Critique of Linear Connectivity Methods in Functional Magnetic Resonance Imaging

Catherine E. Davey

B.E (Elec, Hons)

A thesis submitted in total fulfilment of the requirements

for the degree of Doctor of Philosophy

December, 2011

Department of Electrical and Electronic Engineering

The University of Melbourne, Australia

Abstract

Functional magnetic resonance imaging (fMRI) garners insight into brain functionality using a non-invasive imaging technique that capitalises on the coupling between neuronal activity and blood oxygen concentration. FMRI connectivity utilises fMRI data to ascertain how spatially remote neuronal regions cooperate to provide functionality in response to a task or when the subject is at 'rest'.

Connectivity methods have traditionally been categorised into data-driven and model-driven methods, also termed functional and effective connectivity, respectively. Data-driven methods conventionally extract whole-brain connectivity maps using contemporaneous correlation-based dependence measures, while model-driven methods examine hypothesised networks founded on a parametric model of regional relationships that typically incorporate contemporaneous and directed influence.

Traditionally the data-driven/model-driven dichotomy has been considered incontrovertible and the two classes regarded as mutually exclusive. However, indistinct categorisation of several methods has prompted a discernable shift in this perception, with recent recognition that the demarkation between modalities is not entirely clear, motivating proposals of a continuous scale from data-driven to model-driven. This thesis scrutinises the validity of the data-driven/model-driven dichotomy by clarifying methodological dependencies between linear techniques drawn from both modalities. In particular, Granger causality measures, which model instantaneous and directed influence, are demonstrated to be remappings of correlation-based metrics. This result highlights the centrality of partial correlation and partial variance in linear connectivity analyses and bridges the gap between data-driven and model-driven connectivity techniques. Consequently, assignment of a linear Gaussian connectivity method into the data-driven/model-driven dichotomy is ultimately a function of node selection and model configuration, rather than a distinction derived from methodological difference.

Linear Gaussian connectivity methods introduce data assumptions, conformance to which is necessary to guarantee integrity of results. A problematic assumption for fMRI is that of sample independence, which requires an absence of autocorrelation in voxel timeseries.

Temporal filtering, employed to address the high noise content of fMRI data, exacerbates the problem by filtering out components of the frequency spectra, necessarily introducing autocorrelation. In this thesis, temporal filtering is shown to induce artificial connectivity by reducing the effective degrees of freedom contributing to a sample correlation estimate. An analytic derivation for the distribution of sample correlation as a function of the complex filter frequency response is detailed, and a correction proposed to re-establish valid statistical inference testing of correlation estimates. The utility of the correction for mitigating artificial correlation is demonstrated both analytically and empirically.

Linear Gaussian models assume stationarity of timeseries, requiring the mean and variance to remain constant with time. FMRI data is known to contain non-stationary processes, typically modelled as time-varying noise variance. In this thesis, a model for non-stationary signal power is introduced to address variability in power across slice acquisition. The distribution of voxel intensity in the presence of non-stationary variance is analytically derived to be a generalised Student's-t distribution. Additionally, the impact of non-stationary signal power on correlation is deduced and the expected value of sample correlation shown to differ to the true correlation from which the signals were generated. A correction is proposed to restore signal stationarity and rectify the change in the expected value of sample correlation.

Declaration

This is to certify that

(i) the thesis comprises only my original work towards the PhD except where indicated in the Preface,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Preface

The material presented in this thesis, submitted for the degree of Doctor of Philosophy, is the culmination of postgraduate study undertaken at the University of Melbourne, Australia, in affiliation with the Howard Florey Institute, and supported by National Information Communication Technology Australia.

The following is a list of publications resulting from work presented in this thesis.

Journal Paper

- C.E. Davey, D.B. Grayden, M. Gavrilescu, G.F. Egan, L.A. Johnston, *The equivalence of linear Gaussian connectivity techniques*, in editorial review with Human Brain Mapping.
- C.E. Davey, D.B. Grayden, G.F. Egan, L.A. Johnston, *Filtering induces correlation in fMRI resting state data*, in preparation for submission to NeuroImage.
- C.E. Davey, D.B. Grayden, L.A. Johnston, *Non-stationary signal power induces correlation in resting state fMRI*, in preparation for submission to IEEE TMI.

Conference Abstracts

- C.E. Davey, D.B. Grayden, G.F. Egan, L. Johnston, *Filtering induces correlation in fMRI resting state data*, Human Brain Mapping, Quebec, 2011.
- C.E. Davey, D.B. Grayden, M. Gavrilescu, G.F. Egan, L.A. Johnston, *On the equivalence of Granger causality and correlation-based connectivity methods*, New Horizons in Human Brain Imaging, Hawaii, December 2010.
- C.E. Davey, D.B. Grayden, M. Gavrilescu, M. Farrell, G.F. Egan, L.A. Johston, *On the applicability of autoregressive models and Granger causality theory in fMRI analyses*, Human Brain Mapping, Melbourne 2008.

Acknowledgments

I would like to express my profound gratitude to my primary supervisor, Dr. Leigh Johnston, without whom this thesis would not exist. I am indebted to Leigh for taking a significant risk in supervising a student whose candidature was almost expired, and for reinvigorating my love of mathematics and engineering by exposing me to the world of linear Gaussian models. Warm thanks also go to Associate Professor David Grayden for providing support and guidance throughout my candidature and imparting a fraction of his intricate understanding of frequency spectra. Without David and Leigh, I would have returned to a corporate workforce rather than being in the rarefied and stimulating atmosphere of academia.

A very special thanks go to my family - my wonderful partners Annie and Rob, and my beautiful children Alex and Grier - for their unwavering support and continued belief in my ability to complete this thesis in the face of all evidence to the contrary. Lastly, thanks to my brother, Chris, for providing realism to my idealistic and unrealistic expectations of brain imaging, to my sister, Belinda, for always being interested, kind and generous, and my mother, Constance, for her immense emotional and financial support.

AIC	Akaike's information criterion
AR	autoregressive
ARX	autoregressive exogenous input
BIC	Bayesian information criterion
BOLD	blood oxygen level dependent
CDGC	conditional directed Granger causality
CGGC	conditional Geweke's Granger causality
CIGC	conditional instantaneous Granger causality
DCM	dynamic causal modelling
DGC	directed Granger causality
EPI	echo-planar imaging
FIR	finite impulse response
fMRI	functional MRI
FPE	final prediction error
FWER	family-wise error rate
GGC	Geweke's Granger causality
GLM	general linear model
GOF	goodness of fit
HQC	Hannan and Quinn's information criterion
HRF	haemodynamic response function
iid	independent and identically distributed
ICA	independent component analysis
IGC	instantaneous Granger causality
LFP	local field potential
LGM	linear Gaussian model
LISREL	Linear Structural Relations
LMC	left hemisphere primary motor cortex
LSE	least squares estimation
MDL	minimum description length
MLE	maximum likelihood estimation
MRI	magnetic resonance imaging
PACF	partial autocorrelation function
PCA	principal component analysis
PET	positron emission tomography
PDGC	partial directed Granger causality
PET	positron emission tomography
RF	radio-frequency
RMC	right hemisphere primary motor cortex
ROI	region of interest
RSS	residual sum of squares

SEM	structural equation modelling
sICA	spatial ICA
sMRI	structural MRI
SNR	signal to noise ratio
SVD	singular value decomposition
tICA	temporal ICA
TR	repetition time
tSNR	temporal SNR
VAR	vector autoregressive

Contents

List of Figures

Chapter 1

Introduction

Magnetic resonance imaging (MRI) is a non-invasive imaging technique that can be utilised to gain insight into brain structure and function, aiding understanding of the brain and providing tools for diagnosis of injury and disease. Structural MRI (sMRI) obtains single, high resolution images allowing for detailed discernment of brain anatomy, while functional MRI (fMRI), introduced subsequent to sMRI in the early 1990s, targets brain function by collecting a series of low resolution images across time. SMRI and fMRI harness MRI technologies differently to garner insights into distinct aspects of the brain, achieving sensitivity to different physiological processes. Briefly, sMRI differentiates between grey matter, white matter and cerebral spinal fluid by capitalising on the differing magnetic signal strengths of hydrogen nuclei in each environment. Blood oxygen level dependent (BOLD) fMRI takes advantage of the coupling between blood oxygen concentration and neural activity, generating a BOLD signal (Ogawa et al., 1990). BOLD fMRI will be referred to as fMRI for simplicity.

The objective of fMRI is to gain insight into brain function. The two primary paradigms from which to understand and examine brain function are termed functional segregation and functional integration. The defining principle of functional segregation is localised functionality; spatially distinct brain regions contain specialised neurons responsible for providing specific brain functionality, and hence a region responds only to inputs associated with its function. This paradigm is supported by examples in which a person suffering from a brain lesion exhibits the loss of the region's associated function (Goltz, 1881). However, functional segregation was found to be insufficient to explain functional organisation since it discounted interactions between brain regions (Brett et al., 2003).

Functional integration denotes the principle of activation distributed across a network of neuronal populations cooperating to provide functionality (Friston et al., 1997). Brain function is provided by a network of neurons and the interactions between them, so that a single region can participate in multiple functions with context dependent specialisation. Brett et al. (2003) argue that the functional role of a region is defined by its connections, rendering functional specialisation relevant only in the context of functional integration. Hence, it becomes necessary to identify which networks are associated with particular cognitive or perception tasks and how regions are coupled in the context of each task (Rogers et al., 2007). Techniques that have been proposed to consider such questions using BOLD data are known as fMRI connectivity methods, and are the subject of this thesis.

The concept of connectivity was proposed to quantitatively measure communication between neuronal regions under specific experimental conditions. Connectivity between brain regions can be determined in response to a task or when the subject is at 'rest', termed activation connectivity and resting state connectivity, respectively. Resting state fMRI connectivity emerged in response to networks being repeatedly identified when the subject was instructed to relax and think of nothing in particular (Biswal et al., 1995; Greicius et al., 2003). This thesis primarily focuses on resting state connectivity, and hence discussion will concentrate

on resting state methodology.

A resting state fMRI connectivity experiment typically proceeds by acquisition of a series of brain image volumes, typically with a sampling rate of $1.3 - 3.3$ seconds (Tong et al., 2011). The voxels in the resulting image volumes usually contain cubes of tissue about $1 - 5\text{mm}^3$ in length on each side (Amaro and Barker, 2006). Following data acquisition, image volumes are typically preprocessed to address issues such as head movement, noise, and acquisition timing, after which a connectivity analysis technique is applied to the resulting dataset to identify functional networks in the brain.

Connectivity methods have traditionally been categorised into data-driven methods and model-driven methods, also termed functional and effective connectivity, respectively. Data-driven connectivity techniques most commonly extract whole-brain correlation or covariance maps, making no *a priori* assumptions regarding connectivity structure, and lacking a notion of directional influence (McKeown and Sejinowski, 1998). Conversely, model-driven methods employ some form of *a priori* knowledge to examine hypothesised networks in the brain via a parametric model that typically incorporates directed influence (Friston et al., 1993). Parameters of the model are estimated from the data and connectivity estimates derived from model parameters.

This chapter provides a review of fMRI connectivity methods, and is organised as follows. Section 1.1 introduces data-driven methods prevalent in the fMRI connectivity literature, while Section 1.2 reviews principal model-driven methods. The methods are reviewed with a focus on resting state connectivity, since it is the primary focus of this thesis. Finally, an overview of the contributions of this thesis is provided in Section 1.3.

1.1 Data-Driven Connectivity

A common feature among many data-driven techniques is the use of a correlation-based measure to extract a whole-brain map. Indeed, a widely used definition of data-driven connectivity asserts that data-driven connectivity evaluates temporal correlations between spatially remote brain regions (Friston et al., 1993). Techniques within the data-driven category are distinguished by the exact form of the correlation-based measure, e.g. univariate versus multivariate, and the mathematical method used to infer connectivity relationships from the whole-brain map.

