/or EEG profiles ($n = 39$). On average, participants not included in the present analyses had higher body

mass index (BMI) values, reported more subjective sleep complaints, and had shorter sleep durations, compared to participants who were included in these analyses ($p-values < 0.01$). These groups did not differ in terms of NREM delta EEG power or high frequency HRV during NREM sleep, age, menopausal status, or percent NREM sleep. The study protocol was approved by each site's institutional review board. Participants gave written informed consent and received compensation for participation.

Ambulatory PSG sleep studies were conducted in participants' homes on the first three nights of the SWAN Sleep Study protocol as previously described [28]. Study staff visited participants in their homes on each sleep study night to apply and calibrate PSG study monitors. Participants slept at their habitual sleep times, and upon awakening in the morning, participants turned off and removed the study equipment.

Participants' apnea-hypopnea index (AHI), assessed by PSG on the first night of the sleep study, was used to quantify sleep disordered breathing (SDB). Participants with an $AHI \geq 15$ were considered to have clinically significant SDB. The self-report Insomnia Symptom Questionnaire (ISQ), a 13 item self-report instrument, was used to identify participants meeting criteria for insomnia based on the American Psychiatric Association's fourth edition of the Diagnostic Statistical Manual (DSM-IV) criteria for insomnia and the American Academy of Sleep Medicines (AASM) Research Diagnostic Criteria (RDS) [3, 15]. The ISQ retrospectively queried participants' chronic sleep disturbances, such as difficulties initiating or maintaining sleep, or experiencing un-refreshing sleep at least 3 nights per week over the past month or longer [48].

A single night of data was used to compute power spectral analysis of the EEG and HRV for each participant given the high short-term temporal stability of whole night measures of EEG delta power and HF-HRV [32]. PSG records were visually scored in 20-second epochs [57]. Fast Fourier Transform (FFT) was employed to derive delta EEG power spectral estimates in 4-second epochs and HF-HRV power spectral estimates in 2-minute epochs. Delta EEG and HF-HRV epochs occurring during NREM sleep were then temporally aligned across the entire sleep period. The bins selected for analysis of EEG and HRV data were consecutive 4-second intervals, corresponding to the non-overlapping spectral estimates of delta EEG power generated by FFT. Missing data were handled in a "paired" fashion; when

4-second bins of EEG data were missing values, the corresponding 4-second bins of HRV data were also considered missing values. Only a portion of an entire 2-minute HRV measurement was discarded, unless the concurrent EEG data were missing for the entire 2-minute interval. Similarly, if a 2-minute epoch of HRV data was a missing value, the simultaneous bins of EEG data were treated as missing values.

2.3 DATA PROCESSING

Absolute delta EEG power was log-transformed and normalized HF-HRV power was square-root-transformed in order to produce approximately normally-distributed values. Analyses were limited to the first 3 NREM periods due to the limited amount of data available for subsequent sleep cycles. Analyses were conducted in relative time as opposed to absolute (clock) time to compensate for inter-individual differences in the length of individual NREM periods. An approach similar to Achermann et al. (2003) [1] was used to compute relative time. First, each participant's NREM "clock" was standardized to take values between $t = -1$ and $t = +1$. Next, the time at which the maximum in delta EEG power occurred was detected for each participant, and this time was designated as $t = 0$. Finally, HRV and EEG data for each participant were linearly interpolated and re-sampled on the new time scale, giving the same number of relative time points per participant ($T = 582$) within each NREM period.

3.0 ANALYZING A FUNCTIONAL CORRELATION

3.1 INTRODUCTION

Before discussing our techniques for analyzing a functional correlation, it helps to describe the data giving rise to the correlation function of interest and the various assumptions that are needed. Some challenges faced when analyzing a time-varying correlation are also introduced. In addition, a brief review of Fisher's correlation transformation is provided below, as it plays a fundamental role in our methods.

Suppose that the variables of interest are pairs of continuous functions $\{X_i(t), Y_i(t)\}$, where $i = 1, \ldots, n$ denotes subject and t denotes time. As an example from our motivating application, $X_i(t)$ and $Y_i(t)$ may represent the delta EEG power and high frequency HRV functions for the ith subject at time t. We assume that the observed data are a random sample from the model

$$\begin{cases} X_i(t) &= \mu_X(t) + \delta_{iX}(t); \\ Y_i(t) &= \mu_Y(t) + \delta_{iY}(t), \end{cases} \tag{3.1}$$

where $EX_i(t) = \mu_X(t)$ and $EY_i(t) = \mu_Y(t)$ are the functional population means for all subjects. The stochastic error processes $\delta_{iX}(t)$ and $\delta_{iY}(t)$ have mean zero and covariance functions given by $\gamma_X(s,t) = \text{cov}\,[\delta_{iX}(s),\, \delta_{iX}(t)]$, $\gamma_Y(s,t) = \text{cov}\,[\delta_{iY}(s),\, \delta_{iY}(t)]$, and $\gamma_{XY}(s,t) = \text{cov}\,[\delta_{iX}(s),\, \delta_{iY}(t)]$. These functions represent the dependence between measurements made on the same subject at different times and are well-defined as long as $EX_i^2(t)$ and $EY_i^2(t)$ are finite for all t. Note that, since observations from different subjects are assumed to be independent, $\text{cov}\,[\delta_{iX}(s),\, \delta_{i'X}(t)] = \text{cov}\,[\delta_{iY}(s),\, \delta_{i'Y}(t)] = \text{cov}\,[\delta_{iX}(s),\, \delta_{i'Y}(t)] = 0$ for all $i \neq i'$.

The autocovariance functions $\gamma_X(\cdot)$ and $\gamma_Y(\cdot)$, along with the cross-covariance function

$\gamma_{XY}(\cdot)$, are used to define the cross-correlation function (CCF) as

$$\rho(s,t) = \frac{\gamma_{XY}(s,t)}{\sqrt{\gamma_X(s,s)\gamma_Y(t,t)}}.$$

The CCF is a measure of the linear association between the processes $X_i(s)$ and $Y_i(t)$, at possibly different values of time s and t, and $-1 \leq \rho(s,t) \leq +1$. When $s = t$, we have $\gamma_X(s,t) = \text{var}\left[X_i(t)\right] = \sigma_X^2(t)$, $\gamma_Y(s,t) = \text{var}\left[Y_i(t)\right] = \sigma_Y^2(t)$, and CCF given by:

$$\rho(t) = \frac{\gamma_{XY}(t)}{\sigma_X(t)\sigma_Y(t)}. \tag{3.2}$$

Notice that $\rho(t)$ in Equation (3.2) is simply the population correlation coefficient between the functions $X_i(t)$ and $Y_i(t)$ (the second time index has been dropped for ease of notation). Analysis of the correlation function $\rho(t)$ is the focus of this chapter of the dissertation.

If one assumes that the random functional variates $\{X_i(t),\, Y_i(t)\}_{i=1}^n$ follow a bivariate normal distribution at each time point t, we may write

$$\begin{pmatrix} X_i(t) \\ Y_i(t) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X(t) \\ \mu_Y(t) \end{pmatrix}, \begin{pmatrix} \sigma_X^2(t) & \rho(t)\sigma_X(t)\sigma_Y(t) \\ \rho(t)\sigma_X(t)\sigma_Y(t) & \sigma_Y^2(t) \end{pmatrix} \right). \tag{3.3}$$

Under this distribution, Pearson's sample correlation coefficient $r(t)$ is an asymptotically unbiased estimator of the population correlation coefficient $\rho(t)$ such that $\sqrt{n}\left[r(t) - \rho(t)\right] \xrightarrow{d} N\left(0, \left[1-\rho^2(t)\right]^2\right)$ [2]. Furthermore, if one makes the very strong assumption of independence over time, i.e., $\gamma_X(s,t) = \gamma_Y(s,t) = \gamma_{XY}(s,t) = 0$ for $s \neq t$, then $r(s) \perp r(t)$ for $s \neq t$.

Even though $r(t)$ is an asymptotically unbiased and consistent estimator of $\rho(t)$, it is not optimal to perform correlation analyses on this scale. Many of the deficiencies of correlation analyses based on the natural scale that are known in classical statistics, namely that large sample asymptotics are slow with heteroscedastic variances that depend on the parameter of interest, are exacerbated in the functional setting where smoothing procedures can be negatively affected by heteroscedastic data. A one-to-one variance stabilizing transformation,

known as Fisher's correlation transformation, is commonly employed. The Fisher transformations and corresponding inverse transformations are given by Equations (3.4) and (3.5), respectively:

$$z(t) = \frac{1}{2}\ln\left[\frac{1+r(t)}{1-r(t)}\right], \qquad \eta(t) = \frac{1}{2}\ln\left[\frac{1+\rho(t)}{1-\rho(t)}\right] \tag{3.4}$$

$$r(t) = \frac{\exp\left[2z(t)\right]-1}{\exp\left[2z(t)\right]+1}, \qquad \rho(t) = \frac{\exp\left[2\eta(t)\right]-1}{\exp\left[2\eta(t)\right]+1} \tag{3.5}$$

It can be shown that $\sqrt{n-3}\left[z(t)-\eta(t)\right] \xrightarrow{d} N(0,1)$ [30], where $z(s) \perp z(t)$ for $s \neq t$ under the strong independence assumption made above. The variance of $z(t)$ is approximately constant and equal to $(n-3)^{-1}$ for large n; i.e., the errors on Fisher's transformed scale are approximately homoscedastic with known variance. As this chapter of the dissertation demonstrates, Fisher's correlation transformation is extremely useful and provides a reasonable framework for the analysis of a functional correlation.

Unfortunately, the assumption that $\gamma_X(s,t) = \gamma_Y(s,t) = \gamma_{XY}(s,t) = 0$ for $s \neq t$ is far from realistic. Measurements made on the same subject at different times are naturally expected to be dependent. Assuming such independence would lead to a biased estimate of the variance function of $z(t)$ (and of $r(t)$), and many of the distributional results given above cannot be directly applied in practice. When the data are sampled at fixed times $\{t_j\}_{j=1}^{T}$, the covariance matrix of $[z(t_1),\ldots,z(t_T)]'$ will have non-zero off-diagonal elements. The distribution of $[z(t_1),\ldots,z(t_T)]'$ becomes extremely complicated, especially if the dependence structure is not known in advance. However, as discussed in later sections, our novel method for analyzing a correlation function circumvents such issues.

We propose that the analysis of a functional correlation be performed on Fisher's transformed scale, as opposed to the original correlation scale, for several reasons. First, the convergence to normality of $z(t)$ is much faster than that of $r(t)$, especially for small sample sizes and more extreme values of the sample correlation. Second, the large-sample properties of $r(t)$ are very sensitive to departures from bivariate normality in the underlying data $\{X_i(t), Y_i(t)\}_{i=1}^{n}$ at each time t, but inference based on $z(t)$ is still reliable when the data are only approximately normally distributed. In addition, $z(t)$ can take any value in $[-\infty,+\infty]$; when mapped back to the original correlation scale, the inverse transformation

guarantees that $r(t) \in [-1, +1]$. This may be important when constructing confidence intervals for $\rho(t)$, as the CLT for $r(t)$ may lead to intervals which extend outside $[-1, +1]$. Lastly, data on the original correlation scale can easily suffer from heteroscedasticity, while Fisher's transformation produces approximately homoscedastic data. Fisher's transformation is especially necessary to ensure homoscedasticity when correlations are fairly high in magnitude, such as those encountered in our motivating study [58] (see Section 4.4 and Figures 5 and 6). As described in the next section, we propose analyzing a functional correlation using a smoothing spline model; however, smoothing spline methods are not efficient when the data are heteroscedastic over time. Thus, analyzing $\eta(t)$ via smoothing splines on Fisher's transformed scale and applying the inverse transformation is an efficient way to analyze $\rho(t)$, but smoothing directly on the original correlation scale is not.

3.2 ESTIMATION OF A FUNCTIONAL CORRELATION

We wish to estimate the time-varying correlation between the variables $X(t)$ and $Y(t)$ measured on the same $i = 1, \ldots, n$ subjects from a homogeneous population. We obtain sample data $\{x_{ij}, y_{ij}\}_{i=1}^{n}$ at design points $\{t_j\}_{j=1}^{T}$. The sample data are assumed to be discrete realizations of the random processes $\{X_i(t), Y_i(t)\}_{i=1}^{n}$ given by Equation (3.1). The processes are assumed to follow a bivariate normal distribution, as depicted in Equation (3.3), at each instantaneous moment t in the continuous interval $[t_1, t_T]$. The sample correlation coefficient

$$r_j = \frac{\sum_{i=1}^{n} \left(x_{ij} - \overline{x}_{\cdot j}\right)\left(y_{ij} - \overline{y}_{\cdot j}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_{ij} - \overline{x}_{\cdot j}\right)^2} \sqrt{\sum_{i=1}^{n} \left(y_{ij} - \overline{y}_{\cdot j}\right)^2}}$$

is computed at each design point t_j, and the Fisher-transformed correlation z_j is obtained using Equation (3.4).

The model considered throughout this section is given by

$$z_j = \eta(t_j) + \varepsilon_j\,, \qquad j = 1, \ldots, T \tag{3.6}$$

where $\{z_j\}_{j=1}^{T}$ are the Fisher-transformed correlations at observation times $t_1, \ldots, t_T$, $Ez_j = \eta(t_j)$, $\eta(\cdot)$ is a smooth function on the continuous interval $[t_1, t_T]$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)'$ is a

mean zero Gaussian random vector with covariance matrix Γ. The off-diagonal elements of Γ are left unspecified, as we do not assume to know the autocovariance function $\gamma_\varepsilon(t_j, t_k) = \text{cov}(\varepsilon_j, \varepsilon_k)$ in advance. We will, however, assume that the diagonal elements of Γ are equal to $(n-3)^{-1}$, since errors on Fisher's transformed scale are homoscedastic with known variance for reasonably large n. For simplicity of technical arguments, we suppose that the observation times are equally spaced and common to all n subjects.

Our method proceeds as follows: To estimate the correlation function $\rho(t)$ on the continuous interval $[t_1, t_T]$, we first estimate the function $\eta(t)$ based on the observed data which follow the model in Equation (3.6). We propose obtaining the estimator $\hat{\eta}_0(t)$ through cubic spline smoothing of the pairs $\{t_j, z_j\}_{j=1}^T$ by minimizing penalized sum-of-squares, with smoothing parameter selected through generalized cross-validation [25]. Finally, the estimated correlation function of interest, denoted by $\hat{\rho}_0(t)$, is obtained using Equation (3.5).

Some technical details of cubic spline smoothing in general are now provided under the model in Equation (3.6), and the final results will help to understand the challenges of confidence interval construction described in the next section. First, one assumes that $\eta(t) \in \mathcal{W}_2^2$, where $\mathcal{W}_2^2$ is a Sobolev space of smooth functions defined by

$$\mathcal{W}_2^2 = \left\{ f(t) : [t_1, t_T] \to \mathbf{R}^1;\ f(t), f'(t) \text{ absolutely continuous}; \int_{t_1}^{t_T} f''(t)^2 dt < \infty \right\}. \quad (3.7)$$

One seeks the optimal estimator $\hat{\eta}(t)$ that fits the observed data well while also being reasonably smooth. For some smoothing parameter $\lambda > 0$, $\hat{\eta}(t)$ is the function that minimizes the penalized sum-of-squares. Formally,

$$\hat{\eta}(t) = \underset{\eta(t) \in \mathcal{W}_2^2}{\text{argmin}} \left\{ (\mathbf{z} - \boldsymbol{\eta})' W (\mathbf{z} - \boldsymbol{\eta}) + \lambda \int_{t_1}^{t_T} \eta''(t)^2 dt \right\}, \quad (3.8)$$

where $\mathbf{z} = (z_1, \ldots, z_T)'$, $\boldsymbol{\eta} = [\eta(t_1), \ldots, \eta(t_T)]'$, and W is a symmetric, positive-definite matrix of weights. The solution $\hat{\eta}(t)$ exists and is an order four spline smooth with knots placed at all observation times (i.e., a cubic spline smooth). Let $\phi_1, \ldots, \phi_{T+2}$ be $(T+2)$ fourth-order B-spline basis functions with knots at $t_1, \ldots, t_T$ and set $\boldsymbol{\phi}(t) = [\phi_1(t), \ldots, \phi_{T+2}(t)]'$. Denote the $T \times (T+2)$ matrix of basis functions as $\Phi = \begin{pmatrix} \boldsymbol{\phi}(t_1)' & \boldsymbol{\phi}(t_2)' & \cdots & \boldsymbol{\phi}(t_T)' \end{pmatrix}$, and represent the $(T+2) \times (T+2)$ penalty matrix R of integrated second derivatives by its

elements $\left\{R_{kl} = \int_{t_1}^{t_T} \phi_k''(t)\phi_l''(t)dt\right\}_{k,l=1}^{T+2}$. The smoothing parameter λ is chosen to minimize some objective criterion; the generalized cross-validation score GCV(λ) is very popular and will be the criterion considered in this proposal. The resulting estimator and its variance can be written as:

$$\hat{\eta}(t) = \boldsymbol{\phi}(t)' \left[\Phi' W \Phi + \lambda R\right]^{-1} \Phi' W \mathbf{z},$$

$$\text{var}\left[\hat{\eta}(t)\right] = \boldsymbol{\phi}(t)' \left[\Phi' W \Phi + \lambda R\right]^{-1} \Phi' W \Gamma W \Phi \left[\Phi' W \Phi + \lambda R\right]^{-1} \boldsymbol{\phi}(t).$$

Because the model in Equation (3.6) assumes that errors are homoscedastic, we can express the covariance matrix as $\Gamma = \sigma^2 P$, where $\sigma^2 = (n-3)^{-1}$, and the diagonal elements $P_{jj} = 1$ for all $j = 1, \ldots, T$. It has been demonstrated that var $[\hat{\eta}(t)]$ will be minimized when $W = P^{-1}$ [72, 39, 36]. Ideally, one uses the weight matrix . In this case, the estimator and its variance can be re-expressed as:

$$\hat{\eta}(t) = \boldsymbol{\phi}(t)' \left[\Phi' P^{-1} \Phi + \lambda R\right]^{-1} \Phi' P^{-1} \mathbf{z}, \tag{3.9}$$

$$\text{var}\left(\hat{\eta}(t)\right) = \sigma^2 \boldsymbol{\phi}(t)' \left[\Phi' P^{-1} \Phi + \lambda R\right]^{-1} \Phi' P^{-1} \Phi \left[\Phi' P^{-1} \Phi + \lambda R\right]^{-1} \boldsymbol{\phi}(t). \tag{3.10}$$

In contrast, since a structure for P is unknown, our estimation method sets $W = I$ to obtain:

$$\hat{\eta}_0(t) = \boldsymbol{\phi}(t)' \left[\Phi' \Phi + \lambda R\right]^{-1} \Phi' \mathbf{z}, \tag{3.11}$$

$$\text{var}\left[\hat{\eta}_0(t)\right] = \sigma^2 \boldsymbol{\phi}(t)' \left[\Phi' \Phi + \lambda R\right]^{-1} \Phi' P^{-1} \Phi \left[\Phi' \Phi + \lambda R\right]^{-1} \boldsymbol{\phi}(t). \tag{3.12}$$

Equation (3.11) and the inverse Fisher transformation of Equation (3.5) are employed to give our estimated correlation function $\hat{\rho}_0(t)$.

