WAVELET ANALYSIS OF BIG DATA CONTAMINATED BY LARGE NOISE IN AN FMRI STUDY OF NEUROPLASTICITY

by

Jiayi Wu

APPROVED BY SUPERVISORY COMMITTEE:

Dr. Sam Efromovich, Chair

Dr. Larry P. Ammann

Dr. Pankaj K. Choudhary

Dr. Frank Konietschke

To my family and professors

WAVELET ANALYSIS OF BIG DATA CONTAMINATED BY LARGE NOISE IN AN FMRI STUDY OF NEUROPLASTICITY

by

JIAYI WU, BS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

STATISTICS

THE UNIVERSITY OF TEXAS AT DALLAS

August 2018

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my family. I would have never taken a single step on the journey to a doctoral study at The University of Texas at Dallas without their understanding and support. I would also like to express my special memory of my grandfather, who educated me with greatest patience when I was very little and has kept influencing me with his kindness, creativeness and enthusiasm.

I would like to express my greatest appreciation to my academic advisor Dr. Sam Efromovich for his insightful and constructive suggestions on the development of this research work as well as my doctoral study and future career. He generously spent time working closely with me. With his guidance, I aquired not only the professional knowledge of nonparametric estimation, but also the ability to identify a problem on a high level and discover subtleties in the analysis. His professionalism, enthusiasm and productiveness have also influenced me.

I would also like to thank my committee members, Dr. Larry P. Ammann, Dr. Pankaj K. Choudhary, Dr. Frank Konietschke, for taking time to review and improve the dissertation.

My grateful thanks are also extended to Dr. Natalia Humphreys, who introduced me to the topics in acturial science, and Dr. Ekaterina Smirnova, who introduced me to the fMRI studies and invited me to give my first presentation at a conference.

I am also grateful to all professors, staff and colleagues at the Department of Mathematical Sciences at The University of Texas at Dallas. The years spent in the department are among the most wonderful experiences I have ever had. The professional knowledge I aquired here, the teaching and research skills I developed here and the friends I made here will be my life-long assets.

April 2018

WAVELET ANALYSIS OF BIG DATA CONTAMINATED BY LARGE NOISE IN AN FMRI STUDY OF NEUROPLASTICITY

Jiayi Wu, PhD
The University of Texas at Dallas, 2018

Supervising Professor: Dr. Sam Efromovich, Chair

Functional magnetic resonance imaging (fMRI) allows researchers to analyze brain activity on a voxel level, but using this ability is complicated by dealing with Big Data and large noise. A traditional remedy is averaging over large parts of the brain in combination with more advanced technical innovations in reducing fMRI noise. In this dissertation a novel statistical approach, based on a wavelet analysis of standard fMRI data, is proposed and its application to an fMRI study of neural plasticity of 24 healthy adults is presented. The aim of the study was to recognize changes in connectivity between left and right motor cortices (the neuroplasticity) after button clicking training sessions. A conventional method of the data analysis, based on averaging images, has implied that for the group of 24 participants the connectivity increased after the training. The proposed wavelet analysis suggests to analyze pathways between left and right hemispheres on a voxel-to-voxel level and for each participant via estimation of the corresponding cross-correlations. This immediately necessitates statistical analysis of large-p-small-n correlation matrices contaminated by large noise. Furthermore, the distributions that we are dealing with in the analysis are neither Gaussian nor sub-Gaussian but sub-exponential. The dissertation explains how the problem may be solved and presents results of a dynamic analysis of the ability of a human brain to reorganize itself for 24 healthy adults. Results show that the ability of a brain to reorganize

itself varies widely even among healthy individuals, and this observation is important for our understanding of a human brain and treatment of brain diseases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Big Data refers to data sets with both large sample size and high dimensionality. Advances in today's technology lowers the price of producing and storing data and thus causes an explosion of information. Big Data brings new statistical and computational challenges as its features differ from that of the traditional small- or median-scale data sets. One classic example of Big Data is the fMRI study, see a review in Fan et al. (2014). An fMRI machine produces 3-D time-course cerebral images by scanning human brains noninvasively. An image contains hundreds of thousands of voxels and each voxel is scanned for hundreds of times. The high-dimension property of fMRI data set turns it into a problem of Big Data.

Another familiar complication of an fMRI study is a large noise. An fMRI machine captures the hemodynamic responses in the active part of brain after a certain stimulus, as well as noise caused by physiological activities such as respiration and heartbeat. We will show shortly that the magnitude of the nuisance physiological activities is almost comparable to that of the signal component of interest, especially during the resting states when there is no stimulus.

In order to simplify analysis and decrease noise, one popular statistical approach is to average observations over a number of voxels of interest. This was the approach used by bioengineers from the UT Southwestern Medical Center in a study of neuroplasticity based on an fMRI experiment involving 24 healthy adults (Tung et al. 2013); technical details of the study will be described shortly. The aim of the study was to recognize changes in connectivity between left and right motor cortices after button clicking training sessions, and the authors used a paired t-test for 24 participants which compared pre-training and post-training Pearson correlations between spatially averaged signals in left and right motor cortices. The conclusion was that the plasticity exists because the (overall over 24 participants) correlation significantly increased. Averaging over a region of interest (ROI) is a

well known and respectful statistical technique which simultaneously solves the Big Data and large noise fMRI problems (Worsley et al. 1992; Lazar 2008; Anderson et al. 2010; Kelly et al. 2011) . At the same time, it would be of interest to understand how the plasticity can be studied on a voxel-to-voxel level and simultaneously over all voxel-to-voxel pathways based on the available data. Note that this approach would allow us to study plasticity for each participant in place of considering them as a sample from a generic participant.

In this work, using the fMRI data from Tung et al. (2013), we propose and then test a new wavelet methodology for the analysis of neural plasticity. Main innovations are: (i) A special multiresolution analysis; (ii) A new procedure for the analysis of noise in wavelet coefficients; (iii) Thresholding analysis of large-p-small-n cross-correlation matrices; (iv) Developing simultaneous confidence bounds for cross-covariance and cross-correlations matrices. As we will see shortly, the new methodology allows us to simultaneously analyze dynamic changes in pathways between all pairs of active voxels. Let us stress that there is an acute interest in solving these problems, see a discussion in Chang and Glover (2009); Weissenbacher et al. (2009); Birn (2012); He and Liu (2012); Murphy et al. (2013); Efromovich and Smirnova (2014); Chen et al. (2017); Marusak et al. (2017); Mill et al. (2017); where further references may be found.

The above mentioned thresholding analysis of large-p-small-n matrices is a new statistical result in the area that have seen a surge in publications, see reviews Fan et al. (2014); Cai (2017). At the same time, so far the main emphasis has been on the study of classical covariance and correlation matrices with the main application being microarrays. In this work, we apply the thresholding analysis on large-p-small-n cross-correlation matrices with fMRI data.

The work's context reflects the above-formulated challenges and tasks. Chapter 2 reviews the concepts of fMRI and neural plasticity, nonparametric regressioin and minimax estimation, wavelets, regression with errors in both variables, and subexponential distribution and

tail bounds. Chapter 3 describes the proposed wavelet methodology. Let us stress that the idea of using a wavelet decomposition for fMRI signals is not new, see a discussion in Lazar (2008); Efromovich and Valdez-Jasso (2010). Further, a wavelet analysis of plasticity was proposed in Efromovich and Smirnova (2014). However there were two major unresolved issues that made the wavelet approach not feasible for a simultaneous analysis of all active voxels. The first one is that known methods of estimation of the noise in wavelet coefficients imply levels that are too large for a feasible simultaneous inference. The second one is that traditional asymptotic simplifications in wavelet analysis require larger sample sizes than the fMRI data provides. As a result, a new theory and methodology are needed for dealing with these challenging statistical issues; they are explained in Chapter 4. Evaluation of the noise model assumptions, simulation results and practical applications of the proposed wavelet methodology for the analysis of neuroplasticity is presented in Chapter 5. Chapter 6 presents another study related to nonparametric analysis of nonstationary asset returns.

As a final note, the results of the dissertation were published in Efromovich and Wu (2018a,b) and Efromovich and Wu (2017).

CHAPTER 2

CONCEPTUAL SETTING AND LITERATURE REVIEW

2.1 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging or functional MRI (fMRI) is a noninvasive technology which measures the human brain functionality by detecting alterations associated with cerebral blood flow. The earlier MRI scanning technology provides a static structural view of organs in the body using strong magnetic fields, radio waves, and field gradients, without requiring people to undergo surgery or to be exposed to ionising radiation. The concept of fMRI is to extend MRI technology to further capture the human brain functionality, not by a direct measure of neuronal activities but through a measure of the blood-oxygen-level dependent (BOLD) level contrast. The cerebral blood flow and neuronal activities are related; the neurons' activation causes a need for energy which could be provided by the oxygen in the blood, and therefore local blood flow to those brain regions increases, and oxygenated blood displaces deoxygenated blood. The fact that the oxygenated and deoxygenated forms of blood could be detected using MRI provides an indirect measure of the brain activities.

An fMRI scanner produces 3-D time-course cerebral images. Each image contains thousands of voxels and each voxel is scanned for hundreds of times, which implies the high-dimension property of fMRI data set (Fan et al. 2014). On the other hand, an fMRI machine captures the hemodynamic responses as well as nuisance physiological activities, which implies the noisiness of fMRI data. These properties bring challenges in the data representation, archiving, and analysis.

2.1.1 Neuroplasticity

Neuroplasticity, also known as *neural plasticity*, is the ability of the human brain to reorgnize itself, even through adulthood. Examples of neuroplasiticity include brain anatomical structure change, brain functional location change, and neural connectivity change. These brain

remodeling could be a result of personal experience, environmental stimuli, thoughts and emotions, or drugs. We are most interested in the *activity-dependent plasticity*, which is the plasticity arising from intrinsic or endogenous activities. For example, practicing one's less dominant hand could cause one to become ambidextrous; physical rehabilitation could reduce neurological dysfunction. Neuroplasticity due to training has significant implications for the development of learning and memory methods, therapies for acquired brain injuries such as physical trauma and stroke, and therapies for brain diseases such as Alzheimer's disease.

Let us introduce some popular approaches that are used to analyze neuroplasticity. Voxel-based morphometry (VBM) (Ashburner and Friston 2000) is a statistical technique that detects the significant differences in brain anatomical structure. VBM is an application of statistical parametric mapping (SPM), which models each voxel as a general linear model and perform hypotheses testing on the model coefficients.

Voxel-Mirrored Homotopic Connectivity (VMHC) is proposed to measure the neural connectivity, where the Pearson correlation of the resting state signals from the mirrored inter-hemispheric voxels are computed (Zuo et al. 2010). However, VMHC can be only used to indicate the synchrony in the neural fluctuations between mirrored brain regions.

The Region of Interest (ROI) analysis is very commonly used in fMRI studies. The idea is to restrict the study to only a small area, called ROI, and average the parameter estimators from all voxels in that ROI. Therefore, the study ends up with one averaged estimator for one subject and these subject-wise estimators are used for a t test. Efromovich and Smirnova (2014) attempted to extend the ROI method to a voxel-level analysis but within a single slice due to the complications caused by high dimensionality and large noise.

In this work, we want to make a further extension of the voxel-level analysis to the whole ROI so that all inter-hemispheric pairs of voxels could be considered under the analysis.

2.1.2 Physiological Noise and Its Correction

Since the discovery of the spontaneous neural activity at rest in the humain brain (Biswal et al. 1995), resting state fMRI has become an active area of investigation. However, one big challenge in exploring the resting state fMRI is its large noise-to-signal ratio. An fMRI scanner captures (spontaneous) BOLD signals as well as physiological activities, like cardiac and respiratory fluctuations. They must be corrected in order to avoid any distortions in the statistical analysis, especially when resting state fMRI are considered where the scale of the nuisance noise is comparable to that of the spontaneous neuronal activity.

Basically, there are two classes of techniques to remove physiological noise from the resting-state BOLD time series; see reviews in Murphy et al. (2013); Birn (2012). The first class of techniques is using external recordings of physiological activities. For instance, facilities like a finger (or toe) pulse oximeter and respirotory bellows are used during the fMRI scanning, making it easy to capture cardiac and respiratory processes. The second class of noise clean-up techniques is purely data-based. That is, these methods use the resting-state fMRI data itself. We next briefly summarize some popular methods from the second class. As we will see, none of these noise reduction approaches is perfect, and therefore this still is an active area of investigation (Birn 2012).

Image-based Physiological Artifacts estimation and Correction Technique (IMPACT) is the most straightforward method which removes the primary cardiac and respiration fluctuations using bandpass filtering (Chuang and Chen 2001). The reason why this works is because the frequencies of research interest in the resting state fMRI data are between 0.01 Hz and 0.1 Hz, while the cardiac and respiratory frequencies are usually greater than 0.25 Hz (Barrett et al. 2009). The disadvantage of this method is that it requires the scanning repetition time (TR) to be short (< 0.5 s), otherwise the nuisance frequencies might be aliased into the lower frequency range.

Under the assumption that the physiological noise is independent of the neural activity-related signal component, an *independent component analysis* (ICA) can be used as a noise reduction strategy. ICA is a commonly used computational method in signal processing, which decomposes a signal into additive subcomponents that are statistically independent. The two popular applications of ICA in fMRI studies are CORSICA (Perlbarg et al. 2007) and PESTICA (Beall and Lowe 2007).

The *nuisance regression* approaches are also commonly used, whose idea is to consider the signal from each voxel as a linear regression model and compute a nuisance regressor which is supposed to be removed from the regression model. One type of the nuisance regression is called *global signal regression* (GSR). The idea of GSR is to remove the brain-wise common signal component from the raw data. That is, the fMRI data is spatially averaged over the whole brain, and the averaged signal is used as a regressor, to be removed, in a linear regression model for the data. The disadvantage of GSR is that the averaged signal contains the neural activity-related component. Its removal could distort the underlying signal of interest. To address this issue, another type of the nuisance regression method is to derive the nuisance regressor from brain regions that are very unlikely to show neural activity-related BOLD signal, like the white matter (WM) and cerebrospinal fluid (CSF) (Birn et al. 2009). Therefore, it is likely that only non-neuronal signal changes have been removed from the raw data. Note that the assumption of these two nuisance regression methods is that the physiological noise is spatially in common for the whole brain.

An extension to the above introduced nuisance regression approaches is to use a localized averaged signal from the WM that is surrounding a voxel of interest as a nuisance regressor. This method has a better reduction in noise, especially for spatially inhomogeneous physiology occurring in the brain (Jo et al. 2010).

The fMRI data collected by Tung et al. (2013) (the data at hand, which is used in practical applications shown in Chapter 5) has been prepocessed by regressing out the time

courses of the whole brain WM and CSF. However, our analysis shows that the physiological noise clean-up approach is not perfect (see Chapter 5), and hence further development is needed for more accurate estimation of noise level.

Thanks to the frequency extraction property of wavelet decomposition, the physiological noise is easily filtered out in our proposed wavelet analysis. On the other hand, the removed physiological component plays an important role in the analysis of the noise level in the empirical wavelet coefficients.

2.2 Series Approach for Nonparametric Regression

The problem of a standard nonparametric regression is to estimate the regression function $f(t), t \in [0,1]$ from the observations $(X_1, Y_1), ..., (X_n, Y_n)$ in the supposed model

$$Y_l = f(X_l) + \tau\epsilon_l, l = 1, 2, ..., n. \tag{2.1}$$

Here $\epsilon_1, ..., \epsilon_n$ are realizations of a random variable (error) ϵ with zero mean and unit variance, and τ is a constant which represents the noise level. The set of values $\{X_1, ..., X_n\}$ is called the *design*. If the design is a set of i.i.d. realizations of a random variable, the corresponding model is called *random design* regression; if the design is a set of fixed values, then the corresponding model is called *fixed design* regression. A fixed design regression model can be simplified as

$$Y_l = f(l/n) + \tau\epsilon_l, l = 1, 2, ..., n. \tag{2.2}$$

Suppose that $f \in L_2[0,1]$ and a system of functions $\{\varphi_0, \varphi_1, ...\}$ is an *orthonormal* basis on $L_2[0,1]$, i.e.,

$$\int_0^1 \varphi_i(x)\varphi_j(x)dx = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}. \tag{2.3}$$

The set of values $\{\theta_0, \theta_1, \ldots\}$ defined as

$$\theta_j = \int_0^1 f(x)\varphi_j(x)dx, \quad j = 0, 1, \ldots,$$

are called the *Fourier coefficients* of the function f with respect to an orthonormal basis $\{\varphi_0, \varphi_1, ...\}$. Then f can be represented as a sum of series

$$f(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x). \tag{2.4}$$

The idea of *orthonormal series estimation* is to approximate f by a truncated (finite) sum of orthonormal series

$$f_J(x) = \sum_{j=1}^{J} \theta_j \varphi_j(x), \tag{2.5}$$

and replace the coefficients θ_j by their estimators. Here J is a positive integer called *cutoff*, which is a *smoothing parameter*, i.e., a parameter whose choice is crucial for establishing the balance between bias and variance. The adaptive choice of J is discussed in *Nonparametric Curve Estimation* (Efromovich 1999a).

The orthonormal series estimation is also known as *projection estmation* in the literature, as equation (2.5) is the projection of f on the linear span of the first J functions of the basis $\{\varphi_0, \varphi_1, ...\}$. Through the projection (2.5), it will be possible to identify the basis elements that play the most important role in describing the function f. The basis elements $\{\varphi_0, \varphi_1, ...\}$ contain complementary properties which allow the decomposition of the underlying components contained in the function f.

The choice of a basis depends on the type of underlying information the researcher wants to identify from the function under inverstigation. The orthonormal bases that are most frequently used in series estimation include the trigonometric bases, the wavelet bases and ploynomial bases. Let us describe two classes of orthonormal bases which will be the mathematical tools we use in our proposed methodology.

A *cosine orthonormal basis* $\{\varphi_0, \varphi_1, ...\}$ on [0, 1] is defined by

$$\varphi_0(x) := 1, \text{ and } \varphi_j(x) := \sqrt{2}\cos(\pi j x), \text{ for } j = 1, 2, ...$$

where $x \in [0, 1]$.

A *wavelet orthonormal basis* $\{\psi_{jk}(x), j, k \in \mathbb{Z}\}$ on $\mathbb{R}$ is defined by scaling and shifting a single *Mother* function $\psi(x)$ as following

$$\psi_{jk}(x) := 2^{j/2}\psi(2^j x - k), \quad j, k \in \mathbb{Z}, \tag{2.6}$$

Here $\psi(x)$ is a sufficiently smooth function with compact support. Then any function $f \in L_2(\mathbb{R})$ can be represented

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \theta_{jk}\psi_{jk}(x), \tag{2.7}$$

we have

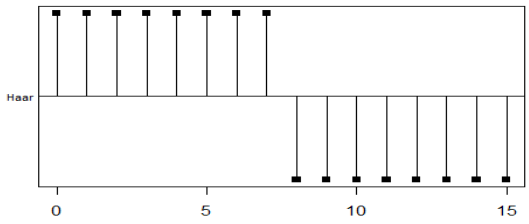
$$\theta_{jk} = \int_{-\infty}^{\infty} f(x)\psi_{jk}(x)dx. \tag{2.8}$$

We refer to $\{\theta_{jk}\}$ as *wavelet coefficients*. The wavelet orthonormal basis on $[0, 1]$ is defined in the similar way only with correction at the interval endpoints.

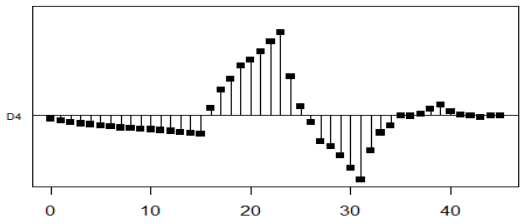
Both cosine basis and wavlet basis are useful mathematical tools. A cosine basis could dissection f with respect to frequency, while a wavelet basis extracts information from f both in the frequency domain and time domain if x is considered as a variable of time. As we can see from the definition of wavelet basis (2.6), the index j, called the *dilation* (scale) index, characterizes the frequency of the wavelet base function and the index k, called the *translation* index, characterizes the position of the wavelet base function in the time domain, i.e., the shift on the x-axis. More details on the wavlet transform will be introduced in the next section.

2.3 Wavelets and Multiresolution Based Discrete Wavelet Transforms

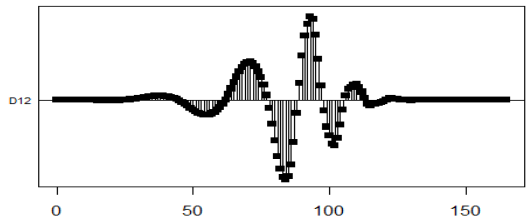
A wavelet is a wave-like function defined on a compact support whose integral is zero. The first wavelet family or basis is the *Haar* wavelet, which was introduced by Alfred Haar in 1910 (Haar 1910). The invention of wavelets by Meyer boosted the popularity of wavelets in applications since 1990s. Other main contributors in the history of wavelets include Stephane Mallat, Ingrid Daubechies and Ronald Coifman.



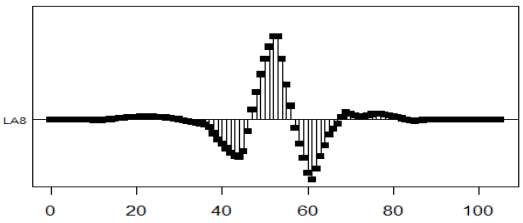
(a) Haar Wavelet



(b) Daubechies 4



(c) Daubechies 12



(d) Symlet 8

Figure 2.1. Examples of wavelet families.

Figure 2.1 presents examples of wavelet families. Figure 2.1 (a) displays the shape of a Haar wavelet. Figure 2.1 (b) and (c) display the shapes of Daubechies 4 and Daubechies 12, respectively. The *Daubechies* wavelets are designed to achieving special properties in vanishing moments and they are widely used in signal processing. Figure 2.1 (d) presents the shape of *Symlet 8*, also known as *Least Asymmetric* 8. The Symlet wavelets are designed by increasing the symmetry of Daubechies wavelets. It is one of the most commonly used wavelets in statistical analysis. The number associated with the wavelet name is an index of its width and smoothness. Larger index yields wider and smoother wavelets. Although the majority of the commonly used wavelets have no closed mathematical expression, such as the famouse Daubechies wavelets, they can be generated numerically using the cascade algorithm.

The definition of a wavelet orthonormal basis, introduced in (2.6),

$$\psi_{jk}(x) := 2^{j/2}\psi(2^j x - k), \quad j,k \in \mathbb{Z},$$

indicates that the entire basis is generated by scaling and translating just one function $\psi(x)$, called the *Mother wavelet*. This means that each base function $\psi_{jk}(x)$ has the same shape as the Mother wavelet $\psi(x)$ but with different scale and different location determined by the two indices j and k. Recall that j is the dilation (scale) index, and k is the translation index; j characterizes the frequency of the basis element which k characterizes the location. Such a system of wavelet functions $\{\psi_{jk}(x), j,k \in \mathbb{Z}\}$ can be shown to be orthonormal. Using this wavelet orthonormal basis, any function $f \in L_2(\mathbb{R})$ can be represented as

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \theta_{jk}\psi_{jk}(x), \qquad \theta_{jk} = \int_{-\infty}^{\infty} f(x)\psi_{jk}(x)dx. \tag{2.9}$$

This is called a *discrete wavelet transform*, as the scale and shift parameters are discrete.

The popularity of wavelets is due to their ability to extract information from a function or a signal in both time and frequency domains. Its mechanism is described as follows. Mathematically, a wavelet coefficient θ_{jk}, defined in (2.9), reflects the correlation between the signal and corresponding wavelet base function ψ_{jk}. A large wavelet coefficient implies the unknow signal contains underlying information of the corresponding frequencey at the corresponding location. The ability to extract underlying information from a signal in both time and frequency domains allows wavelets to study the characteristics of a signal like sudden spikes, frequency alteration, and discontinuities in details. This ability makes them an outstanding mathematical tool for approximation of spatially inhomogeneous curves and images.

A multiresolution analysis was introduced to avoid numerical computational complexity in the discrete wavelet transforms by reducing the number of elements in the wavelet basis and therefore reducing the number of wavelet coefficients, i.e. the corresponding integrals,

to evaluate. The idea of a multiresolution analysis is to introduce an auxiliary function, the so-called *Father* or *Scaling wavelet*, denoted as $\phi(x)$. The Father wavelet is coarser than the Mother wavelet and it is not integrated to zero. Therefore, the Father wavelet is used to extract the slowly varying information, or trend, from a signal f. Note that the multiresolution analysis is not applicable to all discrete wavelet transforms.

We next introduce the multiresolution based discrete wavelet transform. Let $\phi(x)$ be a Father wavelet and let $\psi(x)$ be a Mother wavelet. For any positive integer j_0, known as the *number of multiresolution components*, the system of functions on $\mathbb{R}$ defined by

$$\phi_{j_0k}(x) := 2^{j_0/2}\psi(2^{j_0}x - k), \;\; k \in \mathbb{Z}, \quad \text{and} \quad \psi_{jk}(x) := 2^{j/2}\psi(2^j x - k), \;\; j \geq j_0, k \in \mathbb{Z}, \tag{2.10}$$

forms an orthogonal basis, which generates the same space as the one which is spanned by the system of functions defined in (2.6). For any function $f \in L_2(\mathbb{R})$, the *wavelet multiresolution expansion* of f is

$$f(x) = \sum_{k\in\mathbb{Z}} s_{j_0k}\phi_{j_0k}(x) + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} \theta_{jk}\psi_{jk}(x), \tag{2.11}$$

where $s_{j_0k} = \int_{-\infty}^{\infty} f(x)\phi_{j_0k}(x)dx$ and $\theta_{jk} = \int_{-\infty}^{\infty} f(x)\psi_{jk}(x)dx$ are called *wavelet coefficients*. The set of wavelet coefficients is called the *discrete wavelet transform* (DWT) of the function $f(x)$.

Additional background on wavelets and wavelet transform can be found in Vidakovic and Mueller (1994); Hernández and Weiss (1996); Mohlenkamp and Pereyra (2008).

2.4 Empirical Wavelet Coefficients and Thresholding Estimation

The wavelet multiresolution expansion introduced in (2.11) is defined for a continuous function $f \in L_2(\mathbb{R})$. However, in practical applications, a signal is only observed and recorded at certain discrete time points, which means we only obtain a discrete signal or a sample of the unobservable continuous signal. Therefore, the two indices j and k in an empirical

wavelet transform will not exceed certain levels that are specified by the observed signal length. Moreover, the observations of a signal are usually not pure but contaminated by noise. That is, in practice, we are dealing with problems that could be described using a regression model as defined in (2.2), where f is the underlying unobservable signal of interest, n is the observed signal length, ϵ_l is the noise or random component, and Y_l is the observation at time l/n. This fact induces a corresponding random component in the empirical wavelet coefficients, and shortly we will discuss this random component.