The absence of *a priori* assumptions regarding network structure renders data-driven methods robust to violation of these assumptions. Less prior knowledge is relied upon for region selection, and the connectivity analysis is exhaustive, searching through all brain voxels (Rogers et al., 2007). However, this also renders data-driven methods both spatially and computationally expensive.

Use of correlation-based measures imposes certain assumptions on the data, such as sample (i.e. serial) independence that requires an absence of signal autocorrelation, stationarity, constant mean and variance, and linearity; these assumptions limit dependence between voxels to be linear (Hamilton, 1994). However, a lack of *a priori* assumptions regarding influence type, e.g. instantaneous or directed and network structure, is frequently cited as restricting correlation-based measures to instantaneous influence (Friston et al., 1993; Sato et al., 2006b; Rogers et al., 2007).

To date, numerous data-driven techniques have been proposed, but there is no consensus on which methods are the most effective at capturing connectivity (Rogers et al., 2007). The principal data-driven methods are now reviewed.

1.1.1 Seed-Voxel Correlation Maps

Correlation maps were introduced to neuroimaging by Horwitz et al. (1992), where correlation refers to instantaneous correlation. Exhaustively estimating correlation between all voxel pairs requires calculating a correlation matrix proportional to the square of the number of voxels - unfeasible in the absence of a supercomputer. To resolve this issue, it has become standard practice to identify a seed voxel or timeseries via some alternative means, such as an accompanying activation analysis (Sarty, 2006, p.146). Correlation is then evaluated between the seed timeseries and all brain voxels. The resulting matrix grows linearly with the number of voxels and is equal in size to the image volumes so that it can be displayed as a parametric image with an intuitive visual interpretation (Rogers et al., 2007). The most significant drawback of this approach is that connectivity results unavoidably depend on the choice of seed.

Identifying the seed timeseries is critical and may proceed using one of several available options. Frequently, an fMRI activation analysis is employed in which the highest activating voxel in the map is identified (Beckmann et al., 2005). Alternatively, if the objective is to determine correlation with a particular region of interest, active voxels within the region can be reduced to a single timeseries by finding the mean or using a dimension reduction algorithm (Cordes et al., 2000). For activation data, another approach is to create a seed timeseries by convolving a block design of the task with the haemodynamic response function (HRF) (McKeown and Sejinowski, 1998).

After creating a seed timeseries, pairwise correlation is estimated between the seed and all brain voxels using Pearson's correlation coefficient (Rogers et al., 2007). Voxels significantly correlated with the seed are identified via significance testing. Since the variance of sample correlation depends on the unknown true correlation (Fisher, 1915), correlation estimates are typically transformed using a variance stabilising procedure. This can be achieved via Fisher's z-transformation (Fisher, 1915) or by transforming the correlation estimate to a Student's-t statistic (Rahman, 1968). Note that these analytic statistical inference tests

assume Gaussian signals and are not robust to distributional assumptions for small sample sizes (Kowalski, 1972). For short non-Gaussian timeseries, empirical resampling procedures are required. Using an appropriate statistical inference test, the probability of a seed-voxel correlation value being obtained is determined and all voxels with probability above some predefined threshold are identified as being functionally connected with the seed (McKeown and Sejinowski, 1998).

Cordes et al. (2000) employed seed-voxel correlation to identify functional connectivity, generating an empirical null distribution for sample correlation to avoid making assumptions about the temporal distribution of voxel intensity. To generate an empirical null distribution, the phase of each constituent temporal frequency of the seed timeseries was randomised and a correlation estimate between it and all brain voxels recorded. Since the power spectrum contains no phase information, this process did not change the power spectrum of the original timeseries. The empirical null distribution was then used to threshold the correlation map and identify regions functionally connected with the seed. Despite the authors concerns regarding non-Gaussianity of signals, it was unrecognised in the article that correlation is an optimal dependence measure for Gaussian signals only.

Generating an empirical null distribution addresses the problem of unknown voxel distribution, but it does not resolve the dependence of sample correlation on the true correlation. In calculating an empirical null distribution from voxel correlation estimates, many different true correlation values will be present. However, to generate an empirical null distribution it is crucial that the proportion of null cases is large - at least 90% according to Efron (2010, p.96). If this condition is not met, the false positive rate will not coincide with the stated confidence interval.

The standard seed-voxel correlation method can be extended to an iterative technique, in which the first correlation map (calculated using the seed-voxel correlation procedure) is used to identify a second region/voxel to then be used as the seed in a second iteration. This technique can expose functional connections that would otherwise have gone unnoticed (Rogers et al., 2007).

Correlation estimates tested using Fisher's z-transformation or a Student's t-test typically assume sample independence. Temporal filtering is a conventional preprocessing technique used in fMRI connectivity, which necessarily causes violation of the sample independence assumption by introducing autocorrelation to the signals. Friston et al. (1994) considered the problem of sample dependence in the context of an fMRI activation analysis, introducing the notion that the null distribution of sample correlation depends on the extent of autocorrelation in the constituent signals. The proposed correction to the effective degrees of freedom contributing to z-score estimates is valid only for uncorrelated signals, and is estimated using the spectral densities of the constituent signals, necessitating an estimate for each signal pair. Bullmore et al. (2000) addressed the issue of sample dependence in

correlation estimates by modelling autocorrelation using a first order autoregressive process and extracting a white residual using a generalised least squares approach (Bullmore et al., 1996b). The variance ratio of the residual to the original timeseries was then used to weight the timeseries length to obtain an estimate of the effective degrees of freedom. The efficacy of this approach depends heavily on the composition of autocorrelation in the timeseries.

The popularity of seed-voxel correlation is primarily due to the effectiveness garnered from a comparatively simply approach (Rogers et al., 2007). However, it has significant limitations. The use of pairwise correlation does not permit removing mediated influence or regressing out known sources of artificial correlation. Functional content is accepted to comprise less than 10% of the signal; it cannot be assumed that the remaining time-varying processes are uncorrelated with the task and they may skew the task-related correlation scores (McKeown and Sejinowski, 1998), rendering results vulnerable to correlations in the noise.

It is frequently argued that a significant limitation to correlation techniques is an inability to determine the intrinsic structure of the data. Lack of temporal precedence information in instantaneous correlation prohibits discovering directional or causal relationships. It is possible to obtain directional information using lagged correlation, although this remains insufficient to conclude causality since the influence may be mediated by a third, unmodelled voxel or exogenous input (Rykhlevskaia et al., 2005).

Seed-voxel correlation methods are sensitive to regionally specific haemodynamic responses that change the shape and/or temporal delay of a regional response. For example, disparate delays in haemodynamic response means that, despite true dependence in the signal of interest, seed-voxel correlation will be unable to identify the connection if it is lagged. This concern pertains to connectivity analysis of activation data only (Sun et al., 2004).

1.1.1.1 Corrections for multiple comparisons

Typically, an fMRI data set contains tens of thousands of voxels. Identifying voxels actively connected with the seed timeseries, therefore, requires tens of thousands of significance tests. Since the number of false positives identified is linearly proportional to the number of tests conducted, it becomes unacceptably high in whole-brain fMRI connectivity (Poldrack et al., 2011, p.116). For example, a significance test with a confidence interval of 95%, applied to $10,000$ voxels, gives an expected 500 false positives. Spatial smoothing exacerbates the problem because false positives may appear to belong to a cluster, providing further visual evidence of regional connectivity. Consequently, the probability of identifying one or more false positives - the family-wise error rate (FWER) - is very high. The Bonferroni correction (Bonferroni, 1935) was introduced to control the FWER in an environment in which a large number of identical tests are performed.

The Bonferroni correction states that if, N independent significance tests are performed

simultaneously with a desired statistical significance of α, then to maintain the FWER, a statistical significance of $\frac{\alpha}{N}$ should be used for each test (Abdi, 2007).

A crucial issue when applying the Bonferroni correction is a requirement of independence between the multiple tests. FMRI data contains spatial correlations among the voxels so the number of independent tests is generally smaller than the number of voxels. Consequently, the Bonferroni correction is usually too severe in an fMRI environment, which can result in true activations not being identified (Sarty, 2006, p.66).

False discovery rate (FDR) is another method designed to address the multiple comparisons problem, introduced by Benjamini and Hockberg (1995). They proposed to control the expected proportion of false positives among significant hypotheses only, rather than among all hypothesis tests (i.e. FWER), as for the Bonferroni correction. For example, an FDR of 5% implies that 5% of the statistically significant tests are expected to be false positives. FDR is less conservative than the Bonferroni correction, prioritising avoiding true negatives over avoiding false positives (Efron, 2010).

FDR was first introduced to neuroimaging by Genovese et al. (2002). Simulation data was used to compare use of the Bonferroni correction and FDR for addressing the multiple comparisons problem in the analysis of fMRI data containing thousands of voxels and, therefore, thousands of hypothesis tests. Using simulation data the authors showed that the Bonferroni correction controlled the inclusion of false positives but, in doing so, also excluded many active voxels, while FDR missed few active voxels but included many more false discoveries. Genovese et al. (2002) proposed the use of FDR for the additional information provided about the proportion of false positives in a connectivity map. Junning et al. (2008) used FDR to correct thresholds when identifying connectivity via graphical conditional dependence methods. However, they did not account for spatial dependence in their adjustment.

1.1.2 Multiple correlation

Wang and Xia (2007) introduced a multiple correlation technique in which correlation between a voxel and a set of seed voxels is calculated. Correlation between the seeds themselves does not inflate the dependence measures; only correlation between the seeds and the voxel contributes to the dependence estimate. The benefit of this technique is the ability to consider multiple seeds and visually assess them on the same image whilst accounting for correlations between the seeds themselves.

Wang and Xia (2009) introduced a technique they termed partial multiple correlation, in which multiple correlation between each voxel and the set of seed voxels was calculated with the effect of certain stimuli removed. They used simulation data to demonstrate that partial multiple correlation is superior in its ability to adjust for stimulus-locked events.

1.1.3 Partial Correlation

Seed-voxel correlation estimates pairwise connectivity without regard for potential influence from other voxels or exogenous inputs. Conversely, partial correlation measures the degree of linear dependence between two timeseries after removing the influence of a third signal or set of signals. It has obvious application to fMRI connectivity, in which mediated influence - i.e. two regions jointly influenced by a third - may be present, and voxel timeseries are known to be correlated with physiological noise processes (Biswal et al., 1995). Partial correlation enables mediated influence or influence from known noise processes to be regressed out prior to estimating correlation.

The effectiveness of partial correlation is limited by sample size, which restricts the number of estimable parameters and, therefore, the number of regions that can be partialled out. Consequently, it is not possible to calculate partial correlation between two voxels while accounting for potential mediated influence from all other brain voxels. This restriction requires a reduced set of voxel or regional timeseries to be selected *a priori*, thus requiring prior knowledge. Hence, partial correlation is not a purely data driven technique. However, its application to instantaneous influence (Salvador et al., 2005a) and membership in correlation-based methodologies depicts a data-driven method. Consequently, Marrelec et al. (2006) argued that partial correlation bridges the gap between data-driven and model-driven connectivity.

Marrelec et al. (2006) applied partial correlation to fMRI data by selecting regions related to the motor network, which they assumed would be activated by the motor task completed by their subjects. They further employed model assumptions for significance testing by supposing that the signal at each time point is multivariate Gaussian. Significance testing is then performed via Bayesian inference with an empirical null distribution. Their results find that the premotor cortices play a central role in the connectivity network. Note that only instantaneous linear dependence was considered.