3.3 CONFIDENCE INTERVALS FOR A FUNCTIONAL CORRELATION

The utility of an estimated functional correlation is largely determined by the accuracy of its estimated variability. As previously mentioned, the covariance structure of the data

$\{X_i(t), Y_i(t)\}_{i=1}^n$, and subsequently the covariance structure of $(z_1, \ldots, z_T)'$, is quite complicated in any realistic situation. The functions $\gamma_X(\cdot)$, $\gamma_Y(\cdot)$ and $\gamma_{XY}(\cdot)$ defined in Section 3.1 for the observed data will not be known in practice, and assuming parametric forms could badly misrepresent the truth. Even if parametric models for the covariance functions of the observed data were close to the truth, Fisher's nonlinear correlation transformation will most likely give rise to a large sample distribution that is not tractable or analytically representable. Furthermore, if one were to first compute $\{z_1, \ldots, z_T\}$, assume a form for P and then use smoothing spline results to estimate the variability in $\hat{\eta}(t)$, a misspecification of P could lead to very misleading inference. The following sections describe some difficulties inherent in constructing spline-based confidence intervals for a functional correlation and our novel method which circumvents such issues.

3.3.1 Bayesian Confidence Intervals for a Functional Correlation

Spline-based confidence intervals are typically referred to as Bayesian "confidence intervals," as there exists a well-established connection between smoothing spline models and Bayesian models. Constructing Bayesian CI's is by far the most popular approach for inference in the smoothing spline framework. The smoothing spline estimator $\hat{\eta}(t)$ can be shown to be equivalent to a Bayes estimate of $\eta(t)$ when one assumes that $\eta(t)$ is a sample function from a certain Gaussian prior distribution [69]. In the Bayesian framework, one can construct Gaussian confidence limits for $\eta(t)$ based on the posterior variances of $\hat{\eta}(t_1), \ldots, \hat{\eta}(t_T)$. Bayesian CI's have a certain frequentist interpretation referred to as "across the function": when restricting the $(100 - \alpha)\%$ Bayesian CI's to the observation times $t = t_1, \ldots, t_T$, approximately $(100 - \alpha)\%$ of the T intervals will cover the true values of $\eta(t)$. These are not exactly "point-wise" confidence intervals in the conventional sense; there will typically be under-coverage where $\eta(t)$ is rough and over-coverage where $\eta(t)$ is smooth [26]. When the smoothing parameter is selected via generalized cross-validation, the value of σ^2 in Equation (3.10) is estimated by

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{T} [z_j - \hat{\eta}(t_j)]^2}{T - \text{trace}\left[\Phi \left(\Phi' P^{-1} \Phi + \lambda R\right)^{-1} \Phi' P^{-1}\right]}, \tag{3.13}$$

and $\hat{\text{var}}[\hat{\eta}(t)]$ is easily obtained by substituting $\hat{\sigma}^2$ for σ^2 in the expression for $\text{var}[\hat{\eta}(t)]$. The form of a Bayesian confidence interval is quite simple; for instance, a 95% Bayesian CI for $\eta(t)$ is given by

$$\hat{\eta}(t) \pm 1.96\sqrt{\hat{\text{var}}[\hat{\eta}(t)]}. \tag{3.14}$$

Four potential approaches for obtaining confidence intervals for $\eta(t)$ are described below. The main problem inherent in these methods has been alluded to in previous sections: we do not know the true structure of P!

1. Perform a spline smooth under the independence assumption and simply use Equations (3.11) and (3.12) to construct Bayesian confidence intervals. Due to the gross misspecification of P, the estimate of $\text{var}[\hat{\eta}(t)]$ will be far from the truth, and this method will lead to poor inference.
2. Perform the procedure above to obtain Bayesian confidence intervals. Attempt to circumvent the specification of P through the construction of "modified" Bayesian CI's: vary the value of $\hat{\sigma}^2$ until $(100-\alpha)\%$ of the observed $z_1, \ldots, z_T$ values are contained in the "modified" CI's at times $t_1, \ldots, t_T$. However, such confidence intervals will not have an "across the function" interpretation, or any reasonable interpretation at all.
3. Attempt to estimate the entire covariance matrix Γ based on residuals from a preliminary spline smooth assuming independence to obtain $P = \hat{\sigma}^{-2}\hat{\Gamma}$, perform a secondary spline smooth assuming the estimated P, and lastly construct Bayesian confidence intervals. However, this approach is seriously flawed; among other reasons, it would require a substantial number N of replicated curves from the same n subjects under the same N time-varying conditions. Such an estimate of Γ would be computed as $\hat{\Gamma} = (N-1)^{-1}E'E$, where E is the $N \times T$ matrix of residuals. Γ contains $T(T+1)/2$ elements to be estimated, and it is unlikely that we would ever have a large enough N to do this accurately [54]. In fact, we will only have $N = 1$ reliable replications in any realistic situation. Furthermore, even if reliable replications were available, the diagonal elements of $\hat{\Gamma}$ would never be equal in practice. This is necessary for the homoscedasticity assumption of our functional correlation model.
4. Assume that the process generating the errors is stationary and construct Bayesian con-

fidence intervals. Such an approach would greatly reduce the number of covariance parameters to be estimated. For instance, an AR(1) process with parameter ϕ might be assumed, in which case the correlation matrix P would have the form:

$$P(\phi) = \begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{T-1} \\ \phi & 1 & \phi & \cdots & \phi^{T-2} \\ \phi^2 & \phi & 1 & \cdots & \phi^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \cdots & 1 \end{pmatrix}.$$

One might attempt to estimate ϕ based on the residuals $\{z_j - \hat{\eta}_0(t_j)\}_{j=1}^T$ from a preliminary spline smooth under the independence assumption, and then use $P(\hat{\phi})$ in a subsequent spline smooth to obtain $\hat{\eta}(t)$ and var $[\hat{\eta}(t)]$. Alternatively, one might attempt to estimate ϕ while simultaneously performing the spline smooth by employing a mixed effects smoothing spline model [70, 71, 75, 7, 40]. Similarly, one could assume more complex stationary processes such as AR(p), MA(q), or ARIMA(p,d,q) and use the described approaches. However, these methods are only feasible when there are just a few extra parameters to estimate. Estimating a P that is close to the truth will most likely require too many extra parameters under the stationarity assumption, especially when faced with complicated data such as Fisher transformed correlations.

These various approaches based on Bayesian confidence interval construction are fundamentally flawed and are expected to give unreliable or uninterpretable estimates of var $[\hat{\eta}(t)]$. Thus, we advocate a novel bootstrap-based method for constructing confidence intervals for a correlation function.

3.3.2 Bootstrap-based Confidence Intervals for a Functional Correlation

Parametric estimation of var $[\hat{\eta}(t)]$ is not reliable or recommended when the complex covariance structure of $(z_1, \ldots, z_T)'$ is not known in advance. The challenges discussed above suggest that a nonparametric procedure for estimating confidence intervals for $\rho(t)$ may be most appropriate to preserve the covariance structure of the data. The well-known bootstrap, introduced by Efron [16, 17], provides an excellent framework for robust and reliable

inference for a correlation function. The bootstrap should not be viewed as a last-resort; it is often a dependable first choice when one is not confident in making parametric assumptions about the data. Efron demonstrated that the bootstrap method is preferable to other resampling schemes. Among other advantages in our setting, a Monte Carlo evaluation of $\hat{\text{var}}\,[\hat{\eta}(t)]$ based on random bootstrap samples of the observed data converges to the nonparametric maximum likelihood estimate of $\text{var}\,[\hat{\eta}(t)]$ [17].

Bootstrap-based inference for Pearson correlations has been extensively explored in the non-functional setting, leading to some debate about its relative merits compared to parametric methods [12, 55, 65, 18, 56]. As Efron points out in this discussion, bootstrap and parametric methods provide nearly equivalent inferences about correlations when parametric assumptions are very close to the truth. Furthermore, to quote Efron: "the bootstrap is not intended to be a substitute for precise parametric results but rather a way to reasonably proceed when such results are unavailable" [18]. The unavailability of precise parametric results is exactly the challenge we face when constructing confidence intervals for a functional correlation, and thus we employ the bootstrap in our work.

Our novel procedure for constructing point-wise confidence intervals for $\rho(t)$, based on the large sample Gaussian distribution of Fisher transformed correlations, is carried out through the following steps. For simplicity of notation below, denote the pairs of observed functions as $\mathbf{U}_i = \mathbf{U}_i(t) = [X_i(t),\, Y_i(t)]'$, where $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are independent with a common distribution F. Let F_n be the empirical distribution function of $\mathbf{U}_1, \ldots, \mathbf{U}_n$ that puts mass $\frac{1}{n}$ on each $\mathbf{U}_i$.

1. Based on the data $\mathbf{U}_1, \ldots, \mathbf{U}_n$ observed at times $t_1, \ldots, t_T$, obtain the estimator $\hat{\eta}_0(t)$ through cubic spline smoothing of the pairs $\{t_j, z_j\}_{j=1}^T$ using the proposed method of Section 3.2.

2. Let $\mathbf{U}_1^*, \ldots, \mathbf{U}_n^*$ be i.i.d. samples from F_n (i.e., random sampling with replacement from the set $\{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$). The sample $\{\mathbf{U}_1^*, \ldots, \mathbf{U}_n^*\}$ is then used to calculate the sample correlation coefficient r_j^* at each time t_j, and Fisher's correlation transformation is employed to obtain z_j^*, $j = 1, \ldots, T$. Finally, Step 1 is repeated to calculate a bootstrap estimator $\hat{\eta}_1^*(t)$ through the cubic spline smoothing of the pairs $\left\{t_j, z_j^*\right\}_{j=1}^T$.

3. Repeat Step 2 independently a large number B of times to obtain the bootstrap estimators $\{\hat{\eta}_b^*(t)\}_{b=1}^B$.
4. Let $\hat{\eta}_\cdot^*(t) = \frac{1}{B}\sum_{b=1}^B \hat{\eta}_b^*(t)$, and compute the sample variance of the B estimators:

$$\hat{\text{var}}\left[\hat{\eta}^*(t)\right] = \frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{\eta}_b^*(t) - \hat{\eta}_\cdot^*(t)\right]^2. \tag{3.15}$$

5. At a fixed time point t, the standard percentile-based $(1-\alpha/2)\%$ bootstrap confidence limits for $\eta(t)$ would be given by the α and $(1-\alpha/2)$ quantiles of $\{\hat{\eta}_1^*(t), \ldots, \hat{\eta}_B^*(t)\}$. However, since the large sample distribution of Fisher transformed correlations is symmetric and Gaussian, we instead compute the $(1-\alpha)\%$ bootstrap CI for $\eta(t)$ as

$$\hat{\eta}_0(t) \;\pm\; z_{1-\alpha/2}\cdot\sqrt{\hat{\text{var}}\left[\hat{\eta}^*(t)\right]}\,, \tag{3.16}$$

   where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. For example, using our method, a 95% point-wise CI for $\eta(t)$ is

$$\left\{\hat{\eta}_0(t) - 1.96\cdot\sqrt{\hat{\text{var}}\left[\hat{\eta}^*(t)\right]}\,,\; \hat{\eta}_0(t) + 1.96\cdot\sqrt{\hat{\text{var}}\left[\hat{\eta}^*(t)\right]}\right\}. \tag{3.17}$$

6. Lastly, if we denote the lower and upper $(1-\alpha)\%$ confidence limits for $\eta(t)$ as $L_{\eta(t)}$ and $U_{\eta(t)}$, respectively, then the $(1-\alpha)\%$ point-wise confidence interval for $\rho(t)$ is obtained using the inverse Fisher transformation of Equation (3.5):

$$\left[\frac{\exp\left[2L_{\eta(t)}\right]-1}{\exp\left[2L_{\eta(t)}\right]+1}\,,\; \frac{\exp\left[2U_{\eta(t)}\right]-1}{\exp\left[2U_{\eta(t)}\right]+1}\right] \tag{3.18}$$

   We note that the total number of unique bootstrap samples is $B = \begin{pmatrix} 2n-1 \\ n \end{pmatrix}$. In practice, though, one does not use all $B = \begin{pmatrix} 2n-1 \\ n \end{pmatrix}$ samples; for example, a sample size as small as $n = 10$ subjects gives a maximum of $B = 92378$. There is no established rule for how large B should be in the functional correlation setting. We suggest that B be chosen large enough such that $\hat{\text{var}}\left[\hat{\eta}^*(t)\right]$ stabilizes to a constant (within a reasonable precision level). Since $\lim_{B\to\infty}\hat{\text{var}}\left[\hat{\eta}^*(t)\right]$ is the nonparametric MLE of $\text{var}\left[\hat{\eta}(t)\right]$, and because $\text{var}\left[\hat{\eta}(t)\right]$ is not constant over time, B must be chosen large enough such that $\hat{\text{var}}\left[\hat{\eta}^*(t)\right]$ is

stable at each time t. This recommendation may appear to give extremely large values of B; however, in small empirical studies, we found $B \approx 1000$ to be quite adequate. In fact, for our motivating application, $B = 250$ was large enough to stabilize the estimated variance; larger values of B did not noticeably change $\hat{\text{var}}\,[\hat{\eta}^*(t)]$ at each time t.

3.4 APPLICATION: DELTA EEG POWER AND HF-HRV

The first goal of our motivating study [58] addressed the time-varying correlation between delta EEG power and HF-HRV in midlife women during individual NREM periods. Within a given NREM period, delta EEG power and HF-HRV values for each of the $n = 197$ participants were obtained at each of the $T = 582$ relative time points. The novel method described in Section 3.2 was employed to estimate the functional correlation between these two variables within each NREM period. Point-wise 95% confidence intervals were constructed for the correlation functions by applying the novel bootstrap-based method of Section 3.3.2, using $B = 250$ random bootstrap samples. In the results that follow, $\hat{\rho}_1(t)$, $\hat{\rho}_2(t)$ and $\hat{\rho}_3(t)$ denote the estimated correlation functions during NREM-1, NREM-2 and NREM-3, respectively.

The functional correlations between delta EEG power and HF-HRV during the first three NREM periods for the full sample are depicted in Figure 1. Data appear in relative time as opposed to absolute (clock) time in order to account for inter-individual differences in NREM period length. On the relative time scale, $t = 0$ designates the time at which the maximum in delta EEG power occurs for all participants, represented in the figure by dotted vertical lines. The top row displays the mean delta EEG power profile (natural-log-transformed) as a function of relative time. The middle row displays the mean normalized HF-HRV profile (square-root-transformed) as a function of relative time. The bottom row reveals the time-varying correlation between delta EEG power and HF-HRV; solid lines represent the estimated time-varying correlation functions, and the shaded areas represent point-wise 95% confidence intervals for the correlation functions. A correlation is deemed point-wise significant if its 95% confidence interval does not include zero.

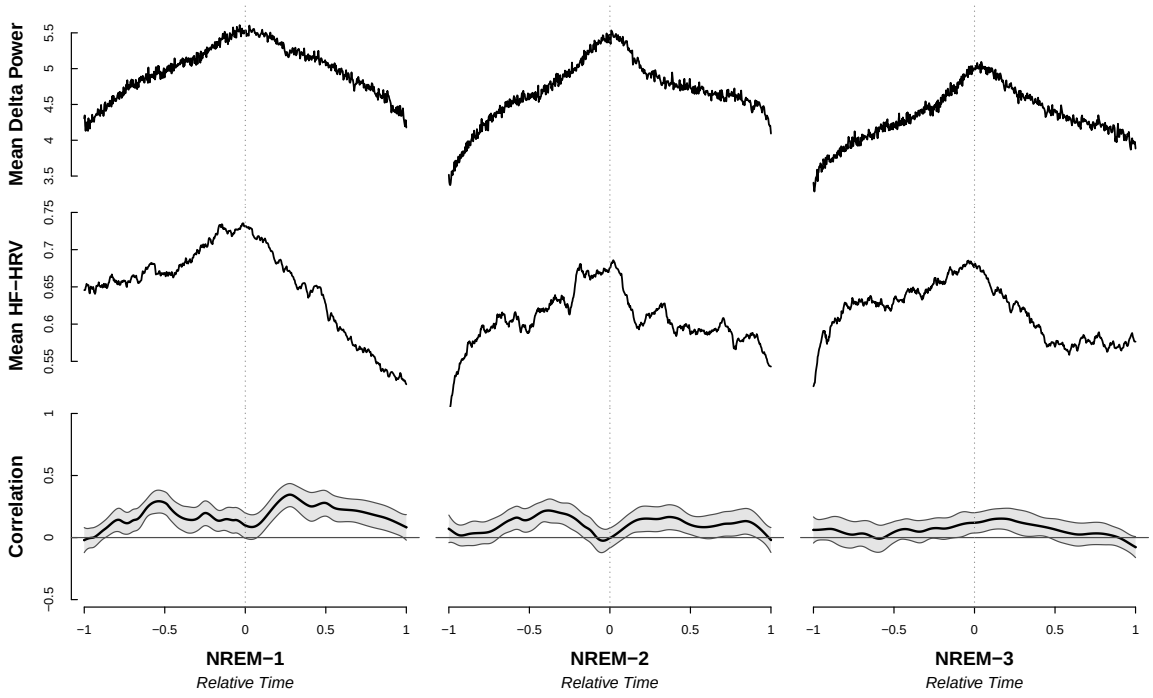

Figure 1: Delta EEG power and HF-HRV: Whole Sample of Participants

How are delta EEG power and HF-HRV correlated over time within NREM periods for the sample as a whole? As shown in Figure 1, the estimated correlation profile for delta EEG power and HF-HRV during NREM-1 is bimodal; $\hat{\rho}_1(t)$ increases from zero at sleep onset to a local maximum of $\hat{\rho}_1 = +0.29$ with 95% CI $(0.20, 0.38)$. The estimated correlation function then decreases in magnitude until $t = 0$, where $\rho_1(t)$ is no longer significant. Following $t = 0$, the estimated functional correlation between these two physiological parameters increases to $\hat{\rho}_1 = +0.35$ with 95% CI $(0.25, 0.43)$ and remains significantly positive until the end of the first NREM period, $\hat{\rho}_1(t)$ approaches zero. During NREM-1, $\rho_1(t)$ is point-wise significant for 86% of the time before the peak in delta power and 90.8% of the time after peak delta power.

During NREM-2, the time-varying correlation between delta EEG power and HF-HRV also has two peaks; $\hat{\rho}_2(t)$ rises to a maximum value $\hat{\rho}_2 = +0.22$ with 95% CI $(0.12, 0.31)$ and subsequently decreases to zero when the maximum in delta EEG power occurs. After $t = 0$, $\hat{\rho}_2(t)$ increases to $\hat{\rho}_2 = +0.17$ with 95% CI $(0.06, 0.26)$ and decreases back to zero by the end of NREM-2. The amount of time during which the correlation function is significantly positive decreased from NREM-1 to NREM-2; $\rho_2(t)$ is point-wise significant for 52.3% of time before $t = 0$ and 71.4% of time following $t = 0$. Qualitatively, $\hat{\rho}_2(t)$ crudely follows the same pattern as $\hat{\rho}_1(t)$, but the overall magnitude of $\hat{\rho}_2(t)$ is lower. In addition, the peaks in $\hat{\rho}_2(t)$ are much broader compared to $\hat{\rho}_1(t)$, and the drop in $\hat{\rho}_2(t)$ near $t = 0$ is much sharper.