Due to the algorithm design of the numerical wavelet transform, most wavelet transform toolkits require an equidistantly sampled time series of a *dyadic* length, that is $n = 2^J$ for some positive integer J, as the algorithm input. Under the assumption that n is divisible by 2^{j_0}, where j_0 is the number of multiresolution components (or *scales*) predetermined by the researcher, the multiresolution based wavelet expansion for a time series is defined by the following

$$Y_l = \sum_{k=1}^{n/2^{j_0}} \tilde{s}_{j_0 k}\phi_{j_0 k}(l/n) + \sum_{j=1}^{j_0}\sum_{k=1}^{n/2^j} \tilde{\theta}_{jk}\psi_{jk}(l/n). \tag{2.12}$$

Here, $\{Y_l, l = 1, ..., n\}$ is the observed time series. The system of functions $\{\phi_{jk}(x) := 2^{-j/2}\phi(2^{-j}x - k), k = 1, \ldots, n/2^{j_0}; \psi_{jk}(x) := 2^{-j/2}\psi(2^{-j}x - k), j = 1, \ldots, j_0, k = 1, \ldots, n/2^j\}$ forms an orthogonal wavelet basis, where $\phi(x)$ is called the Father wavelet, and $\psi(x)$ is called the Mother wavelet. The index j is called the dilation (scale) index which characterizes the frequency of the wavelet base function, and the index k is called the translation index which characterizes the shift of the wavelet base function on the x-axis. Here, $\{\tilde{s}_{j_0 k}, k = 1, ..., n/2^{j_0}; \tilde{\theta}_{jk}, j = 1, ..., j_0, k = 1, ..., n/2^j\}$ are traditionally referred to as *empirical wavelet coefficients* (Johnstone and Silverman 1997; Efromovich 1999a; Nason 2008). Note that we use the notation $\tilde{s}_{j_0 k}$ and $\tilde{\theta}_{jk}$ with *tilde* to emphasize that these wavelet coefficients contain random components inherited from the observations $\{Y_l, l = 1, ..., n\}$.

On the other hand, the underlying signal of interest has the corresponding wavelet expansion given by

$$f(l/n) = \sum_{k=1}^{n/2^{j_0}} s_{j_0 k}\phi_{j_0 k}(l/n) + \sum_{j=1}^{j_0}\sum_{k=1}^{n/2^j} \theta_{jk}\psi_{jk}(l/n). \tag{2.13}$$

Here $\{s_{j_0 k}, k = 1, ..., n/2^{j_0}; \theta_{jk}, j = 1, ..., j_0, k = 1, ..., n/2^j\}$ are the underlying wavelet coefficients.

Under mild conditions, the fixed design regression model (2.2) can be converted by the wavelet tranform (2.12) and (2.13) into the Gaussian sequence model with respect to the wavelet coefficients,

$$\tilde{\theta}_{jk} = \theta_{jk} + n^{-1/2}\tau\xi_{jk}, \tag{2.14}$$

where $\{\xi_{jk}, j = 1, ..., j_0, k = 1, ..., n/2^j\}$ are i.i.d. Gaussian noise components with zero mean and unit varance and τ is the standard deviation of the noise in the original regression model (2.2). Note that wavelet transform has a "decorrelating" feature. That is, even though the observed signal may be strongly autocorrelated, the wavelet transform will very often yield much less correlated wavelet coefficients (Johnstone and Silverman 1997).

The underlying noiseless wavelet coefficients are usually a sparse representation of the underlying signal f, that is, only a small number of the wavelet coefficients are effectively nonzero compared to the signal length. One way of estimating a sparse set of underlying wavelet coefficients is to "threshold" the empirical wavelet coefficients by comparing them with the Gaussian noise background. This estimation approach is called *wavelet thresholding*, where each empirical wavelet coefficient is compared with a threshold whose value depends on the background noise level, so that we can decide whether the nonzero value of the empirical wavelet coefficient is due to a corresponding feature of the underlying signal or it is purely a result of the background noise.

There are two types of thresholding functions that are commonly used. The *hard thresholding function* is defined by

$$h_H(\tilde{\theta}, \lambda) = \begin{cases} \tilde{\theta}, & \text{if } |\tilde{\theta}| > \lambda, \\ 0, & \text{if } |\tilde{\theta}| \leq \lambda, \end{cases} \tag{2.15}$$

where $\tilde{\theta}$ is the empirical wavelet coefficient, and λ is the threshold level. The idea of hard thresholding is to "kill" all empirical wavelet coefficients if they are less than the threshold level. It keeps the amplitude of the signal but hurts the smoothness. In contrast to hard thresholding, the *soft thresholding function* "shrinks" the empirical wavelet coefficients by λ towards zero,

$$h_S(\tilde{\theta}, \lambda) = \begin{cases} \tilde{\theta} - \lambda, & \text{if } \tilde{\theta} > \lambda, \\ \tilde{\theta} + \lambda, & \text{if } \tilde{\theta} < -\lambda, \\ 0, & \text{if } |\tilde{\theta}| \leq \lambda. \end{cases} \tag{2.16}$$

A commonly used threshold is

$$\lambda = n^{-1/2}\tau\sqrt{2\ln(n)}, \tag{2.17}$$

called "Universal" threshold. According to our notation, τ is the noise level of the original signal (2.2) while $n^{-1/2}\tau$ is the corresponding noise level in the empirical wavelet coefficients (2.14). A threshold proportional to $\sqrt{2\ln(n)}$ is considered to be "conservative", because

$$\lim_{n\to\infty} P\left(\max_{1\leq i\leq n} |\xi_i| > \sqrt{2\ln(n)}\right) = 0,$$

where $\xi_1, ..., \xi_n$ are, not neccessarily independent, standard Gaussian random variables. Thus, if the sample size is large, it is very unlikely that the nonzero empirical wavelet coefficients due to pure noise will "survive" the threshold. In that sense, the "Universal" thresholding provides "noise-free" estimation.

To apply any denoising procedure, such as the "Universal" thresholding (2.17), we need to know the standard deviation τ of the noise in empirical wavelet coefficients (2.14). A standard method, used by wavelet denoising algorithms and recommended in the literature, is to assume that the wavelet transform on the finest first scale is purely a result of random noise and the standard deviation $n^{-1/2}\tau$ could be estimated using the finest scale via a robust procedure, say the sample median, which is then rescaled into the empirical standard deviation. One approach is to use a robust estimator such as

$$\hat{\tau} = n^{1/2} mad(\theta_{1k}, k = 1, ..., n/2^j)/0.6745. \tag{2.18}$$

Here, *mad* represents *median absolute deviation*, which is defined as the median of the absolute deviations from the data's median,

$$mad(X_1, ..., X_n) = median_{i\in\{1,...,n\}}(|X_i - median_{j\in\{1,...,n\}}(X_j)|),$$

where $X_1, ..., X_n$ are realizations of a distribution. The factor 0.6745 in the formula (2.18) is approximately $\Phi^{-1}(3/4)$ ($\Phi(\cdot)$ is the c.d.f. of a standard normal distribution), and it is chosen for calibration with the normal distribution, i.e., $E\{mad(X_1, ..., X_n)/0.6745\} = \sigma$ if $X_1, ..., X_n$ are normally distributed with standard deviation σ. Other estimates of τ could be used; see Ogden (1996); Vidakovic (1999); Efromovich (1999a,b); Nason (2008).

2.5 Minimax Risk and Rate Optimal Estimation

A nonparametric estimation problem is characterized by the following three components:

(i) A nonparametric class of estimands, Θ. For instance, we will discuss a problem of estimation of large sparse cross-covariance matrices in Chapter 4, where we introduce a class of sparse $p_1 \times p_2$ matrices

$$\mathcal{U}_q(s_0(p_1, p_2)) := \Big\{\{\sigma_{rl}\} : \quad \max_{r\in\{1,2,...,p_1\}} \sum_{l=1}^{p_2} (\sigma_{rr}\sigma_{ll})^{(1-q)/2} |\sigma_{rl}|^q \le s_0(p_1, p_2),$$

$$\max_{r,l} |\ln(\sigma_{rr}/\sigma_{ll})| \leq C_0 < \infty, \;\; q \in [0,1)\Big\}. \tag{2.19}$$

(ii) A family of probability measures $\{P_\theta, \theta \in \Theta\}$.

(iii) A distance $d(\cdot,\cdot)$ which is used to quantify the loss. For instance, the l_1-norm for a $p_1 \times p_2$ matrix $A = (a_{ij})_{p_1 \times p_2}$ is defined by

$$||A||_1 := \max_{i \in \{1,2,...,p_1\}} \sum_{j=1}^{p_2} |a_{ij}|,$$

and the corresponding distance between two $p_1 \times p_2$ matrices A and B is defined by

$$d(A,B) := \max_{i \in \{1,2,...,p_1\}} \sum_{j=1}^{p_2} |a_{ij} - b_{ij}|.$$

An important concept of performance measure in nonparametric estimation is the so-called *minimax*. The idea of minimax estimation is to choose the estimator which minimize the risk in the worst case scenario. We will describe this concept in a general framework. Suppose there are n observations $X_1, ..., X_n$. A statistic $\hat{\theta}_n := \hat{\theta}_n(X_1, ..., X_n)$ is introduced as an estimator of θ. The *maximum risk* of $\hat{\theta}_n$ on (Θ, d) is defined as

$$r_w(\hat{\theta}_n) := \sup_{\theta \in \Theta} E_\theta[w(\psi_n^{-1} d(\hat{\theta}_n, \theta))].$$

Here, $\{\psi_n, n = 1, 2, ...\}$ is a positive sequence and $\psi_n \to 0$ as $n \to \infty$. $w(\cdot)$ is a monotone increasing function mapping from $[0,\infty)$ to $[0,\infty)$, and $w(0) = 0$. A commonly used example of $w(\cdot)$ is $w(u) = u^p, p > 0$. The *minimax risk* on (Θ, d) is defined by

$$R_n := \inf_{\hat{\theta}_n} r_w(\hat{\theta}_n), \tag{2.20}$$

where the infimum is over all estimators $\hat{\theta}_n$ based on a sample of size n. The positive sequence $\{\psi_n, n = 1, 2, ...\}$ is called an *optimal rate of convergence* of estimators on (Θ, d) if there exists two positive constants C and c such that

$$\limsup_{n \to \infty} R_n \leq C \quad \text{and} \quad \liminf_{n \to \infty} R_n \geq c.$$

An estimator $\tilde{\theta}_n^*$ is called *rate optimal* on (Θ, d) if there exists a constant $C' > 0$ such that

$$r_w(\tilde{\theta}_n^*) = \sup_{\theta \in \Theta} E_\theta[w(\psi_n^{-1} d(\tilde{\theta}_n^*, \theta))] \leq C',$$

where $\{\psi_n, n = 1, 2, ...\}$ is the optimal rate of convergence. See Tsybakov (2009) for an introduction to nonparametric estimation.

2.6 Regression with Errors in Both Variables

A "standard" regression model assumes no measurement error in the explanatory variables. If this assumption is violated, then the "standard" estimator, i.e., the ordinary least squares (OLS) estimator, is no longer consistent. That is, the estimators do not tend to the true parameter values as the sample size increases to infinity. Therefore, the "standard" regression model is generalized to take the measurement errors in the explanatory variables into account. Such model is called *regression with errors in variables* (EIV), also known as the *measurement error model*. In a regression model with EIV, the predictor is considered as a random variable just like the response. Estimation and statistical inference of this model are very different from that of the "standard" regression model.

A regression with EIV model can be described using latent variables in general. Let Y be the response variable and X be the explanatory variable, both of which are random variables. It is assumed that there is a latent variable Y^* associated with Y and a latent variable X^* associated with X such that there exists some functional relationship $g(\cdot)$ between the two latent variables X^* and Y^*. The variables Y and X can be observed; suppose $\{(X_i, Y_i), i = 1, ..., n\}$ are i.i.d. observations from the random variables (X, Y). Here, $\{(X_i, Y_i), i = 1, ..., n\}$ can be considered as the noisy observations of a latent sequence $\{(X_i^*, Y_i^*), i = 1, ..., n\}$ which

are not observed. Using these notations, a regression with EIV model is defined as follows

$$\begin{cases} X_i = X_i^* + \epsilon_i, \\ Y_i = Y_i^* + \delta_i, \\ Y_i^* = g(X_i^*|\theta), \end{cases} \tag{2.21}$$

where θ is the parameter of the functional relationship, ϵ_i and δ_i are independent random components.

There are two approaches to modeling the latent sequence $X_1^*, ..., X_n^*$. They can be considered as unknown constants in which case the corresponding model is called a *functional model*. Or they can be regarded as a random sequence and the corresponding model is called a *structural model*. We will only introduce the estimation of simple linear models here.

A *simple linear functional relationship model* can be described as

$$\begin{cases} Y_i = \alpha + \beta X_i^* + \delta_i, & \delta_i \overset{i.i.d.}{\sim} N(0, \sigma_\delta^2), \\ X_i = X_i^* + \epsilon_i, & \epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2), \end{cases} \tag{2.22}$$

where δ_i and ϵ_i are independent random components. Remember that X_i^* are constants in a functional relationship (2.22). Thus, $E(X_i) = X_i^*$ and the functional relationship can be represented as $E(Y_i) = \alpha + \beta E(X_i)$. For a *simple linear structural relationship*, the model (2.21) can be simplified into

$$\begin{cases} Y_i = \alpha + \beta X_i^* + \delta_i, & \delta_i \overset{i.i.d.}{\sim} N(0, \sigma_\delta^2), \\ X_i = X_i^* + \epsilon_i, & \epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2), \\ X_i^* \overset{i.i.d.}{\sim} N(x^*, \sigma_{x^*}^2), \end{cases} \tag{2.23}$$

where δ_i and ϵ_i are independent random components, and they are also independent of X_i^*.

Under the assumption that $\sigma_\epsilon^2 = \lambda\sigma_\delta^2$ where λ is known, the maximum likelihood estimates for the coefficients α and β are same for both functional relationship model and structural

relationship model,

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \text{ and } \hat{\beta} = \frac{-(S_{xx} - \lambda S_{yy}) + \sqrt{(S_{xx} - \lambda S_{yy})^2 + 4\lambda S_{xy}^2}}{2\lambda S_{xy}}, \tag{2.24}$$

where $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2, S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$. When $\lambda = 1$, the MLE estimators agree with the *orthogonal least squares* estimators. The MLE estimators for σ_ϵ^2 are different for these two models. For a functional relationship model,

$$\hat{\sigma}_\epsilon^2 = \frac{\lambda}{2n(1 + \lambda\hat{\beta}^2)} \sum_{i=1}^{n}(y_i - (\hat{\alpha} + \hat{\beta}x_i))^2. \tag{2.25}$$

Note that $\hat{\sigma}_\epsilon^2$ is not consistent, and $\hat{\sigma}_\epsilon^2$ is convergent in probability to $\frac{1}{2}\sigma_\epsilon^2$. For a structural relationship model, the variance estimators are

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\left(S_{xx} - \frac{S_{xy}}{\hat{\beta}}\right), \ \hat{\sigma}_\delta^2 = \frac{\hat{\sigma}_\epsilon^2}{\lambda}, \text{ and } \hat{\sigma}_{x^*}^2 = \frac{1}{n}\frac{S_{xy}}{\hat{\beta}}.$$

Note that these estimators are all consistent. See details in Casella and Berger (2002).

Finally, note that the variance of X_i is different in these two models. In the functional relationship model, $Var(X_i) = \sigma_\epsilon^2$, while in the structural relationship model, $Var(X_i) = \sigma_\epsilon^2 + \sigma_{x^*}^2$

2.7 Subexponential Distributions and Tail Bounds

One commonly used heavy-tailed distribution family is the subexponential family. A random variable X with mean $\mu = E(X)$ is called *subexponential* with parameters (ν^2, b) if there exists constants $\nu > 0$ and $b > 0$ such that

$$E\{\exp(\lambda(X - \mu))\} \leq \exp\left(\frac{\nu^2\lambda^2}{2}\right), \text{ for all } |\lambda| < \frac{1}{b}. \tag{2.26}$$

The most familiar example of subexponential distributions is the chi-square distribution. Suppose X follows chi-square distribution with one degree of freedom and therefore $E(X) = 1$. With some algebra, we can show that

$$E\{\exp(\lambda(X - 1))\} = \frac{\exp(-\lambda)}{\sqrt{1 - 2\lambda}} \leq \exp(\frac{4\lambda^2}{2}), \text{ for all } |\lambda| < \frac{1}{4}.$$

This implies that the chi-square random variable X is subexponential with parameters $(4, 4)$.

In statistical inference, it is always of interest to obtain probability bounds on the tails of a distribution. Let us introduce the *subexponential tail bounds*. If a random variable X is subexponential with parameters (ν^2, b), then

$$P(X \geq \mu + t) \leq \begin{cases} \exp\left(-\frac{t^2}{2\nu^2}\right), & \text{if } 0 \leq t \leq \frac{\nu^2}{b}, \\ \exp\left(-\frac{t}{2b}\right), & \text{if } t > \frac{\nu^2}{b}, \end{cases} \tag{2.27}$$

where $\mu = E(X)$. We can see from (2.27) that when t is small, a subexponential distribution and a Gaussian distribution have the same probability bounds; when t is large, a subexponential distribution has a larger tail bound than a Gaussian distribution. The similar probability bounds apply on the left tail

$$P(X \leq \mu - t) \leq \begin{cases} \exp\left(-\frac{t^2}{2\nu^2}\right), & \text{if } 0 \leq t \leq \frac{\nu^2}{b}, \\ \exp\left(-\frac{t}{2b}\right), & \text{if } t > \frac{\nu^2}{b}, \end{cases} \tag{2.28}$$

We next describe the statistical properties of a sum of independent sub-exponential random variables. Suppose $X_1, ..., X_n$ are independent random variables where each X_i is subexponential (ν_i^2, b_i) with mean μ_i. Then $\sum_{i=1}^n X_i$ is subexponential with (ν'^2, b'), where $\nu'^2 = \sum_{i=1}^n \nu_i^2$ and $b' = \max_{1 \leq i \leq n} b_i$. Together with formula (2.27) and (2.28), this implies that

$$\max\left(P(\sum_{i=1}^n (X_i - \mu_i) \leq -t), P(\sum_{i=1}^n (X_i - \mu_i) \geq t)\right) \leq \begin{cases} \exp\left(-\frac{t^2}{2\nu'^2}\right), & \text{if } 0 \leq t \leq \frac{\nu'^2}{b'}, \\ \exp\left(-\frac{t}{2b'}\right), & \text{if } t > \frac{\nu'^2}{b'}. \end{cases} \tag{2.29}$$

CHAPTER 3

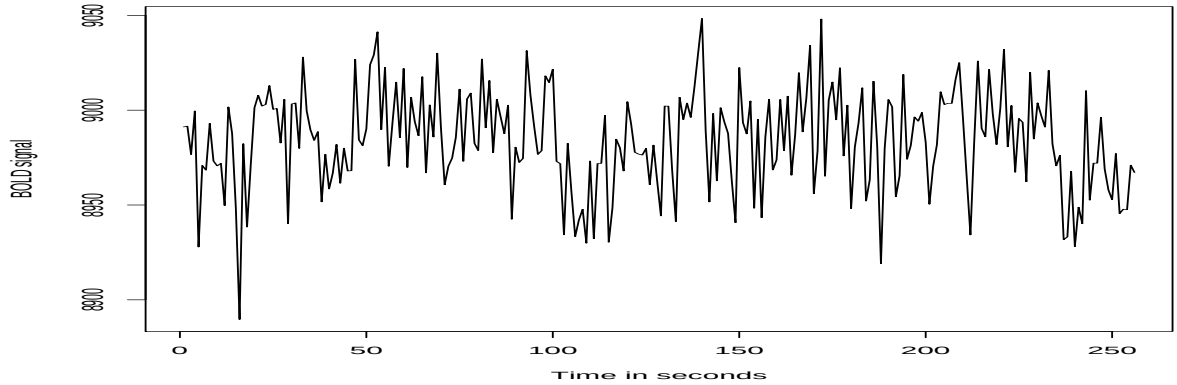
WAVELET METHODOLOGY FOR NEURAL PLASTICITY

3.1 Methodology

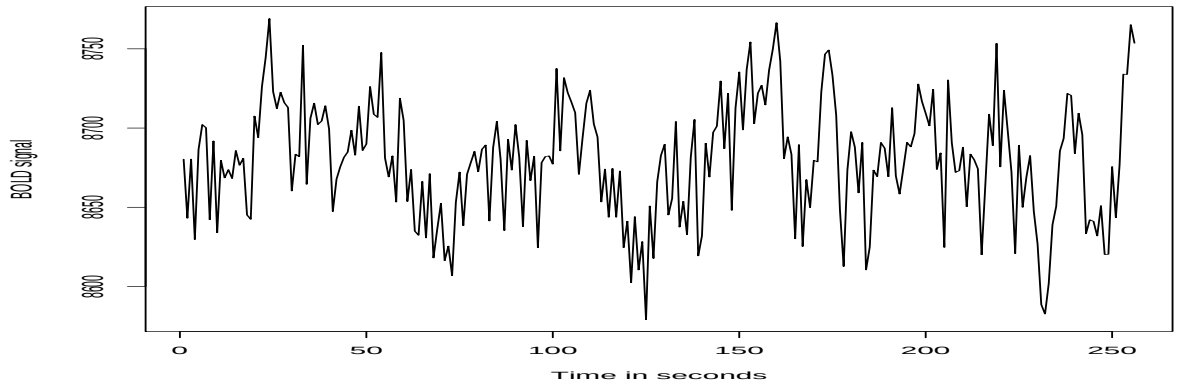
Let us consider a general neural plasticity study, where our aim is to investigate the alterations in the inter-hemispheric neural connectivity due to a certain training. Suppose the experiment involves three stages: pre-training, training (or task), and post-training. The pre-training and post-training states are also known as *resting states*. Suppose the fMRI data in the region of interest (ROI) is available. The ROI in the right and left hemispheres contain p_1 and p_2 voxels, respectively. The fMRI data is collected every second at each voxel and the sample size (signal length) during each scan run is n. The above is also a summary of the fMRI data in Tung et al. (2013) and we will apply our proposed method on this data set in Chapter 5. A detailed description of the experiment in Tung et al. (2013) can be found in Chapter 5. Examples of observed resting-state signals are presented in Figure 3.1.

In particular, during the motor task designed in Tung et al. (2013), a subject is asked to use his right thumb to press a button. Left motor cortex is responsible for this movement, and hence we may expect to see pronounced BOLD signals in some voxels in the left hemisphere. If there is a connectivity between the right and left hemispheres, then we may see a similar activity among voxels in the right hemisphere. We would like to understand how neurons in left and right hemispheres of each participant are connected and work together during two resting and four training runs. In this work, our aim is to define and quantify this connectivity, suggest methods of its estimation and then develop feasible methods for statistical inference. In what follows it is convenient to first describe main steps of the proposed statistical analysis and then justify them.

We say that a right-hemisphere's voxel is connected to and activated by a left-hemisphere's voxel if components of BOLD signals, representing neural activity, resemble each other. We



(a) Pre-training



(b) Post-training

Figure 3.1. Available BOLD signals from one voxel during pre- and post-training resting runs.

refer to such signal components as *BOLD-components of interest*. Since BOLD signals depend on the blood-oxygen-level, they contain components that have nothing to do with the neural activities. For example, they may be affected by the cardiac activity, respiration and stochastic noise in the background of the MRI scanner. Our approach to addressing these problems is described next.

The **first step** should be a procedure for extracting a BOLD-component of interest from a BOLD signal. This will be done using a special wavelet decomposition and a special

denoising procedure proposed for the fMRI data; see Section 3.2 and Section 3.3 . We will show examples of BOLD-components of interest in Section 3.4. The **second step** is to quantify a relationship between a pair of BOLD-components of interest in left and right hemispheres. This is done using the definition of cross-covariance and cross-correlation; see Section 3.4. In the **last step**, for the data at hand, it is proposed to estimate cross-correlation matrices for possible pathways between left and right hemispheres for the 6 runs of fMRI experiments with 24 participants. This requires estimation of and inference for 6 matrices with dimension 1000×1000 ($p_1 \times p_2$) for each participant in the fMRI study.

3.2 Wavelet Decomposition of an fMRI Signal

The available fMRI data were preprocessed using the standard software AFNI. Physiological fluctuations in the signal time course were removed by regressing out time courses of the whole brain white matter and cerebral spinal fluid (Tung et al. 2013). However, as we will see in Chapter 5, this preprocessing is not perfect and we still observe physiological and MRI induced components in the BOLD signals. In this and next Sections, it is explained how a proposed wavelet denoising method deals with this issue.

For a particular voxel, the BOLD signal is observed every second with 300 observations for the pre- and post-training fMRI runs, and 340 observations for the training runs. Wavelet analysis is based on a dyadic number n of observations, and we choose $n = 256 = 2^8$ observations for each run. Further, we may rescale signals for all runs to the unit interval $[0, 1]$. Then an observed continuous BOLD signal $\tilde{Y}(t)$, $t \in [0, 1]$ for an fMRI run may be written as

$$\tilde{Y}(t) = \sum_{k=1}^{n/2^{j_0}} \tilde{s}_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=1}^{j_0} \sum_{k=1}^{n/2^j} \tilde{\theta}_{jk} \psi_{jk}(t), \tag{3.1}$$

where j_0 is the number of multiresolution components (or scales). We use five wavelet scales, i.e., $j_0 = 5$, in our application. Here $\phi_{j_0 k}(x) = 2^{-j_0/2} \phi(2^{-j_0} x - k)$, $k = 1, \ldots, n/2^{j_0}$ and

$\psi_{jk}(x) = 2^{-j/2}\psi(2^{-j}x - k)$, $j = 1, \ldots, j_0$, $k = 1, \ldots, n/2^j$, where $\phi(t)$ and $\psi(t)$ are called scaling (Father) and wavelet (Mother) functions in the wavelet analysis, and $t \in [0, 1]$. Recall that Father functions are coarser than Mother functions and are used to extract the slowly varying trend in the $\tilde{Y}$. Here $\{\tilde{s}_{j_0k}, k = 1, \ldots, n/2^{j_0}\}$ and $\{\tilde{\theta}_{jk}, j = 1, \ldots, j_0, k = 1, \ldots, n/2^j\}$ are empirical wavelet coefficients which can be obtained via a cascade algorithm supported by all wavelet software packages (Vidakovic 1999; Lazar 2008; Nason 2008).

Now we are in a position to explain how the wavelet expansion (3.1) allows us to define the BOLD-component of interest which reflects either a spontaneous neural activity during pre- or post-training runs or a hemodynamic response due to a stimulus during training runs (the motor signal). Each scale j in a wavelet multiresolution decomposition corresponds to a specific frequency. In our case, because a signal is observed every second, the corresponding frequency of each scale is 2^{-j} Hz. As a result, the frequency of the first scale is 0.5 Hz, the second is 0.25 Hz, the third is 0.125, etc. During rest runs, traditional bandpass for the BOLD-component of interest is 0.01-0.1 Hz, while hemodynamic responses during training runs last 15-20 seconds and the stimuli occur every 27-32 seconds (Lazar 2008; Birn 2012; Tung et al. 2013). Table 3.1 illustrates the detailed relationship among the wavelet scales, the freqencies and the signal components. As a result, the observed BOLD-component of interest $\tilde{B}(t)$ may be written as a combination of scales 3, 4 and 5, namely

$$\tilde{B}(t) = \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \tilde{\theta}_{jk}\psi_{jk}(t). \tag{3.2}$$

The low-frequency component $\sum_{k=1}^{n/2^{j_0}} \tilde{s}_{j_0k}\phi_{j_0k}(t)$ in the wavelet expansion (3.1) represents the fMRI background trend and is not of interest to us. The frequency of physiological fluctuations is greater than 0.25 Hz (Barrett et al. 2009), and this is why these fluctuations are captured by the two finest scales. While we are not interested in the signals on the two finest scales, their analysis will be the pivotal part of the proposed wavelet denoising procedure discussed in the next section.