Tana et al. (2008) employed partial correlation to resolve mediated influence in temporal lobe epilepsy. They noted that, in the presence of a linear Gaussian model assumption, Fisher's z-transform of partial correlation has an asymptotically normal distribution and, hence, analytic statistical inference testing was employed. Tana et al. (2008) asserted that partial correlation is unable to discern directed influence so that any causal relationships in the resulting network structure must be inferred.

Marrelec et al. (2005) introduced conditional correlation to fMRI connectivity to prevent indirect, or mediated, connections from being erroneously labelled as direct connections. They found that conditional correlation successfully classified indirect connections as insignificant. A linear Gaussian model was assumed, though the equivalence of partial correlation and conditional correlation in this context was not acknowledged.

1.1.4 Principal Component Analysis

Principal component analysis (PCA) is a method originally developed to reduce the dimensionality of a data set by extracting a set of orthogonal components, each with a variance measure describing the amount of signal variability for which it accounts (Pearson, 1901). It was first introduced to neuroimaging by Moeller and Strother (1991) to identify functional connectivity in positron emission tomography (PET) data. Singular value decomposition (SVD), applied in the fMRI context, is used to decompose timeseries datasets into three components (Friston and Büchel, 2007): eigentimeseries (patterns in time), eigenimages (patterns in space), and eigenvalues (the variance an eigentimeseries accounts for). The PCA and SVD methods are equivalent when applied to a covariance matrix of zero-meaned timeseries – as in fMRI connectivity – and can be referred to collectively as PCA (Moler, 2004, p.290). Note that the dataset is assumed to be jointly Gaussian distributed, a necessary assumption to guarantee orthogonality of components.

Employing PCA to identify functionally connected networks in fMRI data requires calculating the full cross-correlation matrix, and applying the PCA algorithm to this matrix to extract principal components. Since the storage and computation requirements for generating the full cross-correlation matrix are prohibitive, a more tractable method has fortunately been developed for the decomposition (Worsley et al., 2005). In the context of PET datasets, Friston et al. (1993) employed a recursive PCA algorithm when examining connectivity.

The principal components of the cross-correlation matrix are interpreted as connectivity networks. Each component can be thresholded to identify active voxels within each network, although there is no specific statistical interpretation to such a threshold. Worsley et al. (2005) compared PCA and correlation for identifying fMRI connectivity and found PCA to be more adept at recognising widely spread networks, whereas correlation was more effective at identifying localised networks.

Since fMRI typically generates a timeseries for thousands of brain voxels. whole-brain connectivity is a high-dimensional problem. Approximating the data with principal components can therefore be a very useful tool for dimensionality reduction. PCA is frequently used as a preprocessing step in fMRI connectivity analyses, utilising its dimension reducing properties prior to employing another connectivity tool to identify connectivity networks (Zhong et al., 2009; Zhou et al., 2009b). However, using the tool in such a manner can pose a problem in a low SNR environment such as fMRI. Given that the fMRI-related variations in the signal account for only a small percentage of the total signal changes, it cannot be assumed that connectivity networks will be the dominant components emerging from the analysis; therefore, there is a risk that these patterns will be removed (Thomas et al., 2002). Additionally, brain connectivity networks do not necessarily occur in orthogonal spatial patterns so a connectivity network may be split across multiple orthogonal components (Le and Hu, 1995) and it can be difficult to apply a neurophysiological interpretation to the eigenimages

(Rogers et al., 2007). More critically, because of interactions between brain systems, it may be physically impossible for the functional networks to coexist orthogonally (Friston et al., 1993).

1.1.5 Spatial Independent Component Analysis

Independent component analysis (ICA) was first introduced to fMRI by McKeown and Sejinowski (1998) to detect task activation. The method identifies maximally independent components, and the choice to maximise independence temporally or spatially creates two versions; temporal ICA (tICA) and spatial ICA (sICA), respectively. SICA is predominantly used in fMRI, primarily because the spatial dimension is significantly larger that the temporal dimension (Calhoun et al., 2001). However, there has been at least one instance of tICA being applied to fMRI data (Biswal and Ulmer, 1999). Given the predominance of sICA in fMRI connectivity it will be the focus of this discussion.

SICA examines the whole dataset to identify maximally independent, spatial component maps with associated time courses for each map. Each voxel is allocated a weighting for each component to indicate the extent to which the voxel timeseries is modulated by the component's timeseries. A fundamental premise of sICA is that errors in the data are Gaussian, and the non-normal trends in the data contain the information of interest. Consequently, sICA assumes that all components have a non-Gaussian distribution (Hyvärinen and Oja, 2000). This differs to the PCA assumption of Gaussian components. SICA algorithms exploit the non-Gaussianity assumption by employing a measure of non-Gaussianity to extract components.

McKeown and Sejinowski (1998) justified the use of spatially independent components to detect task activation by asserting that regions activated by the task will be completely unrelated to regions activated by artifacts such as head movement; the processes are independent and, therefore, the spatial realisations will be statistically independent. Londei et al. (2006) motivated the non-Gaussianity assumption by arguing that the spatial components of interest will be localised in space and, therefore, non-Gaussian.

PCA (Section 1.1.4) considers only second order relationships between voxels since orthgonality of components is assumed in the context of jointly Gaussian timeseries. The sICA algorithm includes higher order connectivity in the results so that non-linear connectivity will be detected (Yang and Rajapakse, 2004).

The sICA method requires each component map and its corresponding timeseries to be determined; since both need to be estimated from the dataset without prior information, this is known as a blind source separation problem. A benefit of the PCA algorithm is that orthogonal components can be ordered by the proportion of variance they account for using the eigenvalues. SICA does not have this property because of being a blind source

separation problem - it is possible to double a component map and halve the magnitude of it's time course with no net change. Consequently, it is not possible to determine the relative significance of each spatial component (Hyvärinen and Oja, 2000), though significance of voxels *within* component maps can be determined. McKeown and Sejinowski (1998) addressed this problem by determining a component's contribution to be the root mean square of the timeseries generated solely by the component. By similar reasoning, the sign factor of the component maps can not be determined. Furthermore, resulting component timeseries are not necessarily orthogonal so that the sum of variances explained by individual components will not necessarily equal the variance of the original dataset (McKeown and Sejinowski, 1998). Hyvärinen and Oja (2000) managed this problem by assuming that the sum of component variances was one.

McKeown and Sejinowski (1998) consistently found during activation analysis that, of the component maps generated by sICA, only one was found to have a timeseries highly correlated with the experimental design. This component map was then converted to a map of z-values and thresholded to identify active voxels within each map. To adopt this procedure for connectivity analyses, Yang and Rajapakse (2004, p.28) modified the definition of data-driven connectivity to "brain systems within which the different voxels or regions are activated by the same dependency of temporal variation in the complete statistical sense". They then employed the sICA technique but declared the active voxels within a component map to be connected rather than active. Yang and Rajapakse (2004) utilised only the independent component maps with the highest kurtosis scores, representing the greatest departure from normality. They considered these maps to be the most likely to have neural origins. They then considered active voxels within these maps to be connected. A similar procedure was followed by van de Ven et al. (2004) except the resulting maps were not tested for normality. Ma et al. (2007) compared correlation and sICA and found that sICA was effective at removing structured data, for which the performance of correlation was significantly worse.

1.1.6 Mutual Information

The entropy of a random variable is a measure of the uncertainty associated with the variable. Mutual information is an entropy based measure of dependence between two variables, but is more generalised than correlation because it incorporates both linear and nonlinear dependence. Mutual information between two variates depicts the extent of the information shared between them. Consequently, mutual information is zero only when two variables are independent.

Mutual information has an obvious application to fMRI connectivity because the measure is able to identify nonlinear dependence between pairs of stochastic variables. However, several issues need to be resolved first, including application of the measure to timeseries

and, crucially, estimation of the joint and marginal densities of timeseries.

Mutual information between two Gaussian processes was shown to be a function of coherence (Granger and Hatanaka, 1964). This equivalence was exploited by Salvador et al. (2007) to introduce a technique in which mutual information is approximated by an expression of coherence derived in the frequency domain. The expression calculates mutual information between clusters of brain regions, where each cluster may contain an arbitrary number of regions. The timeseries in a cluster are assumed stationary and jointly normal. The marginal and joint densities are then expressed via the spectral and cross-spectral densities which, using the joint normality assumption, simplify to an easily estimable, analytic form. The resulting mutual information is unbounded so a normalisation is suggested to ensure a range of $[0, 1]$. Salvador et al. (2007) also include an extension that decomposes *conditional* mutual information into summable frequency expressions, where conditional mutual information is the mutual information between two random variables when a third is known and accounted for. Salvador et al. (2007) showed that a disadvantage to their proposed method was the existence of strong correlation between mutual information and cluster size.

Saito and Harashima (1981) applied mutual information to evaluate directed influence by utilising a measure of nonlinear causality proposed by Granger (1969). Hinrichs et al. (2006) introduced the concept to fMRI connectivity, applying the technique in such a way as to incorporate only the immediately previous time point, similar to an AR(1) model. To estimate the underlying probability distributions, a technique proposed by Moon et al. (1995) was utilised. Since no analytic expression for the distribution of the necessary mutual information measures exist, empirical null distributions were generated using resampling procedures. Hinrichs et al. (2006) evaluated the technique on simulated autoregression data with both linear and nonlinear coefficients. They found that, while correlation-based measures could identify only linear connectivity, their mutual information based measure successfully identified both linear and nonlinear connectivity and correctly identified the direction of influence.

1.2 Model-Driven Connectivity

Model-driven connectivity, also termed effective connectivity, refers to methods that assume fMRI data can be described by a specific data model detailing the types of connections found, e.g. linear or non-linear, and the noise and signal dynamics. Connectivity is interpreted from parameters of the model, which are estimated using the BOLD dataset. Sample size must be sufficiently large such that sample statistics utilised for parameter estimation are indicative of true (i.e. population) statistics. Most model driven methods employ prior knowledge to postulate a specific connectivity network, including constituent regions and the presence or absence of specific connections between regions. In this case, only parameters of the specified

connections are estimated from the data and much of the data may remain unused.

A key feature distinguishing model-driven methods from data-driven methods is the inclusion of directed influence in the proposed models. Estimating directed influence between a complete set of regions enables causality to be inferred (McIntosh and Gonzalez-Lima, 1994). This distinction motivates the definition of model-driven connectivity as "the influence one neural system exerts over another" (Friston, 1994, p.57).

Imposing assumptions regarding the data model and network structure enables model-driven techniques to draw stronger conclusions, including causality. However, the expense of stronger assumptions is that validity of results depends on correctness of the assumptions. Invalid model assumptions may erroneously identify connections, and omitting an active region from the hypothesised network may lead to spurious or absent connectivity (Waldorp et al., 2011). Importantly, many model-driven methods only allow for a hypothesised network to be endorsed or discredited, precluding an exhaustive search through the dataset and rendering the result dependent on choice of regions and modelled connections.

Use of a single multivariate data model imposes a limit on the number of input timeseries to ensure the system is not under-determined. Consequently, it is often desirable to represent multiple voxel timeseries within a region using a single timeseries. Regions can be specified via several different methods, such as employing prior anatomical knowledge or by defining a region of interest (ROI) around each active voxel in an accompanying activation analysis. Once all active voxels within a region have been identified they can be summarised by calculating the mean timeseries or employing a dimension reduction technique, such as PCA, to extract a single timeseries that explains the most variability within the region (see Section 1.1.4). This is justified because voxels within a region are assumed to have similar timecourses (Kriegeskorte et al., 2008).