By the third NREM period, the magnitude of the estimated time-varying correlation between delta EEG power and HF-HRV drops dramatically. $\hat{\rho}_3(t)$ is unimodal and only point-wise significant for 11% of the time before $t = 0$ and 46.2% of the time after $t = 0$. A mere maximum of $\hat{\rho}_3 = +0.15$ with 95% CI $(0.07, 0.23)$ is attained following the peak in delta EEG power.

4.0 COMPARING INDEPENDENT CORRELATION FUNCTIONS

4.1 INTRODUCTION

Our novel method for analyzing a single correlation function enables a researcher to explore and gain valuable insight into the time-varying relationship between two variables measured on the same subjects. Results from our motivating study [58] indicate that the correlation between delta EEG power and HF-HRV varies substantially over time within NREM periods. This finding has the potential clinical implication that, for an accurate assessment of cardiac autonomic tone during NREM sleep based on a discrete HRV epoch, the time at which you sample the epoch makes a difference. In addition, insomnia and sleep disordered breathing may have marked effects on the sleep-HRV correlation profiles. How do we detect a significant difference between the time-varying correlation functions from different types of people? We need a way to formally test whether correlation functions differ between independent groups of subjects.

The comparison of correlations in the non-functional data setting has been explored by many [13, 14, 49, 44, 53, 52]. When comparing correlations from independent samples of subjects, the work becomes easier, as the dependence between two correlations at any given time is known to be zero. In the functional data setting, though, no formal hypothesis test for the overall equivalence of two functional correlations exists. In developing a formal testing procedure to compare independent correlation functions, there are several challenges to consider:

1. Nonparametric curve estimation, such as spline smoothing, reduces variability at the expense of introducing some bias. We would like a test procedure which circumvents the

bias caused nonparametric curve estimation.

2. When dealing a functional correlation, the errors are correlated over time; a test that takes the correlation structure of the data into account is necessary.

3. Many existing procedures for functional data are based on the difference in mean functions from two independent sets of curves; however, we do not have replicates of correlation functions to form sets of curves. When testing for equivalence of two functional correlations, we need a procedure that can handle a single curve per group.

4. It is desirable to have a procedure that incorporates dimensionality reduction, as testing too many dimensions accumulates large stochastic noise and decreases the ability to detect a significant difference.

Much work has been done in the functional data setting to test for the equivalence of two nonparametric functions, and we explored various routes in an attempt to gain insight into testing correlation functions. Liu and Wang (2004) [41] reviewed several hypothesis testing procedures for smoothing spline models, including approximate LMP, GML, and GCV tests. Some methods can be extended to compare two curves. However, these methods are sensitive to the independent errors assumption and would have to be modified to handle correlated data. In addition, they are all subject to the bias induced by spline smoothing.

Guo (2002) [27] generalized the GML test to the mixed effects smoothing spline analysis of variance (SSANOVA) model, which is a more suitable for correlated data. One can parameterize such a model to make the difference between two curves a functional component, and a likelihood ratio test for the significance of that functional component can be performed. But, the bias inherent in spline smoothing still remains an issue. Bootstrap-based and L^2-based tests for the equivalence of independent curves were proposed by Zhang et al. (2010) [74]. These methods assume that curves are observed without noise, an assumption that may be far from the truth. Furthermore, the methods require replicates of curves.

A new two-sample test procedure for functional data was introduced by Hall and Van Keilegom (2007) [29]. The procedure actually tests for the equality of distributions of curves, as opposed to the equality of mean functions. The procedure utilizes nonparametric curve estimation, although the authors claim that the bias due to smoothing is minimal. However,

the asymptotic theory is based on the assumption that the number of replicated curves goes to infinity, and the test cannot even be performed with ony one replicate per group.

Lastly, Fan and Lin (1998) [20] provided a formal test for equivalence of two sets of curves based on the adaptive Neyman test introduced by Fan [19] in 1996. Their procedure does not use any nonparametric curve estimation, and it incorporates dimensionality reduction. Correlated errors can be accounted for, assuming the errors are stationary. However, a direct implementation of their procedures requires many curve replications. Nevertheless, the idea behind their test is quite attractive, as it avoids most of the challenges enumerated above.

To provide a procedure for testing the equivalence of two independent correlation curves, we developed a novel adaptive Neyman test, motivated by the ideas of Fan and Lin [20]. Our testing procedure does not inherit bias caused by nonparametric curve estimation, is able to handle dependent errors, automatically incorporates dimensionality reduction, and is adapted to the single curve per independent group setting provided when analyzing functional correlations.

4.2 METHOD

4.2.1 The Adaptive Neyman Hypothesis Test

Before discussing our novel hypothesis test for equivalence of two independent functional correlations, a review of the general adaptive Neyman test is provided in this section. This review follows the discussions in Fan (2006) [19], Fan and Lin (2008) [20], and Darling and Erdös (1956) [11]. More technical details and proofs can be found in those works.

Suppose $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I)$ is an n-dimensional normal random vector. A test of $H_0 : \boldsymbol{\mu} = 0$ versus $H_1 : \boldsymbol{\mu} \neq 0$ gives the maximum likelihood ratio test statistic $\|\mathbf{X}\|^2 = X_1^2 + \ldots + X_n^2$, which combines all n components of $\mathbf{X}$ into a single test statistic. Under the null hypothesis, $\|\mathbf{X}\|^2 \sim \chi_n^2$. For large n, the CLT gives the large sample distribution of $\|\mathbf{X}\|^2$ under the null hypothesis as $\|\mathbf{X}\|^2 \sim N(n, 2n)$; under the alternative hypothesis, the large sample distribution is $\|\mathbf{X}\|^2 \sim N\left(n + \|\boldsymbol{\mu}\|^2, 2n + 4\|\boldsymbol{\mu}\|^2\right)$. If we suppose that $\|\boldsymbol{\mu}_1\|^2 = o(n)$, the

power of the large sample test at the alternative $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ is approximately

$$1 - \Phi\left(\frac{z_{1-\alpha} - \|\boldsymbol{\mu}_1\|^2/\sqrt{2n}}{\sqrt{1 + 2\|\boldsymbol{\mu}_1\|^2/n}}\right) \approx 1 - \Phi\left(z_{1-\alpha} - \frac{1}{\sqrt{2n}}\|\boldsymbol{\mu}_1\|^2\right).$$

Even if $\|\boldsymbol{\mu}_1\|^2 \to \infty$, if $\|\boldsymbol{\mu}_1\|^2 = o(\sqrt{n})$, then the power of the test approaches the significance level α. This demonstrates that substantial noise can build up and cause the power to diminish when one tests too many dimensions of $\mathbf{X}$ [19, 20].

If one has prior knowledge that large contributions to $\|\boldsymbol{\mu}\|^2$ mainly come from the first m elements of $\boldsymbol{\mu}$, then it makes sense to consider a lower-dimensional problem and use the smaller vector $\mathbf{X}^{(m)} = (X_1, X_2, \ldots, X_m)'$. Neyman (1937) [46] proposed testing the m-dimensional subproblem, hence why it is referred to as the "Neyman test." To perform the Neyman test, one would use the standardized test statistic $\frac{1}{\sqrt{2m}}\sum_{j=1}^{m}\left(X_j^2 - 1\right)$. The power of the Neyman test at the alternative $\boldsymbol{\mu}_1^{(m)} = (\mu_{11}, \mu_{21}, \ldots, \mu_{m1})'$ is approximately equal to

$$1 - \Phi\left(z_{1-\alpha} - \frac{1}{\sqrt{2m}}\sum_{j=1}^{m}\mu_{j1}^2\right). \tag{4.1}$$

However, two issues arise: (1) m is typically not known in advance, so an appropriate estimator $\hat{m}$ must be found; subsequently, (2) the asymptotic distribution of $\frac{1}{\sqrt{2\hat{m}}}\sum_{j=1}^{\hat{m}}\left(X_j^2 - 1\right)$ under H_0 will not be the standard normal distribution. To help tackle these issues, we turn to an important theorem proven by Darling and Erdös [11] which solves, in an asymptotic form, the classical optional stopping problem. The theorem is stated as follows: Let $Y_1, Y_2, \ldots$ be independent random variables with mean 0, variance 1, and a uniformly bounded absolute third moment. Let $S_m = Y_1 + Y_2 + \ldots + Y_m$ and $U_n = \max_{1 \le m \le n} \frac{S_m}{\sqrt{m}}$. Then for $-\infty < t < \infty$:

$$\lim_{n\to\infty} \Pr\left\{U_n < \sqrt{2\log\log n} + \frac{\log\log\log n}{2\sqrt{2\log\log n}} + \frac{t}{\sqrt{2\log\log n}}\right\} = \exp\left(\frac{-e^{-t}}{2\sqrt{\pi}}\right). \tag{4.2}$$

The usefulness of this result will become clear in a few more steps.

Considering the approximate power of the Neyman test given in Equation (4.1), Fan [19] suggested the estimator

$$\hat{m} = \arg\max_{1 \le m \le n}\left\{\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(X_j^2 - 1\right)\right\}.$$

Notice that $\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(X_j^2-1\right)$ is an unbiased estimator of $\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\mu_{j1}^2$ in Equation (4.1). This leads to the *adaptive* Neyman test statistic

$$T_{AN}^{*}=\frac{1}{\sqrt{2\hat{m}}}\sum_{j=1}^{\hat{m}}\left(X_j^2-1\right)=\max_{1\le m\le n}\left\{\frac{1}{\sqrt{2m}}\sum_{j=1}^{m}\left(X_j^2-1\right)\right\}. \tag{4.3}$$

Compare the form of T_{AN}^{*} to the quantity $U_n=\max_{1\le m\le n}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}Y_j$ considered in the theorem by Darling and Erdos (1956) and appearing in Equation (4.2). Under the null hypothesis, $X_j^2\overset{\text{iid}}{\sim}\chi_1^2$ for $j=1,\ldots,m$, and hence $Y_j=\frac{1}{\sqrt{2}}\left(X_j^2-1\right)$ are independent random variables with mean 0, variance 1, and a uniformly bounded third moment. Now the connection between T_{AN}^{*} and U_n is clear: they are equivalent. The asymptotic distribution of the adaptive Neyman test statistic T_{AN}^{*} under the null hypothesis is the same extreme value distribution given by Equation (4.2). Fan [19] uses the test statistic

$$T_{AN}=\sqrt{2\log\log n}T_{AN}^{*}-\left[2\log\log n+0.5\log\log\log n-0.5\log(4\pi)\right] \tag{4.4}$$

for convenience. The finite sample distribution of T_{AN} under H_0 was calculated by simulation and reported in Fan and Lin [20], and the asymptotic distribution is given by

$$\lim_{n\to\infty}\Pr\left\{T_{AN}\le x\right\}=\exp\left[-\exp(-x)\right]\ ,-\infty<x<\infty. \tag{4.5}$$

4.2.2 Formal Test of Equivalence of Two Independent Correlation Functions

Having provided some necessary ingredients for an adaptive Neyman test, we now propose our novel hypothesis test of equivalence of two independent functional correlations. Suppose we have two samples of subjects: Sample 1 consists of n_1 subjects, Sample 2 consists of n_2 subjects, and the two samples are independent. The goal is to formally test whether the correlation functions from the two independent samples are equal. Consider the models

$$z_{1t}=\eta_1(t)+\varepsilon_{1t}\,,\qquad t=1,\ldots,T\,, \tag{4.6}$$

$$z_{2t}=\eta_2(t)+\varepsilon_{2t}\,,\qquad t=1,\ldots,T\,, \tag{4.7}$$

where z_{1t} and z_{2t} are Fisher-transformed correlations from Sample 1 and Sample 2, respectively, observed at times $t = 1, \ldots, T$. For each sample, the Fisher-transformed correlations are calculated from the observed bivariate data using the methods of Chapter 3. Since observation times are equally spaced and common to all subjects in both samples, we use the time index $t = 1, \ldots, T$ for ease of notation. In these models, $Ez_{1t} = \eta_1(t)$ and $Ez_{2t} = \eta_2(t)$, where $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are continuous functions on the interval $[1, T]$.

It is clear that a single transformed correlation z_{1t} (or z_{2t}) at any instantaneous time t asymptotically follows a univariate normal distribution. Joint normality of $\{z_{1t}\}_{t=1}^{T}$ (or $\{z_{2t}\}_{t=1}^{T}$) is not crucial because we will be using the Fourier transform of the data to carry out the test in the frequency domain. To ease technical arguments, we assume that the stochastic errors $\{\varepsilon_{1t}\}_{t=1}^{T}$ and $\{\varepsilon_{2t}\}_{t=1}^{T}$ are mean zero, stationary linear Gaussian processes, and the two processes are independent of each other; i.e., $\varepsilon_{1t} \perp \varepsilon_{2s}$ for all t and s. Let $\gamma_1(s,t) = \text{cov}(\varepsilon_{1s}, \varepsilon_{1t})$ and $\gamma_2(s,t) = \text{cov}(\varepsilon_{2s}, \varepsilon_{2t})$ be the autocovariance functions of ε_1 and ε_2, respectively. Since the errors are stationary, the autocovariance functions depend on s and t only through their difference $|t - s|$; we may write $\gamma_1(s,t) = \gamma_1(h)$ and $\gamma_2(s,t) = \gamma_2(h)$, where the lag $h = t - s$. In addition, the errors are homoscedastic: $\gamma_1(0) = \text{var}(\varepsilon_{1t})$ and $\gamma_2(0) = \text{var}(\varepsilon_{2t})$ are constants for all t. For purposes of spectral estimation, we require an absolute summability condition for the autocovariance functions:

$$\sum_{h=-\infty}^{\infty} |h|\,|\gamma_1(h)| < \infty \quad \text{and} \quad \sum_{h=-\infty}^{\infty} |h|\,|\gamma_2(h)| < \infty. \tag{4.8}$$

We denote the correlation functions of Samples 1 and 2 as $\rho_1(t)$ and $\rho_2(t)$, respectively. The goal is to formally test

$$H_0 : \rho_1(t) = \rho_2(t) \text{ for all } t \in [1, T], \qquad H_1 : \rho_1(t) \neq \rho_2(t) \text{ for some } t \in [1, T]. \tag{4.9}$$

Because $\rho(t)$ and $\eta(t)$ are one-to-one functions of each other, we can perform an equivalent test on the Fisher transformed scale:

$$H_0 : \eta_1(t) = \eta_2(t) \text{ for all } t \in [1, T], \qquad H_1 : \eta_1(t) \neq \eta_2(t) \text{ for some } t \in [1, T]. \tag{4.10}$$

To perform our adaptive Neyman test, we must transform the data $\{z_{1t}\}_{t=1}^{T}$ and $\{z_{2t}\}_{t=1}^{T}$ using the Fourier transform. This crucial step is performed for two reasons: (1) Salient signals

in the time domain data are captured mostly by the low-frequency Fourier components. Consequently, it provides the necessary prior that most of the important information lies in the first m dimensions of the data, allowing an adaptive Neyman test to be employed. (2) Correlated, stationary errors in the time domain are transformed into approximately independent Gaussian errors in the frequency domain [6, 61]. Independence is a necessary assumption for the distribution in Equation (4.2) to hold [11]. The frequency domain data is then used to perform the adaptive Neyman test.

Unlike many frequency domain tests which only consider the amplitude or power at each frequency, our procedure treats the real and imaginary parts of the Fourier transform at each frequency as separate components. Thus, our procedure takes advantage of information contained in both amplitudes and phases. Furthermore, many test procedures use an ANOVA statistic to detect differences in the power spectrum at each frequency, leading to a large number of test statistics. Corrections for multiple comparisons, such as the Bonferroni adjustment, must be used to preserve a family-wise significance level. In contrast, our procedure combines separate test statistics at different frequencies to give an overall powerful test. These features make the adaptive Neyman test more appealing than other procedures which are commonly employed in the frequency domain [20].

Our novel adaptive Neyman test for equivalence of independent correlation functions is carried out through the following steps:

1. Denote the discrete Fourier transform (DFT) of $\{z_{1t}\}_{t=1}^{T}$ by $d_1(\omega_j)$, and denote the DFT of $\{z_{2t}\}_{t=1}^{T}$ by $d_2(\omega_j)$. Compute $d_1(\omega_j)$ and $d_2(\omega_j)$ using the definitions

$$d_1(\omega_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{1t} e^{-2\pi i \omega_j t} \qquad d_2(\omega_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{2t} e^{-2\pi i \omega_j t}, \tag{4.11}$$

   and evaluate them at the Fourier frequencies $\omega_j = j/T$ for $j = 0, 1, \ldots, [T/2]$. Note that $d_1(0)$, $d_2(0)$, $d_1(1/2)$ and $d_2(1/2)$ are real numbers; at all other Fourier frequencies, $d_1(\omega_j)$ and $d_2(\omega_j)$ are complex numbers.
2. Let $\{Z_{1k}\}_{k=1}^{T}$ and $\{Z_{2k}\}_{k=1}^{T}$ be sets of the Fourier transformed data. Specifically, set $Z_{11} = d_1(0)$, $Z_{12} = \mathrm{Re}\left[d_1(\omega_1)\right]$, $Z_{13} = \mathrm{Im}\left[d_1(\omega_1)\right]$, $Z_{14} = \mathrm{Re}\left[d_1(\omega_2)\right]$, $Z_{15} = \mathrm{Im}\left[d_1(\omega_2)\right], \ldots$ such that the real and imaginary parts of $\{d_1(\omega_j)\}_{j=0}^{[T/2]}$ are separate elements of $\{Z_{1k}\}_{k=1}^{T}$. Arrange the elements of $\{Z_{2k}\}_{k=1}^{T}$ in the same way.

3. Using the models (4.6) and (4.7), we have the following relations:

$$d_1(\omega_j) = d_{\eta_1}(\omega_j) + d_{\varepsilon_1}(\omega_j) \qquad d_2(\omega_j) = d_{\eta_2}(\omega_j) + d_{\varepsilon_2}(\omega_j), \tag{4.12}$$

where $d_{\eta_1}(\omega_j)$, $d_{\eta_2}(\omega_j)$, $d_{\varepsilon_1}(\omega_j)$ and $d_{\varepsilon_2}(\omega_j)$ are the DFTs of $\eta_1(t)$, $\eta_2(t)$, ε_{1t} and ε_{2t}, respectively. Since ε_{1t} and ε_{2t} are mean zero, stationary linear Gaussian processes whose autocovariance functions satisfy Equation (4.8), the real and imaginary parts of $\{d_{\varepsilon_1}(\omega_j)\}_{j=0}^{[T/2]}$ and $\{d_{\varepsilon_2}(\omega_j)\}_{j=0}^{[T/2]}$ are approximately uncorrelated errors by Theorem C.4 of [61]. Using Theorem C.7 of [61], central limit theory gives that $\left\{\left(\mathrm{Re}\left[d_{\varepsilon_1}(\omega_j)\right], \mathrm{Im}\left[d_{\varepsilon_1}(\omega_j)\right]\right)'\right\}_{j=0}^{[T/2]}$ are asymptotically independent 2×1 normal vectors, and $\mathrm{Re}\left[d_{\varepsilon_1}(\omega_j)\right]$ and $\mathrm{Im}\left[d_{\varepsilon_1}(\omega_j)\right]$ are asymptotically independent for all j. The same results hold for $d_{\varepsilon_2}(\omega_j)$. Thus, the real and imaginary parts of $\{d_{\varepsilon_1}(\omega_j)\}_{j=0}^{[T/2]}$ and $\{d_{\varepsilon_2}(\omega_j)\}_{j=0}^{[T/2]}$ are approximately independent Gaussian errors.