Table 3.1. The frequency and corresponding component of an fMRI signal on each wavelet scale.

wavelet scale	frequency	resting	task
D1	0.5 Hz	cardiac	cardiac
D2	0.25 Hz	respiration	respiration
D3	0.125 Hz	component of interest	hemodynamic responses
D4	0.0625 Hz		
D5	0.03125 Hz		

3.3 Noise Model

Empirical wavelet coefficients $\tilde{\theta}_{jk}$ in the observed BOLD-component of interest $\tilde{B}(t)$, defined in (3.2), are contaminated by noise. Namely, if $B(t)$ is the underlying BOLD-component of interest and

$$B(t) = \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \theta_{jk} \psi_{jk}(t), \tag{3.3}$$

then the underlying wavelet coefficients θ_{jk} and the empirical $\tilde{\theta}_{jk}$ are related as follows,

$$\tilde{\theta}_{jk} = \theta_{jk} + n^{-1/2} \tau \xi_{jk}. \tag{3.4}$$

Here ξ_{jk} is a standard Gaussian noise (zero mean and unit variance normal random variable) and τ is the standard deviation of the wavelet noise. The interested reader can find a discussion of the model (3.2) – (3.4) in Johnstone and Silverman (1997); Efromovich (1999a); Nason (2008), and the model was tested in Efromovich and Valdez-Jasso (2010) for the same type of fMRI experiment.

As a result, to restore the underlying BOLD-component of interest $B(t)$, we need to develop a feasible denoising procedure. There are a number of denoising procedures proposed in the wavelet literature that allow us to estimate $B(t)$ based on $\tilde{B}(t)$, but they all

require knowledge of the parameter τ in (3.4). A standard estimation procedure, used by wavelet software and recommended in the literature, is to assume that a signal on the finest first scale is a pure stochastic noise. Its standard deviation is then estimated via a robust procedure, say by the sample median, and then rescaled into the empirical standard deviation. The corresponding procedure is called *mad* which was used for wavelet analysis of fMRI data in Efromovich and Valdez-Jasso (2010) and Efromovich and Smirnova (2014). A detailed description of the *mad* procedure can be found in (2.18). While this approach was successful in Efromovich and Valdez-Jasso (2010), the outcome of Efromovich and Smirnova (2014) was less encouraging. The source of the difference will be explained shortly. Other methods of estimating τ proposed in the literature by Ogden (1996); Vidakovic (1999); Efromovich (1999a,b); Nason (2008) were also explored and the outcomes were similar to the *mad* procedure.

To address the issue of estimation of parameter τ (the standard deviation of the noise in empirical wavelet coefficients), let us first explain the nature of the noise. First of all, following Johnstone and Silverman (1997); Efromovich (1999a); Nason (2008), let us recall that in a classical (and ideal) theoretical model, we are dealing with observations of a regression model

$$Y_l = f(l/n) + \tau\zeta_l, \quad l = 1, \dots, n,$$

where ζ_l are independent standard normal random variables and $f(t)$ is a regression function of interest. Then, if θ_{jk} are wavelet coefficients of $f(t)$, then the corresponding empirical wavelet coefficients $\tilde{\theta}_{jk}$ calculated from Y_l, $l = 1, \dots, n$, satisfy the expression (3.4) with ξ_{jk} independent standard normal random variables. Keeping this in mind, the only chance for the *mad* procedure to be consistent is if on the first scale all underlying wavelet coefficients

$$\{\theta_{1k}, k = 1, ..., n/2^j\}$$

are zero. In other words, the assumption is that the first scale contains only white Gaussian noise and no deterministic component.

Keeping this in mind, let us return to the wavelet expansion (3.1). If the first scale of the wavelet decomposition contains some physiological components, then the *mad* procedure may overestimate the parameter τ in (3.4). And indeed, despite the above-discussed denoising procedures, it is shown in Efromovich and Valdez-Jasso (2010) that the first and second scales may contain physiological components due to respiratory (breathing) and cardiac (heart beats) activities since this is what allows the oxygen to be delivered to brain neurons. Further, cardiac and respiratory signals occur on frequencies larger than 0.25 Hz (Barrett et al. 2009), and this is why they may be present on the two finest scales. In Efromovich and Valdez-Jasso (2010) a unique fMRI experiment is discussed when BOLD signals are measured every 50 milliseconds so that no physiological signal is present on the finest scale. A discussion of physiological noise in fMRI can be found in Birn (2012); Chang and Glover (2009); He and Liu (2012); Murphy et al. (2013); Weissenbacher et al. (2009).

This explanation motivates the following idea of estimating τ. Denote by

$$\tilde{Y}'_l := \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \tilde{\theta}_{jk} \psi_{jk}(l/n)$$

the sum of first and second wavelet scales for a voxel of interest, and by

$$Y'_l := \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \theta_{jk} \psi_{jk}(l/n)$$

the corresponding underlying deterministic component, $l = 1, \ldots, n$. According to the above-presented explanation, the deterministic component may be explained by physiological factors and hence it should be close to the deterministic component

$$X'_l := \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \theta'_{jk} \psi_{jk}(l/n)$$

of a nearby voxel. More precisely it is reasonable to assume that

$$Y'_l = \beta X'_l, \tag{3.5}$$

(note that signals on the first and second scales are integrated to zero so there may be only the scale/slope term β in the relation). As a result, we may regress the observed $\tilde{Y}' := \{\tilde{Y}'_l, l = 1, 2 \ldots, n\}$ on the observed $\tilde{X}' := \{\tilde{X}'_l, l = 1, 2, \ldots, n\}$. Note that discrete wavelet transform is a linear operator. Therefore, relationship (3.5) holds for the corresponding wavelet coefficients and we can describe the problem as a regression model with errors in variable (EIV),

$$\begin{cases} \tilde{\theta}_{jk} = \beta\theta'_{jk} + n^{-1/2}\tau\xi_{jk}, & \xi_{jk} \overset{i.i.d.}{\sim} N(0,1), \\ \tilde{\theta}'_{jk} = \theta'_{jk} + n^{-1/2}\tau\xi'_{jk}, & \xi'_{jk} \overset{i.i.d.}{\sim} N(0,1), \end{cases} \tag{3.6}$$

where $j = 1, 2, k = 1, \ldots, n/2^j$, and ξ_{jk} and ξ'_{jk} are independent. Here, we assume the underlying wavelet coefficients $\{\theta'_{jk}, j = 1, 2, \ k = 1, \ldots, n/2^j\}$ are fixed constants, and therefore we are dealing with a functional relationship model. Then we estimate the standard deviation τ of the regression noise using maximum likelihood estimation,

$$\hat{\tau}_{MLE} = \sqrt{\frac{2}{3(1+\hat{\beta}^2)} \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} (\tilde{\theta}_{jk} - \hat{\beta}\tilde{\theta}'_{jk})^2}. \tag{3.7}$$

Here, $\hat{\beta}$ is the MLE of the coefficient β,

$$\hat{\beta} = \frac{-(S'_{xx} - S'_{yy}) + \sqrt{(S'_{xx} - S'_{yy})^2 + 4S'^2_{xy}}}{2S'_{xy}}, \tag{3.8}$$

where

$$S'_{xx} = \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} (\tilde{\theta}'_{jk})^2, S'_{yy} = \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} (\tilde{\theta}_{jk})^2, S'_{xy} = \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \tilde{\theta}'_{jk}\tilde{\theta}_{jk}. \tag{3.9}$$

Note that the noise model (3.6) is a *functional relationship model*, where $\hat{\tau}^2_{MLE}$ is convergent in probability to $\frac{1}{2}\tau^2$ (Casella and Berger, 2002). On the other hand, the number of empirical wavelet coefficients that are used to calculate the MLE is $\frac{3}{4}n$ (which is 192 in our case). Therefore, the proposed estimator of τ is

$$\hat{\tau} = \sqrt{2}\hat{\tau}_{MLE} = \sqrt{\frac{4}{3(1+\hat{\beta}^2)} \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} (\tilde{\theta}_{jk} - \hat{\beta}\tilde{\theta}'_{jk})^2}. \tag{3.10}$$

The assumption of this approach is that two "nearby" voxels have similar physiological fluctuations. Hence, we select a "nearby" voxel from those that adjoin the voxel of interest. In order to avoid the slice timing issue and to align with the assumption, it is suggested to choose a "nearby" voxel as the one with largest correlation among the 3×3 vicinity within the same slice. Let us also note that the choice of a larger vicinity, tested in simulations and on real data, made no difference. The above is the recommended approach for estimation of parameter τ.

Now, we will show that expressions (3.7) and (3.8) are maximum likelihood estimations of τ and β, respectively, given the model (3.6). First, we write the likelihood function,

$$L\left(\beta, \theta_{jk}, \tau \middle| (\tilde{\theta}_{jk}, \tilde{\theta}'_{jk}), j = 1, 2, k = 1, \ldots, n/2^j\right)$$

$$= \frac{1}{(2\pi)^{\frac{3}{4}n}} \frac{n^{\frac{3}{4}n}}{\tau^{\frac{3}{2}n}} \exp\left\{ -\sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \frac{(\tilde{\theta}'_{jk} - \theta_{jk})^2 + (\tilde{\theta}_{jk} - \beta\theta_{jk})^2}{2n^{-1}\tau^2} \right\}. \tag{3.11}$$

Second, let us maximize the likelihood with respect to θ_{jk}, $j = 1, 2, k = 1, \ldots, n/2^j$, which is equivalent to minimizing $\sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \{(\tilde{\theta}'_{jk} - \theta_{jk})^2 + (\tilde{\theta}_{jk} - \beta\theta_{jk})^2\}$. The function is convex and differentiable, and thus by taking derivatives, for each (j, k), we obtain

$$\theta^*_{jk} = \frac{\tilde{\theta}'_{jk} + \beta\tilde{\theta}_{jk}}{1 + \beta^2}. \tag{3.12}$$

Substitution of (3.12) into (3.11) gives

$$\max_{\theta_{jk}} L\left(\beta, \theta_{jk}, \tau \middle| (\tilde{\theta}_{jk}, \tilde{\theta}'_{jk}), j = 1, 2, k = 1, \ldots, n/2^j\right)$$

$$= \frac{1}{(2\pi)^{\frac{3}{4}n}} \frac{n^{\frac{3}{4}n}}{\tau^{\frac{3}{2}n}} \exp\left\{ -\frac{1}{2\tau^2 n^{-1}} \left(\frac{1}{1 + \beta^2} \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} (\tilde{\theta}_{jk} - \beta\tilde{\theta}'_{jk})^2 \right) \right\}. \tag{3.13}$$

Maximization of the likelihood (3.13) with respect to β is equivalent to minimization of

$$\frac{1}{1 + \beta^2} \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} (\tilde{\theta}_{jk} - \beta\tilde{\theta}'_{jk})^2 = \frac{1}{1 + \beta^2} (S'_{yy} - 2\beta S'_{xy} + \beta^2 S'_{xx}). \tag{3.14}$$

Here S'_{xx}, S'_{yy}, and S'_{xy} are defined in (3.9). Note that (3.14) is differentiable with respect to β and it tends to S'_{xx} as β approaches positive or negative infinity. The derivative of (3.14) with respect to β is

$$-\frac{2\beta}{(1+\beta^2)^2}(S'_{yy}-2\beta S'_{xy}+\beta^2 S'_{xx})+\frac{1}{1+\beta^2}(-2S'_{xy}+2\beta S'_{xx}). \tag{3.15}$$

Equating (3.15) to zero and simplifying the equation, we obtain

$$\beta^2 S'_{xy}+\beta(S'_{xx}-S'_{yy})-S'_{xy}=0. \tag{3.16}$$

Solving quadratic equation (3.16) for β, we obtain two possible values of β,

$$\beta_{\pm}=\frac{-(S'_{xx}-S'_{yy})\pm\sqrt{(S'_{xx}-S'_{yy})^2+4S'^2_{xy}}}{2S'_{xy}}. \tag{3.17}$$

Substitution of both solutions for β into (3.14), respectively, gives the following two results

$$S'_{xx}-\frac{\sqrt{(S'_{xx}-S'_{yy})^2+4(S'_{xy})^2}}{1+\beta_+^2} \quad \text{and} \quad S'_{xx}+\frac{\sqrt{(S'_{xx}-S'_{yy})^2+4(S'_{xy})^2}}{1+\beta_-^2}. \tag{3.18}$$

Obviously, the first term in (3.18) is the minimum value that is attained by (3.14), and hence $\hat{\beta}=\beta_+$ defined in (3.8) minimizes (3.14). Finally, we want to maximize

$$\max_{\beta,\theta_{jk}} L\left(\beta,\theta_{jk},\tau\middle|(\tilde{\theta}_{jk},\tilde{\theta}'_{jk}), j=1,2, k=1,\ldots,n/2^j\right)$$

$$=\frac{1}{(2\pi)^{\frac{3}{4}n}}\frac{n^{\frac{3}{4}n}}{\tau^{\frac{3}{2}n}}\exp\left\{-\frac{1}{2\tau^2 n^{-1}}\left(\frac{1}{1+\hat{\beta}^2}\sum_{j=1}^{2}\sum_{k=1}^{n/2^j}(\tilde{\theta}_{jk}-\hat{\beta}\tilde{\theta}'_{jk})^2\right)\right\}. \tag{3.19}$$

with respect to τ ($\tau>0$), which is equivalent to maximizing

$$t^{\frac{3}{4}n}\exp\left\{-\frac{nt}{2}\left(\frac{1}{1+\hat{\beta}^2}\sum_{j=1}^{2}\sum_{k=1}^{n/2^j}(\tilde{\theta}_{jk}-\hat{\beta}\tilde{\theta}'_{jk})^2\right)\right\}, \tag{3.20}$$

with respect to t ($t>0$). Note that (3.20) is differentiable with respect to t and it tends to 0 as t approaches 0 and positive infinity. The derivative of (3.20) is

$$\frac{3}{2}-t\left(\frac{1}{1+\hat{\beta}^2}\sum_{j=1}^{2}\sum_{k=1}^{n/2^j}(\tilde{\theta}_{jk}-\hat{\beta}\tilde{\theta}'_{jk})^2\right). \tag{3.21}$$

Equating (3.21) to zero, we obtain the MLE of t by

$$\hat{t}_{MLE} = \frac{3/2}{\frac{1}{1+\hat{\beta}^2}\sum_{j=1}^{2}\sum_{k=1}^{n/2^j}(\tilde{\theta}_{jk} - \hat{\beta}\tilde{\theta}'_{jk})^2}. \tag{3.22}$$

Due to the invariance property of MLEs,

$$\hat{\tau}_{MLE} = \sqrt{\frac{2}{3(1+\hat{\beta}^2)}\sum_{j=1}^{2}\sum_{k=1}^{n/2^j}(\tilde{\theta}_{jk} - \hat{\beta}\tilde{\theta}'_{jk})^2}. \tag{3.23}$$

3.4 Quantifying the Plasticity

Our second step in the proposed methodology is to quantify a relationship between a pair of BOLD-components of interest in left and right hemispheres. Figure 3.2 shows us four diagrams with pairs of BOLD-components of interest; the solid black and dashed red lines show BOLD-components of interest in the left and right hemispheres respectively (colored figures can be found in the online supplementary materials Efromovich and Wu (2018a)). The method of obtaining these signals will be explained shortly. The top diagram (a) exhibits signals of interest with the left voxel being from the 3rd slice and the right one from the 6th slice; hence, the time of delay between them is 57 ms (the slice timing issue is due to the mechanism of a MRI scanner where the human brain is scanned slice by slice sequentially). Despite the fact that voxels are from different slices and, of course, from different hemispheres, the BOLD-components of interest are remarkably similar. In information theory a (normalized) cross-correlation is a popular measure of similarity between two continuous signals. Namely, for continuous functions $f(t)$ and $g(t)$, $t \in [0, T]$, the (normalized) cross-covariance (at lag zero) is defined as

$$\sigma_{fg} := \int_0^T [f(t) - \int_0^T f(\tau)d\tau/T][g(t) - \int_0^T g(\tau)d\tau/T]dt,$$

and then the corresponding cross-correlation is defined as

$$\rho_{fg} := \sigma_{fg}/[\int_0^T (f(t) - \int_0^T f(\tau)d\tau/T)^2 dt \int_0^T (g(t) - \int_0^T g(\tau)d\tau/T)^2 dt]^{1/2}.$$

Note that these definitions resemble classical covariance and correlation for two random variables.

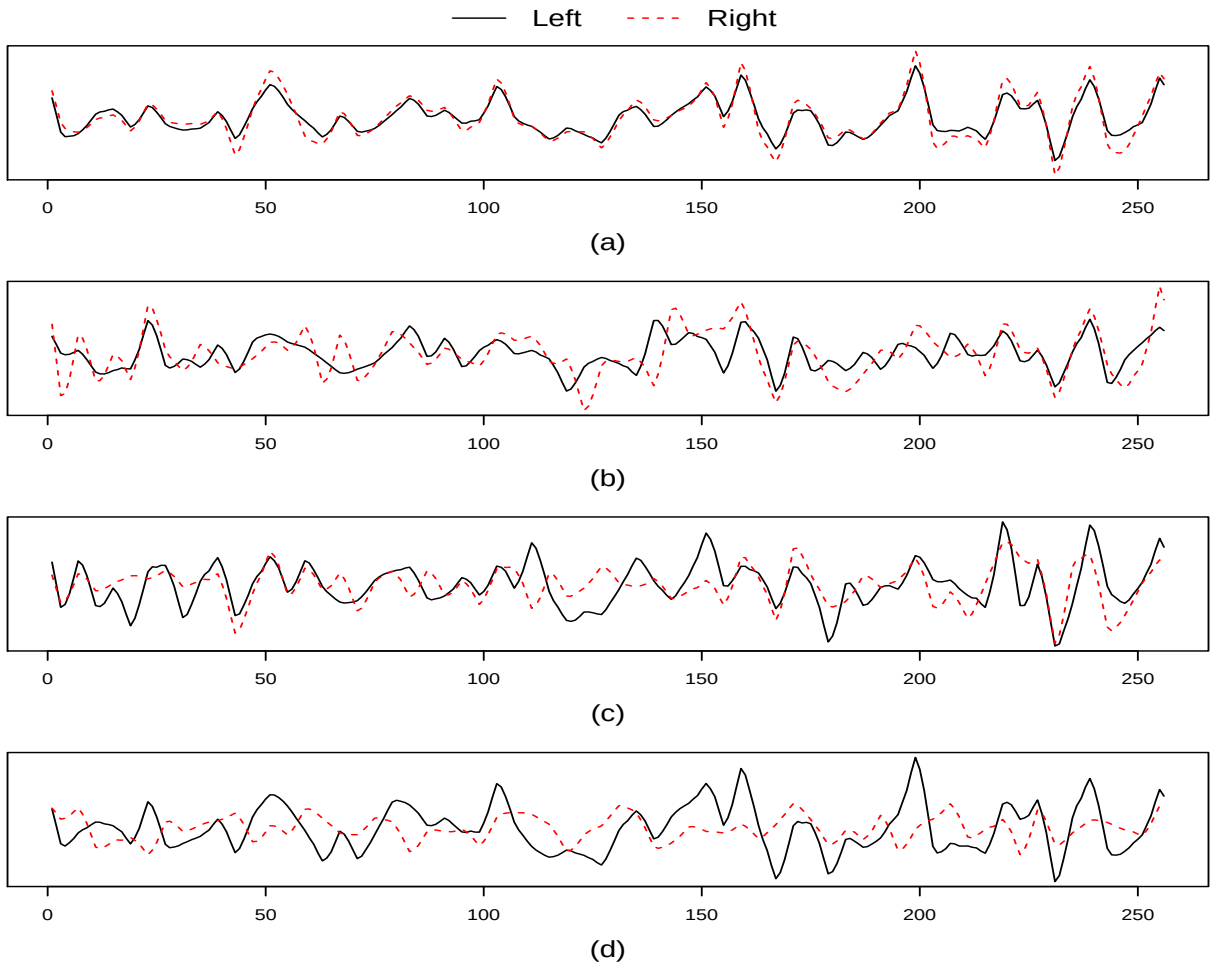


Figure 3.2. BOLD-components of interest for four pairs of voxels from left and right hemispheres.

Now consider the diagrams in Figure 3.2. In the top diagram (a) we see a pronounced similarity between signals in the two different hemispheres, and the corresponding cross-correlation is 0.96. Diagram (b) shows us a pair of signals for voxels from the same slice. The signals exhibits some similarities (look at the right tails and larger peaks in the left tail and in the middle) but overall the resemblance is not perfect and worse than in the

top diagram. The cross-correlation is equal to 0.61 and this quantifies the decrease in the similarity. Diagram (c) again exhibits BOLD-components of interest for voxels from the same slice and hence there is no delay between them. While we can see here some similarities during several intervals of time, overall the resemblance is worse than in the two top diagrams. The latter is reflected by the cross-correlation equal to 0.51. Finally, the bottom diagram (d) presents a pair of signals for the left voxel from the 5th slice and the right voxel from the 3rd slice. This yields the delay time of 38 ms, but taking this into account cannot change our opinion that the signals do not resemble each other. This visual analysis is supported by the value 0.015 of the cross-correlation.

Figure 3.2 presents an example of how a quantitative scale for the degree of a relationship between neurons in left and right hemispheres may be proposed, and we will use cross-correlation to quantify the relationship. Our analysis of a number of similar diagrams indicates that values of cross-correlation larger than 0.75 indicate that there is an active pathway between two voxels while values larger than 0.6 indicate that it is reasonable to believe that a pathway is at least periodically active.

Now, we are in a position to describe how to calculate the cross-covariance and cross-correlation based on the proposed wavelet decomposition of a fMRI signal. We begin with a convenient notation. Consider an rth voxel and an lth voxel in the right and left hemispheres, respectively, and assume that we are interested in pairs (r, l) from a set $\mathcal{N}$. In what follows indexes r and l explicitly indicate that we are dealing with voxels from the right and left hemispheres. Then, following (3.3) and (3.4) , denote by $B_r(t)$ and $B_l(t)$ underlying BOLD-components of interest,

$$B_r(t) = \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \theta_{r,jk} \psi_{jk}(t), \quad B_l(t) = \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \kappa_{l,jk} \psi_{jk}(t),$$

(note the new notation for wavelet coefficients) and also introduce the corresponding empirical wavelet coefficients (statistics)

$$\tilde{\theta}_{r,jk} = \theta_{r,jk} + n^{-1/2} \tau_r \xi_{r,jk}, \quad \tilde{\kappa}_{l,jk} = \kappa_{l,jk} + n^{-1/2} \nu_l \eta_{l,jk}.$$

Recall that $\xi_{r,jk}$ and $\eta_{l,jk}$ are independent standard normal random variables, and τ_r and ν_l are the standard deviations of the noise. If no confusion may occur, then we may skip indexes r and l and simply use $\theta_{jk} := \theta_{r,jk}$, $\nu := \nu_l$, etc. to simplify visualization of formulas.

Using orthogonality of wavelet functions, the Parseval identity, and the fact that the wavelet functions are integrated to zero, we can write for the cross-covariance between $B_r(t)$ and $B_l(t)$,

$$\sigma_{rl} := \int_0^1 B_r(t)B_l(t)dt = \sum_{j=3}^{5}\sum_{k=1}^{n/2^j} \theta_{jk}\kappa_{jk}. \tag{3.24}$$

Then the natural unbiased estimate of the cross-covariance is the plug-in estimate

$$\tilde{\sigma}_{rl} = \sum_{j=3}^{5}\sum_{k=1}^{n/2^j} \tilde{\theta}_{jk}\tilde{\kappa}_{jk} \tag{3.25}$$

Now, we introduce a cross-correlation between BOLD-components of interest in rth and lth voxels,

$$\rho_{rl} := \frac{\int_0^1 B_r(t)B_l(t)dt}{[\int_0^1 B_r^2(t)dt \int_0^1 B_l^2(t)dt]^{1/2}} =: \frac{\sigma_{rl}}{[\sigma_r\sigma_l]^{1/2}}. \tag{3.26}$$

Then, similarly to (3.25), the plug-in estimate may be used. Here,

$$\sigma_r := \int_0^1 B_r^2(t)dt = \sum_{j=3}^{5}\sum_{k=1}^{n/2^j} \theta_{jk}^2 \text{ and } \sigma_l := \int_0^1 B_l^2(t)dt = \sum_{j=3}^{5}\sum_{k=1}^{n/2^j} \kappa_{jk}^2. \tag{3.27}$$

Using the fact $E\{\tilde{\theta}_{jk}^2\} = \theta_{jk}^2 + n^{-1}\tau_r^2$ and $E\{\tilde{\kappa}_{jk}^2\} = \kappa_{jk}^2 + n^{-1}\nu_l^2$, we propose the corresponding plug-in estimates

$$\tilde{\sigma}_r := \sum_{j=3}^{5}\sum_{k=1}^{n/2^j}(\tilde{\theta}_{jk}^2 - n^{-1}\hat{\tau}^2), \text{ and } \tilde{\sigma}_l := \sum_{j=3}^{5}\sum_{k=1}^{n/2^j}(\tilde{\kappa}_{jk}^2 - n^{-1}\hat{\nu}^2). \tag{3.28}$$

We restrict attention to voxels with large power in the BOLD component of interest relative to the wavelet noise, for example, $\sum_{j=3}^{5}\sum_{k=1}^{n/2^j}\tilde{\theta}_{jk}^2 > \hat{\tau}_r/4$ and $\sum_{j=3}^{5}\sum_{k=1}^{n/2^j}\tilde{\kappa}_{jk}^2 > \hat{\nu}_l/4$. Our analysis of the fMRI data has shown that no BOLD signal of interest is observed in voxels not satisfying this assumption. This remark allows us to avoid a complex denominator in (3.26). The plug-in estimator for the cross-correlation (3.26) is

$$\tilde{\rho}_{rl} = \frac{\tilde{\sigma}_{rl}}{[\tilde{\sigma}_r\tilde{\sigma}_l]^{1/2}}. \tag{3.29}$$

CHAPTER 4

NONPARAMETRIC ESTIMATION AND STATISTICAL INFERENCE FOR CROSS-COVARIANCE AND CROSS-CORRELATION MATRICES IN AN FMRI STUDY

We begin with the analysis of the randomness in estimators of cross-covariance and cross-correlation that were proposed in (3.25) and (3.29). Recall that we have the following relationship between empirical wavelet coefficients (statistics) and the underlying wavelet coefficients (parameters),

$$\tilde{\theta}_{r,jk} = \theta_{r,jk} + n^{-1/2}\tau_r\xi_{r,jk}, \quad \tilde{\kappa}_{l,jk} = \kappa_{l,jk} + n^{-1/2}\nu_l\eta_{l,jk}. \tag{4.1}$$

Also recall that $\xi_{r,jk}$ and $\eta_{l,jk}$ are independent standard normal random variables, and τ_r and ν_l are the standard deviations of the noise. If no confusion may occur, then we may skip indexes r and l and simply use $\theta_{jk} := \theta_{r,jk}$, $\nu := \nu_l$, etc. to simplify visualization of formulas.