1.2.1 Structural Equation Modelling

Structural equation modelling (SEM) has been in use since the 1920s but was not introduced to neuroimaging until 1994 when McIntosh et al. (1994a) applied it to PET data. SEM employs a linear regression model containing instantaneous samples to identify directed connectivity. The method does not search for connectivity but rather confirms whether or not the sample data supports a proposed connectivity model (McIntosh et al., 1994b).

On introducing SEM to fMRI connectivity, McIntosh and Gonzalez-Lima (1994, p.10) asserted that 'a path coefficient is the expected change in activity of one region given a unit change in the region influencing it'. However, they argued that there is no such general rule for physical interpretation of neural pathway connections because each region has different physical characteristics defining its neural response. Therefore, isolated coefficients have no absolute meaning, but rather meaning is extracted when comparing a *change* in coefficients

across tasks or groups.

To employ SEM, the fMRI dataset must adhere to certain assumptions. Most notably, timeseries are assumed stationary while samples must be serially independent and noise processes are typically required to conform to a multivariate Gaussian distribution. Sample size must be sufficient to avoid inaccurate or improper solutions, such as non-convergence of the optimisation procedure (Hoyle, 1995). Boomsma (1985) found that sample sizes of 100 or more were sufficient, though Tanaka (1987) argued that the appropriateness of sample size is governed by the number of unknown variables being estimated and the parameter estimation method employed. Benter and Chou (1987) recommended 15 samples per unknown and claimed that if the data are well behaved (e.g. no outliers) and conform to all assumptions it is sufficient to acquire a minimum of 5 samples per unknown parameter.

SEM requires a hypothesised connectivity network with constituent regions identified and every pairwise connection anticipated. SEM is a directional algorithm so a path must be included for each direction of expected influence. It is important to ensure that the proposed network is not under-determined. For example, it would not be possible to have only two regions with each region proposed as influencing the other since this introduces two unknown path coefficients to be calculated from a single sample covariance between the two regions.

The SEM model is over-determined by design so that, following system specification, the best-fit path coefficients are determined via an iterative estimation algorithm such as maximum likelihood estimation (MLE), typically employing a Gaussian noise assumption. Iteration terminates when parameters are found to suitably fit the data, as determined by a prespecified value of a goodness of fit (GOF) index, or when some predetermined maximum number of iterations is reached. A GOF index compares the sample data covariance matrix with the model covariance matrix in such a way as to give a scalar 'distance' measure (Hoyle, 1995). Parameter estimation methods employ GOF indices compatible with their distributional assumptions. Each index introduces different qualities to the SEM method, such as robustness of MLE to violations in a multivariate Gaussian assumption (Hoyle, 1995).

A well-known issue with iterative optimisation algorithms is the potential for returning a local, rather than global, minimum. SEM implementations, such as the Linear Structural Relations (LISREL) software package (Jreskog and Srbom, 1996), manage this risk by using a non-iterative least squares estimation to initialise starting values, ensuring that starting values are in the vicinity of the global minimum, and thereby reducing the chance of converging to local minima.

On completion of iterative optimisation the best-fit model is used for hypothesis testing to determine how accurately the model describes the data. The null hypothesis asserts that the sample and model covariance matrices are equal, which is tested using the GOF index, rendering the significance test dependent on the distribution of the GOF index. The scaled GOF index introduced by Hu et al. (1992) converges to a χ^2 variate for large sample sizes

so that the null hypothesis is often tested using a χ^2 GOF test (Hoyle, 1995). However, convergence requires the distributional assumptions and large sample assumptions to hold; use of the χ^2 test is known to be sensitive to violations of these assumptions. To address this sensitivity, Hu et al. (1992) introduced a scaled version of the test statistic that better approximates a χ^2 distribution when these assumptions are violated. Jreskog and Srbom (1996) introduced an exploratory mode, initiated when a best-fit model is rejected, that revises the model in such a way as to improve the GOF index.

The SEM assumption of sample independence is known to be violated by BOLD data (Zarahn et al., 1997). Hypothesis testing with the χ^2 GOF index scales the index by sample size to generate a test statistic that is asymptotically χ^2 distributed; this scaling is based on the assumption of independence. For data containing autocorrelation between the samples, the scale factor must be adjusted down. If this remains unaccounted for, the test statistic will be scaled too high, thus increasing the false positive rate (Horwitz, 2005). Bullmore et al. (2000) addressed this issue by estimating the number of independent observations in the principal task-related timeseries via a PCA analysis.

When McIntosh and Gonzalez-Lima (1994) introduced SEM to fMRI connectivity, they recognised the potential problem in restricting unknown variables to ensure an overdetermined system, worsened by the inclusion of both residual variances and regression coefficients as unknowns. McIntosh and Gonzalez-Lima (1994) proposed to fix the residual variances to allow for more unknown path coefficients and/or a better ratio of samples to unknown parameters. However, Gonzalez-Lima and McIntosh (1994) showed that the value assigned to the residual variance impacts the goodness of fit test, increasing the chance of error. Conversely, Goncalves et al. (2001) later found that the choice of fixed residual variance had little impact on connectivity results.

Marrelec et al. (2009) compared partial correlation with SEM. They evaluated the methods via simulation data generated from a large-scale, neurobiologically realistic network using a method proposed by Horwitz (2005). Neurons and interactions were modelled by randomly modifying connections during the stimulus trial periods and adding noise asynchronous to the trial periods to half of each region's neurons (Horwitz, 2005). Marrelec et al. (2009) found that partial correlation and SEM produced very similar results, with all directional links identified by SEM having an undirected counterpart in the partial correlation graph depicting connections between regions.

Bullmore et al. (2000) claimed that hypothesis testing a proposed SEM model using a GOF index is problematic because the χ^2 GOF index is asymptotic and fMRI timeseries length may be insufficient to employ the asymptotic result, with sample dependence potentially further reducing the effective degrees of freedom contributing to the index estimate. They instead proposed an exploratory procedure beginning with a null network (all path coefficients are 0) and adding paths according to a modification index. Since additional connections

never worsen the GOF, they include a penalty for complexity, implemented in the form of the Akaike's information criterion (AIC). To avoid an underdetermined system, they prohibit feedback loops, whereby a path is going in both directions between two regions. An output of the exploratory phase is a bootstrapped distribution for the fit index, which is then employed to determine whether the proposed model is within specified confidence intervals.

Nonlinear terms can be incorporated into SEM models by including nonlinear combinations of a single variable or nonlinear interactions between pairs of variables (Kenny and Judd, 1984). Büchel and Friston (1997b) used interaction terms generated by the product of two regional timeseries to evaluate connectivity in visual pathways using fMRI data. Correlation between an interaction timeseries and its constituents must first be removed by regressing the product of the timeseries onto the individual signals, extracting the orthogonal component of the interaction timeseries. The interaction term is then input to the SEM algorithm as a regular regional timeseries. However, because the maximum likelihood optimisation procedures require normal variates, a different estimation procedure must be employed, such as generalised least squares (Kenny and Judd, 1984).

Zhuang et al. (2005) employed an exploratory version of SEM that began by utilising an activation analysis to extract five ROIs and prohibiting improbable networks, leaving 16384 feasible networks. Each potential network was optimised to determine best-fit path coefficients and generate an associated GOF index. The network best explaining the data was identified by comparing the GOF indices from all networks. The authors found that the best-fit model was repeatable across subjects. This technique is obviously limited in scope as it is not scalable with number of regions; the exponential growth in possible networks with each new region will render the exhaustive analysis intractable very quickly.

Zhuang et al. (2008) applied SEM in a semi-exploratory manner by proposing six feasible connectivity networks for a motor task. The networks were created from a combination of prior anatomical knowledge and the results of an activation study. Each network contained three or four regions assumed known and a single unknown region. Brain voxel timeseries were substituted into the unknown region in succession and a GOF index calculated with an accompanying t-value to determine significance of the voxel's estimated path coefficient in the model. If the resulting SEM model had a sufficiently high GOF index and the voxel's path was deemed significant, then both were saved and the results displayed on a model index map. The model index map was colour coded to illustrate the model each active voxel belonged to, as well as the GOF measure for the model. This technique appears open to misinterpretation in the sense that two voxels on the model index map may both be colour coded red, offering a visual impression that the voxels are connected when in fact they were never assessed as part of a common network. Rather, the red colouring of both voxels indicates that when each voxel was separately substituted into a model it passed some fit criterion. Additionally, because each proposed model contained three or four regions of

interest, a voxel's colour cannot be easily interpreted to mean the voxel is connected with a particular region but rather that the voxel was found to be an element of a network.

Storkey et al. (2007) introduced a method that learns a SEM network structure by interpreting the SEM model as the equilibrium activity of a Gaussian Bayesian network. They introduced uniform priors over network structures and independent priors for connections (i.e. priors assuming voxel independence), and then employed empirical methods to generate the posterior distribution. Noise variance priors were found to have no impact on results and were hence excluded. Simulation results showed that the structure with the maximum posterior probability was similar to the true structure, but frequently included an incorrect edge. The conclusion was, therefore, that learning SEM structures is informative but should be used as a guide only.

A combined approach was introduced by de Marco et al. (2009), in which independent component analysis (ICA) was used to determine the model to hypothesise for a SEM analysis. The motivation behind using ICA as a precursor to SEM was to exhaustively search the entire dataset to identify functionally connected regions. However, since ICA precludes analytic methods to identify significant connections, SEM can be employed to complete this task. Marrelec et al. (2007) employed a similar strategy, but utilised partial correlation as a precursor to SEM instead of ICA. Since partial correlation cannot be used to exhaustively search the entire dataset for connections, they assumed region selection was completed *a priori* by some other means, which is not discussed. Partial correlation was chosen for its capacity to identify contemporaneously connected regions whilst resolving mediated influence. Marrelec et al. (2007) claimed that the additional partial correlation phase is invaluable because of the limited number of regions in a SEM analysis. They also proposed using partial correlation to validate the resulting SEM structure.

The use of partial correlation may assist in establishing which path coefficients to include and exclude in SEM, but it should be recognised that partial correlation is also limited in the number of possible regions in the network for the same reasons as SEM; namely, to ensure the model is overdetermined. However, partial correlation employed to examine instantaneous influence will be less compromised than SEM owing to its unidirectional connections; SEM is bidirectional and hence requires two path coefficients to discover the relationship between two regions, whilst partial correlation requires only one.

1.2.2 Autoregressive Modelling

autoregressive (AR) models, introduced by Yule (1927), are timeseries models with the distinct property of being driven by noise; a zero variance noise process generates a zero signal. At each time point, an auto (i.e. self) regressive signal is a weighted sum of past signal values plus a white, Gaussian noise component. The number of lagged samples included in the sum is dictated by the order, p, of the AR model, written AR(p). Model coefficients are

determined via least squares estimation (LSE) or MLE, equivalent for the linear Gaussian model (LGM) class.

Behaviour of an AR(p) process can be examined via spectral analysis, specifically by inspecting the denominator of its rational transfer function (Zhang et al., 2006, p.191). Poles are the roots of the denominator and govern the dominant frequencies in the signal. An AR(p) process will have p non-zero complex poles; if any has magnitude greater than one, the AR process will be unstable and signal magnitude will diverge to infinity over time with probability one. The magnitude of each pole determines the significance of its contribution and its location in the unit circle determines the frequency over which it dominates. Stable AR models are wide-sense stationary, implying that both the mean and covariance of the timeseries exist and are independent of time.

AR models can be generalised to vector autoregressive (VAR) models (Fuller, 1996, p.75), which are multivariate models in which linear dependence between signals is estimated after regressing out lagged influence from a vector of regions (Greene, 2008, p.693). The noise process is multivariate Gaussian with no lagged linear dependence between constituents, although they may be correlated contemporaneously.