4. Since Z_{1k} and Z_{2k} are approximately normally distributed, we obtain following new models:

$$Z_{1k} \sim N\left(F_1(k), \sigma_{1k}^2\right), \qquad k = 1, \ldots, T, \tag{4.13}$$

$$Z_{2k} \sim N\left(F_2(k), \sigma_{2k}^2\right), \qquad k = 1, \ldots, T. \tag{4.14}$$

$F_1(k)$ and $F_2(k)$ contain the real and imaginary parts of $d_{\eta_1}(\omega_j)$ and $d_{\eta_2}(\omega_j)$, arranged in the same manner as Z_{1k} and Z_{2k} in Step (2). $F_1(k)$ and $F_2(k)$ are frequency domain representations of $\eta_1(t)$ and $\eta_2(t)$; they contain the same information about the correlation functions. Z_{1k} and Z_{2k} are independent for all k, and the variance functions σ_{1k}^2 and σ_{2k}^2 are derived from the spectral densities $f_{\varepsilon_1}(\omega)$ and $f_{\varepsilon_2}(\omega)$ of the stationary processes ε_{1t} and ε_{2t}.

5. Because we do not know the spectral densities $f_{\varepsilon_1}(\omega)$ and $f_{\varepsilon_2}(\omega)$ in advance, they must be estimated. According to Property P4.6 in [61], any spectral density can be approximated arbitrarily close by the spectrum of an AR process. Thus, we represent the process ε_{1t} as an AR(p) process with the form

$$\varepsilon_{1t} = \sum_{k=1}^{p} \phi_k \varepsilon_{1t-k} + w_t, \tag{4.15}$$

where w_t is white noise with variance σ_w^2. We use a generalized least squares procedure to estimate the coefficients $\{\phi_k\}_{k=1}^p$ and variance σ_w^2, where the order p is selected by minimizing the Bayesian Information Criterion (BIC). A generalized linear model is to the Fisher transformed correlations $\{z_{1t}\}_{t=1}^T$ using k Fourier basis functions and an AR(p) error structure, and we obtain the BIC. This is repeated over a large grid of values for k and p, and the model producing the lowest BIC was selected. We obtain the "optimal" order p, the estimated AR coefficients $\hat{\phi}_1, \ldots, \hat{\phi}_p$ and the estimated variance $\hat{\sigma}_w^2$. The spectral density of ε_{1t} is then estimated by

$$\hat{f}_{\varepsilon_1}(\omega) = \frac{\hat{\sigma}_w^2}{\left|1 - \sum_{k=1}^p \hat{\phi}_k e^{-2\pi i \omega k}\right|^2} \tag{4.16}$$

The procedure is then repeated using $\{z_{2t}\}_{t=1}^T$ to obtain $\hat{f}_{\varepsilon_2}(\omega)$.

6. Having obtained $\hat{f}_{\varepsilon_1}(\omega)$ and $\hat{f}_{\varepsilon_2}(\omega)$, the variance function σ_{1k}^2 described in Step (4) is estimated by

$$\hat{\sigma}_{1k}^2 = \begin{cases} \hat{f}_{\varepsilon_1}(0) & k = 1 \\ \hat{f}_{\varepsilon_1}(\omega_{[k/2]})/2 & k \neq 1,\ T \text{ odd} \\ \hat{f}_{\varepsilon_1}(\omega_{[k/2]})/2 & k \neq 1 \neq T,\ T \text{ even} \\ \hat{f}_{\varepsilon_1}(\omega_{T/2}) & k = T,\ T \text{ even} \end{cases} \tag{4.17}$$

and $\hat{\sigma}_{2k}^2$ is obtained in the same way using $\hat{f}_{\varepsilon_2}(\omega)$.

7. Consider the hypotheses

$$H_0 : F_1(k) = F_2(k) \text{ for all } k = 1, \ldots, T, \quad H_1 : F_1(k) \neq F_2(k) \text{ for some } k = 1, \ldots, T. \tag{4.18}$$

Testing these is equivalent to testing the hypotheses in (4.9) or (4.10); the only difference is that these hypotheses are tested in the frequency domain, as opposed to in the time domain. Both domains contain the same information; rejecting the null in (4.18) gives the conclusion that $\rho_1(t) \neq \rho_2(t)$.

8. For $k = 1, \ldots, T$, define the standardized difference as

$$D_k = \frac{Z_{1k} - Z_{2k}}{\sqrt{\hat{\sigma}_{1k}^2 + \hat{\sigma}_{2k}^2}}. \tag{4.19}$$

Under H_0, $D_k \sim N(0,1)$, and $D_k^2 \sim \chi_1^2$. Since Z_{1k} and Z_{2k} in models (4.13) and (4.14) are independent for all k, the random variables $\{D_k^2\}_{k=1}^T$ are independent with mean 1 and variance 2. Thus, $\sum_{k=1}^m \left(D_k^2 - 1\right)/\sqrt{2}$ is the sum of m independent random variables with mean 0, variance 1, and uniformly bounded absolute third moment. Thus, we can use the adaptive Neyman procedure to test H_0.

9. Let the adaptive Neyman test statistic be

$$T_{AN}^* = \max_{1 \le m \le c_T} \left\{ \frac{1}{\sqrt{2m}} \sum_{k=1}^{m} \left(D_k^2 - 1\right) \right\}, \tag{4.20}$$

where c_T is some constant tending to infinity with $c_T \le T$, and define the standardized adaptive Neyman test statistic as

$$T_{AN} = \sqrt{2 \log\log c_T}\, T_{AN}^* - \left[2 \log\log c_T + 0.5 \log\log\log c_T - 0.5 \log(4\pi)\right]. \tag{4.21}$$

Comparing Equation (4.21) with Equation (4.4) reveals that the test statistics have the exact same form. We reject H_0 when T_{AN} is too large. The finite sample distribution of T_{AN} under H_0, calculated by simulation and reported in Fan and Lin (2008) [20], can be used to perform this test. As an example, the critical value for $\alpha = 0.05$ and $c_T = 100$ is 3.90; we reject H_0 when $T_{AN} \ge 3.90$.

Theorem 3 in Fan and Lin (1998) [20] provides valuable information. If the condition (4.8) holds and certain regularity conditions are met, then the asymptotic distribution of T_{AN} under H_0 is given by: $\Pr(T_{AN} < x) \to \exp\left[-\exp\left(-x\right)\right]$ as $T \to \infty$.

The conclusion to be drawn from the theorem is this: the effect of stationary errors on the null distribution is asymptotically negligible, and the impact of a chosen variance estimator, such as ours given in (4.17), is also asymptotically negligible. It is worth noting that the quantity c_T was introduced to allow one to perform their own dimensionality reduction before the adaptive Neyman test automatically does it. For instance, if one believes that only the first $k = 100$ components contain useful information and that the remaining Fourier

components are noise, then the power of the adaptive Neyman test may be further improved by setting $c_T = 100$. However, unless one has very good information to support such a belief, the value of c_T should be equal to or very close to T; otherwise, the test may lose its claimed significance level.

4.3 SIMULATIONS

In order to evaluate the performance of our adaptive Neyman test of equivalence of independent functional correlations, empirical significance and power calculations were implemented using simulations. Bivariate time-series data were randomly generated to mimic measurements from two independent samples of subjects. The data were simulated to give rise to known correlation functions which we observe with noise.

To investigate the empirical significance level of our test, we considered the following correlation function common to both samples of subjects, which is displayed in Figure 2:

$$\rho_1(t) = \rho_2(t) = \tanh\left[0.55\sin^2\left(2\pi t/T\right) - 0.1\right].$$

For power calculations, we used very similar correlation functions for the two samples of subjects, as shown in Figure 3:

$$\rho_1(t) = \tanh\left[0.6\sin^2\left(2\pi t/T\right) - 0.1\right],\ \rho_2(t) = \tanh\left[0.5\sin^2\left(2\pi t/T\right) - 0.1\right].$$

Three sets of balanced sample sizes were considered: $n = 25$, $n = 50$, and $n = 100$. Three values of T were investigated: $T = 100$, $T = 200$, and $T = 500$. In all 9 settings, the simulated bivariate normal data $[X_{ij}(t), Y_{ij}(t)]'$ for subject i in group j at fixed time t have the following properties:

- $X_{ij}(t) = 0.5 + 0.3\sin\left(\pi t/T - 0.3\right) + \delta_{ijX}(t)$
- $Y_{ij}(t) = 5 + 2.3\sin^3\left(\pi t/T - 0.1\right) + \delta_{ijY}(t)$
- $\delta_{ijX}(t)$ is a Gaussian AR(1) process with parameter $\phi = 0.5$ and variance 0.02
- $\delta_{ijY}(t)$ is a Gaussian AR(1) process with parameter $\phi = 0.5$ and variance 0.4
- $\mathrm{corr}\left[X_{ij}(t), Y_{ij}(t)\right] = \rho_j(t)$.

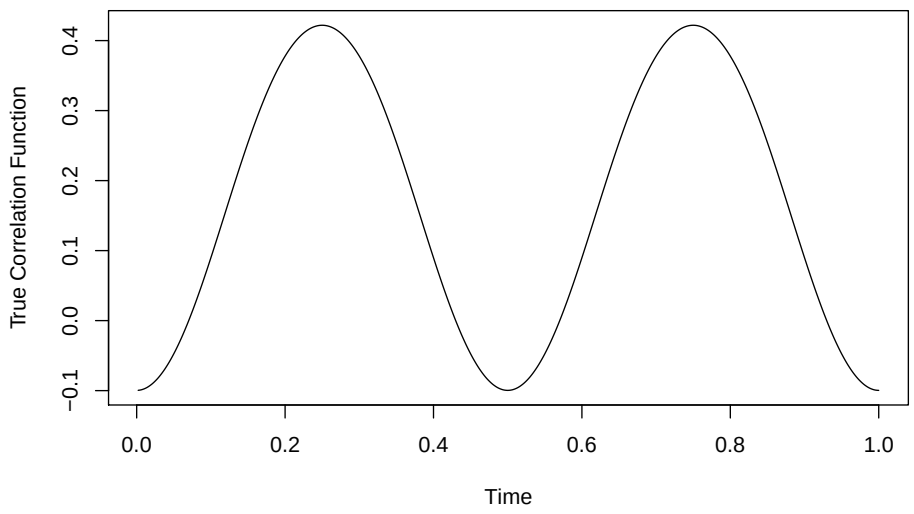

Figure 2: Independent Samples: Correlation function for evaluating empirical significance

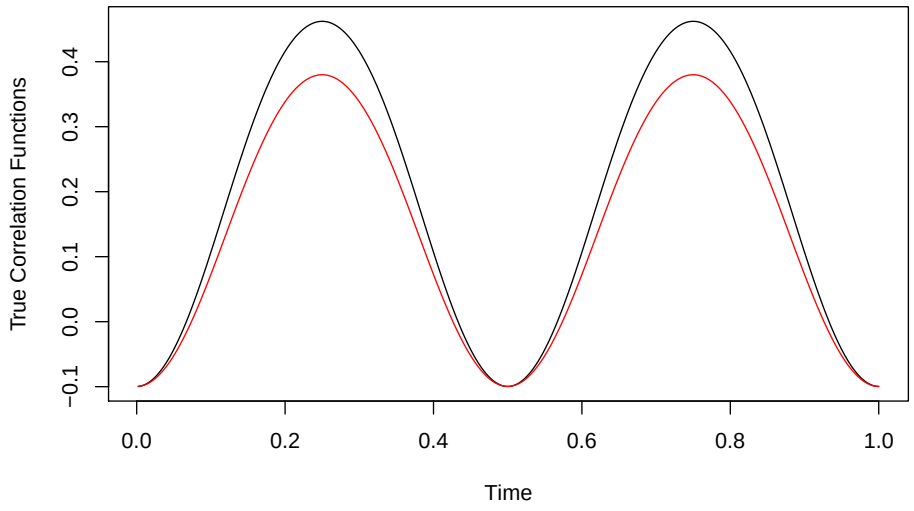

Figure 3: Independent Samples: Correlation functions for evaluating empirical power

A sample realization of $r_1(t)$ and $r_2(t)$ for $T = 200$ and $n = 50$ subjects per group, in the case where $\rho_1(t) \neq \rho_2(t)$, is shown in Figure 4. Black points are values of $r_1(t)$, and red points are values of $r_2(t)$; it is very difficult to see any difference in the underlying correlation functions based on the observed data.

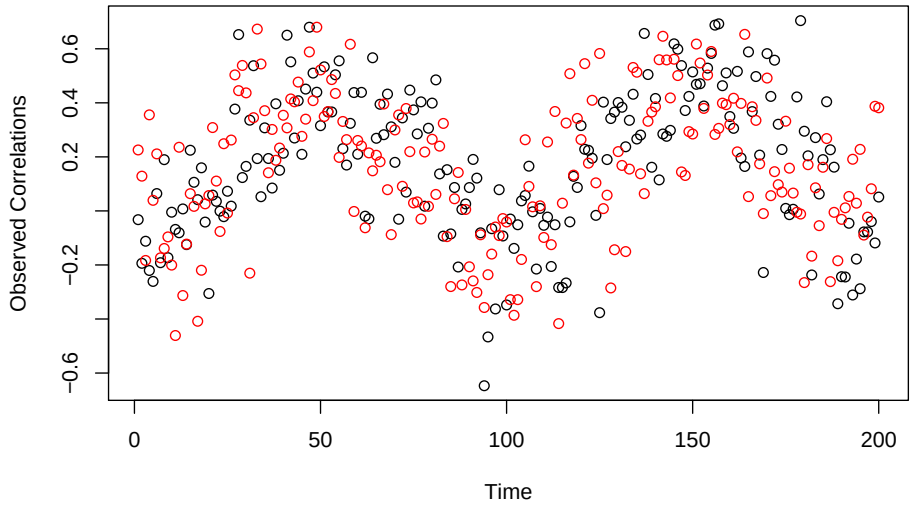

Figure 4: Independent Samples: Example of simulated data

In each setting, $N = 1000$ simulations were run, and our adaptive Neyman procedure was employed to test for equality of the independent correlation functions. Tests were performed at the $\alpha = 0.05$ level of significance. For comparison purposes, we also used two types of Bayesian simultaneous confidence bands based on penalized spline estimators to test for equality of correlation functions. The first was developed by Tatyana Krivobokova [37], and we refer to her method as "TK" in the simulation results. The second confidence band approach was introduced by Ciprian Crainiceanu [10], and we abbreviate his method as "CC." Lastly, we call our own testing procedure "AN." The results of the empirical significance and power calculations are given in Tables 1 and 2, respectively.

In 7 out of 9 settings of this simulation study, our adaptive Neyman test achieved a Type I error rate that was lower than the Type I error rates of the other two procedures.

Table 1: Comparing Independent Correlation Functions: Empirical Significance Results

		$T = 100$	$T = 200$	$T = 500$
$n = 25$	**AN**	0.130	0.069	0.035
	CC	0.097	0.108	0.187
	TK	0.110	0.118	0.196
$n = 50$	**AN**	0.127	0.053	0.036
	CC	0.089	0.127	0.215
	TK	0.111	0.144	0.216
$n = 100$	**AN**	0.101	0.061	0.038
	CC	0.107	0.093	0.197
	TK	0.122	0.118	0.202

Table 2: Comparing Independent Correlation Functions: Empirical Power Results

		$T = 100$	$T = 200$	$T = 500$
$n = 25$	**AN**	0.346	0.423	0.797
	CC	0.213	0.322	0.685
	TK	0.253	0.348	0.683
$n = 50$	**AN**	0.534	0.731	0.982
	CC	0.355	0.599	0.925
	TK	0.401	0.643	0.928
$n = 100$	**AN**	0.809	0.968	1.000
	CC	0.627	0.909	0.999
	TK	0.670	0.928	0.999

The empirical significance level of our test was markedly closer to the desired $\alpha = 0.05$ level for $T = 200$ and $T = 500$ than the other two tests. The CC and TK procedures based on simultaneous confidence bands had particularly poor performance for $T = 500$; all of their empirical significance levels were approximately 0.20. In contrast, the Type I error rate of our procedure was slightly lower than 0.05 for $T = 500$. The adaptive Neyman test performed fairly well at $T = 200$, while the empirical levels of the other two methods were roughly twice as large as the desired level. All three procedures had roughly the same Type I error rate for $T = 100$; our method outperformed the other two at $n = 100$, but not by much. Virtually all of their lowest empirical significance levels were achieved at the smallest value of T considered. In most cases, the Type I error rates of the CC and TK procedures grew larger as T increased; the empirical significance level of the adaptive Neyman test, however, inflated as T decreased for all three values of n considered.

With respect to power calculations, our procedure outperformed the other two in every case. Fan's motivation for choosing the estimator $\hat{m}$ in Equation (4.3) was based on the large sample power of the Neyman test; the observed power of our adaptive Neyman testing procedure provides some evidence that Fan's choice of estimator paid off. However, the high power of our procedure is not as impressive as it appears for low values of T in light of the fairly large Type I error rates. Nevertheless, this simulation study demonstrates that our adaptive Neyman test for equivalence of independent correlation functions is quite powerful and performs very well even when faced with a small n and true correlation functions that are extremely close to each other, particularly for large T.

4.4 APPLICATION: DELTA EEG POWER AND HF-HRV

The second aim of our motivating study [58] addressed whether time-varying relationships between delta EEG power and HF-HRV in midlife women differ as a function of sleep disordered breathing and insomnia. A total of 32 participants in our study met criteria for clinically significant sleep disordered breathing (SDB), 25 participants in our study met criteria for insomnia, and 146 participants did not exhibit symptoms of insomnia or SDB (i.e.,

146 participants were considered non-disorder controls). As a preliminary step in addressing the second goal of our study, the correlation functions during the first three NREM periods of sleep were analyzed separately for the SDB group, insomnia group, and the control group. Figures 5 and 6 show the correlation profiles for SDB participants and insomnia participants, respectively. Figure 7 reveals the correlation profiles for non-disorder controls.

In these figures, data appear in relative time as opposed to absolute (clock) time in order to account for inter-individual differences in NREM period length. On the relative time scale, $t = 0$ designates the time at which the maximum in delta EEG power occurs for all participants, represented in the figure by dotted vertical lines. The top row displays the mean delta EEG power profile (natural-log-transformed) as a function of relative time. The middle row displays the mean normalized HF-HRV profile (square-root-transformed) as a function of relative time. The bottom row reveals the time-varying correlation between delta EEG power and HF-HRV; solid lines represent the estimated time-varying correlation functions, and the shaded areas represent point-wise 95% confidence intervals for the correlation functions. A correlation is deemed point-wise significant if its 95% confidence interval does not include zero.

Does the time-varying relationship between delta EEG power and HF-HRV differ as a function of sleep apnea? As shown in Figures 5 and 7, compared to controls, the estimated correlation function for participants with sleep apnea is stronger and higher in magnitude for virtually the entire first NREM period, reaching a maximum of $\hat{\rho} = +0.62$ with 95% CI $(0.46, 0.73)$ near $t = -0.5$. During NREM-2, the correlation function appears higher for participants with sleep apnea compared to controls, although a formal test is needed to confirm whether the difference is statistically significant. During NREM-3, the functional correlation between delta EEG power and HF-HRV looks stronger in participants with clinically significant SDB compared to non-disorder controls, particularly before $t = 0$.