Substitution of (4.1) into the estimator of cross-covariance defined in (3.25) gives

$$\tilde{\sigma}_{rl} = \sum_{j=3}^{5}\sum_{k=1}^{n/2^j}\tilde{\theta}_{jk}\tilde{\kappa}_{jk}$$

$$= \sum_{j=3}^{5}\sum_{k=1}^{n/2^j}(\theta_{jk} + n^{-1/2}\tau\xi_{jk})(\kappa_{jk} + n^{-1/2}\nu\eta_{jk})$$

$$= \sigma_{rl} + n^{-1/2}\sum_{j=3}^{5}\sum_{k=1}^{n/2^j}(\tau\kappa_{jk}\xi_{jk} + \nu\theta_{jk}\eta_{jk} + n^{-1/2}\tau\nu\xi_{jk}\eta_{jk}) =: \sigma_{rl} + n^{-1/2}\zeta_{jk}. \tag{4.2}$$

Note that the variance of ζ_{jk} is

$$v_{rl}^2 := \sum_{j=3}^{5}\sum_{k=1}^{n/2^j}[\tau^2\kappa_{jk}^2 + \nu^2\theta_{jk}^2 + n^{-1}\tau^2\nu^2]. \tag{4.3}$$

Similarly, $\tilde{\sigma}_r$ and $\tilde{\sigma}_l$ can be written as following,

$$\tilde{\sigma}_r := \sum_{j=3}^{5}\sum_{k=1}^{n/2^j}(\tilde{\theta}_{jk}^2 - n^{-1}\tau^2)$$

$$= \sigma_r + n^{-1/2} \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \left(2\tau\theta_{jk}\xi_{jk} + n^{-1/2}\tau^2(\xi_{jk}^2 - 1) \right), \tag{4.4}$$

and

$$\tilde{\sigma}_l := \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} (\tilde{\kappa}_{jk}^2 - n^{-1}\nu^2)$$

$$= \sigma_l + n^{-1/2} \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \left(2\nu\kappa_{jk}\eta_{jk} + n^{-1/2}\nu^2(\eta_{jk}^2 - 1) \right). \tag{4.5}$$

Finally, let us identify the randomness component in the estimator of cross-correlation,

$$\tilde{\rho}_{rl} = \frac{\tilde{\sigma}_{rl}}{[\tilde{\sigma}_r \tilde{\sigma}_l]^{1/2}}. \tag{4.6}$$

From Taylor's formula

$$\frac{1}{\sqrt{x}} = \frac{1}{\sqrt{x_0}} - \frac{1}{2}(x - x_0)\frac{1}{x_0^{3/2}} + R(x, x_0) \tag{4.7}$$

is used where $x > 0$, $x_0 > 0$ and $|R(x, x_0)| \leq (1/4)(x - x_0)^2[\min(x, x_0)]^{-5/2}$. Application of (4.7) to (4.6) gives

$$\tilde{\rho}_{rl} = \tilde{\sigma}_{rl} \left[\frac{1}{\sqrt{\sigma_r \sigma_l}} - \frac{1}{2} \frac{\tilde{\sigma}_r \tilde{\sigma}_l - \sigma_r \sigma_l}{\sigma_r^{3/2} \sigma_l^{3/2}} + \hat{R} \right] \tag{4.8}$$

where $|\hat{R}| \leq \frac{(\tilde{\sigma}_r \tilde{\sigma}_l - \sigma_r \sigma_l)^2}{4[\min(\tilde{\sigma}_r \tilde{\sigma}_l, \sigma_r \sigma_l)]^{5/2}}$. Skipping smaller in order terms to simplify formulae, consider

$$\hat{\rho}_{rl} := \rho_{rl} + n^{-1/2} \frac{1}{\sqrt{\sigma_r \sigma_l}} \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} (\tau\kappa_{jk}\xi_{jk} + \nu\theta_{jk}\eta_{jk} + n^{-1/2}\tau\nu\xi_{jk}\eta_{jk})$$

$$+ n^{-1/2} \frac{\sigma_{rl}}{2\sigma_r^{3/2}\sigma_l^{1/2}} \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \left(2\tau\theta_{jk}\xi_{jk} + n^{-1/2}\tau^2(\xi_{jk}^2 - 1) \right)$$

$$+ n^{-1/2} \frac{\sigma_{rl}}{2\sigma_r^{1/2}\sigma_l^{3/2}} \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \left(2\nu\kappa_{jk}\eta_{jk} + n^{-1/2}\nu^2(\eta_{jk}^2 - 1) \right). \tag{4.9}$$

After some reorganization of the terms in (4.9), we obtain

$$\hat{\rho}_{rl} = \rho_{rl} + n^{-1/2}\rho_{rl} \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} \Big\{ \tau(\kappa_{jk}\sigma_{rl}^{-1} + \theta_{jk}\sigma_r^{-1})\xi_{jk} + \nu(\theta_{jk}\sigma_{rl}^{-1} + \kappa_{jk}\sigma_l^{-1})\eta_{jk}$$

$$+ \frac{1}{2}n^{-1/2}\tau^2\sigma_r^{-1}(\xi_{jk}^2 - 1) + \frac{1}{2}n^{-1/2}\nu^2\sigma_l^{-1}(\eta_{jk}^2 - 1) + n^{-1/2}\tau\nu\sigma_{rl}^{-1}\xi_{jk}\eta_{jk} \Big\}. \tag{4.10}$$

4.1 Thresholding Estimation of Large-p-Small-n Cross-covariance Matrices

Estimation of the cross-covariance matrices is a large-p-small-n problem, and so regularization is required. A thresholding estimator is proposed and its minimax property is shown in Theorem 1.

Let us comment on relations (4.2) and (4.3). It is a tradition in an asymptotic wavelet analysis (as $n \to \infty$) to ignore asymptotically negligible terms. Here this would imply

$$\tilde{\sigma}^*_{rl} = \sigma_{rl} + n^{-1/2} \sum_{j \in N_n} \sum_{k=1}^{n/2^j} (\tau \kappa_{jk} \xi_{jk} + \nu \theta_{jk} \eta_{jk}) \tag{4.11}$$

in place of (4.2) , and

$$[v^*_{rl}]^2 := \sum_{j \in N_n} \sum_{k=1}^{n/2^j} (\tau^2 \kappa^2_{jk} + \nu^2 \theta^2_{jk}) \tag{4.12}$$

in place of (4.3). Here N_n is the set of scales corresponding to frequencies of the BOLD signal.

Then, similarly to recent results on estimation of sparse covariance matrices (see a comprehensive review Cai (2017)), a number of interesting theoretical results that shed light on asymptotic estimation may be established for the cross-covariance. Let us formulate one such result which is of a particular interest for the problem at hand. Define a class of sparse $p_1 \times p_2$ matrices by

$$\mathcal{U}_q(s_0(p_1,p_2)) := \Big\{ \{\sigma_{rl}\} : \max_{r \in \{1,2,\ldots,p_1\}} \sum_{l=1}^{p_2} (\sigma_r \sigma_l)^{(1-q)/2} |\sigma_{rl}|^q \le s_0(p_1,p_2),$$

$$\max_{r,l} |\ln(\sigma_r/\sigma_l)| \le C_0 < \infty, \ \ q \in [0,1) \Big\}. \tag{4.13}$$

In what fiollows $I(\cdot)$ is the indicator function.

Theorem 1. *Assume that random variables ξ_{jk} and η_{jk} in (4.11) are standard normal for all $p_1 + p_2$ studied voxels. Consider a hard-thresholding oracle-estimate of the $p_1 \times p_2$ cross-covariance matrix $\{\sigma_{rl}\} := \{\sigma_{rl}, r = 1, \ldots, p_1, l = 1, \ldots, p_2\}$,*

$$\hat{\sigma}_{rl} := I(|\tilde{\sigma}^*_{rl}| > \sqrt{2 \ln(p_1 p_2)}\, n^{-1/2} v^*_{rl}) \tilde{\sigma}^*_{rl}, \tag{4.14}$$

where $\tilde{\sigma}^*_{rl}$ and v^*_{rl} are defined in (4.11) *and* (4.12)*, respectively. Then this hard-thresholding is rate minimax over the class $\mathcal{U}_q(s_0(p_1,p_2))$ and a minimax risk with the matrix l_1-norm as a loss function. That is, for some finite constant $C>0$,*

$$\sup_{\{\sigma_{rl}\}\in\mathcal{U}_q(s_0(p_1,p_2))} E\Big\{\max_{r\in\{1,2,\ldots,p_1\}}\sum_{l=1}^{p_2}|\hat{\sigma}_{rl}-\sigma_{rl}|\Big\}\le Cs_0(p_1,p_2)[\ln(p_1p_2)n^{-1}]^{(1-q)/2}. \tag{4.15}$$

Remark 1. The assumption of Theorem 1 about normality of the noise in empirical coefficients is a standard one. See a discussion in Efromovich (1999a,b); Lazar (2008); Valdez-Jasso (2010).

Remark 2. The threshold proposed in (4.14) is very similar to the commonly-used "Universal" threshold described in (2.17) with only the noise level being replaced by $n^{-1/2}v^*_{rl}$, the standard deviation of the estimator $\tilde{\sigma}^*_{rl}$.

Proof of Theorem 1. Set $\lambda:=\sqrt{2\ln(p_1p_2)}$ and write,

$$\sum_{l=1}^{p_2}|\hat{\sigma}_{rl}-\sigma_{rl}|$$

$$=\sum_{l=1}^{p_2}|I(|\sigma_{rl}|>n^{-1/2}\lambda v^*_{rl})\tilde{\sigma}_{rl}-(I(|\sigma_{rl}|>n^{-1/2}\lambda v^*_{rl})+I(|\sigma_{rl}|\le n^{-1/2}\lambda v^*_{rl}))\sigma_{rl}|$$

$$\le\sum_{l=1}^{p_2}I(|\sigma_{rl}|>n^{-1/2}\lambda v^*_{rl})|\tilde{\sigma}_{rl}-\sigma_{rl}|+\sum_{l=1}^{p_2}I(|\sigma_{rl}|\le n^{-1/2}\lambda v^*_{rl})|\sigma_{rl}|. \tag{4.16}$$

Now we consider the two terms in (4.16) in turn. For the first term we study separately cases of large and small $|\tilde{\sigma}_{rl}-\sigma_{rl}|$. Write,

$$\sum_{l=1}^{p_2}I(|\sigma_{rl}|>n^{-1/2}\lambda v^*_{rl})|\hat{\sigma}^*_{rl}-\sigma_{rl}|$$

$$\le\sum_{l=1}^{p_2}[I(|\sigma_{rl}|>n^{-1/2}\lambda v^*_{rl})|\tilde{\sigma}_{rl}-\sigma_{rl}|[I(|\tilde{\sigma}_{rl}-\sigma_{rl}|<\lambda v^*_{rl}n^{-1/2})+I(|\tilde{\sigma}_{rl}-\sigma_{rl}|\ge\lambda v^*_{rl}n^{-1/2})]$$

$$\le\sum_{l=1}^{p_2}[I(|\sigma_{rl}|>n^{-1/2}\lambda v^*_{rl})\lambda v^*_{rl}n^{-1/2}$$

$$+\sum_{l=1}^{p_2} I(|\sigma_{rl}| > n^{-1/2}\lambda v^*_{rl}) I(|\tilde{\sigma}_{rl} - \sigma_{rl}| \geq \lambda v^*_{rl} n^{-1/2}) |\tilde{\sigma}_{rl} - \sigma_{rl}|. \tag{4.17}$$

Combining the first sum in (4.17) with the second sum in (4.16) we get

$$\sum_{l=1}^{p_2} [I(|\sigma_{rl}| > n^{-1/2}\lambda v^*_{rl})\lambda v^*_{rl} n^{-1/2} + I(|\sigma_{rl}| \leq n^{-1/2}\lambda v^*_{rl})|\sigma_{rl}|]$$

$$\leq \sum_{l=1}^{p_2} [(\lambda v^*_{rl} n^{-1/2})^{1-q} |\sigma_{rl}|^q + (\lambda v^*_{rl} n^{-1/2})^{1-q} |\sigma_{rl}|^q$$

$$\leq (\lambda n^{-1/2})^{1-q} [\max_{r,l} v^*_{rl} [\sigma_r \sigma_l]^{1/2} \sum_{r=1}^{p_2} (\sigma_l \sigma_r)^{(1-q)/2} \sigma_{rl}|^q \leq [\lambda n^{-1/2}]^{1-q} s_0(p_1, p_2). \tag{4.18}$$

For the second term in (4.17) we can write,

$$E\{\max_{r \in \{1,2,\ldots,p_1\}} \sum_{l=1}^{p_2} I(|\tilde{\sigma}_{rl} - \sigma_{rl}| > \lambda v^*_{rl} n^{-1/2}) |\tilde{\sigma}_{rl} - \sigma_{rl}|\}$$

$$\leq C n^{-1/2} \sum_{r=1}^{p_1} \sum_{l=1}^{p_2} E\{I(|\tilde{\sigma}_{rl} - \sigma_{rl}| > \lambda v^*_{rl} n^{-1/2}) |\tilde{\sigma}_{rl} - \sigma_{rl}|\}. \tag{4.19}$$

Recall that $\tilde{\sigma}_{rl} - \sigma_{rl}$ is normally distributed. Further, using $xe^{-x^2/2} = de^{-x^2/2}/dx$ we get

$$\int_{|x| > \sqrt{2\ln(p_1 p_2)}} |x| e^{-x^2/2} dx \leq C e^{-\ln(p_1 p_2)} = C(p_1 p_2)^{-1}. \tag{4.20}$$

Application of (4.20) to (4.19) gives the desired result.

Remark 3. The proof shows that a similar result may be established for any known distribution of $\tilde{\sigma}_{rl} - \sigma_{rl}$ with a corresponding modification in evaluation of the expectations in (4.19).

Theorem 1 sheds a new light on estimation of and inference for cross-covariances and cross-correlations. We may conclude that, even if n is large but the fMRI noise is still relatively large (implying that values v^*_{rl} are large), no feasible *simultaneous* inference, like simultaneous confidence intervals, may be obtained for all pairs of voxels. This is why very accurate estimation of the level of noise is paramount.

Further, there is another issue that must be considered. For many settings the available sample size $n = 256$ would be considered as a large one that fits the classical wavelet

asymptotic theory. This is no longer the case here due to the following. Since the BOLD-component of interest "occupies" scales 3, 4 and 5, then (4.3) implies that (compare with (4.12))

$$v_{rl}^2 := \sum_{j=3}^{5} \sum_{k=1}^{n/2^j} [\tau^2 \kappa_{jk}^2 + \nu^2 \theta_{jk}^2] + (7/32)\tau^2 \nu^2. \tag{4.21}$$

Note that the term $(7/32)\tau^2\nu^2$ may be comparable with the double sum in (4.21) and hence it cannot be ignored. This issue will be addressed in the next section.

4.2 Theory of Simultaneous Analysis of Cross-covariance Matrices

In this and the following sections, without introducing ambiguity, we replace the double index jk by the single index i to simplify the formulae. Also, recall the definition of a subexponential random variable: a random variable X with mean $\mu = E(X)$ is called *subexponential* with parameters (ν^2, b) if there exists constants $\nu > 0$ and $b > 0$ such that

$$E\{\exp(\lambda(X - \mu))\} \leq \exp\left(\frac{\nu^2 \lambda^2}{2}\right), \text{ for all } |\lambda| < \frac{1}{b}. \tag{4.22}$$

Express (4.2) as

$$\tilde{\sigma}_{rl} = \sigma_{rl} + n^{-1/2} \sum_i (\tau \kappa_i \xi_i + \nu \theta_i \eta_i + n^{-1/2} \tau \nu \xi_i \eta_i) =: \sigma_{rl} + n^{-1/2} \sum_i X_i, \tag{4.23}$$

where

$$X_i := \tau \kappa_i \xi_i + \nu \theta_i \eta_i + n^{-1/2} \tau \nu \xi_i \eta_i. \tag{4.24}$$

The aim of this section is to prove the following theoretical result.

Theorem 2. *Assume that ξ_i and η_i in (4.23) are independent standard normal variables. Introduce a parameter $b \in (0, 1)$. Then the following statements hold for X_i:*

(i) The random variable X_i, defined in (4.24), is subexponential with parameters $(A_i, \frac{1}{B})$, where

$$A_i := A_i(b) = -\frac{\ln(1-b)\tau^2\nu^2}{bn} + \frac{2b\tau\nu\kappa_i\theta_i + \tau^2\kappa_i^2 + \nu^2\theta_i^2}{1-b} \text{ and } B := B(b) = \frac{\sqrt{bn}}{\tau\nu}. \tag{4.25}$$

(ii) The sequence of $X_1, \ldots, X_N$ with corresponding A_i and B being defined in (4.25) *are independent and satisfy the following inequality,*

$$\max\Big(P(\sum_{i=1}^{N} X_i \le -t), P(\sum_{i=1}^{N} X_i \ge t)\Big)$$

$$\le \exp\Big\{-\frac{t^2}{2A}\Big\} I(0 \le t \le AB) + \exp\Big\{-\frac{tB}{2}\Big\} I(t \ge AB), \;\; A := A(b) = \sum_{i=1}^{N} A_i(b). \quad (4.26)$$

(iii) For any $\alpha \in (0,1)$, $t(\alpha)$ such that

$$\max\Big(P(\sum_{i=1}^{N} X_i \le -t(\alpha)), P(\sum_{i=1}^{N} X_i \ge t(\alpha))\Big) \le \alpha, \quad (4.27)$$

can be obtained by

$$t(\alpha) := t(\alpha, b_\alpha) = \frac{2\ln(\alpha^{-1})\tau\nu}{(b_\alpha n)^{1/2}}, \quad (4.28)$$

where b_α is the solution to the equation

$$A(b_\alpha)[B(b_\alpha)]^2 = 2\ln(\alpha^{-1}). \quad (4.29)$$

To shed some light on function $t(\alpha)$, note that as $n/(\tau\nu)^2$ increases, b_α becomes proportional to $\ln(\alpha^{-1})$, and then in its turn $t(\alpha)$ becomes proportional to $\sqrt{\ln(\alpha^{-1})}$.

(iv) For any $\alpha \in (0,1)$, the solution $t'(\alpha)$ to the equation

$$(4/B(b_\alpha))[t'(\alpha) + (2/B(b_\alpha))]e^{-t'(\alpha)B(b_\alpha)/2} = \alpha, \quad (4.30)$$

where b_α is defined in (4.29)*, satisfies*

$$E\Big\{|\sum_{i=1}^{N} X_i| I(|\sum_{i=1}^{N} X_i| > t'(\alpha))\Big\} \le \alpha. \quad (4.31)$$

Remark 4. We obtained exponential inequalities (4.26) and (4.27) for the probability of large deviations as well as the corresponding inequality (4.31) for the first moment. In particular, (4.27) may be used to obtain $1 - \alpha_0$ level confidence intervals and confidence

lower bounds by choosing $\alpha = \alpha_0/2$ and $\alpha = \alpha_0$, respectively. This concludes the procedure of inference for a cross-covariance.

Proof of Theorem 2. (i) A direct calculation shows that

$$E\{\exp(tX_i)\} = E\{\exp\{t(\tau\kappa_i\xi_i + \nu\theta_i\eta_i + n^{-1/2}\tau\nu\xi_i\eta_i)\}\}$$

$$= \frac{1}{[1 - n^{-1}\tau^2\nu^2t^2]^{1/2}} \exp\left\{\frac{2n^{-1/2}\tau^2\nu^2\kappa_i\theta_i t^3 + (\tau^2\kappa_i^2 + \nu^2\theta_i^2)t^2}{2[1 - n^{-1}\tau^2\nu^2t^2]}\right\} \quad \text{for } t^2 < \frac{bn}{\tau^2\nu^2}. \tag{4.32}$$

Introducing a parameter $b \in (0, 1)$ and considering the following inequality

$$\frac{1}{[1 - x^2]^{1/2}} \leq \exp\{ax^2\} \quad \text{for } 0 < x^2 \leq b < 1, \tag{4.33}$$

which is equivalent to

$$a \geq -\frac{\ln(1 - x^2)}{2x^2} \quad \text{for } 0 < x^2 \leq b < 1.$$

Since the function $-\frac{\ln(1-y)}{2y}$ for $0 < y < 1$ is increasing in y, the minimum a such that (4.33) holds is $a = -\frac{\ln(1-b)}{2b}$. Using

$$\frac{1}{[1 - x^2]^{1/2}} \leq \exp\left\{-\frac{\ln(1 - b)}{2b}x^2\right\} \quad \text{for } 0 < x^2 \leq b < 1,$$

we conclude that

$$\frac{1}{[1 - n^{-1}\tau^2\nu^2t^2]^{1/2}} \leq \exp\left\{-\frac{\ln(1 - b)\tau^2\nu^2t^2}{2bn}\right\} \quad \text{for } t^2 < \frac{bn}{\tau^2\nu^2}. \tag{4.34}$$

Now we plug (4.34) in the right side of (4.32) and get after simplifications,

$$E\{\exp\{t(\tau\kappa_i\xi_i + \nu\theta_i\eta_i + n^{-1/2}\tau\nu\xi_i\eta_i)\}\}$$

$$\leq \exp\left\{t^2\Big[-\frac{\ln(1 - b)\tau^2\nu^2}{2bn} + \frac{2b\tau\nu\kappa_i\theta_i + \tau^2\kappa_i^2 + \nu^2\theta_i^2}{2(1 - b)}\Big]\right\}$$

$$=: \exp\left\{\frac{t^2A_i}{2}\right\}, \quad \text{for } 0 \leq t < \frac{b^{1/2}n^{1/2}}{\tau\nu} =: B. \tag{4.35}$$

Note that $A_i = A_i(b)$ and $B = B(b)$ depend on parameter b. From (4.35) we conclude that the random variable X_i, defined in (4.24), is sub-exponential with parameters $(A_i, \frac{1}{B})$.

(ii) Considering (4.23), in our wavelet analysis of cross-covariances, we are dealing not with a single sub-exponential X_i but with a sum of independent sub-exponential variables. Note that the independency in X_i's is due to the independency in ξ_i's and η_i's. Hence, let us consider a sequence of independent random variables $X_1, \ldots, X_N$ with corresponding A_i and B being defined in (4.25). Then Theorem 3.16 in Petrov (1975), together with (4.35), allow us to get the following exponential inequality,

$$\max\Big(P(\sum_{i=1}^{N} X_i \le -t), P(\sum_{i=1}^{N} X_i \ge t)\Big)$$

$$\le \exp\Big\{-\frac{t^2}{2A}\Big\} I(0 \le t \le AB) + \exp\Big\{-\frac{tB}{2}\Big\} I(t \ge AB), \;\; A := \sum_{i=1}^{N} A_i. \tag{4.36}$$

Note that (4.36) gives us an explicit upper bound that may be used in statistical analysis of large deviations.

(iii) Next, suppose that we would like to find a threshold $t(\alpha, b)$ such that the right side of (4.36) is equal to $\alpha \in (0, 1]$. Inequality (4.36) implies that

$$t(\alpha, b) = [2\ln(\alpha^{-1})A]^{1/2} I(\alpha \ge e^{-AB^2/2}) + [2\ln(\alpha^{-1})/B] I(\alpha < e^{-AB^2/2}). \tag{4.37}$$

The function $A(b)[B(b)]^2$ monotonically increases from zero to infinity as b increases from zero to 1. Hence, there exists a solution b_α of the equation

$$A(b_\alpha)[B(b_\alpha)]^2 = 2\ln(\alpha^{-1}). \tag{4.38}$$

This, together with (4.37), imply that

$$t(\alpha) := t(\alpha, b_\alpha) = \frac{2\ln(\alpha^{-1})\tau\nu}{(b_\alpha n)^{1/2}}. \tag{4.39}$$

To shed some light on function $t(\alpha)$, note that as $n/(\tau\nu)^2$ increases, the b_α becomes proportional to $\ln(\alpha^{-1})$, and then in its turn $t(\alpha)$ becomes proportional to $\sqrt{\ln(\alpha^{-1})}$. We now may conclude that

$$\max\Big(P(\sum_{i=1}^{N} X_i \le -t(\alpha)), P(\sum_{i=1}^{N} X_i \ge t(\alpha))\Big) \le \alpha. \tag{4.40}$$

(iv) Further, recall Remark 3, and then let us explain how we can get a corresponding exponential inequality for $\sum_{i=1}^{N} X_i$. Using elementary

$$2\int_t^\infty xe^{-xB/2}dx = (4/B)[t + (2/B)]e^{-tB/2}, \tag{4.41}$$

introduce a $t'(\alpha)$ such that (in what follows $B := B(b_\alpha)$)

$$(4/B)[t'(\alpha) + (2/B)]e^{-t'(\alpha)B/2} = \alpha. \tag{4.42}$$

Note that $t'(\alpha) > t(\alpha)$ and $t'(\alpha)/t(\alpha) \to 1$ as $\alpha \to 0$. Then the useful property of $t'(\alpha)$ is that

$$E\Big\{|\sum_{i=1}^{N} X_i| I(|\sum_{i=1}^{N} X_i| > t'(\alpha))\Big\} \le \alpha. \tag{4.43}$$

This completes the proof of Theorem 2.

4.3 Theory of Simultaneous Analysis of Cross-correlation Matrices

Consider (4.10) and introduce a new notation X_i',

$$\begin{aligned}\hat{\rho}_{rl} = \rho_{rl} + n^{-1/2}\rho_{rl}\sum_i \Big\{&\tau(\kappa_i\sigma_{rl}^{-1} + \theta_i\sigma_r^{-1})\xi_i + \nu(\theta_i\sigma_{rl}^{-1} + \kappa_i\sigma_l^{-1})\eta_i \\ &+\frac{1}{2}n^{-1/2}\tau^2\sigma_r^{-1}(\xi_i^2 - 1) + \frac{1}{2}n^{-1/2}\nu^2\sigma_l^{-1}(\eta_i^2 - 1) + n^{-1/2}\tau\nu\sigma_{rl}^{-1}\xi_i\eta_i\Big\} \\ &=: \rho_{rl} + n^{-1/2}\rho_{rl}\sum_i X_i'.\end{aligned} \tag{4.44}$$

To make calculations more transparent, introduce the following notations,

$$\begin{aligned}A := \tau(\kappa_i\sigma_{rl}^{-1} + \theta_i\sigma_r^{-1}),\ \ a := \nu(\theta_i\sigma_{rl}^{-1} + \kappa_i\sigma_l^{-1}), \\ B := \frac{1}{2}n^{-1/2}\tau^2\sigma_r^{-1},\ \ b := \frac{1}{2}n^{-1/2}\nu^2\sigma_l^{-1}, \\ c := n^{-1/2}\tau\nu\sigma_{rl}^{-1}.\end{aligned} \tag{4.45}$$

Then we can write down X_i' as (we may skip index i whenever no confusion occurs)

$$X_i' = A\xi + a\eta + B(\xi^2 - 1) + b(\eta^2 - 1) + c\xi\eta. \tag{4.46}$$

Theorem 3. *Assume that ξ_i and η_i in (4.44) are independent standard normal variables. Also introduce a parameter $\beta \in (0,1)$. Then the following statements hold for X_i':*

(i) The random variable X_i', defined in (4.46), is subexponential with parameters $(A_i', \frac{1}{B'})$, where

$$B' := B'(\beta) = \frac{-(B+b) + \sqrt{(B+b)^2 + \beta(c^2 - 4Bb)}}{c^2 - 4Bb}, \tag{4.47}$$

and

$$A_i' := A_i'(\beta) = c^2 - 4Bb + \frac{(B')^2(c^2-4Bb)^2}{2(1-\beta)}$$

$$+ \frac{2(B+b) + a^2 + A^2 + 2B'[(B+b)(c^2-4Bb) - a^2B - A^2b + Aac]_+}{1-\beta}, \tag{4.48}$$

and $[x]_+ := \max(x, 0)$ denotes the positive part of x.