Another variant of AR models is the autoregressive exogenous input (ARX) model (Jensen et al., 2007, p.540). The current signal value of a timeseries in an autoregressive exogenous input (ARX) model can depend on its own past as well as the past of a known, exogenous input. ARX models can be combined with VAR models to include a multivariate signal model with exogenous inputs.

Theoretically, an AR(p) model has zero coefficients for all lags longer than p. However, in practice this is rarely the case and thus application of AR models to fMRI connectivity requires establishing model order prior to parameter estimation. The most common class of technique employed to identify model order is founded on stipulating a compromise between minimising complexity and minimising error, with error designated by the residual sum of squares (RSS) (McQuarrie and Tsai, 1998, p.24). Complexity, in this case, is defined by model order because the number of coefficients increases linearly with p. Since there is a guaranteed reduction in RSS with increasing complexity, this class of methods penalise order, the exact details of which differ for the various order selection criteria and depend on the paradigm from which it was derived, such as information theory – AIC (Akaike, 1973) and Hannan and Quinn's information criterion (HQC) (Hannan and Quinn, 1979) – or Bayesian inference – Bayesian information criterion (BIC) (Schwarz, 1978). An order selection criterion that is not a member of this class is the partial autocorrelation function (PACF), which instead uses partial autocorrelation coefficient values to identify model order (Fuller, 1996, p.62). In theory, lags not incorporated in the AR model have partial autocorrelation coefficients equal to zero. Lag orders with partial autocorrelation values that are significantly different from zero are, therefore, considered part of the model (Bullmore et al., 1996a).

Initially AR models were introduced to fMRI analysis by Bullmore et al. (1996a), who established that fMRI activation data analysed using a general linear model (GLM) produced residuals that were not independent and Gaussian as required, but rather contained significant autocorrelation. The result was realised using a single timeseries produced by averaging voxels from the occipital cortex. Bullmore et al. (1996a) concluded that the residual timeseries was most effectively modelled as an AR(1) process.

Tagaris et al. (1997) employed AR models in an fMRI activation analysis and examined the residual for conformance to the assumptions of whiteness and Gaussianity. They employed the PACF to establish that 18.9% of voxels had significant autocorrelation remaining in the residuals, whereas 76.5% had white noise residuals (4.6% of voxel timeseries were not white but could not be modelled using an AR model).

Following the seminal paper from Bullmore et al. (1996a) establishing autocorrelation in fMRI activation residuals, AR models began to see increasing application in fMRI connectivity analyses. Temporal precedence information facilitates estimation of directional influences while the possibility for multivariate models addresses concerns regarding mediated influence. Harrison et al. (2003) applied VAR models to connectivity by interpreting regression coefficients as connection strengths. They introduced an additional bilinear term, generated by calculating the product of two voxel timeseries, or a voxel timeseries and an input, depending on the questions being explored, and orthogonalising the result with the constituent signals. The orthogonalised timeseries was incorporated in the model as an exogenous input and coefficients estimated via least squares.

Use of regression coefficients to designate connection strength by Harrison et al. (2003) appears open to misconception since relative signal magnitudes remain unaccounted for. Consider the case in which the predictor and target signals have a correlation coefficient of one. If the target signal has magnitude ten times greater than the predictor signal, the VAR coefficient will be 10, whilst if the target has magnitude ten times smaller than the predictor, the coefficient will be 0.1. Thus, coefficients are susceptible to relative signal size as well as strength of connectivity. Moreover, VAR models containing multiple lags utilise multiple coefficients to describe the relationship between two regions. In this case, Harrison et al. (2003) suggested using a coefficient table, which may be cumbersome to interpret.

Harrison et al. (2003) employed a Bayesian evidence technique introduced by Penny and Roberts (2002) to simultaneously determine model parameters and model order. They ascribed a Gaussian prior with zero mean to VAR coefficients. Model order was identified by setting coefficients values to the mean of the posterior distribution and assessing each coefficient individually to determine whether it is 95% confidence interval crossed zero. Significance of connectivity was established by ascertaining whether the zero vector of comparable size crossed the 95% confidence interval of the vector of coefficients describing directional influence. Harrison et al. (2003) identified ROIs via an activation analysis, and timeseries

within each ROI were reduced to a single dimension by extracting the principle eigentime-series. The authors found that using samples sizes less than 250 occasionally resulted in an incorrect order estimate. Stability of the resulting AR models was not assessed, and hence it is possible that the proposed network model is divergent.

A single AR model cannot be used to exhaustively search an fMRI dataset for connectivity since the ratio of timeseries length to number of brain voxels is very small. Consequently, not all voxels can be simultaneously incorporated into a connectivity analysis so that the problem of spurious causality is not entirely mitigated. To address this limitation, prior knowledge can be used to select regions of interest or, to circumvent the need for prior knowledge, a sparse VAR model can be utilised in which the autoregression coefficient matrix consists primarily of zeros. Valdes-Sosa et al. (2005) proposed such a method for fMRI connectivity, in which regression coefficients were calculated using least squares, altered to include a penalty term. Valdes-Sosa et al. (2005) evaluated and compared several penalty terms in the context of fMRI connectivity. Each penalty operates slightly differently but they all utilise the concept of down-weighting small coefficients to zero to ensure a sparse coefficient matrix. Valdes-Sosa et al. (2005) simplified the problem by assuming an AR(1) model, motivated by results of the application of sparse VAR matrices using Bayesian techniques (Valdes-Sosa, 2004).

Kim et al. (2007) introduced a joint SEM-VAR modelling technique to model both instantaneous connectivity (via the SEM model) and directed connectivity (via the VAR model). The covariance matrix for the SEM-VAR model was determined via a maximum likelihood procedure and then the model was estimated and interpreted through standard SEM procedures. For simplicity, VAR(1) models were assumed. The necessity for the joint SEM-VAR technique by Kim et al. (2007) is not clear since VAR models have the capacity to model contemporaneous causality in conjunction with lagged causality, without the need for SEM.

Gates et al. (2009) introduced a unified SEM and VAR method in which they provided a mathematical transformation from distinct SEM and VAR models to a single model capable of identifying both contemporaneous and lagged influence. Initially expressing a signal as a weighted sum of current and past values, the model was then rearranged to express the signal as a sum of exclusively past values with the coefficient vector dependent on both contemporaneous and past influence. An automated model search was performed using LISREL's Lagrange multiplier test (Gates et al., 2009) in which the Lagrange multiplier uses a χ^2 test to determine if the change in GOF is significantly altered by the inclusion/exclusion of a path. Again, the imperative for the proposal by Gates et al. (2009) is not clear given VAR models have the capacity to identify significant instantaneous and directed influence. Furthermore, model search may not be optimised if a model of sufficient fit is found and thus the search terminated. Furthermore, the SEM coefficients used for directional influence become indicative of instantaneous influence in the joint model.

1.2.3 Granger Causality

The conventional axiom of causality is that cause must precede effect. Granger (1969) used this general notion of temporal precedence to formally define causality in the context of time series and detailed a specific application of the concepts to AR models. Granger argued that if a source timeseries causally influences a target then knowledge of the source's lagged samples reduces uncertainty of the target, where uncertainty is estimated using unexplained signal variance. Geweke (1982) extended Granger's concept of causality to AR models by decomposing linear dependence into measures of instantaneous and directed dependence, termed instantaneous Granger causality (IGC) and directed Granger causality (DGC), respectively. Total linear dependence between two time series was then defined as the sum of two directed influence measures plus an instantaneous influence measure.

AR models (Chapter 1.2.2) are integral to Geweke's Granger causality (GGC) measures and, therefore, application of the measures is valid only when assumptions of an AR model are met. AR models are themselves a subclass of the LGM class and thus inherit LGM assumptions in addition to those associated uniquely with the AR subclass. Therefore, stationarity of voxel timeseries is required and noise processes must be independent and identically distributed (iid) Gaussian. The model must be stable and model order must be established via an order selection criterion. Validity of GGC measures necessarily depends on conformance to these assumptions and correct identification of model order.

Significance testing GGC estimates in the context of Gaussian noise typically proceeds with scaling the measure by the degrees of freedom contributing to its estimation, producing a χ^2 variate (Geweke, 1982). The constituent instantaneous influence measure, IGC, and directed influence measures, DGC, can be evaluated for significance individually or the total linear dependence can be assessed.

GGC was used to identify directed neuronal interactions in neurophysiological data in the late 1990s (Bernasconi and Konig, 1999), though it was not applied to fMRI connectivity until Goebel et al. (2003) evaluated its potential as a model-driven fMRI connectivity technique. The main issue considered by Goebel et al. (2003) was the feasibility of Granger causality concepts in a low sampling rate and low signal to noise ratio (SNR) environment. This was assessed by simulating local field potential (LFP) signals, representative of the net neuronal activation contributed by all neurons within some given radius. The simulated LFP signals were convolved with a sample haemodynamic response function (HRF), modelled as a gamma function, and superimposed with white Gaussian noise (WGN). The result was down-sampled to reflect the low BOLD sampling rate. Goebel et al. (2003) utilised GGC measures to identify Granger causality and found that lagged influence embedded in LFP simulations was identified using GGC measures although, as expected, the influence emerged partly as directed dependence with a component of it erroneously appearing as instantaneous influence owing to the down-sampling. AR model order was set to the minimum of

several order selection criteria, including AIC, BIC and HQC, while significance testing was performed using resampling procedures.

Goebel et al. (2003) performed a whole-brain connectivity analysis with GGC by calculating a seed signal and estimating GGC measures between the seed and all brain voxels. Consequently, no *a priori* region selection was required - a point of departure from typical model-driven methods. While Goebel et al. (2003) assessed the feasibility of estimating directed influence from down sampled signals, they did not examine BOLD data for compliance with AR assumptions, and, therefore, did not fully assess the validity of GGC measures for fMRI connectivity. Furthermore, calculation of pairwise connectivity disqualifies the method from resolving mediated influence.

Roebroeck et al. (2005) extended the work of Goebel et al. (2003) by further examining the feasibility of GGC measures in fMRI connectivity. Roebroeck et al. (2005) demonstrated that true directed influence from source to target can appear as influence from the target to the source; this is a consequence of the low sampling rate of BOLD relative to the rate of neuronal activity. In lieu of this phenomenon, Roebroeck et al. (2005) proposed reducing the three GGC dependence measures to two: IGC and the difference between the two DGC measures, with sign indicating direction of influence. The most significant drawback is an inability to recognise feedback between two voxels since the difference between two significant directed measures may be insignificant unless they are sufficiently disparate in size. Roebroeck et al. (2005) evaluated the proposal using the simulation data described in Goebel et al. (2003). They found that as the simulated BOLD sampling rate decreased, the minimum model order required to capture influence decreased. This is intuitive since the interval between lags increases and some of the dynamics are lost in down-sampling. A less intuitive finding was an increase in estimated model order for down-sampled data relative to the model order used to generate the LFP data. An explanation proffered was that the AR model for the down sampled data was modelling the HRF dynamics.

Abler et al. (2005) applied the method proposed in Roebroeck et al. (2005) to experimental data acquired under conditions for which there exists consensus regarding connectivity structure. The data was acquired using a long repetition time (TR) since an objective of the paper was to assess sensitivity of GGC measures to long sample times. The resulting connectivity structure concurred with the literature and illustrated the efficacy of GGC measures in long TR environments.

Deshpande et al. (2010) delineated a "correlation-purged" technique in which a zero lagged term was added to the regression component to render the DGC measures "cleansed of zero-lag correlation" (Deshpande et al., 2010, p.1993). They noted that the relative down-sampling of BOLD signals to neuronal signals causes lagged influence to leak into instantaneous influence and further claimed that use of the correlation-purged GGC measures will rectify the leak. Need for the correlation-purged model is not clear since DGC is sensitive

only to directed influence; existence of instantaneous influence between signals does not modify the estimate of directed influence using DGC (Geweke, 1982).