To formally test whether these observed differences are statistically significant, we employ the adaptive Neyman test for independent samples. For NREM-1, the standardized adaptive Neyman test statistic is $T_{AN} = 26.39$, giving $p - value < 0.00001$. For NREM-2, the adaptive Neyman test gives $T_{AN} = 2.77$ and $0.05 \leq p - value \leq 0.10$ For the last NREM period, the adaptive Neyman test gives $T_{AN} = 7.73$ and $p - value \approx 0.005$. Thus, even

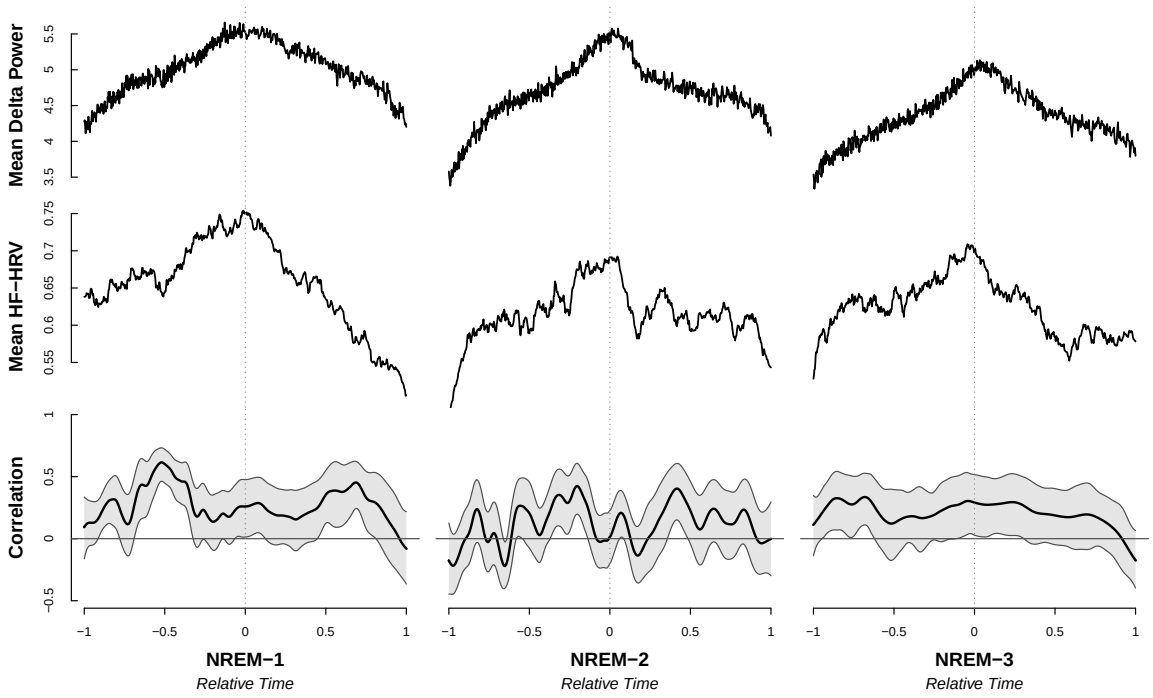

Figure 5: Delta EEG power and HF-HRV: Sleep disordered breathing (SDB) participants

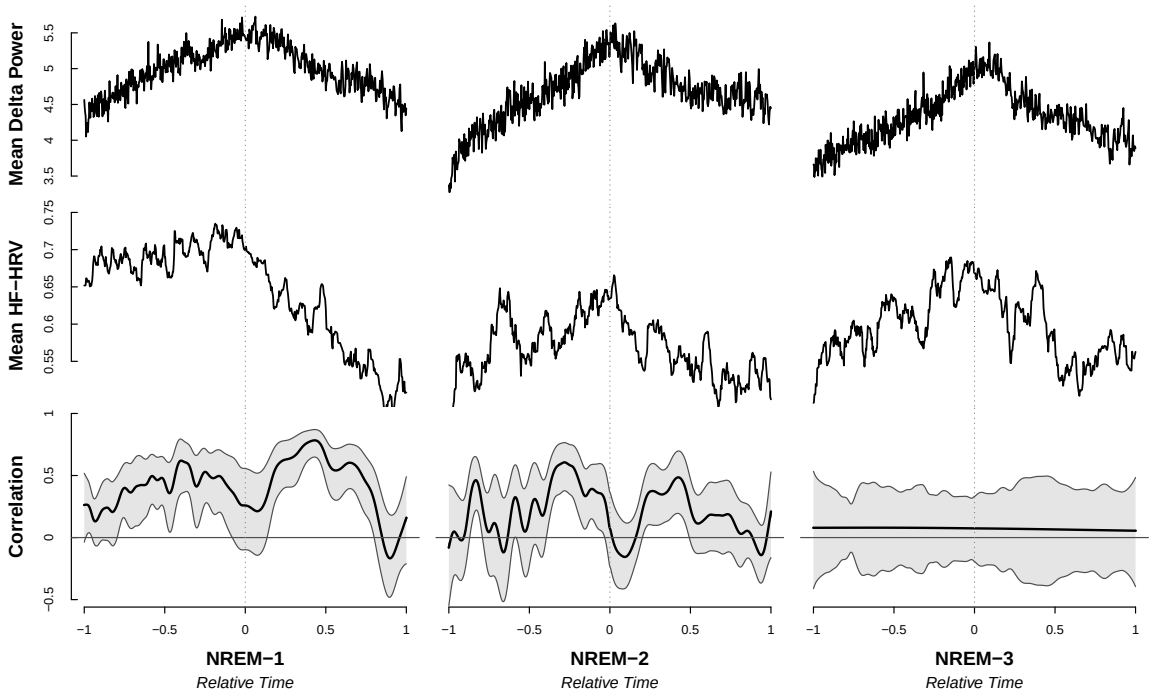

Figure 6: Delta EEG power and HF-HRV: Insomnia participants

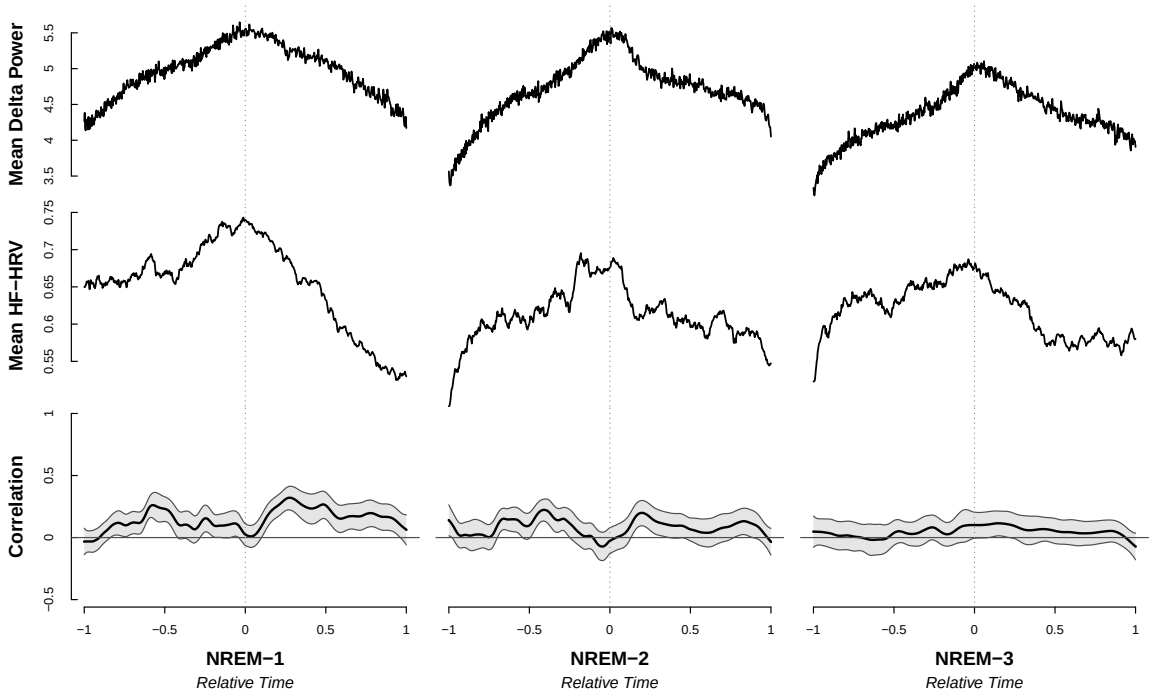

Figure 7: Delta EEG power and HF-HRV: Non-disorder control participants

if we correct for multiple comparisons, the difference in correlation functions is extremely statistically significant during NREM-1 and very significant during NREM-3, evincing that SDB participants have significantly higher time-varying correlations than control participants during those periods, while the difference in correlation functions is not statistically significant during NREM-2.

Does the relationship between delta EEG power and HF-HRV differ as a function of self-reported insomnia? As shown in Figures 6 and 7, the functional correlation between delta EEG power and HF-HRV during the first NREM period reaches an estimated maximum of $\hat{\rho} = +0.62$ with 95% CI $(0.35, 0.79)$ before $t = 0$ and an estimated maximum of $\hat{\rho} = +0.78$ with 95% CI $(0.65, 0.87)$ after $t = 0$ in the insomnia group, while the estimated correlation function for controls attains maximum values $\hat{\rho} = +0.26$ and $\hat{\rho} = +0.32$ before and after $t = 0$, respectively. The correlation between delta EEG power and HF-HRV remains noticeably stronger for the insomnia group compared to non-insomnia controls during NREM-2. The estimated correlation function reaches a maximum of $\hat{\rho} = +0.60$ with 95% CI $(0.38, 0.77)$ before $t = 0$ and a maximum of $\hat{\rho} = +0.49$ with 95% CI $(0.19, 0.70)$ after $t = 0$ in the self-reported insomnia group. In contrast, the maximum correlation attained in non-insomnia controls is $\hat{\rho} = +0.22$. Self-reported insomnia appears unrelated to the time-varying relationship between delta EEG power and HF-HRV during NREM-3.

Formal comparisons using the adaptive Neyman procedure give the test statistic $T_{AN} = 292.10$ with $p - value \ll 0.00001$ for NREM-1 and $T_{AN} = 46.91$ with $p - value < 0.00001$ for NREM-2. For the last NREM period, the adaptive Neyman test gives $T_{AN} = -2.41$ and $p - value > 0.5$. The dynamic coupling between delta EEG power and HF-HRV is extremely significantly stronger in participants with insomnia compared to non-insomnia controls during NREM-1 and NREM-2, but there is no significant difference between insomnia participants and controls during NREM-3.

5.0 COMPARING DEPENDENT CORRELATION FUNCTIONS

5.1 INTRODUCTION

In our motivating study [58], the methods presented in Chapter 3 were used to analyze the time-varying correlation between delta EEG power and HF-HRV during individual periods of NREM sleep in the sample of midlife women as a whole. Qualitative comparisons suggest that the time-varying correlation between delta EEG power and HF-HRV in our whole sample of midlife women may differ across different NREM periods of sleep. In order to conclude statistically significant differences, though, a formal test of equivalence must be employed.

One might be tempted to use our novel adaptive Neyman test for equivalence of independent correlation functions. But, that procedure would not be appropriate for testing whether the delta EEG-HRV correlation function changes significantly across different periods of NREM sleep, as the samples for the different NREM periods consist of data from the same group of subjects. As such, correlations will be dependent both within and between samples, and a procedure to test for equivalence of dependent functional correlations is required. However, in order to compare correlated functional correlations, a new methodology must be developed, as none exists.

Even in the non-functional data setting, testing for equality of two simple population correlation coefficients can be a complicated issue when the two sample correlation coefficients are computed from a single set of individuals. Traditional testing procedures for independent samples must be adjusted, or new methods must be constructed, when testing with dependent samples [13, 14, 49, 44, 53, 52]. Numerous tests have been proposed to compare correlated correlations between two pairs of variables measured on the same sample of subjects; however, their performances vary greatly based on many factors. To illustrate

some of the challenges faced when comparing correlation coefficients from a single sample, we will discuss some very simple testing procedures given by [13, 14].

Dunn and Clark [13, 14] describe several large sample testing procedures which assume a four-variate normal distribution for the observed data and employ Fisher's correlation transformation. For most procedures, the test statistic has the simple form $d = \sqrt{\frac{n-3}{2(1-c)}}\,(z_1 - z_2)$, where z_1 and z_2 are observed Fisher-transformed correlations from a common sample consisting of n subjects, and c is the asymptotic correlation coefficient between z_1 and z_2. The statistic $d \sim N\left(\sqrt{\frac{n-3}{2(1-c)}}\,(\eta_1 - \eta_2)\,, 1\right)$ with $Ez_1 = \eta_1$ and $Ez_2 = \eta_2$; the null hypothesis $H_0 : \eta_1 = \eta_2$ and the standard normal distribution are used to test whether the two population correlation coefficients are equal. Some of the tests explored by the authors are mentioned below:

1. *"Best" test.* The value of c is known and simply plugged into the test statistic d.
2. *Independent test.* It is assumed that $c = 0$; in other words, the test is carried out as if z_1 and z_2 were obtained from independent samples, each consisting of n subjects.
3. *Two-contrast test.* No assumptions about c are made. The sample is divided into two equal parts, and both z_1 and z_2 are computed from each half-sample. The four Fisher-transformed correlations are used to construct two univariate tests, and the acceptance region for the overall test is the intersection of the acceptance regions for the two univariate tests.
4. *Sample estimate test.* No assumptions about c are made. Instead, the value of c is estimated using the observed four-variate data, and this estimate of c is simply plugged into the test statistic d.

Asymptotic power curves were calculated for the "best," independent and two-contrast tests. Small sample ($n = 26$) power was obtained by simulation for the two-contrast and sample estimate tests. The power of the tests were evaluated at various values of c and δ, where $\delta = |\eta_1 - \eta_2|$.

Out of all the approaches for which asymptotic power was considered, the "best" test is the most powerful while maintaining the desired significance level for all values of c, as would be expected. However, the value of c is virtually never known, so the "best" test should not

be considered. When c is positive, the independent test is asymptotically more powerful than the two-contrast test for large values of δ, while the reverse is true for small values of δ. But, as c approaches zero from above, the independent test is asymptotically more powerful for all δ. When c is negative, the independent test has high asymptotic power; however, it does maintain its claimed level of significance. Thus, if one is willing to assume that c is positive and very small, then the independent test may actually be an appropriate choice for large n. Without any knowledge of c, though, the two-contrast test is recommended for large sample sizes, even though it has low power.

For small n, the sample estimate test may seem attractive due to its high power for most values of c and δ, but it is unacceptable due to its highly variable significance level. The two-contrast test has fairly low power for small n, although it does maintain its claimed significance level. Thus, the two-contrast test is recommended for small sample sizes. We conclude that the performance of tests of equality of simple population correlation coefficients may depend highly on sample size and unknown parameters when the correlations are calculated on the same subjects. Furthermore, without prior knowledge of the dependence between the sample correlation coefficients, one may have to settle for a test with low power in order to preserve the significance level.

5.2 METHOD

In developing a new methodology for comparing dependent functional correlations, we start with our adaptive Neyman procedure for comparing independent functional correlations. Similar to the construction of tests for correlated correlations in the non-functional data setting, we modify our adaptive Neyman test procedure for independent samples to allow for dependence between the two samples. Our method does not assume that the correlation structure of observed correlations is known in advance; it is estimated using the observed data. Briefly, when comparing two correlated functional correlations, we treat the observed Fisher-transformed correlations as a two-dimensional vector process $\boldsymbol{z}_t$. A novel state-space model for the observed vector process $\boldsymbol{z}_t$ is fit through maximum likelihood estimation to

quantify the error structure of $\boldsymbol{z}_t$ parametrically, under some very reasonable assumptions. The spectral matrix $f(\omega)$ of $\boldsymbol{z}_t$ encapsulates the correlation structure of observed Fisher-transformed correlations in the frequency domain, and it is estimated using the MLE's of parameters in the state-space model.

5.2.1 Formal Test of Equivalence of Two Dependent Correlation Functions

Suppose we have two samples of subjects: Sample 1 consists of n_1 subjects, Sample 2 consists of n_2 subjects, and the two samples are dependent. Often, the two samples consist of data from the same group of subjects observed under different conditions or over different time periods. To formally test for equivalence of correlation functions from the two samples, the univariate models (4.6) and (4.7) for independent samples are not applicable because the observed Fisher-transformed correlations are correlated over time both within and between samples. Consider the bivariate model

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \eta_1(t) \\ \eta_2(t) \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \qquad t = 1, \ldots, T, \tag{5.1}$$

where z_{1t} and z_{2t} are Fisher-transformed correlations from Sample 1 and Sample 2, respectively, observed at times $t = 1, \ldots, T$. We calculate the Fisher-transformed correlations for each sample using the methods of Chapter 3 and assume that observation times are equally spaced and common to all subjects in both samples. As in the previous chapter, $Ez_{1t} = \eta_1(t)$ and $Ez_{2t} = \eta_2(t)$, where $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are continuous functions on the interval $[1, T]$. Unlike before, though, the error terms ε_{1t} and ε_{2s} are not independent for all times t and s.