(ii) The sequence of $X_1', \ldots, X_N'$ with corresponding A_i' and B' being defined in (4.47) and (4.48) are independent and satisfy the following inequality,

$$P(\sum_{i=1}^{N} X_i' \geq \lambda)$$

$$\leq \exp\Big\{ -\frac{\lambda^2}{2A'} \Big\} I(0 \leq \lambda \leq A'B') + \exp\Big\{ -\frac{\lambda B'}{2} \Big\} I(\lambda \geq A'B'), \;\; A' := \sum_{i=1}^{N} A_i'. \tag{4.49}$$

(iii) Based on the inequality (4.49), for any $\alpha \in (0,1)$, an optimal $t(\alpha)$ such that

$$\max\Big(P(\sum_{i=1}^{N} X_i' \leq -t(\alpha)), P(\sum_{i=1}^{N} X_i' \geq t(\alpha)) \Big) \leq \alpha, \tag{4.50}$$

can be obtained using the following formula

$$t(\alpha) := t(\alpha, \beta_\alpha) = \frac{2\ln(\alpha^{-1})}{B'(\beta_\alpha)}, \tag{4.51}$$

where β_α is the solution to the equation

$$A'(\beta_\alpha)[B'(\beta_\alpha)]^2 = 2\ln(\alpha^{-1}). \tag{4.52}$$

(iv) Based on the inequality (4.49)*, for any* $\alpha \in (0,1)$*, the solution to the following equation with respect to* $t'(\alpha)$

$$(4/B'(\beta_\alpha))[t'(\alpha) + (2/B'(\beta_\alpha))]e^{-t'(\alpha)B'(\beta_\alpha)/2} = \alpha, \tag{4.53}$$

with β_α *defined in* (4.52)*, has the following property,*

$$E\Big\{|\sum_{i=1}^{N} X_i'| I(|\sum_{i=1}^{N} X_i'| > t'(\alpha))\Big\} \leq \alpha. \tag{4.54}$$

Proof of Theorem 3. (i) Following (4.32) - (4.35), we are evaluating $E\{e^{\lambda X_i'}\}$ where

$$\lambda := tn^{-1/2}|\rho_{rl}|, \;\; t > 0. \tag{4.55}$$

We are doing this in two steps. The first step is to write the expectation via conditional expectation

$$E\{e^{\lambda X_i'}\} = E\{\exp(\lambda[A\xi + B(\xi^2 - 1) - b])E\{\exp(\lambda(a\eta + b\eta^2 + c\xi\eta))|\xi\}\} \tag{4.56}$$

and calculate the conditional expectation $E\{\exp(\lambda(a\eta + b\eta^2 + c\xi\eta))|\xi\}$. Let us do the latter. Write,

$$\begin{aligned} &E\{\exp(\lambda(a\eta + b\eta^2 + c\xi\eta))|\xi\} \\ &= \frac{1}{\sqrt{2\pi}} \int \exp\left\{-\frac{y^2}{2} + \lambda b y^2 + \lambda(a + c\xi)y\right\} dy \\ &= \frac{1}{\sqrt{1 - 2\lambda b}} \exp\left\{\frac{\lambda^2(a + c\xi)^2}{2(1 - 2\lambda b)}\right\}, \quad \lambda < 1/(2b). \end{aligned} \tag{4.57}$$

Step two is to plug (4.57) into (4.56). We get after some algebra,

$$\begin{aligned} E\{e^{\lambda X_i'}\} &= E\left\{\exp(\lambda[A\xi + B(\xi^2 - 1) - b])\frac{1}{\sqrt{1 - 2\lambda b}} \exp\left\{\frac{\lambda^2(a + c\xi)^2}{2(1 - 2\lambda b)}\right\}\right\} \\ &= \frac{1}{\sqrt{W}} \exp\left\{-\lambda(B + b) + \frac{\lambda^2 a^2}{2(1 - 2b\lambda)} + \frac{[(1 - 2b\lambda)\lambda A + \lambda^2 ac]^2}{2(1 - 2b\lambda)W}\right\}, \quad W > 0, \end{aligned} \tag{4.58}$$

where

$$W := 1 - 2\lambda(B+b) - \lambda^2(c^2 - 4Bb) = (1-2b\lambda)(1-2B\lambda) - \lambda^2 c^2. \tag{4.59}$$

Note that the Cauchy–Schwarz inequality yields $c^2 - 4Bb \geq 0$ and hence $W > 0$ implies the restriction $\lambda < 1/(2b)$ imposed in (4.57). The expression on the right side of (4.59) allows us to simplify (4.58) and get

$$E\{e^{\lambda X_i'}\} = W^{-1/2} \exp\left\{-\lambda(B+b) + \frac{\lambda^2(a^2+A^2) - 2\lambda^3(a^2B + A^2b - Aac)}{2[1-2\lambda(B+b) - \lambda^2(c^2-4Bb)]}\right\}. \tag{4.60}$$

Now, we introduce a parameter $\beta \in (0,1)$ such that

$$0 \leq 2\lambda(B+b) + \lambda^2(c^2 - 4Bb) \leq \beta < 1. \tag{4.61}$$

Note that the corresponding λ imply $W > 0$ and we get the restriction on considered λ,

$$0 \leq \lambda \leq \frac{-(B+b) + \sqrt{(B+b)^2 + \beta(c^2-4Bb)}}{c^2 - 4Bb} =: \lambda^*. \tag{4.62}$$

Clearly if (4.62) holds then $W \geq 1 - \beta$.

Now recall the inequality

$$\frac{1}{\sqrt{1-u}} \leq \exp\left(\frac{u}{2} + \frac{u^2}{4(1-u)}\right), \quad 0 \leq u < 1. \tag{4.63}$$

This inequality allows us to evaluate $W^{-1/2}$ by an exponential function if we set $u = 2\lambda(B+b) + \lambda^2(c^2 - 4Bb)$. Write,

$$\begin{aligned} E\{e^{\lambda X_i'}\} &\leq \exp\Big\{\lambda(B+b) + \frac{1}{2}\lambda^2(c^2-4Bb) + \frac{[2\lambda(B+b) + \lambda^2(c^2-4Bb)]^2}{4[1-2\lambda(B+b) - \lambda^2(c^2-4Bb)]} \\ &\qquad -\lambda(B+b) + \frac{\lambda^2(a^2+A^2) - 2\lambda^3(a^2B + A^2b - Aac)}{2[1-2\lambda(B+b) - \lambda^2(c^2-4Bb)]}\Big\} \\ &= \exp\Big\{\frac{1}{2}\lambda^2(c^2-4Bb) + \frac{[2\lambda(B+b) + \lambda^2(c^2-4Bb)]^2}{4[1-2\lambda(B+b) - \lambda^2(c^2-4Bb)]} \\ &\qquad + \frac{\lambda^2(a^2+A^2) - 2\lambda^3(a^2B + A^2b - Aac)}{2[1-2\lambda(B+b) - \lambda^2(c^2-4Bb)]}\Big\} \end{aligned}$$

$$= \exp\Big\{\frac{1}{2}\lambda^2(c^2-4Bb) + \frac{4\lambda^2(B+b)+4\lambda^3(B+b)(c^2-4Bb)+\lambda^4(c^2-4Bb)^2}{4[1-2\lambda(B+b)-\lambda^2(c^2-4Bb)]}$$
$$+ \frac{\lambda^2(a^2+A^2)-2\lambda^3(a^2B+A^2b-Aac)}{2[1-2\lambda(B+b)-\lambda^2(c^2-4Bb)]}\Big\}, \quad \text{for } 0 \le \lambda \le \lambda^*. \tag{4.64}$$

We may continue evaluation of (4.64) and write using (4.55) that

$$E\{e^{\lambda X_i'}\} \le \exp\{\lambda^2 \frac{A_i'}{2}\}, \quad 0 \le \lambda \le \lambda^* =: B', \tag{4.65}$$

where

$$A_i' := c^2 - 4Bb + \frac{(\lambda^*)^2(c^2-4Bb)^2}{2(1-\beta)}$$
$$+ \frac{2(B+b)+a^2+A^2+2\lambda^*[(B+b)(c^2-4Bb)-a^2B-A^2b+Aac]_+}{1-\beta}, \tag{4.66}$$

and $[x]_+ := \max(x, 0)$ denotes the positive part of x. Note that $A_i' = A_i'(\beta)$ and $B' = B'(\beta)$ depend on parameter β. From (4.65) we conclude that the random variable X_i' is sub-exponential with parameters $(A_i', \frac{1}{B'})$.

(ii) The independency in X_i''s is due to the independency in ξ_i's and η_i's. Hence, let us consider a sequence of independent random variables $X_1', \ldots, X_N'$ with corresponding A_i' and B' being defined in (4.47) and (4.48). Then Theorem 3.16 in Petrov (1975), together with (4.65), allow us to obtain the inequality in (4.49). Note that (4.49) gives us an explicit upper bound that may be used in statistical analysis of large deviations.

(iii) – (iv) can be obtained by following the same steps (4.37) –(4.43).

This completes the proof of Theorem 3.

Finally, we are going to summarize the proposed procedure as an algorithm. Let us review the Holm procedure for simultaneous inference about pathways with cross-correlation exceeding a given threshold (in our particular case the considered threshold is 0.6). While it is more complicated than the Bonferroni procedure (the latter will simply use $1 - \alpha/(p_1 p_2)$ individual levels of confidence where α is the family-wise error rate), in our case the Holm procedure yields a larger number of active pathways due to a large range in underlying cross-correlations.

To introduce the Holm procedure, recall that a confidence lower bound for a cross-correlation may be constructed by inverting a corresponding one-tailed test. Suppose that we have p-values $\hat{p}_1, \ldots, \hat{p}_s$ of s individual tests. Denote ordered from the smallest to the largest p-values by $\hat{p}_{(1)}, \ldots, \hat{p}_{(s)}$ and associated hypotheses by $H_{(1)}, \ldots, H_{(s)}$. Suppose the family-wise error rate (FWER) is α. Holm procedure begins by determining whether the test that looks most significant should be rejected. Suppose the family-wise error rate (FWER) is α. Let $\alpha_i = \frac{\alpha}{s-i+1}$. Then Holm procedure uses the following criteria:

If $\hat{p}_{(1)} \geq \alpha_1$, accept all hypotheses.

Otherwise, for $r = 1, \ldots, s$, reject $H_{(1)}, \ldots, H_{(r)}$ if $\hat{p}_{(1)} < \alpha_1, \ldots, \hat{p}_{(r)} < \alpha_r$.

In the simultaneous inference about cross-correlations, an individual test for a particular ρ_{rl} is

$$H_0 : \rho_{rl} \leq \rho_{H_0} \quad \text{versus} \quad H_a : \rho_{rl} > \rho_{H_0}. \tag{4.67}$$

In our particular application we use $\rho_{H_0} = 0.6$. To simplify the following formulas we may skip index rl and write $\rho := \rho_{rl}$. The p-value for the above-considered individual test is

$$\begin{aligned} \text{p-value}_{rl} &= P(\hat{\rho}_{theoretical} > \tilde{\rho}_{observed} | H_0) \\ &= P(\rho_{H_0} + n^{-1/2}\rho_{H_0} \sum X_i' > \tilde{\rho}_{observed} | H_0) \\ &= P(\sum X_i' > n^{1/2}[\tilde{\rho}_{observed}/\rho_{H_0} - 1] | H_0). \end{aligned} \tag{4.68}$$

Note that only cases with $\tilde{\rho}_{observed} > \rho_{H_0}$ are of interest and in what follows only these cases are analyzed. Using inequality (4.49) we continue evaluation of the p-value,

$$\begin{aligned} &\text{p-value}_{rl} \\ &\leq \exp\Big\{ - \frac{n(\tilde{\rho}_{observed} - \rho_{H_0})^2}{2A'(\beta)\rho_{H_0}^2} \Big\} I\big(0 \leq \tilde{\rho}_{observed} - \rho_{H_0} \leq n^{-1/2}\rho_{H_0} A'(\beta)B'(\beta)\big) \\ &+ \exp\Big\{ - \frac{n^{1/2}(\tilde{\rho}_{observed} - \rho_{H_0})B'(\beta)}{2\rho_{H_0}} \Big\} I\big(\tilde{\rho}_{observed} - \rho_{H_0} \geq n^{-1/2}\rho_{H_0} A'(\beta)B'(\beta)\big). \end{aligned} \tag{4.69}$$

Algorithm 1 Simultaneous Inference for Cross-Correlations

procedure PATHWAYINDICATOR($\{\tilde{\theta}_{r,jk}, \tilde{\kappa}_{l,jk}\}$, $\{\hat{\tau}_r, \hat{\nu}_l\}$, α, ρ_{H_0})
  pValue ← `initialize a` $p_1 \times p_2$ `matrix by setting each element as 1`
  for `r from 1 to` p_1 **do**
    for `l from 1 to` p_2 **do**
      $\tilde{\sigma}_{rl} \leftarrow \sum_{jk} \tilde{\theta}_{jk}\tilde{\kappa}_{jk}$
      $\tilde{\sigma}_r \leftarrow \sum_{jk}(\tilde{\theta}^2_{jk} - n^{-1}\hat{\tau}^2)$
      $\tilde{\sigma}_l \leftarrow \sum_{jk}(\tilde{\kappa}^2_{jk} - n^{-1}\hat{\nu}^2)$
      $\tilde{\rho}_{rl} \leftarrow \hat{\sigma}_{rl}/(\tilde{\sigma}_r^{1/2}\tilde{\sigma}_l^{1/2})I(\tilde{\sigma}_r > \hat{\tau}/4)I(\tilde{\sigma}_l > \hat{\nu}/4)$
      if $\tilde{\rho}_{rl} < \rho_{H_0}$ **then break**
      end if
      $\beta^* \leftarrow$ argmin of (4.69) with parameters being replaced by their corresponding etimators
      pValue$[r, l] \leftarrow$ right-hand side of (4.69) with $\beta = \beta^*$ and with parameters being replaced by their corresponding etimators
    end for
  end for
  `pathwayIndicator` ← HolmProcedure(pValue, α)
  return `pathwayIndicator`
end procedure

By choosing $\beta = \beta^* \in (0, 1)$ which minimizes the right side of (4.69), we get an expression (upper bound) for the individual p-value which is used in the Holm procedure.

Algorithm 1 summarizes the proposed procedure. In our practical applications, we will use $\alpha = 0.05$ and $\rho_{H_0} = 0.6$. The assumption that a neural pathway is at least periodically active when the cross-correlation between the corresponding the BOLD-components of interest is greater than 0.6, is based on a visual analysis explained in Section 3.4.

CHAPTER 5

NUMERICAL ANALYSIS AND PRACTICAL APPLICATIONS FOR NEURAL PLASTICITY

In this Chapter, we will first give a detailed description of the neural plasticity experiment and the available fMRI data set in Section 5.1. The concepts of a hemodynamic response will be introduced in Subsection 5.1.1, the understanding of which is very helpful for the evalution of noise model and for the design and analysis of the simulation experiment (see Section 5.2). Then, the proposed wavelet methodology is applied on the fMRI data at hand. Informative figures and the corresponding discussions are given in Section 5.3. Finally, a conclusion of this fMRI neural plasticity study is made in Section 5.4.

5.1 fMRI Neural Plasticity Experiment

The neural plasticity experiment, conducted on twenty-four healthy right-handed adult volunteers (participants) by the University of Texas Southwestern Medical Center, includes three scan states:

(1) A five-minute **pre-training** fMRI run when a volunteer was instructed to just look at a white crosshair. Due to the five minute time and the repetition time (TR) of 1 second, 300 observations of the BOLD signal were obtained for each voxel.

(2) A twenty-three-minute **motor task** period when a volunteer was asked to press by a right-hand thumb a button three times when the color of the crosshair changed. The color change occurred every 27-32 seconds randomly and there were 40 stimuli. The data for the motor task period was divided and recorded into 4 parts and 340 observations of the BOLD signal were obtained for each part (This is due to the computer memory limitation).

(3) A five-minute **post-training** fMRI run when a volunteer was instructed to just look at a white crosshair, and 300 observations of the BOLD signal were obtained for each voxel.

The experiment was performed on a 3T system and the data involves 11 slices. All technical details can be found in Tung et al. (2013).

Data, available for statistical analysis, is as follows. For 1000 voxels in each hemisphere (located in the motor cortices), we have time series of BOLD signals for: **pre-training** (resting) run with 300 observations; four **motor-task** parts with 340 observations per part; **post-training** (resting) run with 300 observations. Note that the MRI machine scans the brain slice by slice. The original raw data contains 54 slices and the signals from the whole brain. A standard procedure in an fMRI study is to extract signals within the ROI (motor cortices in our study) and work with the selected data only. Remember that the data at hand contains signals from 11 slices. Also note that the slices are scanned successively, which therefore implies the slice timing issue. However, the delay between each slice scan is negligible if compared to the low frequency signal component of our interest. Some illustrations can be found in the descriptions of Figure 3.2 in the Section 3.4. Hereafter, slice timing will not be considered.

Because of the large number of interhemispheric voxel-pairs and large fMRI noise, the solution proposed in Tung et al. (2013) is based on averaging images in each hemisphere, calculating cross-correlations for average images in pre- and post-training runs, and then using a paired t-test for the 24 participants. This testing procedure supported a conjecture about the increase in cross-correlation after training sessions among the 24 participants and, as a result, supported the conjecture about the brain plasticity. Efromovich and Smirnova (2014) made an attempt to analyze plasticity via a wavelet approach, however using a standard procedure of denoising has allowed the authors to analyze only a single slice with large signal-to-noise ratio.

We would like to investigate the alterations in the inter-hemispheric neural connectivity due to the motor task on a voxel-to-voxel level throughout the entire ROI. This implies the statistical analysis of 1000×1000 (inter-hemispheric) cross-covariance and cross-correlation

matrices during each scan using $n = 256$ observations, which is considered as a large-p-small-n problem. The technical solutions to this Big Data challenge are presented in the Chapter 3 and Chapter 4. In this Chapter, evaluations of the model and analysis of the real data will be given.

5.1.1 Hemodynamic Response

Hemodynamic response is a spontaneous body reaction to physical activities or environmental stimuli by delivering oxygenated blood to the functioning tissues or neurons in order to supply energies. Therefore, neural activities are associated with the distribution of cerebral blood flow. The fact that oxygenated and deoxygenated forms of blood are detectable with fMRI provides an indirect measure of the hemodynamic response.

Figure 5.1 presents a standard *hemodynamic response function* (HRF) from a single voxel due to a single stimulus. The x-axis represents the time in seconds. The y-axis represents the percentage of the BOLD signal change. Therefore, $y = 0$ is the baseline level of a BOLD signal. It is supposed that a stimulus occur at time zero. According to Figure 5.1 there is a delay of approximately 2 seconds before any change in the BOLD signal is observed. At the beginning of the change, the HRF increases gradually (using about 4 seconds) and reaches it peak around 6 seconds after the stimulus. Then, the HRF decreases slowly to the lowest point which is slightly below the baseline level of a BOLD signal and then slowly recovers to the baseline level. The observed whole process of the hemodynamic response (or the BOLD signal change) due to a single stimulus can last for approximately 15 to 20 seconds, depending on the type of stimulus.

The knowledge of the HRF also sheds light on the motor task design in the fMRI plasticity experiment conducted by Tung et al. (2013). Recall that during the motor task period, the stimulus (button clicking following the cross-hair color change) occurred every 27 to 32 seconds randomly. Note that the time gap between two successive stimuli is greater than

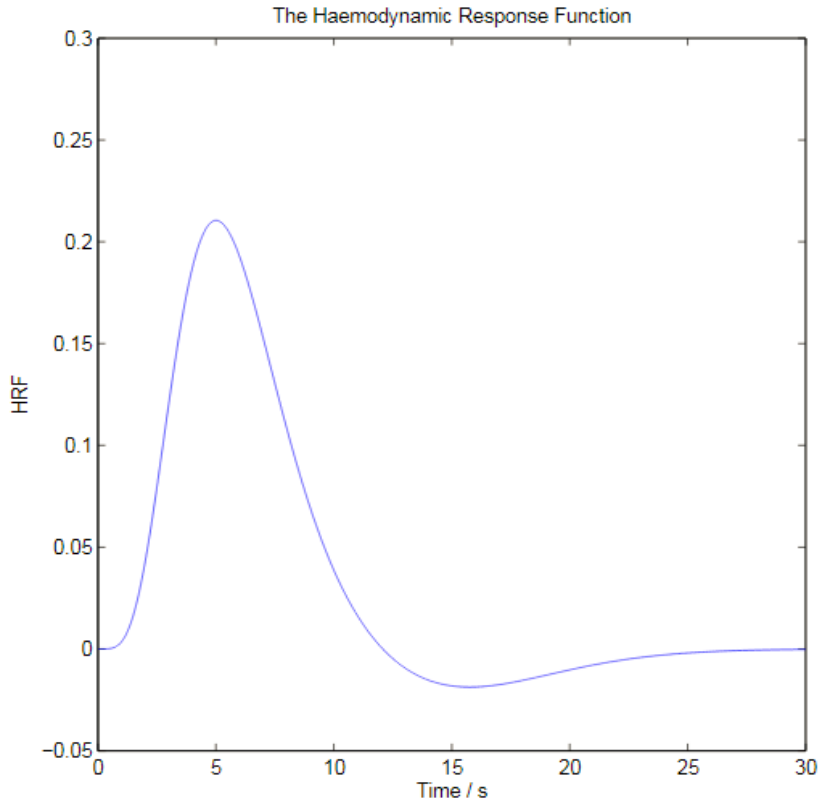


Figure 5.1. Hemodynamic response function.

the support of the an HRF. Therefore, overlapping of two HRF can be avoided and we are expecting to observe approximately 8 to 9 HRF in the BOLD signal every $n = 256$ seconds during the motor task.

5.2 Evaluation of Noise Model and Simulations

Recall the noise model proposed in Section 3.3, where we assume that the physiological factors might be present on two finest scales in the wavelet decomposition of a BOLD signal and the physiological activity in two neighboring voxels are similar. Based on these assumptions, a simple functional relationship regression with EIV model is built in (3.6). In this section, we will evaluate this noise model.

For each voxel, consider its 3×3 vicinity in the same slice (to avoid slice timing issue). Consistent with the notations used in the Section 3.3, let

$$\tilde{Y}'_l := \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \tilde{\theta}_{jk} \psi_{jk}(l/n)$$

be the sum of first and second wavelet scales in an empirical wavelet decomposition of an observed BOLD signal from a voxel of interest, and let

$$\tilde{X}'_l := \sum_{j=1}^{2} \sum_{k=1}^{n/2^j} \tilde{\theta}'_{jk} \psi_{jk}(l/n)$$

be the counterpart from a voxel in its 3×3 vicinity. The correlations between a signal component $\tilde{Y}'$ for a particular voxel and the signal components $\tilde{X}'$ for all voxels within its 3×3 vicinity are computed. Figure 5.2 shows a set of histograms, based on all voxels in both hemispheres of a participant, for weakest correlations between a signal $\tilde{Y}'$ for a particular voxel and signals $\tilde{X}'$ for all other voxels from the same slice within its 3×3 vicinity. Each histogram corresponds to one of the 6 runs of the studied fMRI experiment for a single patient, so we can visualize the dynamics of changes. Figure 5.3 shows similar histograms only now for the strongest correlations. We clearly see that signals on the finest scales are correlated, and hence this fact may be used in estimation of the parameter τ.