A combination technique was introduced by Londei et al. (2006) involving sICA as a preprocessing step, proceeded by application of GGC measures to identify connectivity between sICA networks. SICA was used to extract independent spatial components and associated time courses; task-related temporal components were identified by correlating them with a reference function drawn from the experimental paradigm. GGC was then used to calculate linear dependence between the primary task-related component and all other independent components. The technique was extended in Londei et al. (2007) in which resampling mechanisms were used to determine the statistical significance of each connection. The technique was also applied to experimental data acquired using a task for which a consensus has been reached in the literature regarding its connectivity network. The sICA-GGC method generated results that concur with the consensus.

Use of the primary task-related component as the connectivity seed is unusual since connectivity is typically applied to the residuals of an activation analysis. Use of activation networks was motivated by the "conjecture that the complexity of the fMRI response is due to several other ongoing brain activities, apart from the one we are detecting, as well as other 'perturbations' of blood flow, such as oscillations of blood pressure and changes in cerebral blood volume" (Londei et al., 2007, p.45). GGC measures then estimate the connectivity between these networks. A potential problem with this combined approach is that sICA identifies networks by their non-Gaussianity which may be in conflict with the Gaussianity assumptions of GGC measures.

Sato et al. (2006a) presented a method to cater for time-varying connectivity using wavelet dynamic VAR models, with the objective of removing the stationarity and linearity assumptions necessary for conventional AR models. Each time-varying AR coefficient was represented by a truncated wavelet expansion using a proposed orthonormal basis function. This transforms the task of estimating AR coefficients to one of estimating wavelet coefficients for which an iterative least squares optimisation was employed. Smoothness of the resulting function was decided by the expected smoothness of the connectivity changes and regulated by resolution of the wavelet expansion (i.e. the point of truncation). Time-dependent connections between regions were determined by VAR regression coefficients at that time. The technique was evaluated using five dimensional VAR(1) simulation data depicting six experimental cycles, each with an identical time-varying connectivity network. Estimated AR coefficients output by the method were shown to closely follow the time-varying coefficients of the simulated data. When applied to experimental, data connectivity was assessed using bivariate models.

The dynamic VAR model has the considerable advantage of being able to track changes in the connectivity structure over time. A potential problem of the method is the number of

coefficients required for relatively short timeseries. Depending on the temporal resolution of the dynamic VAR coefficients, the degrees of freedom contributing to each estimate has the potential to become very small. Use of model orders greater than one in addition to dynamic coefficients will exacerbate the problem. This suggests that a compromise exists between restricting the model's capacity to account for autocorrelation and the benefits of removing stationarity and linearity assumptions. Furthermore, iid assumptions of the noise are still required for wavelet models (Bullmore et al., 2001).

1.2.3.1 Conditional measures of Granger Causality

GGC measures consider pairwise linear dependence and thus do not have the capacity to resolve mediated influence. To address this limitation, Geweke (1984) introduced conditional Geweke's Granger causality (CGGC) measures in which linear dependence between two regions is evaluated in the presence of a third region or vector of regions. The conditional measures introduced by Geweke include conditional instantaneous Granger causality (CIGC) and conditional directed Granger causality (CDGC), which evaluate conditional instantaneous and directed influence, respectively. As in the bivariate case, total conditional linear dependence is determined by summing the conditional instantaneous and directed components.

Chen et al. (2006) applied CGGC measures to resolve mediated influence when evaluating connectivities between neural field potentials. Geweke (1984) delineated spectral representations for the conditional measures, the derivation of which Chen et al. (2006) modified slightly. Unsurprisingly, they found the technique effective at eliminating mediated influence in simulation data generating from VAR models with Gaussian white noise. CGGC results on experimental data were found to concur with the neuroanatomy. Use of CGGC measures requires *a priori* region selection, prohibiting application to whole-brain analysis unless conditioning on exogenous inputs.

Gaglianese et al. (2011) employed the spectral representation of CGGC measures in an fMRI connectivity analysis designed to resolve mediated influence in the motor cortex. They significance tested CGGC estimates using resampling techniques in which AR coefficients were shuffled and resampled 1000 times. This method appears prone to underestimating the number of significant connections because the empirical null distribution is generated from models that potentially contain influence. That is, there is nothing to suggest that reshuffled AR coefficients will be absent of directed influence. If the empirical null distribution is comprised of both null and active connections, the true negative rate will be higher than specified by the confidence interval.

Guo et al. (2008) introduced a partial directed Granger causality (PDGC) measure to resolve erroneous influence derived from unknown, and hence unmodelled, exogenous inputs and/or latent variables. PDGC is very similar to CDGC, differing only in an additional step

to remove any remaining covariance between a target and the conditioned variables after partialling out the conditional signals. They contend that since the noise terms are assumed to contain any unmodelled exogenous processes, the additional step mitigates the potential for such processes to erroneously appear as connectivity. When applied to simulation data with contrived exogenous inputs, PDGC was better able to detect causal relationships than DGC.

1.2.4 Dynamic Causal Modelling

Dynamic causal modelling (DCM) is a model-driven connectivity method proposed by Friston et al. (2003b). It is an input-state-output model that incorporates a generative neuronal model to describe the relationship between neuronal activity and the BOLD signal. This is achieved via the Balloon-Windkessel model that describes the changes in deoxyhaemoglobin (i.e. the haemodynamic response) resulting from neuronal activity (Buxton et al., 1998). DCM differs to other model-driven connectivity methods in that, once the model is determined, output is a deterministic function of inputs. The model was developed to better understand how changes in input perturb the system, and how these perturbations affect connections amongst regions.

Time-varying connectivity is supported by depicting changes in neuronal state and network connectivity after perturbing the system, which is observed until the system enters a stable state. DCM assumes continuous signals so Friston et al. (2003b) defined connectivity in DCM as the rate of change in neuronal activity that one region causes in another. Higher connectivity values, therefore, mean that change in neuronal state in the affected region is faster in comparison to smaller values.

DCM models have two types of inputs. The first input group includes inputs established by the experimental conditions, represented in a format known as the design matrix (Friston et al., 2003b). These inputs affect neuronal states directly and, by design, each such input is assumed to affect only one region. The second group of inputs include those that modulate coupling between regions. These inputs change connectivity between regions, indirectly affecting a region's neuronal state. These inputs are also assumed to modulate only one connection. The output of a DCM model is the measurable BOLD signal.

The models contain five state variables. The first state variable represents neuronal activity in a particular region. The remaining state variables model the haemodynamic response, used to estimate the BOLD signal. This review will focus on the connectivity component of the model and will not describe haemodynamic modelling or estimation of the haemodynamic state variables.

DCM has no exploratory component so that hypothesised networks are either endorsed or rejected (Rogers et al., 2007). Parameter estimation of hypothesised networks proceeds us-

ing Bayesian estimators to calculate posterior densities for unknown parameters. Once the posterior probability densities of these parameters are known, the significance of the parameter can be determined. Prior densities of each coupling parameter matrix are assumed to follow Gaussian distributions with zero mean and variances set to a non-zero value that is identical for each unknown. Parameters are assumed to have independent prior distributions. To determine the posterior density, both the prior distribution and a likelihood are required; the likelihood is obtained via Gaussian error assumptions between the observed and estimated output. As further detailing of this process requires an understanding of the haemodynamic modelling component of DCM, it won't be expanded upon.

Lee et al. (2006) evaluated the performance of DCM on a large-scale simulation network with known connectivity structure. They used a comparatively sophisticated technique to generate the simulation data, modelled on that used by Husain et al. (2004). They found that, for the smaller networks containing three regions, DCM determined the correct connectivity structure. For larger networks containing five regions the connectivity network path weights were not correctly identified, but the correct network shape was identified using the model comparison techniques introduced by Penny et al. (2004a).

Penny et al. (2004b) compared the ability of SEM and DCM to identify connectivity. They found that the models were comparable in performance; both models identified the same connections when used to assess connectivity between visual regions during a visual attention task.

1.3 Overview of contributions

Linear fMRI connectivity analyses typically employ an instance of the LGM class. Data-driven techniques ostensibly lack a data model, although the correlation-based members implicitly employ a LGM (Härdle and Simar, 2007, p.132). The ubiquitous employment of correlation-based methods in the data-driven class render the majority of data-driven methods implicitly founded on a LGM. Each method extracts its connectivity estimate from different components of the model or statistic describing the model. Similarly, members of the model-driven category measuring linear dependence between signals utilise an instance of the LGM class, albeit explicitly.

Linear connectivity methods employing an instance of the LGM class impose several (often implicit) assumptions. Specifically, signals are assumed wide-sense stationary, and noise timeseries are required to be both white and Gaussian. Wide-sense stationarity requires the first and second moments of a timeseries to be invariant with time (Fox, 1947, p.372), while whiteness and Gaussianity require serially independent samples drawn from a Gaussian distribution (Fox, 1947, p.113). Consequently, noise samples are assumed iid. Non-conformance to these assumptions may invalidate results, and is particularly problematic if

a lack of robustness is demonstrated. Extensive use of the LGM class in fMRI connectivity suggests the need to examine fMRI data for conformance and/or robustness to assumption violation. Chapter 3 contains a detailed analysis of the experimental resting state fMRI dataset employed to validate theoretical results contained in this thesis. Conformance to LGM assumptions is examined, in addition to ascertaining the applicability of AR models to fMRI.

The primary contributions of this thesis are presented in Chapters 4-6. Each chapter contains predominantly self-contained works, including a description of the problem motivation and its consideration within fMRI connectivity literature. As such, there is some repetition between chapters. Theoretical notation and background is introduced in Chapter 2 and not reviewed in subsequent chapters.

Stationarity of voxel timeseries is important when applying linear Gaussian connectivity techniques. Granger and Newbold (1974) demonstrated that violation can cause spurious connectivity by altering the distribution of model parameters and, subsequently, dependence measures. Non-stationarity can be construed as a time-varying weighting of intensity samples, rendering samples with large weights more significant in the estimation of connectivity. Noise samples randomly coinciding at a time point with a large weight can skew results by making dependence artificially high. Conversely, if the signal of interest coincides in a sample with low weight, it may be obscured by noise.

Non-stationarity has been modelled in fMRI activation studies in an effort to resolve the loss of integrity associated with an invalid stationarity assumption (Woolrich et al., 2004; Lund et al., 2006; Diedrichsen and Shadmehr, 2005). It is typically modelled as an aspect of the noise process, attributed to noise sources such as subject movement during a task, physiological interference, or hardware imperfections (Wise et al., 1996; Weisskoff, 1996; Barry et al., 2010). For example, Diedrichsen and Shadmehr (2005) employed a linear model with Gaussian noise characterised by time-dependent variance. They used empirical evidence to motivate global application of time-varying weights so that all voxels were ascribed an identical time-varying noise process.

The fMRI connectivity literature contains little discussion regarding non-stationarity of resting state data. Furthermore, although the potential for variable signal intensity has been raised (Stejskal, 1965; Weisskoff, 1996; Kimberg, 2008), non-stationarity of signal power – weighting both signal of interest and noise – has not yet, to the best of our knowledge, been considered. In associating non-stationarity with only the noise process, an implicit assumption of constant signal power is made. However, there is potential for signal power to vary, attributable to inhomogeneous radio frequency (RF) amplification (Tanase et al., 2011), variability in scanner performance (Weisskoff, 1996), and signal dissemination amongst surrounding voxels (Stejskal, 1965). In Chapter 4 we examine stationarity properties of resting state fMRI timeseries via slice-dependent signal power. Slice variance

processes are ascertained as non-stationary, and slices are shown to have dissimilar distributions of signal variance. Since timeseries within a BOLD slice are typically acquired simultaneously, it is reasonable to assume uniform power properties across voxels within a slice. The non-stationary process models variation in signal power both spatially (across slices) and temporally. An analytic derivation for the distribution of voxel intensity, allowing for non-stationary variance, is detailed.