It is not crucial to assume joint normality of all observed Fisher-transformed correlations, for we will be using the Fourier transform to carry out the test in the frequency domain, as in Section 4.2.2. To ease technical arguments, here we assume that $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a mean zero, stationary linear Gaussian vector process. Let the 2×2 autocovariance matrix of $\boldsymbol{\varepsilon}_t$ be $\Gamma_\varepsilon(s,t) = \text{cov}(\boldsymbol{\varepsilon}_s, \boldsymbol{\varepsilon}_t)$. The elements of $\Gamma_\varepsilon(s,t)$ are the cross-covariance functions $\gamma_{11}(s,t) = \text{cov}(\varepsilon_{1s}, \varepsilon_{1t})$, $\gamma_{12}(s,t) = \text{cov}(\varepsilon_{1s}, \varepsilon_{2t})$, $\gamma_{21}(s,t) = \text{cov}(\varepsilon_{2s}, \varepsilon_{1t})$, and $\gamma_{22}(s,t) = \text{cov}(\varepsilon_{2s}, \varepsilon_{2t})$. Under the assumption of joint stationarity, these functions depend on s and t only through

their difference $|s-t|$, so we represent the autocovariance matrix as $\Gamma_\varepsilon(s,t) = \Gamma_\varepsilon(h)$, where the time lag $h = s - t$. Thus, we have

$$\Gamma_\varepsilon(h) = \begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) \end{pmatrix} = \begin{pmatrix} \operatorname{cov}(\varepsilon_{1,t+h}, \varepsilon_{1,t}) & \operatorname{cov}(\varepsilon_{1,t+h}, \varepsilon_{2,t}) \\ \operatorname{cov}(\varepsilon_{1,t}, \varepsilon_{2,t+h}) & \operatorname{cov}(\varepsilon_{2,t}, \varepsilon_{2,t+h}) \end{pmatrix}.$$

In the previous chapter for the adaptive Neyman test pertaining to independent samples, we had implicitly set $\gamma_{12}(h) = \gamma_{21}(h) = 0$. In the context of dependent samples, though, we do not make such an assumption; the stochastic errors $\{\varepsilon_{1t}\}_{t=1}^T$ and $\{\varepsilon_{2t}\}_{t=1}^T$ from the two samples are correlated with each other in some way. Given that the large-sample distribution of an observed Fisher-transformed correlation from a sample of size n is Gaussian with known variance $(n-3)^{-1}$, and continuing to assume that $\boldsymbol{\varepsilon}_t$ is a stationary linear Gaussian vector process, we can write $\Gamma_\varepsilon(0)$ as:

$$\Gamma_\varepsilon(0) = \begin{pmatrix} (n_1-3)^{-1} & \theta\,(n_1-3)^{-\frac{1}{2}}\,(n_2-3)^{-\frac{1}{2}} \\ \theta\,(n_1-3)^{-\frac{1}{2}}\,(n_2-3)^{-\frac{1}{2}} & (n_2-3)^{-1} \end{pmatrix} \equiv \Sigma_\varepsilon\,, \tag{5.2}$$

where $\theta = \operatorname{corr}(\varepsilon_{1t}, \varepsilon_{2t}) = \frac{\gamma_{12}(0)}{\sqrt{\gamma_{11}(0)}\sqrt{\gamma_{22}(0)}}$ is a constant parameter for all t, and $-1 \leq \theta \leq 1$. The notation $\Gamma_\varepsilon(0) \equiv \Sigma_\varepsilon$ is introduced for notational simplicity, since Σ_ε is the time-independent covariance matrix of $\boldsymbol{\varepsilon}_t$. The model (5.1) can be represented a bit more concisely and using a bivariate normal distribution:

$$\boldsymbol{z}_t \sim N_2\left(\boldsymbol{\eta}(t), \Sigma_\varepsilon\right)\,, \qquad t = 1, \ldots, T\,, \tag{5.3}$$

where $\boldsymbol{z}_t = (z_{1t}, z_{2t})'$, and $\boldsymbol{\eta}(t) = [\eta_1(t), \eta_2(t)]'$. Analogous to the independent samples case, we require absolute summability conditions on the auto- and cross-covariance functions for the purpose of spectral estimation:

$$\begin{array}{l} \sum_{h=-\infty}^{\infty} |h|\,|\gamma_{11}(h)| < \infty \\ \sum_{h=-\infty}^{\infty} |h|\,|\gamma_{12}(h)| < \infty \\ \sum_{h=-\infty}^{\infty} |h|\,|\gamma_{21}(h)| < \infty \\ \sum_{h=-\infty}^{\infty} |h|\,|\gamma_{22}(h)| < \infty \end{array}. \tag{5.4}$$

As in the previous chapter, we denote the correlation functions of Samples 1 and 2 by $\rho_1(t)$ and $\rho_2(t)$, respectively. We will formally test

$$H_0 : \rho_1(t) = \rho_2(t) \text{ for all } t \in [1, T], \qquad H_1 : \rho_1(t) \neq \rho_2(t) \text{ for some } t \in [1, T]. \tag{5.5}$$

Because $\rho(t)$ and $\eta(t)$ are one-to-one functions of each other, we may perform an equivalent test on the Fisher-transformed scale:

$$H_0 : \eta_1(t) = \eta_2(t) \text{ for all } t \in [1, T], \qquad H_1 : \eta_1(t) \neq \eta_2(t) \text{ for some } t \in [1, T]. \tag{5.6}$$

Our adaptive Neyman test of equivalence of dependent correlation functions is performed through the following steps:

1. First, we must transform the data $\{\boldsymbol{z}_t\}_{t=1}^T$ using the Fourier transform. Similar to the univariate DFT (4.11), we can perform the vector Discrete Fourier Transform (vector DFT) on the observed Fisher-transformed correlations. Denote the vector DFT of $\boldsymbol{z}_t$ by $\boldsymbol{d}(\omega_j) = [d_1(\omega_j), d_2(\omega_j)]'$. Compute $\boldsymbol{d}(\omega_j)$ using the definition

$$\boldsymbol{d}(\omega_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{z}_t e^{-2\pi i \omega_j t}, \tag{5.7}$$

and evaluate $\boldsymbol{d}(\omega_j)$ at the Fourier frequencies $\omega_j = j/T$ for $j = 0, 1, \ldots, [T/2]$. We may also write the vector DFT in terms of its real and imaginary parts:

$$\boldsymbol{d}(\omega_j) = \boldsymbol{d}_c(\omega_j) - i \cdot \boldsymbol{d}_s(\omega_j). \tag{5.8}$$

The terms $\boldsymbol{d}_c(\omega_j)$ and $\boldsymbol{d}_s(\omega_j)$ are referred to as cosine and sine transforms of $\boldsymbol{z}_t$, respectively, evaluated at frequency ω_j. They may be alternatively computed using the formulas

$$\boldsymbol{d}_c(\omega_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{z}_t \cos(2\pi\omega_j t) \quad \text{and} \quad \boldsymbol{d}_s(\omega_j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{z}_t \sin(2\pi\omega_j t). \tag{5.9}$$

2. Let $\{\boldsymbol{Z}_k\}_{k=1}^T$, where $\boldsymbol{Z}_k = (Z_{1k}, Z_{2k})'$, be vectors of the Fourier transformed data, arranged such that the cosine and sine transforms of the data are separate elements of $\{\boldsymbol{Z}_k\}_{k=1}^T$. Specifically, let $\boldsymbol{Z}_1 = \boldsymbol{d}_c(0)$, $\boldsymbol{Z}_2 = \boldsymbol{d}_c(\omega_1)$, $\boldsymbol{Z}_3 = \boldsymbol{d}_s(\omega_1)$, $\boldsymbol{Z}_4 = \boldsymbol{d}_c(\omega_2)$, $\boldsymbol{Z}_5 = \boldsymbol{d}_s(\omega_2), \ldots$, allowing the real and imaginary components of $\boldsymbol{d}(\omega_j)$ at each frequency to be distinct vectors.

3. Denote the vector DFTs of $\boldsymbol{\eta}(t)$ and $\boldsymbol{\varepsilon}_t$ by $\boldsymbol{d}_\eta(\omega_j)$ and $\boldsymbol{d}_\varepsilon(\omega_j)$, respectively. Due to the linearity of the DFT operator, we have

$$\boldsymbol{d}(\omega_j) = \boldsymbol{d}_\eta(\omega_j) + \boldsymbol{d}_\varepsilon(\omega_j).$$

Using the definitions above, we obtain the following relations:

$$\boldsymbol{d}(\omega_j) = \left[\boldsymbol{d}_{\eta,c}(\omega_j) - i \cdot \boldsymbol{d}_{\eta,s}(\omega_j)\right] + \left[\boldsymbol{d}_{\varepsilon,c}(\omega_j) - i \cdot \boldsymbol{d}_{\varepsilon,s}(\omega_j)\right],$$

$$\boldsymbol{d}(\omega_j) = \left[\boldsymbol{d}_{\eta,c}(\omega_j) + \boldsymbol{d}_{\varepsilon,c}(\omega_j)\right] - i \cdot \left[\boldsymbol{d}_{\eta,s}(\omega_j) + \boldsymbol{d}_{\varepsilon,s}(\omega_j)\right],$$

$$\boldsymbol{d}_c(\omega_j) = \boldsymbol{d}_{\eta,c}(\omega_j) + \boldsymbol{d}_{\varepsilon,c}(\omega_j),$$

$$\boldsymbol{d}_s(\omega_j) = \boldsymbol{d}_{\eta,s}(\omega_j) + \boldsymbol{d}_{\varepsilon,s}(\omega_j),$$

where $\boldsymbol{d}_{\eta,c}(\omega_j)$ and $\boldsymbol{d}_{\eta,s}(\omega_j)$ are the cosine and sine transforms of $\boldsymbol{\eta}(t)$, respectively, and $\boldsymbol{d}_{\varepsilon,c}(\omega_j)$ and $\boldsymbol{d}_{\varepsilon,s}(\omega_j)$ are the cosine and sine transforms of $\boldsymbol{\varepsilon}_t$, respectively. Because $\boldsymbol{\varepsilon}_t$ is a mean zero, stationary linear Gaussian vector process whose cross-covariance functions satisfy (5.4), $\{\boldsymbol{d}_{\varepsilon,c}(\omega_j)\}_{j=0}^{[T/2]}$ and $\{\boldsymbol{d}_{\varepsilon,s}(\omega_j)\}_{j=0}^{[T/2]}$ are sets of approximately uncorrelated vector errors by Theorem C.6 of [61]. Using Theorem C.7 in [61], central limit theory gives that $\left\{\left(\boldsymbol{d}'_{\varepsilon,c}(\omega_j), \boldsymbol{d}'_{\varepsilon,s}(\omega_j)\right)'\right\}_{j=0}^{[T/2]}$ are asymptotically independent 4×1 normal vectors. Thus, $\{\boldsymbol{d}_{\varepsilon,c}(\omega_j)\}_{j=0}^{[T/2]}$ and $\{\boldsymbol{d}_{\varepsilon,s}(\omega_j)\}_{j=0}^{[T/2]}$ are sets of approximately independent Gaussian vector errors.

4. Given the steps above, our data in the frequency domain approximately follows the bivariate normal distribution, and we can write the following new model:

$$\boldsymbol{Z}_k \sim N_2\left(\boldsymbol{F}(k), \Delta_k\right), \qquad k = 1, \ldots, T. \tag{5.10}$$

Here, $\boldsymbol{F}(k) = [F_1(k), F_2(k)]'$ is the frequency domain representation of $\boldsymbol{\eta}(t)$, containing the same information about the correlation functions. The elements of $\{\boldsymbol{F}(k)\}_{k=1}^{T}$ are arranged in the same fashion as $\{\boldsymbol{Z}_k\}_{k=1}^{T}$ in Step (2) above: $\boldsymbol{F}(k)$ is equal to either $\boldsymbol{d}_{\eta,c}(\omega_j)$ or $\boldsymbol{d}_{\eta,s}(\omega_j)$, depending on the value of k. The covariance matrix Δ_k is derived from the spectral matrix $f(\omega)$ of the stationary vector process $\boldsymbol{\varepsilon}_t$.

5. The spectral matrix $f(\omega)$ of the error process $\boldsymbol{\varepsilon}_t$ must be approximated in order to estimate the covariance matrix Δ_k in Step (4) above. Because we assume that $\boldsymbol{\varepsilon}_t$ is a stationary vector process, its spectrum can be approximated arbitrarily closely by the spectrum of a causal vector autoregressive (VAR) process [61]. Thus, we can obtain a parametric spectral estimator by modeling $\boldsymbol{\varepsilon}_t$ as a causal VAR(p) process with the representation

$$\boldsymbol{\varepsilon}_t = \sum_{k=1}^{p} \Phi_k \boldsymbol{\varepsilon}_{t-k} + \boldsymbol{w}_t, \tag{5.11}$$

where $\{\Phi_k\}_{k=1}^p$ are 2×2 transition matrices and $\boldsymbol{w}_t = (w_{1t}, w_{2t})'$ is a vector white noise process with covariance matrix Σ_w. The spectral matrix of such a process is given by

$$f(\omega) = \left[\Phi^{-1}\left(e^{-2\pi i\omega}\right)\right] \Sigma_w \left[\Phi^{-1}\left(e^{-2\pi i\omega}\right)\right]^{\star}, \tag{5.12}$$

where $\Phi^{-1}\left(e^{-2\pi i\omega}\right) = \left[I_2 - \sum_{k=1}^{p} \Phi_k e^{-2\pi i\omega k}\right]^{-1}$, and $\star$ denotes the complex conjugate transpose. A novel state-space model and the Kalman filter are employed to perform maximum likelihood estimation of the matrices $\{\Phi_k\}_{k=1}^p$ and the white noise covariance matrix Σ_w, where the "optimal" VAR order p is determined by minimizing the BIC. The state-space model and estimation procedure are described in the next subsection, so we defer such details for now. Having obtained the MLEs, the parametric spectral estimator is given by

$$\hat{f}(\omega) = \left[\hat{\Phi}^{-1}\left(e^{-2\pi i\omega}\right)\right] \hat{\Sigma}_w \left[\hat{\Phi}^{-1}\left(e^{-2\pi i\omega}\right)\right]^{\star}. \tag{5.13}$$

6. Once $\hat{f}(\omega)$ is obtained, an estimate of the covariance matrix Δ_k in Step (4) may be derived. We first consider the approximate joint distribution of the cosine and sine transforms of $\boldsymbol{\varepsilon}_t$:

$$\begin{pmatrix} \boldsymbol{d}_{\varepsilon,c}(\omega_j) \\ \boldsymbol{d}_{\varepsilon,s}(\omega_j) \end{pmatrix} = (d_{\varepsilon_1,c}(\omega_j), d_{\varepsilon_2,c}(\omega_j), d_{\varepsilon_1,s}(\omega_j), d_{\varepsilon_2,s}(\omega_j))' \sim N_4\left(\mathbf{0}, \Omega(\omega_j)\right). \tag{5.14}$$

The 4×4 covariance matrix $\Omega(\omega_j)$ can be written in terms of 2×2 matrices $C(\omega_j)$ and $Q(\omega_j)$:

$$\Omega(\omega_j) = \frac{1}{2} \begin{pmatrix} C(\omega_j) & -Q(\omega_j) \\ Q(\omega_j) & C(\omega_j) \end{pmatrix}, \tag{5.15}$$

for all $\omega_j \neq 0, \frac{1}{2}$; at the frequencies $\omega_j = 0$ and $\omega_j = \frac{1}{2}$, the covariance matrix is given by $2\Omega(\omega_j)$. The covariance matrix $\Omega(\omega_j)$ of the joint distribution of cosine and sine transforms of $\boldsymbol{\varepsilon}_t$ is related to the spectral matrix $f(\omega_j)$ (i.e., the covariance matrix of the vector DFT $\boldsymbol{d}_\varepsilon(\omega_j)$) by the following equation:

$$f(\omega_j) = C(\omega_j) - i \cdot Q(\omega_j). \tag{5.16}$$

Having obtained the 2×2 complex matrix $\hat{f}(\omega)$, it is evaluated at the Fourier frequencies $\{\omega_j\}_{j=0}^{[T/2]}$, and the real and imaginary parts of the spectral estimator are separated to give $\hat{C}(\omega_j)$ and $\hat{Q}(\omega_j)$. Note that $\frac{1}{2}\hat{C}(\omega_j)$ is the estimated covariance matrix of both $\boldsymbol{d}_{\varepsilon,c}(\omega_j)$ and $\boldsymbol{d}_{\varepsilon,s}(\omega_j)$ for all $\omega_j \neq 0, \frac{1}{2}$ (at the endpoints, the estimated covariance matrix is $\hat{C}(\omega_j)$). Thus, the covariance matrix Δ_k in Step (4) is estimated by:

$$\hat{\Delta}_k = \begin{cases} \text{Re}\left[\hat{f}(0)\right] & k = 1 \\ \frac{1}{2}\text{Re}\left[\hat{f}(\omega_{[k/2]})\right] & k \neq 1,\ T \text{ odd} \\ \frac{1}{2}\text{Re}\left[\hat{f}(\omega_{[k/2]})\right] & k \neq 1 \neq T,\ T \text{ even} \\ \text{Re}\left[\hat{f}(1/2)\right] & k = T,\ T \text{ even} \end{cases}. \tag{5.17}$$

7. Consider the hypotheses

$$H_0 : F_1(k) = F_2(k) \text{ for all } k = 1, \ldots, T, \quad H_1 : F_1(k) \neq F_2(k) \text{ for some } k = 1, \ldots, T. \tag{5.18}$$

Testing these is equivalent to testing the hypotheses in (5.5) or (5.6); the only difference is that these hypotheses are tested in the frequency domain, as opposed to in the time domain. Both domains contain the same information; rejecting the null in (5.18) gives the conclusion that $\rho_1(t) \neq \rho_2(t)$.

8. Using the model (5.10) and $\hat{\Delta}_k$ from (5.17), define the standardized difference as

$$D_k = \frac{Z_{1k} - Z_{2k}}{\sqrt{\hat{\Delta}_{k11} + \hat{\Delta}_{k22} - 2\hat{\Delta}_{k12}}}, \qquad k = 1, \ldots, T. \tag{5.19}$$

Under H_0, D_k approximately follows the standard normal distribution, and $D_k^2 \sim \chi_1^2$. The random variables $\{D_k^2\}_{k=1}^T$ are independent with mean 1 and variance 2. Thus, $\sum_{k=1}^m (D_k^2 - 1)/\sqrt{2}$ is the sum of m independent random variables with mean 0, variance

1, and uniformly bounded absolute third moment. We now have the necessary ingredients to use the adaptive Neyman procedure to test H_0.

9. Let the adaptive Neyman test statistic be

$$T^*_{AN} = \max_{1 \leq m \leq c_T} \left\{ \frac{1}{\sqrt{2m}} \sum_{k=1}^{m} \left(D_k^2 - 1 \right) \right\}, \tag{5.20}$$

where c_T is some constant tending to infinity with $c_T \leq T$, and define the standardized adaptive Neyman test statistic as

$$T_{AN} = \sqrt{2 \log \log c_T} T^*_{AN} - \left[2 \log \log c_T + 0.5 \log \log \log c_T - 0.5 \log(4\pi) \right]. \tag{5.21}$$

We reject H_0 when T_{AN} is too large. The finite sample distribution of T_{AN} under H_0, calculated by simulation and reported in Fan and Lin (1998) [20], can be used to perform the test.

5.2.2 State-Space Model for Dependent Correlation Functions

In the adaptive Neyman test procedure for comparing correlation functions from dependent samples, the spectral matrix $f(\omega)$ of the Gaussian vector error process $\boldsymbol{\varepsilon}_t$ in model (5.1) had to be estimated. The process $\boldsymbol{\varepsilon}_t$ was represented as a causal VAR(p) process, but the procedure used to estimate the VAR parameters $\{\Phi_k\}_{k=1}^p$ and Σ_w was only briefly mentioned. In this section, we describe in detail how such estimates are obtained using state-space modeling and maximum likelihood estimation. As no procedures exist for estimating the bivariate error structure of correlations observed over time from two dependent samples, our state-space approach is truly novel.

State-space models, also known as dynamic linear models (DLMs), are extremely useful in analyzing multivariate time series. Such models consist of two components: a state equation and an observation equation. The state equation represents some process that we do not observe directly; instead, we observe a linearly transformed version of it with noise added, and this observed process is represented by the observation equation. There exists an equivalence between stationary VAR models and stationary state-space models [61]; thus, a state-space formulation of our model (5.1) could be quite useful.