There is one more possibility to shed light on the proposed estimation of the standard deviation of the noise in empirical wavelet coefficients in Section 3.3. Let us consider a wavelet estimate of Efromovich (1999a,b) recommended for fMRI signals. Figure 5.4 presents two outcomes for a particular voxel in the left hemisphere during second training session when we should observe a hemodynamic response due to a finger's clicking stimulus that occurs every 27-32 seconds. The only difference in the two columns of diagrams is that in the left the wavelet estimate uses the *mad* procedure to estimate the level of noise in empirical wavelet coefficients while in the right the proposed regression procedure. The rows of diagrams show (from the top to the bottom): Observed signal and estimate of the underlying signal (the

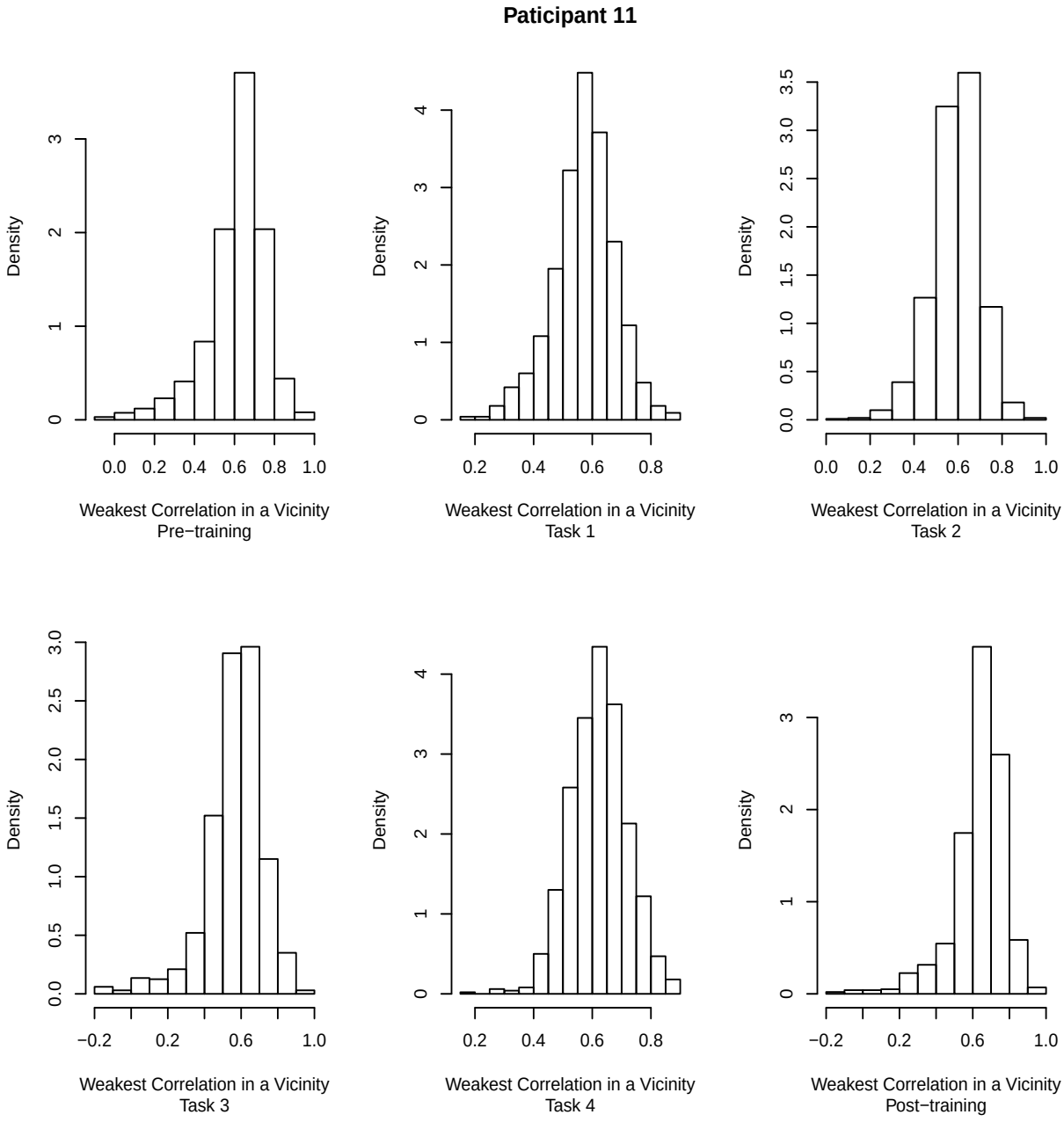


Figure 5.2. Histograms of the weakest correlation between the sums of signals on the first and second scales for a motor-cortex voxel and other voxels in its 3×3 vicinity.

denoised signal); Noise estimated as the sum of signals on the first and second scales minus the denoised signals; Denoised signal on the first scale denoted as iwt1 which stands for

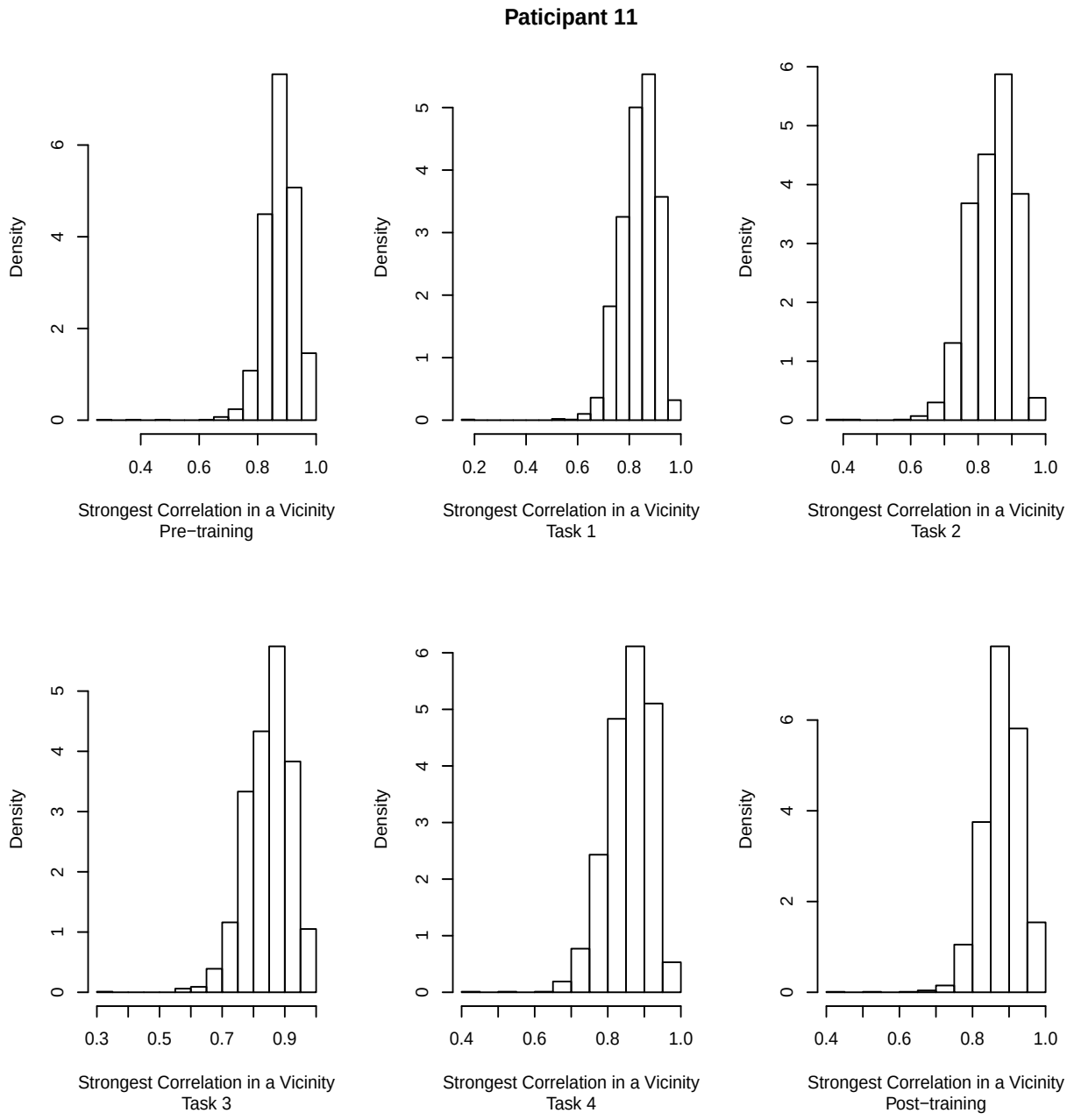


Figure 5.3. Histograms of the strongest correlation between the sums of signals on the first and second scales for a motor-cortex voxel and other voxels in its 3×3 vicinity.

inverse-wavelet-transform for the first scale; Denoised signal on the second scale denoted as

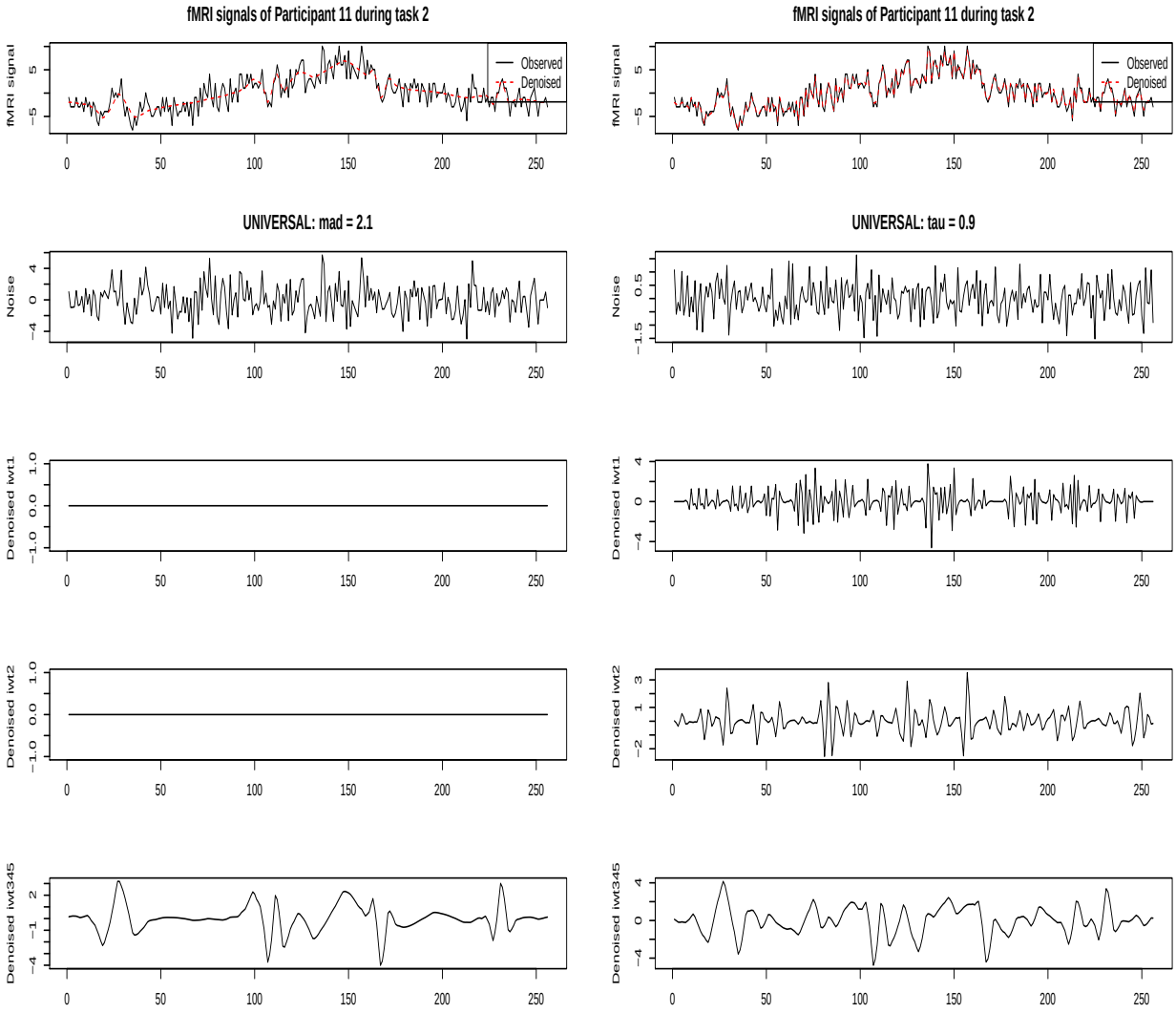


Figure 5.4. Wavelet denoising a signal during a second training scan. Left and right columns show the Universal denoising method of Efromovich (1999a) with the *mad* and the proposed method of estimation of the noise standard deviation.

iwt2; the sum of denoised signals on scales 3, 4, and 5 which is the denoised BOLD-component of interest denoted as iwt345.

The third to fifth rows of Figure 5.4 are of our special interest. Note that the *mad* estimate, equal to 2.1 is too large and it implies no deterministic signals on scales 1 and 2, and what is even more important, we do not see hemodynamic responses every 27-32 seconds

and lasting for about 15-20 seconds (Lazar 2008). (Recall the definition of a hemodynamic response and the shape of a HRF presented in Figure 5.1.) For instance, the time gap between the first two hemodynamic responses is more than 50 s, and this is not reasonable because stimulus occurs every 27-32 seconds. We may conclude that the *mad* clearly hides hemodynamic responses that are clearly observed in the right-bottom diagram.

In order to evaluate the proposed estimate of the standard deviation of the noise in the empirical wavelet coefficients, we design a simulation experiment. Details of the simulation setting are given below.

The simulation study adopts an *additive data generation* model in which three main components can be distinguished: (1) BOLD activation (2) physiological fluctuations (3) white noise. The BOLD activation is simulated by assuming that the hemodynamic response function is double-gamma, stimulus present every 32 seconds and each lasts for 15 seconds. Physiological fluctuation is modeled as sine and cosine functions with different frequencies. White noise is assumed to follow Gaussian distribution. We design the simulation to mimic a real fMRI experiment where TR = 1s, the scan session lasts 5 min which implies 300 observations at each voxel. The above described three components are generated using the R package *neuRosim* (Moerkerke et al. 2015; Welvaert et al. 2011).

A 3 by 3 vicinity of voxels is simulated based on the assumption that the BOLD activation and physiological fluctuations are the same among the 9 voxels but with random white noise. N=10000 simulations of such vicinity are conducted with varied simulation parameters, including frequency of heart beat (freq.heart), frequence of respiration (freq.resp), $\frac{A}{\tau_N}$ and $\frac{\tau_p}{\tau_w}$. Here A is the amplitude of the BOLD activation, τ_p is the standard deviation of the physiological fluctuations, $\tau_w = \tau$ is the standard deviation of white noise and $\tau_N = \sqrt{\tau_w^2 + \tau_p^2}$ is the standard deviation of the additive composition of physiological fluctuations and white noise. Note that $\frac{A}{\tau_N}$ is called *contrast-to-noise ratio* (CNR) (Welvaert and Rosseel 2013).

Table 5.1. Results of simulations. An entry in the Table is written as MdAPE/MdRAE.

		$\frac{\tau_p}{\tau_w}$					
freq.heart	freq.resp	0.5	1	2	0.5	1	2
1.17	0.2	0.05/0.62	0.04/0.33	0.04/0.09	0.05/0.62	0.04/0.33	0.04/0.09
1.17	0.3	0.04/0.41	0.04/0.10	0.04/0.03	0.04/0.40	0.04/0.10	0.04/0.03
1	0.2	0.05/0.58	0.04/0.23	0.04/0.06	0.04/0.57	0.04/0.23	0.04/0.06
1	0.3	0.04/0.23	0.04/0.05	0.04/0.02	0.04/0.23	0.04/0.05	0.04/0.02
		$\frac{A}{\tau_N} = 0.5$			$\frac{A}{\tau_N} = 1$		

Two error measures are computed and presented in Table 5.1. One is the *Median Absolute Percentage Error* (MdAPE), which is defined as,

$$MdAPE = \text{median}_{i=1,\ldots,N}\left(\frac{|\hat{\tau}_w^i - \tau_w|}{\tau_w}\right);$$

another is *Median Relative Absolute Error* (MdRAE), which is defined as,

$$MdRAE = \text{median}_{i=1,\ldots,N}\left(\frac{|\hat{\tau}_w^i - \tau_w|}{|mad^i - \tau_w|}\right),$$

where $\hat{\tau}_w^i$ is the proposed estimate of the standard deviation of the noise in the empirical wavelet coefficients in the ith simulation and mad^i is the standard method implemented in the wavelet softwares. See equations (3.6) - (3.9) for the definition of $\hat{\tau}_w^i$ and see equation (2.18) for the definition of mad^i.

According to Table (5.1), we can see that the proposed estimate is very stable. The MdAPE of the proposed estimate is always 4% to 5% no matter what is the heartbeat frequency, respiratory frequency, CNR or $\frac{\tau_p}{\tau_w}$. By looking into the *mad* procedure, we discover that the *mad* of the finest scale of the wavelet decomposition always overestimates the underlying true standard deviation of the white noise τ_w. By comparing MdRAEs under different

simulation parameters, we can see that the error of the *mad* highly depends on the heartbeat frequency, respiratory frequency and $\frac{\tau_p}{\tau_w}$. The higher the ratio $\frac{\tau_p}{\tau_w}$ is, the more error in *mad*. This is because when TR = 1 s, the finest scale contains some physiological components, which dramatically reduce the performance of *mad* as an estimate for τ_w especially when τ_p is large relative to τ_w.

5.3 Practical Results

The proposed wavelet methodology of simultaneous inference about large cross-correlation matrices allows us to shed new light on the fMRI study described in Section 5.1. In this section three interesting results are presented that illustrate capabilities of the new methodology.

Let us recall some terminology and specific parameters used in the analysis. An inter-hemispheric *neural pathway* is the connectivity between neurons to enable a signal to be sent from one brain hemisphere to another. According to our explanation in Section 3.4, we are saying that there is an *active* pathway between two voxels in different hemispheres if the cross-correlation between the hemodynamic responses exceeds a specific threshold, and in particular threshold 0.6 is chosen. The reason for choosing cross-correlation 0.6 as the threshold is discussed in Section 3.4. Because we estimate pairwise cross-correlations with errors and would like to make a simultaneous conclusion for all possible pairs of voxels, the approach proposed in Chapter 4 is used with the simultaneous 0.95 degree of confidence (confidence coefficient). Note that we are using lower confidence bounds (and not intervals) to define if underlying cross-correlations exceed the threshold 0.6, and then the Holm procedure is used to get the simultaneous 0.95 degree of confidence.

Now we are in a position to present the first result. Figure 5.5 summarizes changes in the number of active inter-hemispheric pairs (pathways) from the pre-training to post-training for each of 24 participants. The two top rows show statistics for patients with increased,

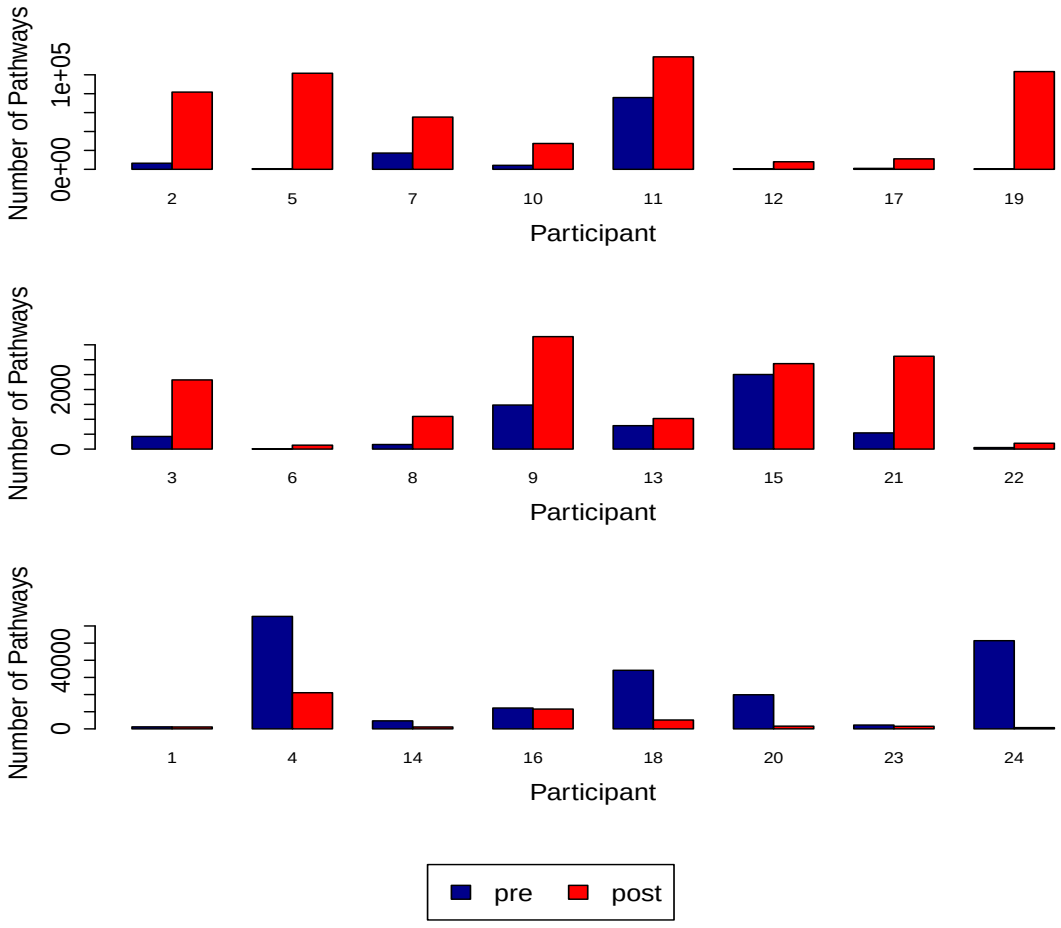
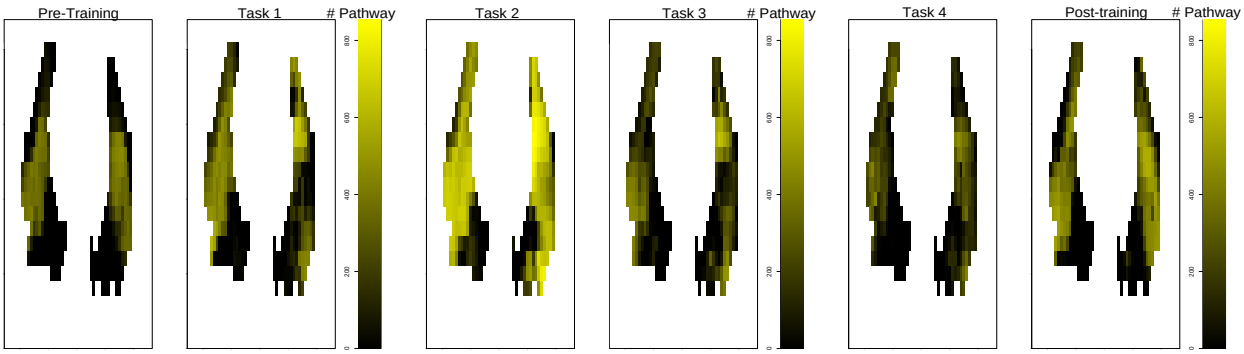


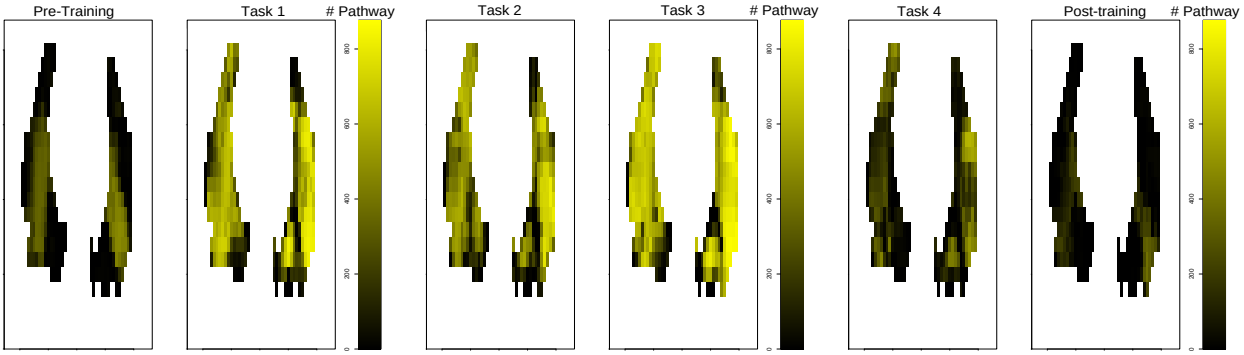
Figure 5.5. Bar chart of the number of active interhemispheric neural pathways for each participant during pre-training and post-training.

from pre- to post-training, number of active pairs (note the different scales used in these rows). These are cases that support the neural plasticity phenomenon. On the other hand, the bottom row present results for 8 participants with decreased number of active pairs. Note the large volatility in the number of active pathways among the 24 young and healthy participants.

Next, let us look at Figure 5.6 which allows us to understand what is going on during 4 training tasks and compare them with pre- and post-training sessions. In other words, we are visualizing the dynamic of active pathways. Figure 5.6 presents the dynamics for



(a) Participant 11



(b) Participant 4

Figure 5.6. Heat map of the motor cortices during each scan as labeled. The color of each voxel represents the number of interhemispheric neural pathways in which the corresponding voxel is involved. Brighter color yields more interhemispheric neural pathways.

Participant 11 and Participant 4. Note that, according to Figure 5.5, the former exhibited the increased number of active pathways from pre-training to post-training while the latter the decrease number. Each row of diagrams in Figure 5.6 presents a so-called *heat map* of the motor cortices in the fifth fMRI slice during each of 6 scans in chronological sequence. The color of each voxel represents the number of active pathways in which the corresponding voxel is involved. In the white-black figure, the darker the color the smaller the number of active pathways, and vise versa the whiter the color the larger this number. More informa-

tive colored figures can be found in the online supplementary material Efromovich and Wu (2018a).

There are several interesting observations. First of all, we see a dramatic increase in active pathways during the initial training tasks (sessions) but then toward the last training session this number decreases. The decrease is especially noticeable for Patient 4 (recall that this is the one whose number of active pathways decreased after training). At the same time, for Patient 11 we observe the increase in activity during the post-training period. Second, we may notice that the brighter color occurs in clusters, and hence active neurons also work together in clusters. Third, for the both patients the yellow regions, seen for the pre-training state, are enlarged and brightened during training states. Fourth, the active regions during Task 4 are shrunk and darkened with respect to the previous training tasks. Fifth, Patient 11 exhibits an interesting pattern in the increased number of active pathways during post-training period with respect to Task 4.

The dynamic analysis of heat maps again confirms our conclusion that brain activity and its reaction to training vary widely among individuals.

Finally, let us show how the proposed methodology allows us to point upon performance of individual voxels. Let us define 4 types of voxels. A voxel is called: (i) "Active" if it is active (has at least one active interhemispheric pathway) during pre- and post-training; (ii) "Inactive" if it has no active interhemispheric neural pathways during pre- and post-training; (iii) "New" if it is not active during pre-training and active during post-training; (iv) "Disappear" if it is active during pre-training and not active during post-training.

Figure 5.7 and Figure 5.8 display the motor cortices for participant 12 and participant 16, respectively, in 11 slices sequentially, where the color of each voxel represents one of the above-defined types of the voxels. (More informative colored figures can be found in the online supplementary material Efromovich and Wu (2018a).) Figure 5.7 and Figure 5.8 allow us to visualize the location of different types of voxels throughout the whole motor

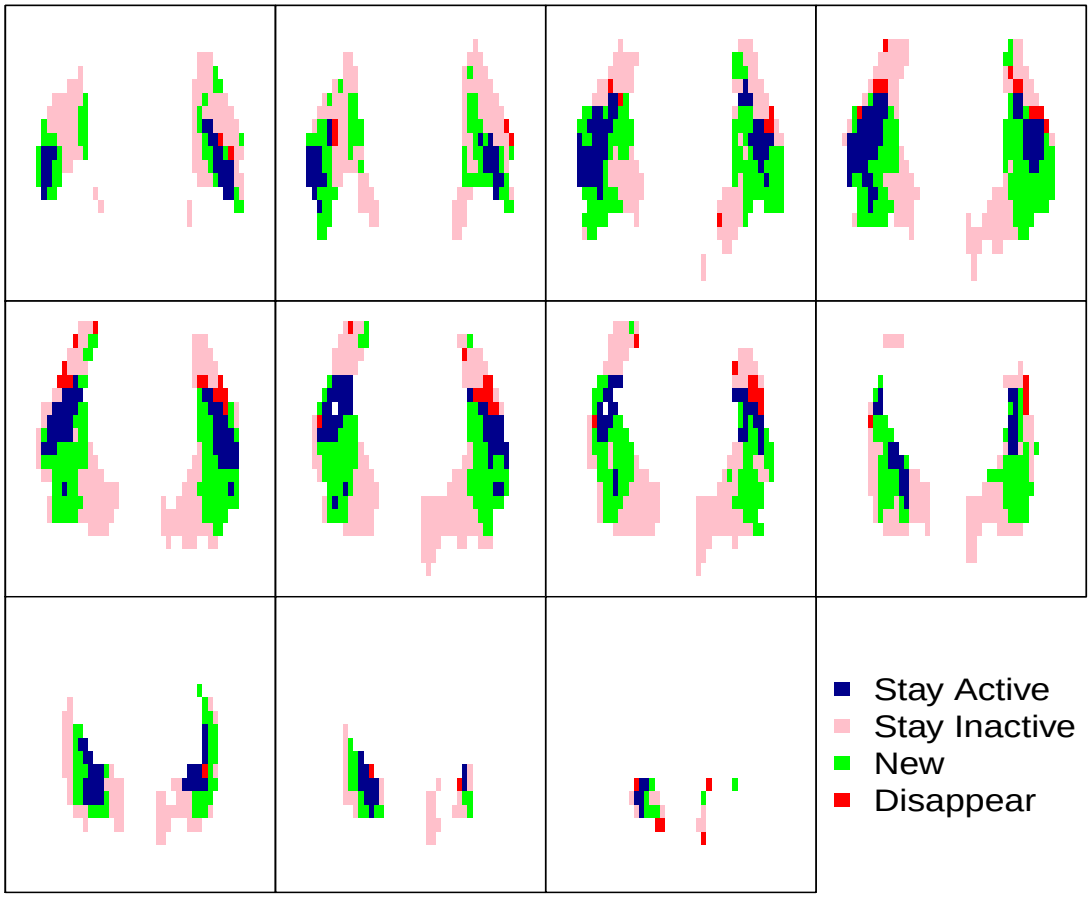


Figure 5.7. Motor cortices of participant 12. The 11 slices are placed sequentially along the rows beginning with the top row.

cortices. For the plasticity phenomenon, the most interesting are "new" voxels, and note that they are primarily located near the boundary of "active" voxels.