The relevance of non-stationarity for fMRI connectivity is in its impact on linear dependence estimates. In Chapter 4 analytic expressions are derived for correlation between non-stationary voxel timeseries, and the expected value of sample correlation is shown to be dependent on the non-stationary variance. A correction is proposed to restore integrity of correlation results, the utility of which is demonstrated both analytically and empirically. The correction is shown to rectify the increase in spurious connectivity introduced by non-stationarity.

BOLD signals are compromised by a myriad of noise sources; the signal of interest comprises less than 10% of signal fluctuation, being instead dominated by physiological noise (Kruger et al., 2001) and scanner noise (Bianciardi et al., 2009). BOLD signal composition strongly motivates the need for preprocessing methods to minimise artificial correlation emanating from noise processes such as synchronised cardiac and respiration (Birn et al., 2006). Temporal filtering is a preprocessing method that aims to remove noise at frequencies that do not coincide with the signal of interest. However, filter specification and implementation vary considerably. For example, Lowe et al. (1998) and Hampson et al. (2002) apply a low-pass finite impulse response (FIR) filter to remove frequencies above 0.08Hz, while Luca et al. (2006) employ a high-pass filter at 0.004Hz implemented using a straight line fitting approach and Liao et al. (2010) preprocess using a bandpass filter at $0.005 - 0.17$Hz by forcing the coefficients of unwanted frequencies to zero.

Chapter 5 examines the impact of temporal filtering on the distribution of sample correlation. Temporal filtering alters the effective degrees of freedom of a signal, which has consequences for the distribution of sample correlation. An analytic expression is derived for the distribution of correlation estimates as a function of the complex frequency response of the filter. Consider an example in which a white signal is ideally low-pass filtered. Partial loss of the frequency spectrum necessarily introduces autocorrelation, rendering temporal samples dependent on previous samples and decreasing the effective degrees of freedom. Crucially, the result is valid for arbitrary true correlation and all filter specifications.

FMRI connectivity heavily depends on statistical testing of dependence measures to identify significant connectivity and disregard weak connectivity. In Chapter 5, analytic distributions of correlation test variates, including Fisher's z-transform and Student's t-test, are described as functions of the complex frequency response of the filter. It is shown that the distributions of correlation test variates acquired from filtered data do not coincide with the

distributions predicated on the independence assumption. Rather, the reduction in effective degrees of freedom, instigated by temporal filtering, increases the sample variance of test variates and, as a consequence, confidence intervals calculated from uncorrected test variates are underestimated, thereby inflating the false positive rate and artificially inducing correlation. The corrected distributions are, therefore, vital to ascertain accurate probabilities of test variates and correctly identify significant connections. Furthermore, the introduced corrections are crucial to enable meaningful comparison of connectivity results acquired from differing temporal filtering procedures.

FMRI connectivity literature contains much discussion regarding the relative attributes of the data-driven and model-driven modalities, and the consequent advantages and disadvantages of each. Data-driven methods are frequently described as correlation-based and incapable of distinguishing directed influence (Friston et al., 1993; Sato et al., 2006b; Rogers et al., 2007); since correlation is a scalar measure of linear connectivity between two timeseries, it is "ambiguous with respect to underlying directed interactions" (Roebroeck et al., 2005, pg. 230). Moreover, employing a data-driven strategy purportedly disqualifies resolving mediated influence since it "does not provide any insight into how...correlations are mediated" (Büchel and Friston, 1997a, p.1). The oft quoted compensation for these deficiencies is an absence of assumptions given that the measures are "a model-free characterization of statistical association between time-series, without any regard for the underlying dependence structure between the components" (Goebel et al., 2003, p.1252). The apparent lack of distributional assumptions and the absence of conjecture regarding the type of dependencies contained in the data implies that results cannot be rendered invalid for infringements. Consequently, data-driven methods are described as drawing relatively weak conclusions from few assumptions (Rogers et al., 2007).

Model-driven connectivity is typically defined as the influence one neural system exerts over another (Friston, 1994; Marrelec et al., 2005; Goebel et al., 2003). This notion of causal influence is attributed to the *a priori* assumptions apropos the data model, specifically preselection of regions (Friston, 2009; Roebroeck et al., 2005; Marrelec et al., 2005), and assumptions regarding the structure of dependence between timeseries, e.g. instantaneous vs. directed. Data models are typically accompanied by additional distributional assumptions, primarily to simplify analytic parameter estimation (Hamilton, 1994, p.117). Model-driven methods are often surmised as having the ability to draw stronger conclusions at the expense of strong assumptions (Rogers et al., 2007).

Data-driven and model-driven connectivity classes have traditionally been considered mutually exclusive. That perception has started to alter with recognition that the demarcation between the two modalities is not entirely clear or, as Lindquist (2011) contends, there is a fluid scale between them. Indeed, use of a common model class - the LGM class - suggests the potential for commonalities. Marrelec et al. (2005) recognised the implicit use of a LGM in correlation-based methods, prompting an analytic assessment of the distinction between

instantaneous connectivity methods across both data-driven and model-driven modalities. They theoretically demonstrated that instantaneous influence measures are remappings of conditional correlation under a linear Gaussian assumption. Directed influence was notably absent from the analysis, as was recognition of equivalence between conditional correlation and partial correlation in the context of a LGM.

Traditionally, a primary driver of the data-driven/model-driven dichotomy has been the apparent enforced absence of directed influence from data-driven methods. However, the fMRI connectivity literature contains some recognition of the capacity for correlation-based measures to discern directed influence. For example, Kimberg (2008, p.4) states, "correlations at a temporal lag are potentially informative (if not decisively so) as to the causal direction of a functional connection". The implications of such claims remain unexamined in the fMRI connectivity literature.

The introduction of partial correlation to fMRI connectivity by Marrelec et al. (2006) further blurred the boundary between data-driven and model-driven modalities because, although partial correlation is correlation-based and hence data-driven, it has the capacity to resolve mediated influence as well as the model-driven attribute of requiring *a priori* region identification to ensure the system is overdetermined. Marrelec et al. (2007, p.1188) claimed that partial correlation can "bridge the gap that has endured between functional and effective connectivity". Granger causality measures can also be seen to bridge this gap since, although they are model-driven in the sense that a data model is assumed, the measures also appear data-driven in amenability to whole-brain analysis using many bivariate VAR models (Goebel et al., 2003).

Indistinct categorisation of both partial correlation and GGC into the data-driven/model-driven dichotomy and the unexplored potential for correlation-based measures to identify directed influence has raised questions regarding the relationship between linear connectivity methods across modalities. The theoretical basis of linear Gaussian connectivity methods for the analysis of fMRI data is examined in Chapter 6, and methodological dependencies between techniques are clarified. In particular, Granger causality connectivity procedures, which describe instantaneous and directed influence between sets of voxel time series, are demonstrated to be remappings of correlation-based metrics. Furthermore, the statistical inference tests applied to Granger causality measures are theoretically proven to be equivalent to inference tests applied to correlation-based metrics.

The equivalence of Granger causality and correlation-based methods, in both metric and test statistic, highlights the centrality of partial correlation and partial variance in linear connectivity analyses, bridges the gap between functional and effective connectivity techniques, and renders linear Granger causality a restatement of traditional data-driven methodologies in the context of brain connectivity studies. Consequently, assignment of a linear Gaussian connectivity method into the data-driven/model-driven dichotomy is ultimately a function

of node selection and model configuration, rather than a distinction derived from methodological difference. Use of many univariate models for whole-brain connectivity analysis, or a single multivariate model requiring *a priori* region selection and designed to resolve mediated influence for a subset of brain voxels or regions, informs the assignment of a linear connectivity method into the data-driven/model-driven dichotomy.

A summary of the organisation of this thesis follows.

- **Chapter 2** provides a brief introduction to BOLD data acquisition and preprocessing before introducing the theoretical background required for presentation of original work, including mathematical definitions of linear connectivity models and associated test variates.
- **Chapter 3** introduces the experimental dataset utilised throughout the thesis to corroborate theoretical results with experimental evidence. Attributes of the data necessary to employ linear Gaussian connectivity models are discussed, with a review of their consideration in the fMRI connectivity literature and an assessment of the experimental dataset in lieu of the stipulated attributes. Issues pertaining uniquely to AR models are introduced with an examination of the legitimacy of AR models for fMRI connectivity.
- **Chapter 4** assesses fMRI data for non-stationarity and contains an analytic evaluation of the impact of non-stationarity on connectivity. A correction is proposed that is shown to restore both signal stationarity and the integrity of correlation estimates. Additionally, the analytic distribution of voxel intensity for non-stationary timeseries is delineated.
- **Chapter 5** examines the impact of temporal filtering on the distribution of sample correlation by deriving an analytic expression for its distribution as a function of the complex frequency response of the temporal filter. Furthermore, the distributions of sample correlation test variates are derived for temporally filtered datasets. The necessity of the corrected distributions is demonstrated empirically using both simulation and experimental data.
- **Chapter 6** unifies linear connectivity methods by expressing model-driven linear Gaussian dependence measures as remappings of data-driven correlational measures. Consequently, assignment into the data-driven/model-driven dichotomy is shown to be a function of model specification and model configuration. Additionally, model-driven Granger causality test variates are shown to be asymptotically equivalent to correlational test variates, and thus Granger causality measures can be seen to offer no new insights into functional connectivity.
- **Chapter 7** provides a conclusion to the thesis, reviewing the major contributions and discussing opportunities for future work.

Chapter 2

Theoretical Background

2.1 Acquiring functional MRI datasets

2.1.1 Principles of functional MRI

Functional MRI (fMRI) is a non-invasive brain imaging technique offering relatively high spatial resolution and poor temporal resolution compared with other imaging modalities, such as electroencephalography and magnetoencephalography (Mulert and Lemieux, 2010, p.453). The technique utilises coupling between blood oxygen concentration and neural activity to generate a blood oxygen level dependent (BOLD) signal (Ogawa et al., 1990). Oxygen is transported throughout the body via haemoglobin in red blood cells. While oxygenated haemoglobin is diamagnetic, its deoxygenated counterpart is paramagnetic, so the two states carry different magnetic signatures (Moonen and Bandettini, 2000). When neurons are stimulated in response to a task, they consume more oxygen, causing nearby blood vessels to experience a comparative reduction in the concentration of oxygenated to deoxygenated haemoglobin (Sarty, 2006, p.79). This reduction persists for approximately one to five seconds after the increase in neural activity, after which arterial oxygenated blood increases to the area, thereby increasing local blood volume, blood flow and oxygenation concentration. The haemodynamics continue to increase for approximately four to five seconds, after which a steady decrease back to baseline is observed, often followed by an undershoot before the baseline is restored. Total oxygen flowing to the area actually exceeds that extracted, so the level of oxygenated blood is higher as a result of neural activity. This process is known as the haemodynamic response, which is described by the haemodynamic response function (HRF) (Poldrack et al., 2011, p.74).

The connection between neural activity and haemodynamics has been recognised since the late 19th century (Roy and Sherrington, 1890). However, it was not until the late twentieth century that sensitivity of tissue type to magnetic resonance was recognised (Lauterbur, 1973). The sensitivity is largely founded on the varying molecular structure of different tissue types and the differing concentrations of hydrogen (protons) (Woodward, 2001). The prevalence of hydrogen atoms in the haemodynamic process enables use of magnetic resonance for detecting haemodynamic change, acting as an indirect indicator of the level of neuronal activity within a region of the brain. A brief overview of the imaging process follows.