Consider the bivariate regression model

$$\boldsymbol{z}_t = \Gamma \boldsymbol{u}_t + \boldsymbol{\varepsilon}_t \,, \tag{5.22}$$

where $\boldsymbol{z}_t$ is an observed two-dimensional vector process, $\boldsymbol{u}_t = (u_{t1}, \ldots, u_{tr})'$ are r regressors which may or may not depend on time, Γ is a $2 \times r$ matrix of regression parameters, and $\boldsymbol{\varepsilon}_t$ is a two-dimensional VAR(p) process. This model could be fit to the observed Fisher-transformed correlations $\boldsymbol{z}_t = (z_{1t}, z_{2t})'$ specified in (5.1) in order to estimate the VAR(p) parameters of the error process $\boldsymbol{\varepsilon}_t$, where $\hat{\Gamma}\boldsymbol{u}_t$ would be an estimate of $[\eta_1(t), \eta_2(t)]'$ for suitably chosen regressors $\{\boldsymbol{u}_t\}_{t=1}^T$. We can fit the bivariate regression model (5.22) by first putting it into state-space form, where the state equation and observation equation are given by

$$\boldsymbol{x}_{t+1} = \Phi \boldsymbol{x}_t + \Psi \boldsymbol{w}_t \,,\ t = 0, 1, \ldots, T \,, \text{ and} \tag{5.23}$$

$$\boldsymbol{z}_t = \Gamma \boldsymbol{u}_t + A\boldsymbol{x}_t + \boldsymbol{w}_t \,,\ t = 1, \ldots, T \,, \tag{5.24}$$

respectively. In the state equation, $\boldsymbol{x}_t$ is a $2p$-dimensional unobserved process, the initial state $\boldsymbol{x}_0 \sim N_{2p}(\boldsymbol{\mu}_0, \Sigma_0)$, Φ is a $2p \times 2p$ matrix, Ψ is a $2p \times 2$ matrix, $\boldsymbol{w}_t \overset{\text{iid}}{\sim} N_2(\boldsymbol{0}, \Sigma_w)$, and $\boldsymbol{w}_t$ is independent of $\boldsymbol{x}_0$. In the observation equation, $\boldsymbol{z}_t$ is 2-dimensional, and $A = [I_2, 0, \cdots, 0]$ is a $2 \times 2p$ matrix. Comparing the regression model to the observation equation, we notice that $\boldsymbol{\varepsilon}_t = A\boldsymbol{x}_t + \boldsymbol{w}_t$ is the VAR(p) error process. The matrices Φ and Ψ are defined as

$$\Phi = \begin{bmatrix} \Phi_1 & I_2 & 0 & \cdots & 0 \\ \Phi_2 & 0 & I_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi_{p-1} & 0 & 0 & \cdots & I_2 \\ \Phi_p & 0 & 0 & \cdots & 0 \end{bmatrix} \quad \text{and} \quad \Psi = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_p \end{bmatrix}. \tag{5.25}$$

Our goals are to determine the "optimal" VAR order p and to estimate $\{\Phi_k\}_{k=1}^p$ and Σ_w based on observed Fisher-transformed correlations. Recall that Σ_ε, the time-independent covariance matrix of $\boldsymbol{\varepsilon}_t$ given by equation (5.2), contains known information; the state-space formulation allows us to use the known information when fitting the model. For simplicity, assume that Samples 1 and 2 consist of the same number of subjects (or the same subjects),

such that $n_1 = n_2 = n$. Then, we have $\Sigma_\varepsilon = (n-3)^{-1} \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}$, where $-1 \leq \theta \leq 1$. For a VAR(1) error process, the white noise covariance matrix can be written as $\Sigma_w = \Sigma_\varepsilon - \Phi_1 \Sigma_\varepsilon \Phi_1'$. Thus, for $p = 1$, we see that Σ_w is a function of 5 parameters: the 4 elements of Φ_1 and θ. If we did not take advantage of the known information about Σ_ε, then Σ_w would be a function of 7 parameters. Furthermore, we would be misrepresenting the true large-sample variance of Fisher-transformed correlations by estimating 3 virtually unrestricted parameters for Σ_ε.

For the general VAR(p), we can represent Σ_w in terms of the matrices Φ and Σ_ε. First, we define the $2p \times 2p$ matrices

$$\tilde{\Sigma}_w = \begin{bmatrix} \Sigma_w & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \text{ and } \tilde{\Sigma}_\varepsilon = \begin{bmatrix} \Sigma_\varepsilon & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{5.26}$$

Then, $\tilde{\Sigma}_w$ and $\tilde{\Sigma}_\varepsilon$ are related by the equation

$$\text{vec}(\tilde{\Sigma}_w) = \left[I_{(2p)^2} - \Phi \otimes \Phi \right] \text{vec}(\tilde{\Sigma}_\varepsilon), \tag{5.27}$$

where *vec* is the stack operator (stacking the elements of a matrix into a vector), and $\otimes$ denotes the Kronecker product. The elements of Σ_w can be extracted from $\text{vec}(\tilde{\Sigma}_w)$ by keeping track of their order.

Notice that Σ_w is a function of Φ and θ; once the estimates $\left\{ \hat{\Phi}_k \right\}_{k=1}^p$ and $\hat{\theta}$ are obtained, we automatically have $\hat{\Sigma}_w$. We can consider the elements of Σ_w to be redundant parameters if $\{\Phi_k\}_{k=1}^p$ and θ already lie in the parameter space. In most other cases of fitting state-space models, Σ_w itself is estimated along with $\{\Phi_k\}_{k=1}^p$, as the elements of Σ_ε are usually unknown. In addition to not having to estimate an extra 2 parameters, our state-space model ensures that the diagonal elements of Σ_ε remain fixed and equal to $(n-3)^{-1}$, the true large-sample variance of Fisher-transformed correlations.

5.2.3 The Kalman Filter and Maximum Likelihood Estimation

Before describing the maximum likelihood estimation of unknown parameters via the Kalman filter for our state-space model, some notation and assumptions are required. Keep in mind that we only observe the Fisher-transformed correlations $\boldsymbol{z}_t$; the state process $\boldsymbol{x}_t$ must be estimated using the observed Fisher-transformed correlations. Our main goal of employing the state-space model is to produce estimators of the parameters driving the unobservable signal $\boldsymbol{x}_t$, given the data $Z_s = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_s\}$ up to time s. For the purpose of estimating the parameters $\{\Phi_k\}_{k=1}^p$ and θ, we deal with times $s \leq t$.

Consider the conditional expectation of $\boldsymbol{x}_t$ given the observed data Z_s up to time s. We define this as:

$$\boldsymbol{x}_t^s = E\left(\boldsymbol{x}_t \mid Z_s\right). \tag{5.28}$$

The corresponding mean-squared error is defined as:

$$P_t^s = E\left[\left(\boldsymbol{x}_t - \boldsymbol{x}_t^s\right)\left(\boldsymbol{x}_t - \boldsymbol{x}_t^s\right)'\right]. \tag{5.29}$$

We assume that the processes are Gaussian; as such, P_t^s is also the conditional error covariance. Thus, we also have:

$$P_t^s = E\left[\left(\boldsymbol{x}_t - \boldsymbol{x}_t^s\right)\left(\boldsymbol{x}_t - \boldsymbol{x}_t^s\right)' \mid Z_s\right]. \tag{5.30}$$

We note that the covariance between $(\boldsymbol{x}_t - \boldsymbol{x}_t^s)$ and Z_s is zero for all t and s. Combined with the Gaussian assumption, $(\boldsymbol{x}_t - \boldsymbol{x}_t^s)$ and Z_s are independent. Thus, the unconditional distribution of $(\boldsymbol{x}_t - \boldsymbol{x}_t^s)$ and the conditional distribution of $(\boldsymbol{x}_t - \boldsymbol{x}_t^s)$ given Z_s are equal. The prediction errors, which we will call the innovations, are defined for $t = 1, \ldots, T$ as:

$$\mathbf{e}_t = \boldsymbol{z}_t - E\left(\boldsymbol{z}_t \mid Z_{t-1}\right) = \boldsymbol{z}_t - A\boldsymbol{x}_t^{t-1} - \Gamma\boldsymbol{u}_t. \tag{5.31}$$

The innovations are Gaussian and independent. Lastly, we define the variance of the innovations for $t = 1, \ldots, T$ as:

$$\Sigma_t = \operatorname{var}\left(\mathbf{e}_t\right) = \operatorname{var}\left[A\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{t-1}\right) + \boldsymbol{w}_t\right] = AP_t^{t-1}A' + \Sigma_w. \tag{5.32}$$

The Kalman filter is a crucial component of parameter estimation for state-space models; it allows us to calculate the innovations and their variances for use in maximum likelihood estimation. We provide the equations that define the Kalman filter for our state-space model without proof (see Property 6.5 in [61] for the general Kalman filter with correlated noise). For our state-space model specified in (5.23) and (5.24), with initial conditions

$$\boldsymbol{x}_1^0 = \Phi\boldsymbol{\mu}_0 \text{ and } P_1^0 = \Phi\Sigma_0\Phi' + \Psi\Sigma_w\Psi', \tag{5.33}$$

the one-step-ahead predictions, for $t = 1, \ldots, T$, are given by

$$\boldsymbol{x}_{t+1}^t = \Phi\boldsymbol{x}_t^{t-1} + K_t\mathbf{e}_t \quad \text{and} \tag{5.34}$$

$$P_{t+1}^t = \Phi P_t^{t-1}\Phi' + \Psi\Sigma_w\Psi' - K_t\Sigma_t K_t', \tag{5.35}$$

$$\text{where} \quad K_t = \left[\Phi P_t^{t-1}A + \Psi\Sigma_w\right]\Sigma_t^{-1}. \tag{5.36}$$

The filtered values of the state and error variance, for $t = 1, \ldots, T$, are given by:

$$\boldsymbol{x}_t^t = \boldsymbol{x}_t^{t-1} + P_t^{t-1}A'\Sigma_t^{-1}\mathbf{e}_{t+1} \quad \text{and} \tag{5.37}$$

$$P_t^t = P_t^{t-1} - P_t^{t-1}A'\Sigma_t^{-1}AP_t^{t-1}. \tag{5.38}$$

The innovations $\mathbf{e}_t = \boldsymbol{z}_t - A\boldsymbol{x}_t^{t-1} - \Gamma\boldsymbol{u}_t$ and innovation variances $\Sigma_t = AP_t^{t-1}A' + \Sigma_w$, for $t = 1, \ldots, T$, are obtained using (5.33)-(5.36). From these, we can construct the innovations form of the likelihood function. To estimate the parameters that specify our state-space model, represented by $\Theta = \{\boldsymbol{\mu}_0, \Sigma_0, \Phi_1, \ldots, \Phi_p, \theta, \Gamma\}$, we minimize the negative log-likelihood function

$$-l(\Theta) = \frac{1}{2}\sum_{t=1}^{T}\log|\Sigma_t(\Theta)| + \frac{1}{2}\sum_{t=1}^{T}\mathbf{e}_t(\Theta)'\Sigma_t(\Theta)^{-1}\mathbf{e}_t(\Theta). \tag{5.39}$$

The minimization of $-l(\Theta)$ begins by setting reasonable initial values for the parameters in Θ. The Kalman filter equations (5.33)-(5.36) are then used to obtain an initial set of innovations $\mathbf{e}_t(\Theta)$ and innovation covariance matrices $\Sigma_t(\Theta)$. Next, one iteration of a Newton-Raphson procedure is run to minimize $-l(\Theta)$, and a new set of estimates for the parameters in Θ are

obtained. The Kalman filter is run again using the updated Θ, and another Newton-Raphson iteration is performed. This is repeated until $-l(\Theta)$ stabilizes within some pre-specified and small amount. The values contained in Θ at the last iteration are the MLE's of the state-space model parameters.

Having provided the necessary ingredients for our state-space model and maximum likelihood estimation using the Kalman filter, we now list the steps to obtain $\left\{\hat{\Phi}_k\right\}_{k=1}^{p}$ and $\hat{\Sigma}_w$, which are needed to estimate the spectral matrix $f(\omega)$ of $\boldsymbol{\varepsilon}_t$:

1. Assume the state-space model (5.23)-(5.24) for the observed Fisher-transformed correlations $\boldsymbol{z}_t$ from Samples 1 and 2.
2. Employ Fourier basis functions as the regressors $\{\boldsymbol{u}_t\}_{t=1}^{T}$ in (5.24), where the parameter matrix Γ contains the unknown coefficients for the basis functions. Note that the number of basis functions does not have to be equal for Samples 1 and 2; this allows each correlation function in $[\eta_1(t), \eta_2(t)]'$ to be most accurately captured by $\hat{\Gamma}\boldsymbol{u}_t$. To account for the different number of basis functions, set the corresponding parameters in Γ equal to zero.
3. Construct a large grid of values for the VAR order p, the number of Fourier basis functions r_1 for Sample 1, and the number of Fourier basis functions r_2 for Sample 2. For each grid point $\{p, r_1, r_2\}$, obtain estimates of $\{\Phi_k\}_{k=1}^{p}$, θ and Γ using the maximum likelihood estimation procedure described above, along with the negative log-likelihood value $-l(\hat{\Theta})$ at the final iteration of the MLE procedure.
4. For each grid point $\{p, r_1, r_2\}$ and its corresponding $-l(\hat{\Theta})$, calculate the Bayesian Information Criterion (BIC) given by

$$BIC(p, r_1, r_2) = 2\left[-l(\hat{\Theta})\right] + (4p + r_1 + r_2 + 1)\left[\ln(T) + \ln(2\pi)\right]. \tag{5.40}$$

   Obtain the estimates $\left\{\hat{\Phi}_k\right\}_{k=1}^{p}$ and $\hat{\theta}$ from the fitted model which gives the smallest BIC and calculate the estimator $\hat{\Sigma}_w$ using equations (5.2), (5.26) and (5.27).
5. Lastly, construct the parametric spectral estimator $\hat{f}(\omega)$ by plugging $\hat{\Sigma}_w$ and $\left\{\hat{\Phi}_k\right\}_{k=1}^{p}$ into equation (5.13).

5.3 SIMULATIONS

Empirical significance and power calculations were implemented using simulations to assess the performance of our novel adaptive Neyman hypothesis test of equivalence of correlated functional correlations. In order to simulate two correlated correlation coefficients that vary over time, four-variate data were randomly generated to mimic four variables measured on the same sample of subjects over time.

More specifically, we simulated Gaussian time-series data $[X_1(t), X_2(t), X_3(t), X_4(t)]'$ giving rise to two correlated but non-overlapping sample correlation coefficients that vary over time with known population correlation functions. We formulated our model for the underlying data using the properties detailed below, and we obtained the relationships governing these properties by extending the methods for simulating simple, non-functional correlations given by Dunn and Clark [13, 14] to our time-varying setting.

When evaluating the empirical significance level, we used the following common correlation function, which is displayed in Figure 8:

$$\rho_{12}(t) = \rho_{34}(t) = \tanh\left[0.55\sin^2\left(2\pi t/T\right) - 0.1\right].$$

When performing power calculations, we used the following very similar correlation functions, which are shown in Figure 9:

$$\rho_{12}(t) = \tanh\left[0.6\sin^2\left(2\pi t/T\right) - 0.1\right],$$

$$\rho_{34}(t) = \tanh\left[0.5\sin^2\left(2\pi t/T\right) - 0.1\right].$$

We set $c = 0.3$ and $\rho_{13}(t) = \rho_{24}(t) = 0.5$, we assumed $\rho_{14}(t) = \rho_{23}(t)$, and we solved for $\rho_{14}(t)$ (and equivalently, for $\rho_{23}(t)$) numerically in terms of all of the other correlation functions. In addition, we set $EX_1(t) = EX_2(t) = EX_3(t) = EX_4(t) = 0$, and $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$.

A sample realization of $r_{12}(t)$ and $r_{34}(t)$ for $T = 200$ and $n = 50$, in the case where $\rho_{12}(t) \neq \rho_{34}(t)$, is shown in Figure 10 below. Black points are values of $r_{12}(t)$, and red points are values of $r_{34}(t)$.

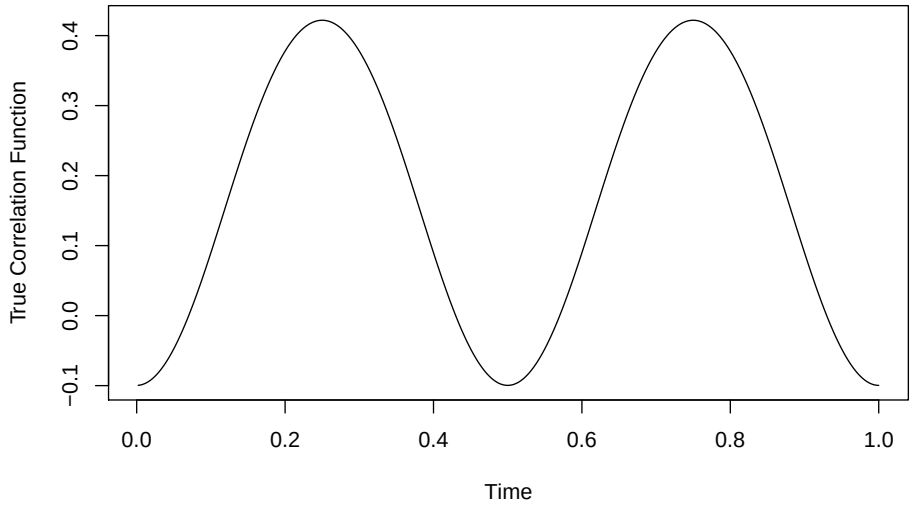

Figure 8: Dependent Samples: Correlation function for evaluating empirical significance

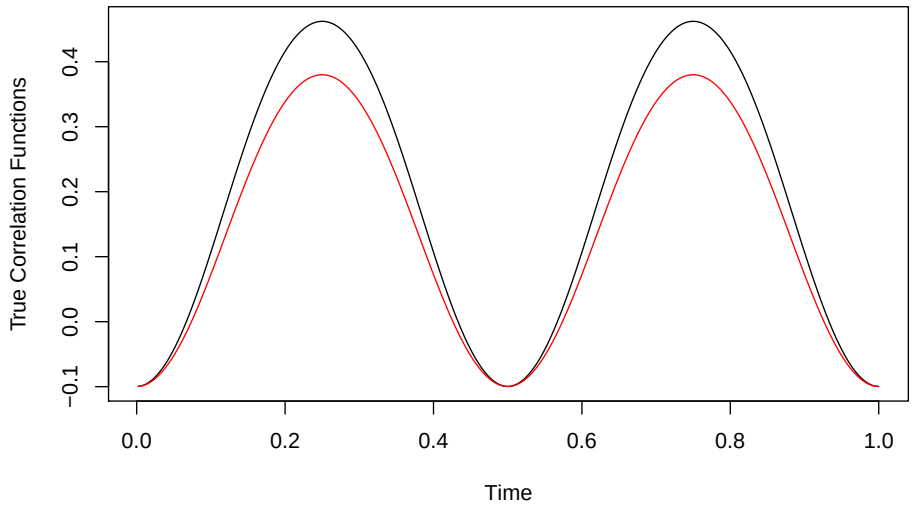

Figure 9: Dependent Samples: Correlation functions for evaluating empirical power

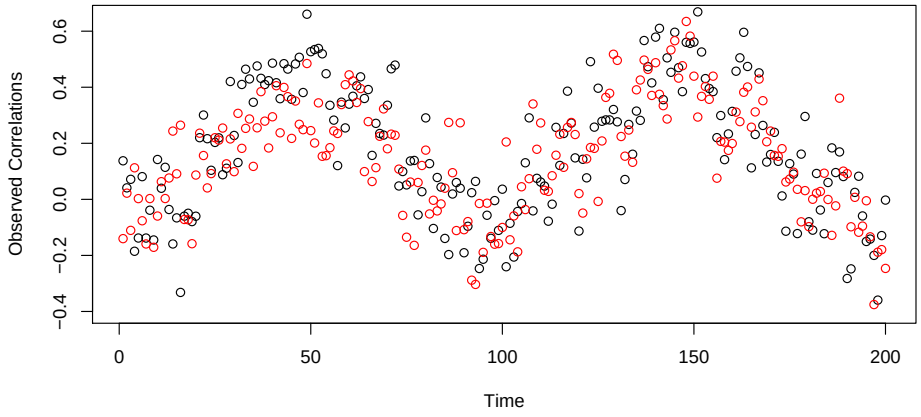

Figure 10: Dependent Samples: Example of simulated data

The data were simulated using three balanced sample sizes: $n = 25$, $n = 50$, and $n = 100$. Three values of T were used: $T = 200$, $T = 350$, and $T = 500$. From the simulated four-variate data, we calculated sample correlations and employed the adaptive Neyman test of equivalence of dependent functional correlations. Tests were performed at the $\alpha = 0.05$ level of significance. When investigating the empirical significance level, we used $N = 3500$ simulation runs for each setting, as the empirical level was slow to converge. For empirical power calculations, we used $N = 1000$ runs for each setting. The results are summarized in Tables 3 and 4.