5.4 Conclusion

Functional magnetic resonance imaging (fMRI) is a powerful tool that may allow us to analyze brain activity on a voxel level. To achieve such a possibility, new wavelet estimation and inference procedures are proposed that allow us to simultaneously analyze BOLD signals at each voxel in the right and left hemispheres. Further, a feasible statistical procedure for

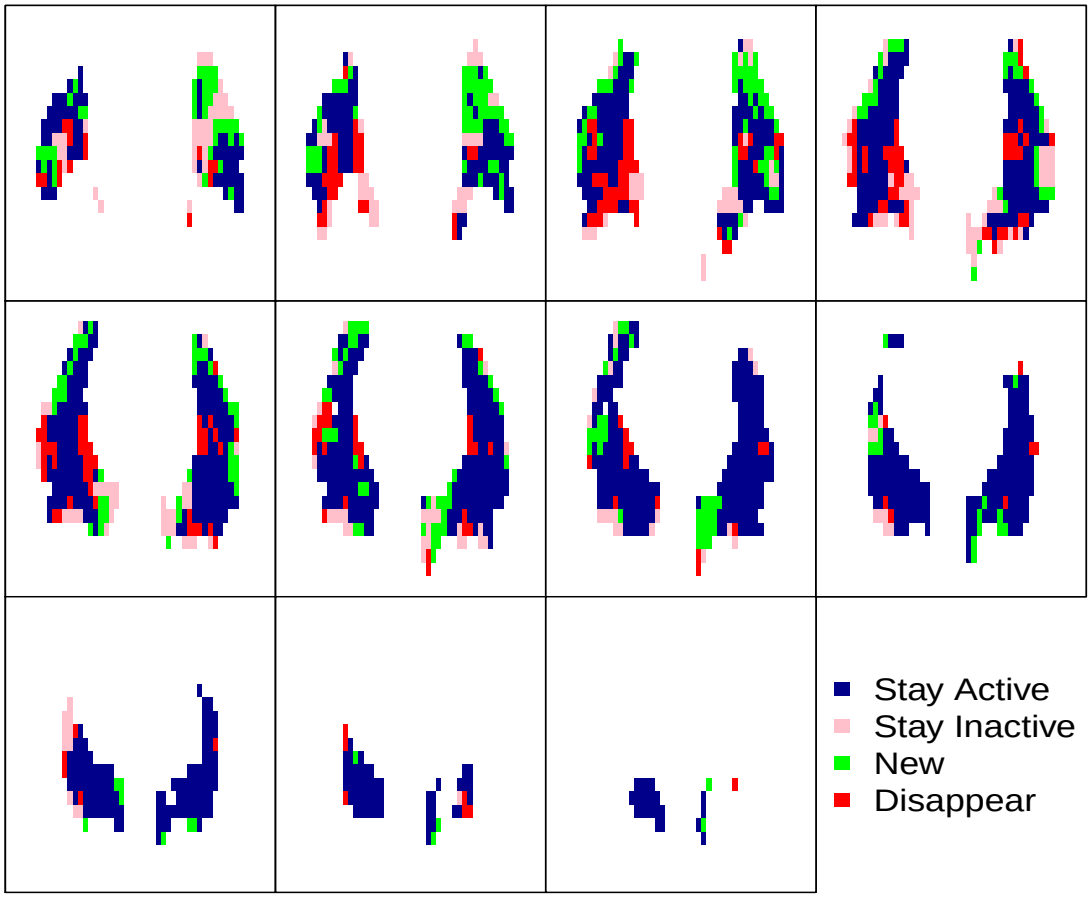


Figure 5.8. Motor cortices of participant 16. The 11 slices are placed sequentially along the rows beginning with the top row.

estimation of and inference about large-p-small-n cross-correlation matrices is developed. The innovative features are the robust procedure of estimation of the level of fMRI noise in empirical wavelet coefficients and the numerical procedure of taking into account sub-Gaussian distributions arising in the analysis of cross-correlation matrices. The proposed statistical methods are supported by the asymptotic theory and tested in a numerical study.

The developed statistical methods have been used for analysis of an fMRI study of neuron plasticity of 24 healthy adults. The aim of the study was to recognize changes in connectivity between left and right motor cortices after button clicking training sessions. A conventional

method of the data analysis, based on averaging images, has implied that for the group of 24 participants the connectivity increased after the training. The proposed wavelet analysis has allowed us to conduct a corresponding statistical analysis for each participant. Namely, for each participant and a simultaneously analysis of all potential neural pathways between left and right hemispheres was conducted via analysis of corresponding cross-correlation matrices.

Obtained results shed a new light on the dynamic of changing the number of active pathways over the time of pre-training, 4 training and post-training sessions. Further, now it is possible to identify pathways that always stay active, become active during training and then remain active during post-training session (the plasticity). The results clearly indicate the ability of a human brain to reorganize itself due to a training, but they also indicate that the ability varies widely even among healthy individuals. The latter observation is important for our understanding of the human brain and possible treatments of brain diseases.

CHAPTER 6

DYNAMIC NONPARAMETRIC ANALYSIS OF NONSTATIONARY ASSET RETURNS AND ITS APPLICATIONS

6.1 Introduction

Construction of many actuarial models is based on the knowledge of the probability distribution of asset returns. Statistical description of financial data often assumes that the returns are independent and identically distributed (Kosta and Stepanova 2015; Rapach and Zhou 2013; Tan and Chu 2012). However, empirical finance has already shown that the asset return time series is subject to data dependence as well as the distribution of returns tend to have heavy tails (Chen and Tang 2005). This paper relaxes this assumption and proposes a new approach based on nonparametric method of analysis.

Conventionally, asset returns are assumed to be identically distributed over time. The classical parametric distribution assumptions for returns include normal distribution, lognormal distribution and non-Gaussian stable distributions. However, each of these distributions has its own disadvantage, which could be crucial for an accurate analysis. Besides the three conventional distributions, other paremetric distribution assumptions of asset returns include the Student t, the skewed Student t, the generalized t, and more sophisticated, autoregressive conditional heteroskedastic (ARCH) or generalized ARCH (GARCH) models (Tan and Chu 2012).

There are also model-free nonparametric attempts in the distribution estimation of financial data. Nonparametric estimation of the distribution of asset returns has two advantages: (i) being free of distributional assumptions on the asset returns, and meanwhile being able to capture tail behavior automatically; (ii) requiring milder assumptions on the dynamics of the asset return time series, which means being free of identical distribution or stationarity assumption on asset returns.

This work is motivated by the above-described advantages of nonparametric approaches. We propose a novel nonparametric estimetor for the distribution of asset returns. The approach can be applied to dependent and nonstationary financial data and it captures the dynamics in the distribution of asset returns over time. The proposed estimation procedure contains two steps. Our **first step** is to denoise the asset price time series and compute the asset returns using the denoised prices. By doing this step, we reduce the noise in returns which is introduced by the sizable unpredictable component in the prices. It is proposed to use an aggregated wavelet estimator for asset prices denoising. Our **second step** is to estimate the *time-varying* probability density of the asset returns. The time-varying density refers to a density function which take two arguments: the value of return and the time. The estimation of the time-varying density is done in an indirect way. It is proposed to consider the asset returns under a nonparametric heteroscedastic regression model, where the probability density of the regression errors is estimated. Then, an density estimator of the asset returns is defined as the transfomed (shifted and rescaled) regression errors density estimator.

The work is organized as follows. Section 6.2 gives a review on the nonparametric estimation. An aggregated Wavelet estimator is described in Section 6.2.1 along with its minimax property. Section 6.2.2 descibes a density estimator of regression errors in a nonparametric heteroscedastic regression model. Section 6.3 presents the propsed methodology and models. Practical applications can be found in Section 6.4.

6.2 Literature Review

6.2.1 Aggregated Wavelet Estimator

Consider a time series

$$Y_l = f(l/n) + \sigma\epsilon_l, \;\; l = 1, 2, \ldots, n. \tag{6.1}$$

where ϵ_l are zero mean and unit variance noise and n is the number of observations (sample size, or signal length). Note that we may refer to the time series as homoscedastic equidistant regression and hence we can consider $f(t)$ as a trend (plus possibly a seasonal component). If there is no prior knowledge of f, it is natural to use an adaptive nonparametric curve estimation. A nonparametric procedure is data-driven and requires neither information about shape or smoothness of f nor tuning of smooting parameters. A variety of adaptive nonparametric procedures is described in Efromovich (1999a).

One way to improve the performance of an estimation is to *aggregate* several estimators in a suitable manner. That is, given a collection of estimates $\{\tilde{f}_1(t), \tilde{f}_2(t), \ldots, \tilde{f}_K(t)\}$, an aggregated estimate is defined as

$$\tilde{f}(t) := \sum_{i=1}^{K} \lambda_i \tilde{f}_i(t),$$

where $\sum_{i=1}^{K} \lambda_i = 1$. We introduce the aggregated wavelet estimator porposed in Efromovich and Valdez-Jasso (2010).

Let us recall the wavlet expansion of a square integrable function f on $[0,1]$. Given wavelet basis $\{\phi_{j_0k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k),\ \ k = 0, 1, \ldots, 2^{j_0} - 1\}$ and $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ \ k = 0, 1, \ldots, 2^j - 1,\ \ j \geq j_0\}$, $x \in [0,1]$, where ϕ and ψ are scaling function and wavelet function, respectively, we have

$$f(x) = \sum_{k=1}^{2^{j_0}} s_{j_0,k}\phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k=1}^{2^j} \theta_{j,k}\psi_{j,k}(x). \tag{6.2}$$

Here, $s_{j_0,k} = \int_0^1 f(t)\phi_{j_0,k}(t)dt$ and $\theta_{j,k} = \int_0^1 f(t)\psi_{j,k}(t)dt$ are wavelet coefficients of f.

Due to the design of the numerical discrete wavelet transform, the input signal to this algorithm must be of a dyadic length. Therefore, let us suppose $n =: 2^J$ for some integer $J > 0$. Then we can obtain n empirical wavelet coefficients $\{\tilde{s}_{j_0,k}, k = 1, \ldots, 2^{j_0}, \tilde{\theta}_{j,k}, k = 1, \ldots, 2^j, j = j_0, \ldots, J-1\}$ through standard discrete wavelet transform of the observed time series $\{Y_l, l = 1, \ldots, n\}$. Empirical wavelet coefficients are essentially statistics, which are

used to construct estimators of underlying wavelet coefficients. Note that empirical wavelet coefficients are noisy. Therefore, the wavelet denoising procedure is critical in building a good estimator.

Let us introduce the SureBlock wavelet denoising proposed by Cai et al. (2009). For wavelet scale $j \in \{j_0, \ldots, J-1\}$, and for translation index $k \in \{(b-1)L_j^* + 1, \ldots, bL_j^*\}$ with some $b \in \{1, 2, \ldots, 2^j/L_j^*\}$, the SureBlock estimator is defined as

$$\hat{\theta}_{j,k} := \left(1 - \frac{\lambda_j^* \sigma^2 n^{-1}}{\sum_{k'=(b-1)L_j^*+1}^{bL_j^*} \tilde{\theta}_{j,k'}^2}\right)_+ \tilde{\theta}_{j,k}. \tag{6.3}$$

The following procedure describes how λ_j^* and L_j^* are selected. First, define a function with two arguments λ and L,

$$\begin{aligned} h_j(\lambda, L) := \sum_{b=1}^{m_j} \Big[& L + n^{-1}\sigma^2[\lambda^2 - 2\lambda(L-2)]\hat{S}_{j,b}^{-2} I(\hat{S}_{j,b}^2 > n^{-1}\sigma^2\lambda) \\ & + (n\sigma^{-2}\hat{S}_{j,b}^2 - 2L) I(\hat{S}_{j,b}^2 \le \sigma^2 n^{-1}\lambda)\Big], \end{aligned} \tag{6.4}$$

where $m_j := 2^j/L$, $\hat{S}_{j,b}^2 := \sum_{k=(b-1)L+1}^{bL} \tilde{\theta}_{j,k}^2$ for $b \in \{1, 2, \ldots, m_j\}$. Then λ_j^* and L_j^* are calculated using the following formula,

$$(\lambda_j^*, L_j^*) := \begin{cases} \text{argmin}_{((L-2)\vee 0) \le \lambda \le 2jL\log(2),\, 1 \le L \le 2^{j/2}} h_j(\lambda, L) & \text{if } \sum_{k=1}^{2^j} (\tilde{\theta}_{j,k}^2 n\sigma^{-2} - 1) > j^{3/2} 2^{j/2} \\ ((1 - j2\log(2)\sigma^2 n^{-1}/\tilde{\theta}_{j,k}^2)_+, 1) & \text{otherwise.} \end{cases}$$

Efromovich and Valdez-Jasso (2010) proposed an estimator which aggregates two known wavelet estimators: SureBlock of Cai et al. (2009) and Universal of Efromovich (1999a). The aggregated wavelet estimator is defined as

$$\hat{f}_A(t) := \hat{f}_{SB,\hat{j}}(t) + \hat{f}_{U,\hat{j}}(t). \tag{6.5}$$

Here

$$\hat{f}_{SB,s} := \sum_{k=1}^{2^{j_0}} \tilde{s}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{s} \sum_{k=1}^{2^j} \hat{\theta}_{j,k}\psi_{j,k}(t) \tag{6.6}$$

is the lower-frequency part of the SureBlock estimator with $\hat{\theta}_{j,k}$ defined in (6.3), and

$$\hat{f}_{U,s}(t) := \sum_{j=s+1}^{J-1} \sum_{k=1}^{2^{2s-j}} \tilde{\theta}_{j,(k)} I(|\tilde{\theta}_{(j,(k))}| > (2^{j-s} \wedge (2\log(n))^{1/2})\sigma n^{-1/2})\psi_{j,(k)}(t) \tag{6.7}$$

is the high-frequency part of the Universal estimator. Here, $\tilde{\theta}_{j,(k)}$ are empirical wavelet coefficients on the jth scale in descending order and $\psi_{j,(k)}(t)$ are corresponding wavelet functions, and

$$\hat{J} := \text{argmin}_{j_0 \le s < J}\{2(2^{s+1} + N_s)\sigma^2 n^{-1} - \sum_{j=j_0}^{s} \sum_{k=1}^{2^j} \tilde{\theta}_{j,k}^2 - \int_0^1 \hat{f}_{U,s}^2(t)dt\}, \tag{6.8}$$

where $N_s = \sum_{j=s+1}^{J-1} \sum_{k=1}^{2^{2s-j}} I(|\tilde{\theta}_{(j,(k))}| > (2^{j-s} \wedge (2\log(n))^{1/2})\sigma n^{-1/2})$.

Efromovich and Valdez-Jasso (2010) proved that $\hat{f}_A(t)$ is rate minimax over a wide class of Besov spaces

$$B_{p,q}^{\alpha}(Q) = \{f : \ (\sum_{k=1}^{2^{j_0}} |s_{j_0,k}|^p)^{1/p} + (\sum_{j \ge j_0} (2^{j(\alpha+1/2-1/p)} (\sum_{k=1}^{2^j} |\theta_{j,k}|^p)^{1/p})^q)^{1/q} \le Q\}$$

with

$$1 \le p, q \le \infty, \quad r \ge \alpha > (4p^{-1} - 2)_+ + 1/2, \quad \frac{2\alpha^2 - 1/6}{1 + 2\alpha} > \frac{1}{p}, \quad Q < \infty.$$

6.2.2 Series Estimation of the Density of Regression Errors

Let us consider a fixed-design heteroscedastic nonparametric regression model

$$Y_l = f(l/n) + \sigma(l/n)\xi_l, \quad l = 1, \ldots, n. \tag{6.9}$$

Here, neither the regression function $f(x)$ nor the scale function $\sigma(x)$ is assumed to be known, $x \in [0, 1]$. $\{\xi_l, l = 1, \ldots, n\}$ are i.i.d. realizations from the zero-mean unit-variance regression error ξ whose probability density is unknown and is of our interest. Let us assume that the error ξ does not take values beyond a known finite interval $[a, a + b]$. Let $\epsilon := (\xi - a)/b$, and let $p_\epsilon(u), u \in [0, 1]$ represent its probability density. Therefore, the probability density

of ξ is $p_\xi(\nu) = b^{-1}p_\epsilon([\nu - a]/b)$, $\nu \in [a, a + b]$. Without any loss of generality, hereafter we consider the estimation of $p_\epsilon(u), u \in [0, 1]$.

The complication of such problem is that we do not have direct observations of the errors ξ or ϵ, which implies an indirect problem setting. In particular, we need to estimate the error density $p(u)$ based solely on n observations $\{Y_l, l = 1, \ldots, n\}$. Efromovich (2005) proposed to use the following procedure: the **first step** is to appropriately calculate the regression residuals; the **second step** is to use the residuals as regression errors proxy and plug them into a Pinsker oracle.

Let us describe how to calculate the proxy residuals. For $n = 1, 2, 3, \ldots$, define $b_n = 4 + \ln\ln(n + 20)$; $n_2 := n - 2n_1$; n_1 is the smallest integer larger than n/b_n; $S := S_n$ is the smallest integer larger than $n^{1/3}$. Consider only sufficiently large n such that $\min(n_1, n_2) > 4$. The observations $\{Y_l, l = 1, \ldots, n\}$ are randomly divided into three mutually exclusive parts; one part is of size n_2 and each of the rest two parts is of size n_1. The two nuisance functions $f(x)$ and $\sigma(x)$ are first estimated using distinct n_1 observations, respectively. Then, $p_\epsilon(u)$ is estimated using the part of n_2 observations. In particular,

$$\hat{f}(x) = \sum_{s=0}^{S} \hat{\kappa}_s \varphi_s(x), \quad \hat{\kappa}_s = n_1^{-1} \sum_{l=1}^{n_1} Y_l \varphi_s(l/n), \tag{6.10}$$

and

$$\hat{\sigma}(x) = [\min(\max(\tilde{\sigma}^2(x), b_n^{-2}), b_n^2)]^{1/2}, \tag{6.11}$$

where

$$\tilde{\sigma}^2(x) = \sum_{s=0}^{S} \hat{\beta}_s \varphi_s(x), \text{ and } \hat{\beta}_s = n_1^{-1} \sum_{l=n_1+1}^{2n_1} (Y_l - \hat{f}(l/n))\varphi_s(l/n).$$

Here $\{\varphi_0 := 1, \varphi_j = \sqrt{2}\cos(\pi j x), \text{ for } j = 1, 2, ...\}$ is a cosine orthonormal basis on $[0, 1]$. Then the residuals that are used as proxy to the standardized regression errors ϵ_l are defined as

$$\hat{\epsilon}_l := \frac{Y_l - \hat{f}(l/n)}{b\hat{\sigma}(l/n)} - \frac{a}{b}, \quad l = n - n_2 + 1, \ldots, n. \tag{6.12}$$

Let us introduce the Pinsker oracle for the regression error density. Suppose we observe the realization $\{\epsilon_1, \ldots, \epsilon_n\}$ of the regression error ϵ. Note that the density $p_\epsilon(u)$ of ϵ is supported on $[0, 1]$. For $k = 1, 2, 3, \ldots$, define $t_k := \ln^{-2}(2 + k)$; $L_k := k^2$; $q_1 := L_1$ and $q_{k+1} := L_k + q_k$; $B_k := \{q_k, q_k + 1, \ldots, q_{k+1} - 1\}$; K is a minimal integer such that $\sum_{k=1}^{K} L_k \geq n^{1/5} b_n$, and recall that $b_n = 4 + \ln \ln(n + 20)$. The Pinsker oracle is defined as

$$\hat{p}_\epsilon(z; \epsilon_1, \ldots, \epsilon_n) := 1 + \sum_{k=1}^{K} \bar{\mu}_k \sum_{j \in B_k} \bar{\theta}_j \varphi_j(z), \quad z \in [0, 1]. \tag{6.13}$$

Here, $\bar{\theta}_j$ are estimates of Fourier coefficients $\theta_j := \int_0^1 p(z)\varphi_j(z)dz$,

$$\bar{\theta}_j := n^{-1} \sum_{l=1}^{n} \varphi_j(\epsilon_l), \tag{6.14}$$

and $\bar{\mu}_k$ are *shrinkage coefficients*,

$$\bar{\mu}_k := \frac{L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2 - n^{-1}}{L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2} I\Big(L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2 > (1 + t_k) n^{-1}\Big). \tag{6.15}$$

Under mild assumptions, the Pinsker oracle, as a data-driven estimator based on n direct observations $\{\epsilon_1, \ldots, \epsilon_n\}$, is minimax for Sobolev and analytic classes (Efromovich 1999a). Therefore the plugged-in Pinsker oracle, which is a solely data-driven estimator, is

$$\hat{p}_\epsilon(z; \hat{\epsilon}_{n-n_2+1}, \ldots, \hat{\epsilon}_n), \tag{6.16}$$

where $\{\hat{\epsilon}_{n-n_2+1}, \ldots, \hat{\epsilon}_n\}$ are defined in (6.12). Note that $n_2 \geq [1 - 3(b_n^{-1} + n^{-1})]n$ and thus using either sample size of n_2 or n implies the same MISE convergence. Efromovich (2005) proved that under a mild assumption on the differentiability of the regression function, scale function and regression error density, the MISE of the plugged-in Pinsker oracle (6.16) satisfies the orale inequality.

Finally, the probability density estimator of ξ, the original regression error, is obtained as

$$\hat{p}_\xi(\nu) = b^{-1} \hat{p}_\epsilon([\nu - a]/b; \hat{\epsilon}_{n-n_2+1}, \ldots, \hat{\epsilon}_n), \quad \nu \in [a, a + b]. \tag{6.17}$$

6.3 Methodology and Models

We are interested in figuring out the dynamics over time in the distribution of the asset returns. As a result, we propose to estimate a time-varying probability density of asset returns using a nonparametric procedure. A time-varying density refers to a density function which take two arguments: the value of return and the time. The proposed procedure will be described in details shortly.

Stock prices inherently contain a sizable unpredictable component, which would cause large deviation in the calculation of stock returns. In particular, consider the time series

$$\tilde{P}(l/n) = P(l/n) + \nu\epsilon_l, \quad l = 1, \ldots, n, \tag{6.18}$$

where $\tilde{P}(t)$ is the observed stock price at time t, $P(t)$ is the underlying price, ν is the volatility (standard deviation) of price $\tilde{P}(t)$, and ϵ_l is a random variable with zero mean and unit variance. Without loss of generality, let us always rescale t to the interval $[0, 1]$.

Our **first step** is to denoise the stock prices. Due to the inhomogeneity of the stock prices, it is natural to use a wavelet denoising. Wavelets is a commonly used mathematical tool for approximation of spatially inhomogeneous curves. Let us consider (6.18) as a homoscedastic equidistant nonparametric regression model. Then we estimate the underlying prices $P(t)$ using the procedure described in Section 6.2.1. That is, the estimator $\hat{P}(t)$ can be calculated by repeating (6.5) – (6.8) with Y_l being replaced by $\tilde{P}(l/n)$ and σ being replaced by ν.

It is proposed to define the stock returns based on denoised prices as following

$$\tilde{R}(t) = \frac{\hat{P}(t) - \hat{P}(t-h)}{\hat{P}(t-h)}, \tag{6.19}$$

where h is a predefined time horizon of interest and $\hat{R}(t)$ is the return between time $t-h$ and t. The definition (6.19) reduces the volatility in returns that is inherted from the stock prices. In this paper, we are interested in the daily return so that the time horizon h is equal to one day.

In the **second step**, we estimate the time-varying probability density of asset returns using a nonparametric series estimator. In particular, consider a time series, defined in a heteroscedastic nonparametric regression setting,

$$\tilde{R}(l/n) := R(l/n) + \sigma(l/n)\xi_l, \quad l = 1, \ldots, n, \tag{6.20}$$

where $R(t)$ is the underlying asset return, $\sigma(t)$ is the volatility in asset return which depends on time, and $\{\xi_l, l = 1, \ldots, n\}$ are i.i.d. realizations from the zero-mean unit-variance regression error ξ. If the support of the probability density of ξ is finite and is known, say $[a, a + b]$, then a nonparametric estimator of the probability density of ξ can be obtained by repeating the procedure (6.10) – (6.17) with Y_l being replaced by $\tilde{R}(l/n)$. However, the support $[a, a + b]$ is unknown in practice and therefore needs an estimation. This is done by using the method proposed in Efromovich (1999a), i.e., estimating $[a, a + b]$ by

$$[2\hat{\xi}_{(1)} - \hat{\xi}_{(2)}, 2\hat{\xi}_{(n)} - \hat{\xi}_{(n-1)}],$$

where $\hat{\xi}_l = \frac{\tilde{R}(l/n) - \hat{R}(l/n)}{\hat{\sigma}(l/n)}$, and $\{\hat{\xi}_{(1)}, \hat{\xi}_{(2)}, \ldots, \hat{\xi}_{(n)}\}$ is the corresponding ordered statistics. Here, $\hat{R}(l/n)$ and $\hat{\sigma}(l/n)$ are the estimators of the regression function and scale funtion in the model (6.20), and they are calculated using (6.10) and (6.11), respectively.

As a result, it is proposed to construct a nonparametric estimator of the probability density of ξ, denoted as $\hat{p}_\xi(x)$, by repeating the procedure (6.10) – (6.17) with Y_l being replaced by $\tilde{R}(l/n)$, a being replaced by $2\hat{\xi}_{(1)} - \hat{\xi}_{(2)}$, and b being replaced by $2\hat{\xi}_{(n)} - \hat{\xi}_{(n-1)} - (2\hat{\xi}_{(1)} - \hat{\xi}_{(2)})$. Finally, the time-varying probability density estimation of return is given by

$$\hat{p}_R(x, t) = \frac{1}{\hat{\sigma}(t)} \hat{p}_\xi \left(\frac{x - \hat{R}(t)}{\hat{\sigma}(t)} \right), \tag{6.21}$$

where the function argument x represents the value of return, the argument t represents the time.

Calculating value at risk (VaR) is one of the most popular applications of the distribution of asset return. VaR is the standard approach to quantifying the exprosure of a financial

asset to the market risk, which is of great importance for risk management. It measures the loss of an asset over a holding period with a given probability. Typically, the holding period ranges from a few hours to a few days, while the probability ranges from 0.001 to 0.1 (Kosta and Stepanova 2015; Taylor 2007). Mathematically, $VaR_t(p)$ is the p-th quantile of the distribution of asset returns, defined by

$$P(\hat{R}(t) \leq VaR_t(p)) = p.$$

In this work, a time-varying VaR series is estimated by the sample percentiles from the realizations generated from the proposed time-varying probability density estimators of asset returns using the acceptance-rejection Monte-Carlo simulation approach.

The acceptance-rejection Monte-Carlo simulation, also known as rejection sampling, is a technique used to generate realizations from a given distribution. This technique is based on the fact that an i.i.d. sample of any random variable can be obtained by selecting uniformly distributed random points on the 2-D plane that is between the random variable's probability density curve and the x-axis. In particular, suppose we would like to obtain a sample from a random variable X with known probability density $p(x), x \in [0, 1]$, and suppose $p(x)$ is bounded and $M = \sup_{0 \leq x \leq 1} p(x)$. Now, repeat the following two steps until a pre-defined sample size is reached:

(i) generating a realization u from uniform$[0, 1]$ and a realization v from uniform$[0, M]$;

(ii) if $v < p(u)$, accepting u as a realization from $p(x), x \in [0, 1]$; otherwise, reject u and return to (i).