Table 3: Comparing Dependent Correlation Functions: Empirical Significance Results

	$T = 200$	$T = 350$	$T = 500$
$n = 25$	0.0588	0.0486	0.0434
$n = 50$	0.0611	0.0366	0.0483
$n = 100$	0.0657	0.0529	0.0443

Table 4: Comparing Dependent Correlation Functions: Empirical Power Results

	$T = 200$	$T = 350$	$T = 500$
$n = 25$	0.722	0.738	0.958
$n = 50$	0.890	0.791	0.974
$n = 100$	1.000	0.909	0.936

The empirical significance levels obtained in all 9 settings are quite satisfactory. $T = 200$ gives the largest Type I error rates, all slightly higher than $\alpha = 0.05$, with the largest level 0.0657 occurring when $n = 100$. In contrast, the empirical significance levels for $T = 500$ are all slightly less than 0.05. The smallest Type I error rate, 0.0366, occurs when $T = 350$ and $n = 50$. The adaptive Neyman test of equivalence of dependent functional correlations also performs very well with respect to power. The power is largest when $T = 200$ and $n = 100$, where a value of 1 is achieved.

5.4 APPLICATION: DELTA EEG POWER AND HF-HRV

The third aim of our motivating study [58] addressed whether the time-varying correlation between delta EEG power and HF-HRV in the whole sample of 197 midlife women changes significantly across different NREM periods. The functional correlation profiles for the three NREM periods of sleep are shown in Figure 1. Qualitatively, the correlation function is bimodal during NREM-1 and NREM-2, with peaks both preceding and following $t = 0$. However, the peaks in the correlation function during NREM-2 seem broader compared to NREM-1, and the drop in correlation near $t = 0$ looks much sharper during NREM-2. During NREM-3, the correlation function is unimodal with a blunted peak following $t = 0$. In addition, the overall magnitude of the functional correlation between delta EEG power and HF-HRV appears to be larger in the first NREM period, compared to the second and

third NREM periods. Overall, the functional correlation between these two physiological parameters does seem to change across the three NREM periods. The largest difference can be seen during NREM-1 compared to NREM-3, while the smallest difference appears to be during NREM-2 compared to NREM-3.

To formally test whether the time-varying correlation between delta EEG power and HF-HRV in full sample significantly differs as a function of NREM period, we use the adaptive Neyman test for dependent samples. When comparing the correlation functions during NREM-1 and NREM-2, the standardized adaptive Neyman test statistic $T_{AN} = 31.00$, and the corresponding $p - value < 0.0001$. The difference in time-varying correlation during NREM-1 compared to NREM-3 is also extremely significant; the formal test gives $T_{AN} = 301.43$ and $p - value < 0.0001$. Somewhat surprisingly, the functional correlation between delta EEG power and HF-HRV changes extremely significantly during NREM-2 compared to NREM-3 as well, as the test statistic $T_{AN} = 115.20$ and the $p - value < 0.0001$. It appears that modest changes in magnitude as well as moderate changes in shape of the time-varying correlation across NREM periods lead us to conclude highly significant differences.

6.0 DISCUSSION

We presented a new methodology for estimation, point-wise inference, and formal comparisons of functional correlations. The utility of our methods was demonstrated by our motivating study of the time-varying correlation between delta EEG power and high frequency heart rate variability during sleep in midlife women [58]. Our estimation technique may be used to model the correlation between two variables measured on a sample of subjects as a continuous function of time, and confidence intervals may be constructed for point-wise inference using our novel bootstrap procedure. Further, beyond these tools for the estimation and inference of a single functional correlation, we developed a new method for the formal hypothesis testing of two functional correlations via adaptive Neyman tests for independent and dependent samples.

As the number of questions one could pose concerning functional correlations is numerous, these formal methodologies are not exhaustive and lead to future work. One area of future work is the analysis of overlapping correlations. There exists two kinds of correlated correlations: overlapping and non-overlapping correlated correlations. To illustrate the difference between the terms "overlapping" and "non-overlapping", consider a four-variate random normal vector $(X_1, X_2, X_3, X_4)'$. The correlations ρ_{12} and ρ_{34} are non-overlapping in the sense that they do not involve a common variable, whereas the correlations ρ_{12} and ρ_{13} are overlapping because they both involve the common variable X_1. The asymptotic correlation between overlapping sample correlation coefficients is different in form than that of non-overlapping sample correlation coefficients, and future work will explore methods for analyzing overlapping correlation functions.

A second area of future research will address local tests for identifying local differences. The tests considered here are global tests and only provide information about the across the

curve equivalence but do not identify areas within the time interval where this difference occurs. Future work will develop procedures that, if the global test concludes differences between groups, identify where this difference occurs. These procedures will have to overcome the challenge of involving an infinite amount of correlated local tests and will need to incorporate local building blocks, as opposed to the global smoothing spline and Fourier blocks used here.

BIBLIOGRAPHY

[1] P. Achermann, D. J. Dijk, D. P. Brunner, and A. A. Borbely. A model of human sleep homeostasis based on EEG slow-wave activity: quantitative comparison of data and simulations. *Brain Research Bulletin*, 31:97–113, 1993.

[2] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, Hoboken, 2003.

[3] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR*. Washington, DC: American Psychiatric Association, 2000.

[4] M. H. Bonnet and D. L. Arand. Heart rate variability: sleep stage, time of night, and arousal influences. *Electroencephalography and clinical neurophysiology*, 102(5):390–396, 1997.

[5] M. H. Bonnet and D. L. Arand. Heart rate variability in insomniacs and matched normal sleepers. *Psychosomatic Medicine*, 60:610–615, 1998.

[6] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. New York: Springer, 2006.

[7] B. A. Brumback and J. A. Rice. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93:961–994, 1998.

[8] F. P. Cappuccio, D. Cooper, L. D'Elia, P. Strazzullo, and M. A. Miller. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *European Heart Journal*, 32(12):1484–1492, 2011.

[9] F. P. Cappuccio, L. D'Elia, P. Strazzullo, and M. A. Miller. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*, 33(5):585, 2010.

[10] C. M. Crainiceanu, D. Ruppert, R. J. Carroll, A. Joshi, and B. Goodner. Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*, 16(2):265–288, 2007.

[11] D. A. Darling and P. Erdös. A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J*, 23(1):143–155, 1956.

[12] M. Dolker, S. Halperin, and D. R. Divgi. Problems with bootstrapping Pearson correlations in very small bivariate samples. *Psychometrika*, 47(4):529–530, 1982.

[13] O. J. Dunn and V. A. Clark. Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, 64:366–377, 1969.

[14] O. J. Dunn and V. A. Clark. Comparisons of tests of equality of dependent correlation coefficients. *Journal of the American Statistical Association*, 66:904–908, 1971.

[15] J. D. Edinger, M. H. Bonnet, R. R. Bootzin, K. Doghramji, C. M. Dorsey, C. A. Espie, A. O. Jamieson, W. V. McCall, C. M. Morin, and E. J. Stepanski. Derivation of research diagnostic criteria for insomnia: report of an American Academy of Sleep Medicine Work Group. *Sleep*, 27(8):1567–1596, 2004.

[16] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.

[17] B. Efron. *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics, 1982.

[18] B. Efron. Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, 104(2):293–296, 1988.

[19] J. Fan. Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, 91(434):674–688, 1996.

[20] J. Fan and S. K. Lin. Test of significance when data are curves. *Journal of the American Statistical Association*, 93:1007–1021, 1998.

[21] P. M. Fuller, J. J. Gooley, and C. B. Saper. Neurobiology of the sleep-wake cycle: sleep architecture, circadian regulation, and regulatory feedback. *Journal of biological rhythms*, 21(6):482–493, 2006.

[22] L. Gallicchio and B. Kalesan. Sleep duration and mortality: a systematic review and meta-analysis. *Journal of Sleep Research*, 18(2):148–158, 2009.

[23] M. Gilman, J. Floras, K. Usui, Y. Kaneko, R. Leung, and T. Bradley. Continuous positive airway pressure increases heart rate variability in heart failure patients with obstructive sleep apnoea. *Clinical Science*, 114:243–249, 2008.

[24] C. Gronfier, C. Simon, F. Piquard, J. Ehrhart, and G. Brandenberger. Neuroendocrine processes underlying ultradian sleep regulation in man. *The Journal of Clinical Endocrinology & Metabolism*, 84(8):2686–2690, 1999.

[25] C. Gu. *Smoothing Spline ANOVA Models*. New York: Springer-Verlag, 2002.

[26] C. Gu and G. Wahba. Smoothing Spline ANOVA with component-wise Bayesian "confidence intervals". *Journal of Computational and Graphical Statistics*, 2:97–117, 1992.

[27] W. Guo. Inference in smoothing spline analysis of variance. *Journal of the Royal Statistical Society: Series B*, 64(4):887–898, 2002.

[28] M. H. Hall, K. A. Matthews, H. M. Kravitz, E. B. Gold, D. J. Buysse, J. T. Bromberger, J. F. Owens, and M. F. Sowers. Race and financial strain are independent correlates of sleep in midlife women: the SWAN sleep study. *Sleep*, 32(1):73–82, 2009.

[29] P. Hall and I. Van Keilegom. Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17(4):1511–1531, 2007.

[30] D. L. Hawkins. Using U statistics to derive the asymptotic distribution of Fisher's Z statistic. *American Statistician*, 43:235–237, 1989.

[31] M. F. Hilton, M. J. Chappell, W. A. Bartlett, A. Malhotra, J. M. Beattie, and R. M. Cayton. The sleep apnoea/hypopnoea syndrome depresses waking vagal tone independent of sympathetic activation. *European Respiratory Journal*, 17(6):1258–1266, 2001.

[32] B. Israel, D. J. Buysse, R. T. Krafty, A. Begley, J. Miewald, and M. H. Hall. Short-term stability of sleep and heart rate variability in good sleepers and patients with insomnia: for some measures, one night is enough. *Sleep*, 35(9):1285–1291, 2012.

[33] F. Jurysta, J. P. Lanquart, V. Sputaels, M. Dumont, P. F. Migeotte, S. Leistedt, P. Linkowski, and P. Van De Borne. The impact of chronic primary insomnia on the heart rate–EEG variability link. *Clinical neurophysiology*, 120(6):1054–1060, 2009.

[34] F. Jurysta, J. P. Lanquart, P. Van De Borne, P. F. Migeotte, M. Dumont, J. P. Degaute, and P. Linkowski. The link between cardiac autonomic activity and sleep delta power is altered in men with sleep apnea-hypopnea syndrome. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, 291(4):R1165–R1171, 2006.

[35] L. Karasulu, P.Ö. Epöztürk, S. N. Sökücü, L. Dalar, and S. Altın. Improving heart rate variability in sleep apnea patients: differences in treatment with auto-titrating positive airway pressure (APAP) versus conventional CPAP. *Lung*, 188(4):315–320, 2010.

[36] R. T. Krafty, P. A. Gimotty, D. Holtz, G. Coukos, and W. Guo. Varying coefficient model with unknown within-subject covariance for analysis of tumor growth curves. *Biometrics*, 64:1023–1031, 2008.

[37] T. Krivobokova, T. Kneib, and G. Claeskens. Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association*, 105(490):852–863, 2010.

[38] D. Liao, X. Li, S. M. Rodriguez-Colon, J. Liu, A. N. Vgontzas, S. Calhoun, and E. O. Bixler. Sleep-disordered breathing and cardiac autonomic modulation in children. *Sleep medicine*, 11(5):484–488, 2010.

[39] X. Lin, N. Wang, A. Welsh, and R. J Carroll. Equivalent kernels of smoothing splines in nonparametric regression for clustered data. *Biometrika*, 91:177–193, 2004.

[40] X. Lin and D. Zhang. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, Series B*, 61:381–400, 1999.

[41] A. Liu and Y. Wang. Hypothesis testing in smoothing spline models. *Journal of Statistical Computation and Simulation*, 74(8):581–597, 2004.

[42] Y. K. Loke, J. W. L. Brown, C. S. Kwok, A. Niruban, and P. K. Myint. Association of obstructive sleep apnea with risk of serious cardiovascular events: A systematic review and meta-analysis. *Circulation: Cardiovascular Quality and Outcomes*, 5(5):720–728, 2012.

[43] E. Lugaresi, F. Provini, and P. Cortelli. Sleep embodies maximum and minimum levels of autonomic integration. *Clinical Autonomic Research*, 11(1):5–10, 2001.

[44] X. Meng, R. Rosenthal, and D. B. Rubin. Comparing correlated correlation coefficients. *Psychological Bulletin*, 111(1):172–175, 1992.

[45] K. Narkiewicz, P. J. H. van de Borne, N. Montano, M. E. Dyken, B. G. Phillips, and V. K. Somers. Contribution of tonic chemoreflex activation to sympathetic activity and blood pressure in patients with obstructive sleep apnea. *Circulation*, 97(10):943–945, 1998.

[46] J. Neyman. Smooth test for goodness of fit. *Scandinavian Actuarial Journal*, 1937(3-4):149–199, 1937.

[47] M. M. Ohayon. Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Medicine Reviews*, 6(2):97–111, 2002.

[48] M. L. Okun, R. T. Krafty, D. J. Buysse, T. H. Monk, C. F. Reynolds, A. Begley, and M. H. Hall. What constitutes too long of a delay? Determining the cortisol awakening response (CAR) using self-report and PSG-assessed wake time. *Psychoneuroendocrinology*, 35(3):460–468, 2010.

[49] I. Olkin and J.D. Finn. Testing correlated correlations. *Psychological Bulletin*, 108(2):330–333, 1990.

[50] H. Otzenberger, C. Gronfier, C. Simon, A. Charloux, J. Ehrhart, F. Piquard, and G. Brandenberger. Dynamic heart rate variability: a tool for exploring sympathovagal balance continuously during sleep in men. *American Journal of Physiology: Heart and Circulatory Physiology*, 275(3):H946–H950, 1998.

[51] H. Otzenberger, C. Simon, C. Gronfier, and G. Brandenberger. Temporal relationship between dynamic heart rate variability and electroencephalographic activity during sleep in man. *Neuroscience Letters*, 229(3):173–176, 1997.

[52] T. E. Raghunathan. An approximate test for homogeneity of correlated correlation coefficients. *Quality and Quantity*, 37(1):99–110, 2003.

[53] T. E. Raghunathan, R. Rosenthal, and D. B. Rubin. Comparing correlated but nonoverlapping correlations. *Psychological Methods*, 1(2):178–183, 1996.

[54] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2nd edition, June 2005.

[55] J. L. Rasmussen. Estimating correlation coefficients: Bootstrap and parametric approaches. *Psychological Bulletin*, 101(1):136–139, 1987.

[56] J. L. Rasmussen. "Bootstrap confidence intervals: Good or bad": Comments on Efron (1988) and Strube (1988) and further evaluation. *Psychological Bulletin*, 104(2):297–299, 1988.

[57] A. Rechtschaffen and A. Kales. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Washington DC: US Government Printing Office, 1968.

[58] S. D. Rothenberger, R. T. Krafty, B. J. Taylor, M. R. Cribbet, J. F. Thayer, D. J. Buysse, H. M. Kravitz, E. D. Buysse, and M. H. Hall. Time-varying correlations between delta EEG power and heart rate variability in midlife women: The SWAN Sleep Study. *Psychophysiology*, (In Press), 2014.

[59] N. Santoro, E. S. Taylor, and K. Sutton-Tyrrell. The SWAN song: Study of Womens Health Across the Nation's recurring themes. *Obstetrics and Gynecology Clinics of North America*, 38(3):417–423, 2011.

[60] S. Schwartz, W. M. Anderson, S. R. Cole, J. Cornoni-Huntley, J. C. Hays, and D. Blazer. Insomnia and heart disease: a review of epidemiologic studies. *Journal of Psychosomatic Research*, 47(4):313–333, 1999.

[61] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. New York: Springer, 3rd edition, 2011.

[62] A. Silvani. Physiological sleep-dependent changes in arterial blood pressure: central autonomic commands and baroreflex control. *Clinical and Experimental Pharmacology and Physiology*, 35(9):987–994, 2008.

[63] F. Sofi, F. Cesari, A. Casini, C. Macchi, R. Abbate, and G. F. Gensini. Insomnia and risk of cardiovascular disease: a meta-analysis. *European Journal of Preventive Cardiology*, 21(1):57–64, 2014.

[64] V. K. Somers, M. E. Dyken, A. L. Mark, and F. M. Abboud. Sympathetic-nerve activity during sleep in normal subjects. *New England Journal of Medicine*, 328(5):303–307, 1993.

[65] M. J. Strube. Bootstrap Type I error rates for the correlation coefficient: An examination of alternate procedures. *Psychological Bulletin*, 104(2):290, 1988.

[66] L. Toscani, P. F. Gangemi, A. Parigi, R. Silipo, P. Ragghianti, E. Sirabella, M. Morelli, L. Bagnoli, R. Vergassola, and G. Zaccara. Human heart rate variability and sleep stages. *The Italian Journal of Neurological Sciences*, 17(6):437–439, 1996.

[67] J. Trinder, J. Kleiman, M. Carrington, S. Smith, S. Breen, N. Tan, and Y. Kim. Autonomic activity during human sleep as a function of time and sleep stage. *Journal of Sleep Research*, 10(4):253–264, 2001.

[68] J. Trinder, J. Waloszek, M. J. Woods, and A. S. Jordan. Sleep and cardiovascular regulation. *Pflügers Archiv-European Journal of Physiology*, 463(1):161–168, 2012.

[69] G. Wahba. Bayesian "confidence intervals" and the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, 45:133–150, 1983.

[70] Y. Wang. Mixed-effects smoothing spline ANOVA. *Journal of the Royal Statistical Society B*, 60:159–174, 1998.

[71] Y. Wang. Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, 60:159–174, 1998.

[72] A. Welsh, X. Lin, and R. J. Carroll. Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. *Journal of the American Statistical Society*, 97:482–493, 2002.

[73] T. Young, P. E. Peppard, and D. J. Gottlieb. Epidemiology of obstructive sleep apnea: a population health perspective. *American Journal of Respiratory and Critical Care Medicine*, 165(9):1217–1239, 2002.

[74] C. Zhang, H. Peng, and J. Zhang. Two samples tests for functional data. *Communications in Statistics: Theory and Methods*, 39(4):559–578, 2010.

[75] D. Zhang, X. Lin, J. Raz, and M. Sowers. Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93:710–719, 1998.