6.4 Results and Applications

Practical results based on the methodology introduced in Section 6.3 will be presented in this section. All the data used in application are downloaded from Yahoo Finance.

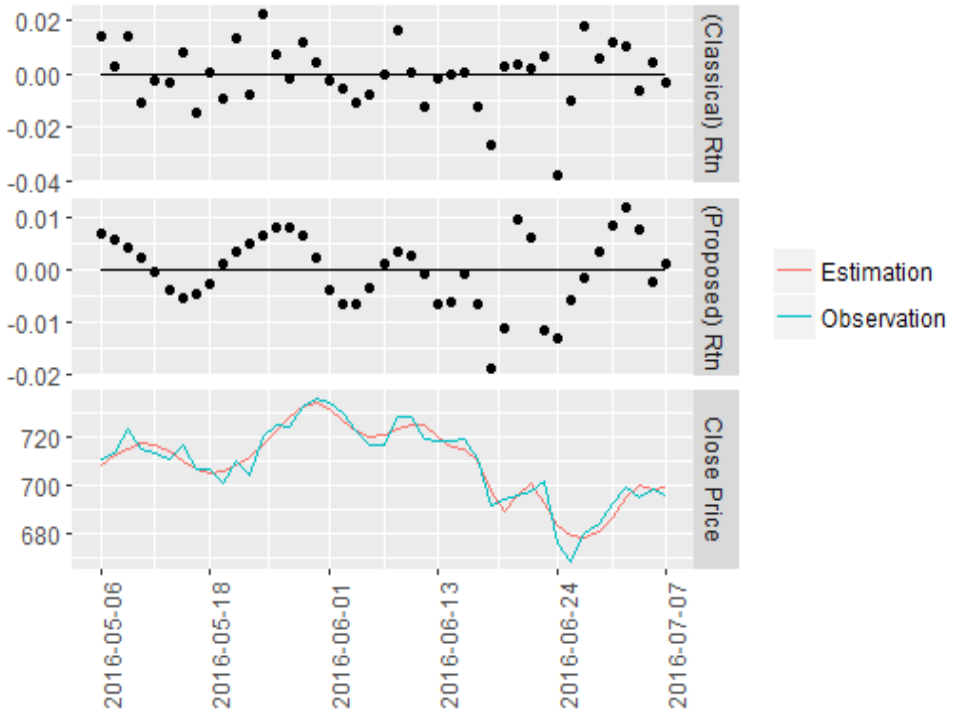


Figure 6.1. Example of 'GOOG'. The figure presents the observed daily prices, denoised prices, returns calculated based on observed daily prices and returns calculated based on the denoised prices.

In Figure 6.1, we use the stock 'GOOG' as an example. The daily adjusted close prices of 'GOOG' from 2016-05-06 to 2016-07-07 are obtained and denoised using the aggregated wavelet estimator described in Section 6.2.1, presented in the bottom subfigure, where is the blue curve represents the observed prices and the red curve represents the wavelet estimation. We can see that the observed price curve is spatially inhomogeneous, while the wavelet estimation curve is a good approximation of the observation but much smoother. The top subfigure presents the daily returns computed using observed stock prices. The middle subfigure presents the daily returns calculated using the denoised prices, that is using the formula (6.19). If we pay attention to the y-axis, we see that the volatility in the return is almost reduced to half after using the proposed calculation method.

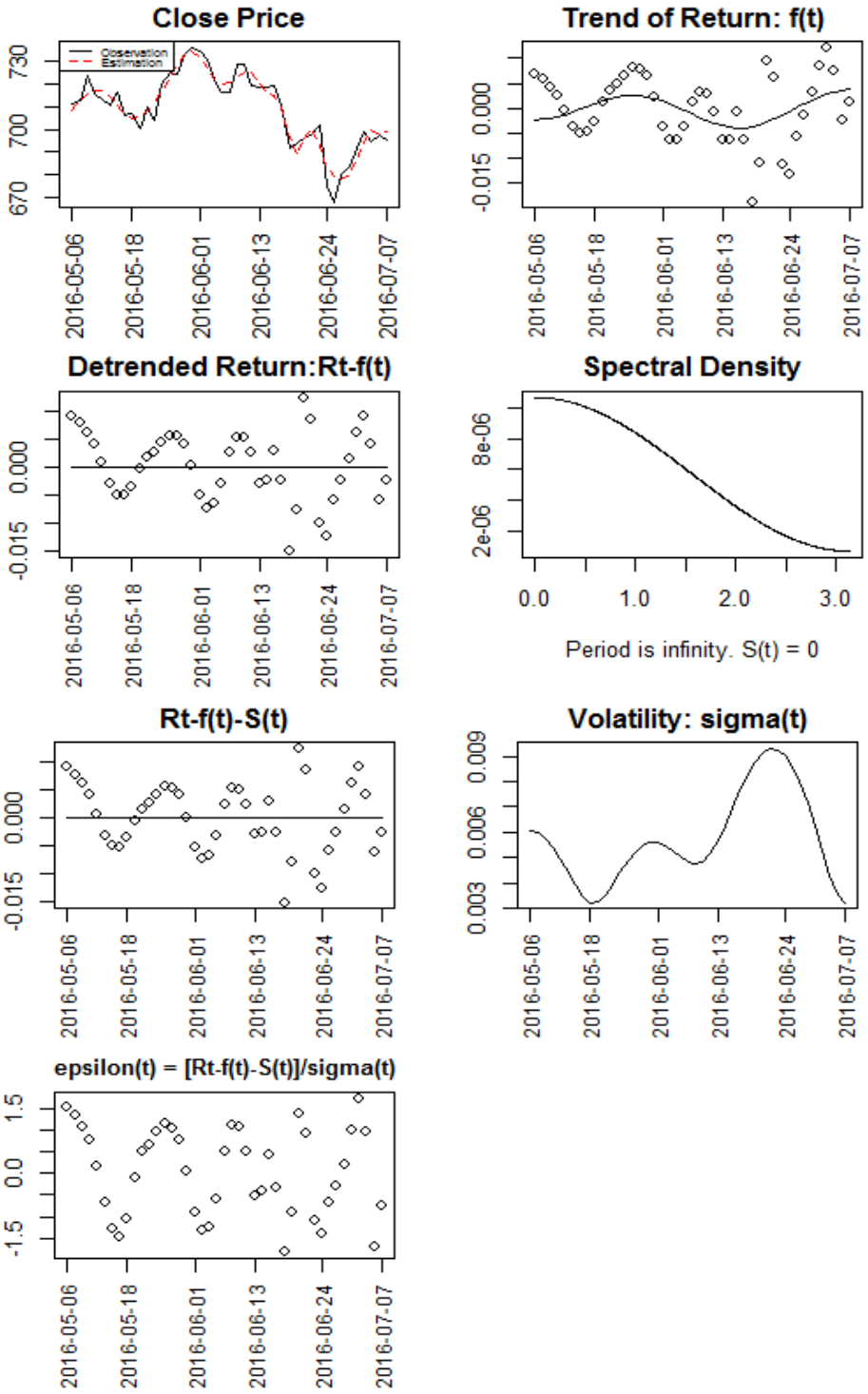


Figure 6.2. Example of 'GOOG'. The figure presents the observed daily prices, denoised prices, and the decompositon procedure of the daily returns (detrending, deseasonalizing and rescaling).

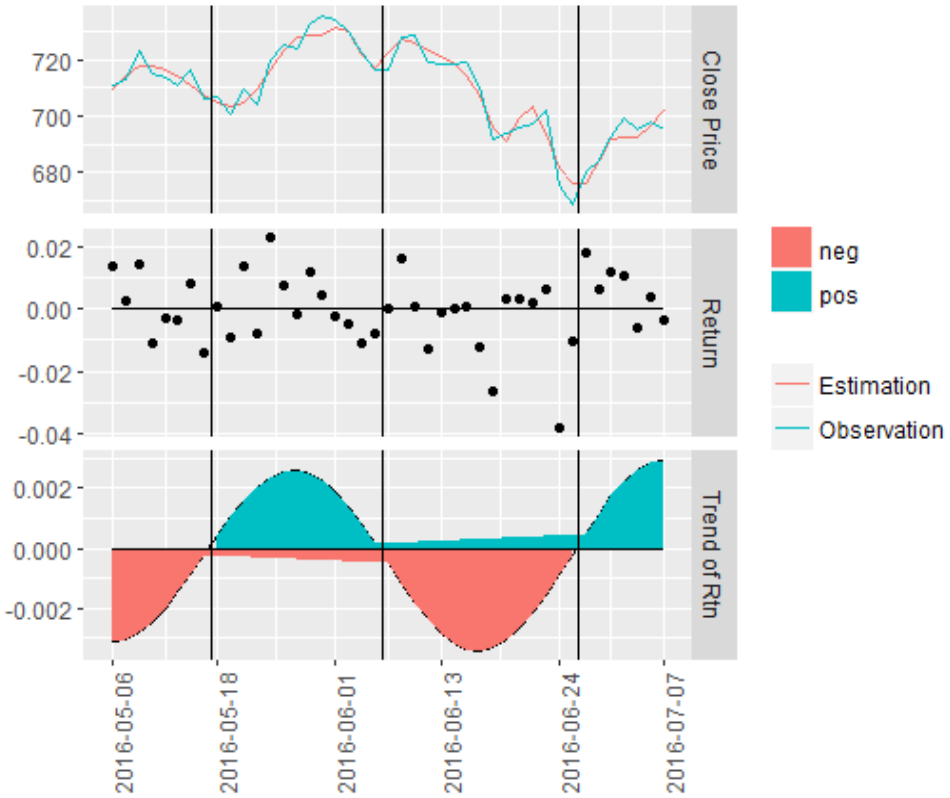


Figure 6.3. Example of 'GOOG'. The figure presents the observed daily prices, denoised prices, returns calculated based on observed daily prices, and the estimation of trend in daily return.

We now consider the returns under the nonparametric regression model (6.20) and present the classical decomposition procedure in Figure 6.2. Looking at the decomposition of returns time series, we notice an interesting phenomenon which can be illustrated by Figure 6.3. The top subfigure in Figure 6.3 presents the observed daily prices, denoised prices. The middle subfigure presents the daily returns computed using observed prices. The bottom subfigure presents the estimator of trend in the returns, which is the same curve as shown in the top-right of Figure 6.2. With the assistance of the three vertical lines in Figure 6.3, we conclude that when the trend of return is positive, the stock price tends to increase; and when the trend of return is negative, the stock price tends to go down. This conclusion is very intuitive but neither trivial or straightforward, because we can hardly gain critical information about

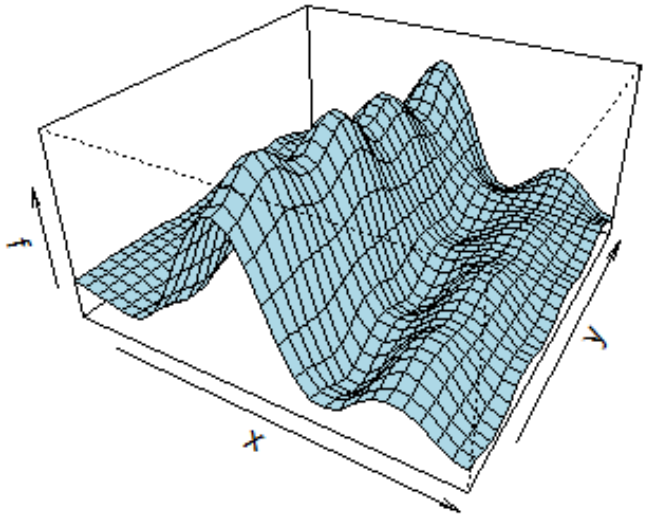


Figure 6.4. Example of 'XOM'. Time-varying probability density function estimation using historical daily prices from 2012-07-03 to 2016-07-27.

stock price by just glancing the dynamics in returns in the middle subfigure. The suggested estimator of the trend of return captures the characteristic of the underlying dynamics in the return and thus makes the relationship between price and return visualizable.

Figure 6.4 presents the estimator of the time-varying probability density function for stock 'XOM' based on the historical daily prices from 2012-07-03 to 2016-07-27. The x-axis corresponds to the value of return, the y-axis corresponds to the time t and z-axis corresponds to the value of the density. With the assistance of the 3-D figure, we can see the estimator captures the dynamics in the distribution of return over time.

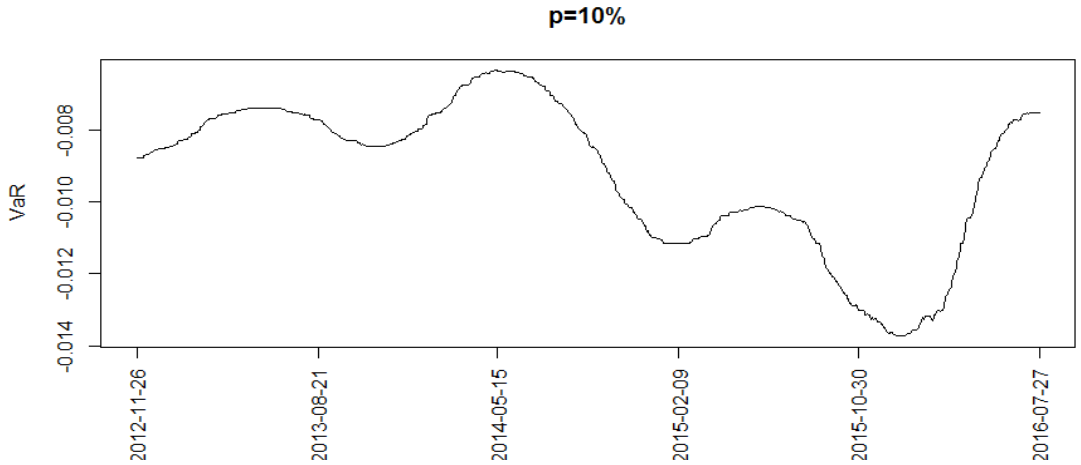


Figure 6.5. Example of 'XOM'. Time-varying VaR from 2012-11-26 to 2016-07-27.

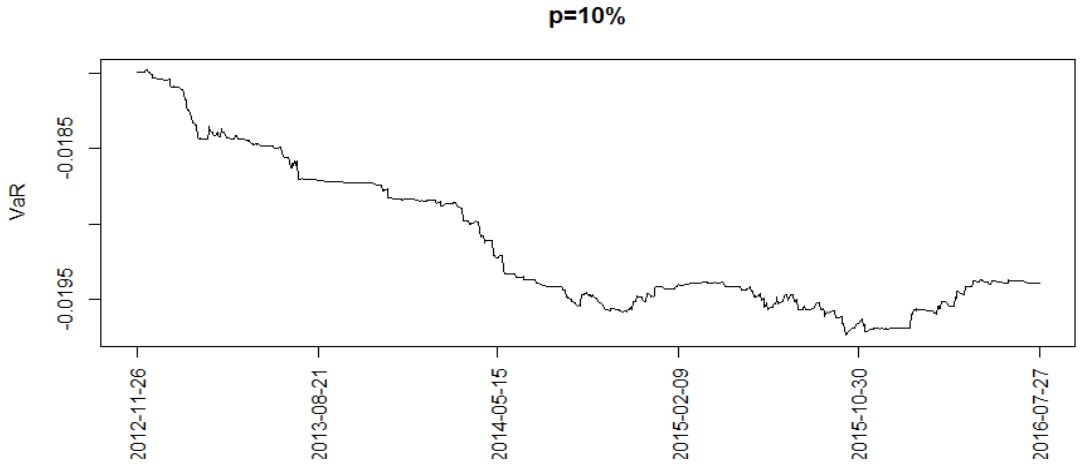


Figure 6.6. Example of 'GOOG'. Time-varying VaR from 2012-11-26 to 2016-07-27.

Figure 6.5 and Figure 6.6 present two examples of the time-varying 10% VaR estimations. The time-varying VaR curves for the two examples, 'XOM' and 'GOOG', have quite different pattern. Investment in 'XOM' turned out to be most risky around the end of the year 2015, during which the 'XOM' investors could lose about 1.4% daily with probability 10%. While investment in 'GOOG' was becoming more and more risky from the end of 2012 to the end of 2014, and after that the risk turned out to be more stable. During the whole year 2015

and the first half year of 2016, the 'GOOG' investors could lose about 1.95% daily with probability 10%.

CHAPTER 7

CONCLUSION

The dissertation presents a novel wavelet approach which solves two biggest challenges in a voxel-to-voxel-level fMRI study of neuroplasticity: large noise and high dimensionality. Neuroplasticity during adulthood is an important discovery in the 20th century, which has many significant implications for the development of learning and memory methods, and therapies for acquired brain injuries such as physical trauma and Stroke, and brain diseases such as Alzheimer's disease. This work aims to develop statistical methodology for investigation of the plasticity in the interhemispheric neural connectivity due to a motor task. The voxel-level neural connectivity is quantified as the cross-correlation between the two neural-activity related fluctuations originating from the corresponding interhemispheric pair of voxels. All interhemispheric voxel-level neural connectivity within the ROI is under investigation. This implies the analysis of large cross-covariance matrices and large cross-correlation matrices. In particular, given the fMRI data at hand, we deal with 1000×1000 inter-hemispheric cross-covariance matrices and large cross-correlation matrices.

It is proposed to perform a discrete wavelet transform on an observed signal from a voxel. Thanks to the frequency extraction property of wavelet decomposition, the frequency range of research interest are captured by the three coarser wavelet scales. However, a feasible denosing procedure is required to restore the underlying neural-activity related fluctuations, since the fMRI data at hand is noisy. Estimation of the noise level is critical in any denoising procedure. The violation of the assumption for the traditional standard deviation estimation of the noise in the empirical wavelet coefficients prevents us from using it, but on the other hand, motivates us to develop a novel noise model based on the fact that the two finest wavelet scales contains physiological fluctuations.

The estimation of the cross-covariance matrices is under a large-p-small-n setting, where regularization is needed. A thresholding estimator is proposed and it is proved to be rate

minimax over a class of sparse matrices. The plugged-in estimators are used for the cross-correlation matrices. An algorithm of the simultaneous analysis of all cross-correlations is proposed, which deals with the complications induced by the subexponential randomness in the plugged-in estimators.

The results obtained from the real data presents the dynamic in the number of neural pathways and localizes the plasticity. The results also indicate that the neuroplasticity varies widely even among healthy individuals. These observations are important for the understanding of human brains and for the developments in brain dysfunctionality therapies.

REFERENCES

Anderson, J. S., Druzgal, T. J., Froehlich, A., DuBray, M. B., Lange, N., Alexander, A. L., Abildskov, T., Nielsen, J. A., Cariello, A. N., Cooperrider, J. R., et al. (2010). Decreased interhemispheric functional connectivity in autism. *Cerebral cortex*, 21(5):1134–1146.

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometrythe methods. *Neuroimage*, 11(6):805–821.

Barrett, K. E., Barman, S. M., Boitano, S., and Brooks, H. (2009). Ganongs review of medical physiology. 23. *NY: McGraw-Hill Medical*.

Beall, E. B. and Lowe, M. J. (2007). Isolating physiologic noise sources with independently determined spatial measures. *Neuroimage*, 37(4):1286–1300.

Birn, R. M. (2012). The role of physiological noise in resting-state functional connectivity. *Neuroimage*, 62(2):864–870.

Birn, R. M., Murphy, K., Handwerker, D. A., and Bandettini, P. A. (2009). fmri in the presence of task-correlated breathing variations. *Neuroimage*, 47(3):1092–1104.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541.

Cai, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annual Review of Statistics and Its Application*, 4:423–446.

Cai, T. T., Zhou, H. H., et al. (2009). A data-driven block thresholding approach to wavelet estimation. *The Annals of Statistics*, 37(2):569–595.

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Chang, C. and Glover, G. H. (2009). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *Neuroimage*, 47(4):1448–1459.

Chen, L. M., Yang, P.-F., Wang, F., Mishra, A., Shi, Z., Wu, R., Wu, T.-L., Wilson, G. H., Ding, Z., and Gore, J. C. (2017). Biophysical and neural basis of resting state functional connectivity: evidence from non-human primates. *Magnetic Resonance Imaging*, 39:71–81.

Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of financial econometrics*, 3(2):227–255.

Chuang, K.-H. and Chen, J.-H. (2001). Impact: Image-based physiological artifacts estimation and correction technique for functional mri. *Magnetic resonance in medicine*, 46(2):344–353.

Efromovich, S. (1999a). *Nonparametric Curve Estimation: Methods, Theory and Applications*. Springer Science & Business Media.

Efromovich, S. (1999b). Quasi-linear wavelet estimation. *Journal of the American Statistical Association*, 94(445):189–204.

Efromovich, S. (2005). Estimation of the density of regression errors. *The Annals of Statistics*, 33(5):2194–2227.

Efromovich, S. and Smirnova, E. (2014). Statistical analysis of large cross-covariance and cross-correlation matrices produced by fmri images. *Journal of Biometrics & Biostatistics*, 5(2):1.

Efromovich, S. and Valdez-Jasso, Z. A. (2010). Aggregated wavelet estimation and its application to ultra-fast fmri. *Journal of Nonparametric Statistics*, 22(7):841–857.

Efromovich, S. and Wu, J. (2017). Dynamic nonparametric analysis of nonstationary portfolio returns and its application to var and forecasting. In *ARCH 2017.1 Proceedings*. 2017 Society of Actuaries.

Efromovich, S. and Wu, J. (2018a). Online supplementary materials for "wavelet analysis of big data contaminated by large noise in an fMRI study of neuroplasticity". https://static-content.springer.com/esm/art%3A10.1007%2Fs11009-018-9626-3/MediaObjects/11009_2018_9626_MOESM1_ESM.pdf.

Efromovich, S. and Wu, J. (2018b). Wavelet analysis of big data contaminated by large noise in an fMRI study of neuroplasticity. *Methodology and Computing in Applied Probability*. https://doi.org/10.1007/s11009-018-9626-3.

Fan, J., Han, F., and Liu, H. (2014). Challenges of big data analysis. *National science review*, 1(2):293–314.

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371.

He, H. and Liu, T. T. (2012). A geometric view of global signal confounds in resting-state functional mri. *Neuroimage*, 59(3):2339–2348.

Hernández, E. and Weiss, G. (1996). *A first course on wavelets*. CRC press.

Jo, H. J., Saad, Z. S., Simmons, W. K., Milbury, L. A., and Cox, R. W. (2010). Mapping sources of correlation in resting state fmri, with artifact detection and removal. *Neuroimage*, 52(2):571–582.

Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the royal statistical society: series B (statistical methodology)*, 59(2):319–351.

Kelly, C., Zuo, X.-N., Gotimer, K., Cox, C. L., Lynch, L., Brock, D., Imperati, D., Garavan, H., Rotrosen, J., Castellanos, F. X., et al. (2011). Reduced interhemispheric resting state functional connectivity in cocaine addiction. *Biological psychiatry*, 69(7):684–692.

Kosta, O. and Stepanova, N. (2015). Efficient density estimation and value at risk using fejér-type kernel functions. *Journal of Mathematical Finance*, 5(05):480.

Lazar, N. (2008). *The statistical analysis of functional MRI data*. Springer Science & Business Media.

Marusak, H. A., Calhoun, V. D., Brown, S., Crespo, L. M., Sala-Hamrick, K., Gotlib, I. H., and Thomason, M. E. (2017). Dynamic functional connectivity of neurocognitive networks in children. *Human brain mapping*, 38(1):97–108.

Mill, R. D., Bagic, A., Bostan, A., Schneider, W., and Cole, M. W. (2017). Empirical validation of directed functional connectivity. *NeuroImage*, 146:275–287.

Moerkerke, Y. R., Verdoolaege, G., and Welvaert, M. M. (2015). Package neurosim.

Mohlenkamp, M. J. and Pereyra, M. C. (2008). *Wavelets, their friends, and what they can do for you*, volume 8. European Mathematical Society.

Murphy, K., Birn, R. M., and Bandettini, P. A. (2013). Resting-state fmri confounds and cleanup. *Neuroimage*, 80:349–359.

Nason, G. (2008). *Wavelet Methods in Statistics with R*. Springer, New York.

Ogden, T. (1996). *Essential Wavelets for Statistical Applications and Data Analysis*. Springer Science & Business Media.

Perlbarg, V., Bellec, P., Anton, J.-L., Pélégrini-Issac, M., Doyon, J., and Benali, H. (2007). Corsica: correction of structured noise in fmri by automatic identification of ica components. *Magnetic resonance imaging*, 25(1):35–46.

Petrov, V. V. (1975). *Sums of independent random variables*. Springer, New York.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In *Handbook of economic forecasting*, volume 2, pages 328–383. Elsevier.

Tan, K. and Chu, M. (2012). Estimation of portfolio return and value at risk using a class of gaussian mixture distributions. *The International Journal of Business and Finance Research*, 6(1):97.

Taylor, J. W. (2007). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, 6(3):382–406.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats*. Springer Series in Statistics. Springer, New York.

Tung, K.-C., Uh, J., Mao, D., Xu, F., Xiao, G., and Lu, H. (2013). Alterations in resting functional connectivity due to recent motor task. *Neuroimage*, 78:316–324.

Valdez-Jasso, Z. A. (2010). *Aggregated wavelet estimation with applications*. PhD dissertation. The University of Texas at Dallas.

Vidakovic, B. (1999). *Statistical modeling by wavelets*. Wiley.

Vidakovic, B. and Mueller, P. (1994). Wavelets for kids. *Instituto de Estadística, Universidad de Duke*.

Weissenbacher, A., Kasess, C., Gerstl, F., Lanzenberger, R., Moser, E., and Windischberger, C. (2009). Correlations and anticorrelations in resting-state functional connectivity mri: a quantitative comparison of preprocessing strategies. *Neuroimage*, 47(4):1408–1416.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., and Rosseel, Y. (2011). neurosim: An r package for generating fmri data. *Journal of Statistical Software*, 44(10):1–18.

Welvaert, M. and Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for fmri data. *PloS one*, 8(11):e77089.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918.

Zuo, X.-N., Kelly, C., Di Martino, A., Mennes, M., Margulies, D. S., Bangaru, S., Grzadzinski, R., Evans, A. C., Zang, Y.-F., Castellanos, F. X., et al. (2010). Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. *Journal of Neuroscience*, 30(45):15034–15043.

BIOGRAPHICAL SKETCH

Jiayi Wu was born in Shanghai, China. She received her B.S. degree in Mathematics and Applied Mathematics from Shanghai Jiao Tong University in July 2013. In August 2013, she entered The University of Texas at Dallas to pursue her Ph.D. degree in Statistics. During her Ph.D. study, she worked as teaching assistant and teaching associate, and did summer internship in Yahoo! Inc.

CURRICULUM VITAE

Jiayi Wu

April, 2018

Contact Information:

Email: jxw133130@utdallas.edu

Education:

B.S., Mathematics and Applied Mathematics, Shanghai Jiao Tong University, China 2013

Teaching Assistant Experience:

Teaching Associate, The University of Texas at Dallas, August 2016 – May 2017, August 2017 – May 2018

Teaching Assistant, The University of Texas at Dallas, August 2013 – May 2016

Working Experience:

Intern, Yahoo! Inc., May 2017 – August 2017

Awards:

First Prize Poster Presentation in the Category of Ph.D. Level - Mathematical Statistics/Biostatistics, Conference of Texas Statisticians (COTS), Southern Methodist University, 2017

Languages:

Shanghainese, Mandarin Chinese, English