Magnetic resonance imaging is based on the properties of spin, the term used for nuclear spin angular momentum (Beutel et al., 2000). The spin of a charged particle, such as a proton, generates a magnetic field and the nucleus exhibits a magnetic moment that depends on the number of protons and neutrons present; a hydrogen nucleus contains a single proton (Hashemi et al., 2010, p.44). When nuclei are not in a magnetic field, there is no preferred orientation of the spins. However, if an external magnetic field is applied in direction z, the classical model of magnetic resonance states that the nuclei tend to align so their magnetic

moments are more oriented in this direction, and they precess, or rotate, around z (Atlas, 2009). In the presence of this external field, all hydrogen protons theoretically precess at the same frequency, which is linearly proportional to field strength and characterised by the Larmor equation (Bloch, 1946). A radio-frequency (RF) pulse applied at the Larmor frequency affects the spins in such a way as to cause the net magnetisation to precess at an angle away from z, where the angle (flip angle) is determined by the duration of the RF pulse (Bloch, 1946). For example, a 90 degree pulse rotates the net magnetisation vector away from z and onto the $x - y$ plane. Consequently, application of an RF pulse creates a transverse magnetisation component. Termination of the RF pulse causes the protons to exponentially recover alignment with the external magnetic field in the z plane (Atlas, 2009). The time constant describing the rate of recovery in the z direction is denoted T_1 (Bandettini, 2006). The time constant describing signal decay on the $x - y$ plane, resulting from dephasing of the precessing protons, is denoted T_2. T_2^* is typically used in place of T_2 to account for signal loss attributable to magnetic field inhomogeneities that cause the protons to spin at slightly different frequencies and weaken the signal (Russell et al., 2002). Inhomogeneities are areas in the magnetic field that do not exactly match the external field strength. For example, a region with slightly higher magnetic field strength will cause protons to precess at a slighter higher frequency, since precession is proportional to field strength. This discrepancy accelerates the dephasing of protons, causing the signal to decay faster (Westbrook, 2010, p.10). It is possible to apply a second RF pulse to rephase protons on the transverse plane, causing an increase in the signal and producing a 'spin' echo (Westbrook et al., 2005). Gradient echos are created using the linear field gradients to force spins in and out of phase. Both T_1 and T_2 (and hence T_2^*) depend on tissue type.

Images generated to enhance T_2 contrast are referred to as T_2 weighted images. Fat has a shorter T_2 time than water so that it generates very little signal in a T_2 weighted image, appearing relatively dark. Conversely, water has a high T_2 constant, producing a stronger signal and therefore appearing relatively bright in a T_2 weighted image. White matter contains fat and therefore appears darker than grey matter in a T_2 weighted image. The haemodynamic response particularly effects T_2 and T_2^* values via the oxygen concentration of the protons (Thulborn et al., 1982), and hence BOLD fMRI data is typically generated using T_2 contrast weighting methods (Atlas, 2009, p.1788).

Numerous magnetic resonance imaging (MRI) acquisition protocols have been developed to optimise configurable parameters according to specific imaging objectives. BOLD fMRI images can be acquired using echo-planar imaging (EPI) (Atlas, 2009). EPI acquires a series of comparably low spatial resolution gradient echo images in order to capture the temporal dynamics of neuronal activity. Each image volume is procured by progressive acquisition of two-dimensional planes, or slices (Bandettini, 2003). Frequency data of a two-dimensional slice is collectively referred to as k-space; k-space is acquired serially, according to one of numerous k-space sampling techniques, and image reconstruction performed via an inverse

Fourier transform of the k-space data. EPI sequences sample k-space in a single RF pulse. The time between RF pulses is referred to as the repetition time (TR) which, for EPI, determines delay between image volumes. A two-dimensional slice can be acquired in 60-100ms and typical sampling rates for image volumes are between 1.3 and 3.3 seconds (Tong et al., 2011).

Sampling rate is determined in part by voxel size. Optimal voxel size is impacted by the signal to noise ratio (SNR), a measure of signal strength compared to to background noise, and partial volume effects, which occur when a voxel spans multiple tissue types (Weibull et al., 2008). When voxels are too small, SNR is limited due to a reduced number of protons contributing to the signal, but as voxel size expands SNR becomes increasingly compromised by partial volume effects, since the risk of spanning multiple tissue types increases (Glover and Krueger, 2002), and physiological noise, which is proportional to signal strength. Glover and Krueger (2002) found the optimal voxel size to be $4mm^3$, while Amaro and Barker (2006) asserted that voxel size should be within the range of $1 - 5mm^3$, and Chen et al. (2003b) argued that optimal voxel size should match the size of activation.

Functional MRI (FMRI) has been revolutionary primarily because it is a non-invasive, *in vivo*, whole-brain imaging method with good spatial resolution (Russell et al., 2002) and temporal resolution sufficient to capture the haemodynamic response (Bandettini, 2003). There are, however, several limitations to the method, primarily deriving from its indirect measurement of neural activity (Johnson and Proctor, 2004). The sampling rate of fMRI is limited by slice acquisition time so that millisecond neuronal activity is being captured on a seconds timescale. Moreover, given the heterogeneous form of the HRF with respect to delay to peak onset, signal magnitude and duration, temporal precedence relationships between neural processes may be destroyed at the haemodynamic level. Bandettini (2009a) asserted that the delay introduced by the HRF varies by up to four seconds, further limiting the temporal resolution of BOLD. Another source of complexity for BOLD signals is the inconsistent vasculature across voxels. For example, large blood vessels generate signals of significantly greater magnitude than capillaries. The coupling between vasculature and the haemodynamic signal limits the questions that can be addressed with fMRI. Consequently, many assumptions are required for a meaningful interpretation of BOLD data (Bandettini, 2009a).

2.1.2 The BOLD Signal

Interpreting the BOLD signal

The fundamental objective of fMRI is to extract information regarding neuronal processes and their interactions. However, extracting neuron-specific information from BOLD data is complicated because the signals depend on factors other than neuronal activity, being composed of complex interactions between neuronal activity, metabolism, blood flow, and

blood volume (Bandettini, 2009b). Since these factors are regionally dependent, it is only meaningful to examine changes in the BOLD signal; BOLD is not considered a quantitative measure (Bandettini, 2006). Moreover, signal changes in BOLD appear to be determined by local field potentials, which provide a measure of the *sum* of local neuronal activity rather than neuronal spiking (Viswanatham and Freeman, 2007). Consequently, a change in the amplitude of the HRF may be attributable to differing durations of neuronal activity rather than an adjustment in the mean level of neuronal activity (Henson and Friston, 2007, p.179). Nonetheless, the BOLD signal is assumed approximately proportional to neuronal activity (Friston et al., 2003a, p.5), and the HRF can be considered a lowpass filter of neuronal activity (Kim et al., 1997).

An important issue for BOLD signal analysis is the signal to noise ratio; BOLD data is significantly compromised by noise, reducing SNR to the extent that the signal of interest comprises between 1% and 10% of signal fluctuation (Howseman et al., 1997), depending on experimental design, and is instead dominated by scanner noise (Bianciardi et al., 2009) and subject noise (Kruger et al., 2001). Scanner noise manifests as either systematic noise, such as baseline drift (Hashemi et al., 2010), or random errors, such as inhomogeneities in the magnetic field that can cause blurring and distortions of the image (Jezzard, 1999). Subject noise derives from between subject differences, describing the intrinsic variability in brain shape and functionality of different subjects, and within subject variability, such as head motion or physiological interference from respiration or cardiac cycles. Physiological noise is particularly problematic because it imposes an upper limit on the temporal SNR (Bodurka et al., 2007), and can induce artificial correlations (Lowe et al., 1998) by triggering widespread synchronised pulsing of blood vessels (Dagli et al., 1999). Head motion has been demonstrated to be correlated with stimulus events, increasing the likelihood that significant activation may derive from motion artifact rather than neural activity (Hajnal et al., 1994).

2.1.2.1 Preprocessing BOLD

BOLD signal composition strongly motivates the need for preprocessing methods to mitigate the adverse impact of noise and, in the context of fMRI connectivity analysis, to minimise artificial correlation emanating from noise, such as synchronised cardiac output and respiration (Birn et al., 2006).

Each two-dimensional slice of an EPI image volume is acquired separately so that the last slice is acquired almost a full TR later than the first. During the delay, changes may occur, such as physiological changes or subject movement (Kiebel et al., 2007). Since connectivity methods assume that all voxels were sampled simultaneously, the preprocessing phase typically commences with a slice timing correction algorithm. Slice timing correction uses interpolation to estimate the signal that would have been obtained if the slices were obtained simultaneously (Strother, 2006). Accuracy of the interpolation is inversely proportional to

TR.

After slice timing correction, a motion correction algorithm is typically employed to correct for a subject's head movement. Head movement can cause the signal from a brain region to shift in and out of a studied voxel (Bannister et al., 2007). This is most problematic for voxels on tissue boundaries that experience significant changes between high and low intensities (Costagli et al., 2009). Typically, motion correction algorithms assume that the head is a rigid body and, therefore, correct only for rotation and translation, requiring a rigid body registration with six degrees of freedom (Costagli et al., 2009). Such methods do not address non-linear effects of head movement, such as the pulsatile motion derived from cardiac and respiratory cycles (Lund et al., 2006), or spin history effects (Friston et al., 1996) in which, unless the TR is much longer than the T_1 of the tissue, the current voxel signal will depend on previous RF pulses and is therefore a function of movement history (Atlas, 2009, p.190).

The high noise environment of fMRI has led to the conventional use of spatial smoothing as a preprocessing step. Spatial smoothing involves blurring the image to reduce noise, with the assumption that noise is averaged out. The activation area must be larger than the blurring area so that the signal of interest is not decimated (Yue et al., 2010). This process, therefore, assumes that random noise has a higher spatial frequency than the signal of interest (Atlas, 2009). Spatial smoothing improves signal characteristics if the assumptions are met but reduces spatial resolution (Yue et al., 2010), vindicated by the pre-existing spatial correlation due to functional similarities of adjacent brain regions (Strother, 2006). However, spatial smoothing can increase partial volume effects and destroy signal of interest (Weibull et al., 2008). The standard spatial smoothing filter is a Gaussian kernel.

The last stage in the preprocessing pipeline is usually temporal filtering, aimed at removing noise concentrated at frequency bandwidths that do not coincide with the BOLD signal of interest. Physiological noise derived from respiratory and cardiac function is concentrated at comparatively high frequencies (> 0.1Hz) (Lowe et al., 1998; Cordes et al., 2001; Thomas et al., 2002), while scanner drift is localised to frequencies below 0.01Hz (Bianciardi et al., 2009). Functional correlations typically subtend frequencies between $0.01 - 0.1$Hz (Demirci et al., 2009; Biswal et al., 1995; Cordes et al., 2001; Salvador et al., 2005a). The most common temporal filter specification utilised in the fMRI connectivity literature is $0.01 -$ 0.08Hz bandpass filtering (Auer, 2008).

This thesis considers BOLD data acquired under resting state conditions so that it is not necessary to remove task activation from signals prior to a connectivity analysis. Connectivity analysis can be performed using a data-driven method or a model-driven method. Data-driven methods typically identify a seed voxel and estimate correlational dependence between the seed and all brain voxels. Model-driven methods employ *a priori* knowledge to identify a subset of brain voxels and regions and calculate dependence between this subset.

In both cases, the estimates of dependence must be inference tested to identify significant connections between the voxels or regions. The theoretical background for the primary data-driven and model-driven linear connectivity methods is now presented, including delineation of models and definitions of dependence measures, with an overview of the analytic inference methods available for significance testing each measure.

2.2 Data Models

2.2.1 Notation

Let $\mathbf{x} = [x_1, \cdots, x_n]'$ and $\mathbf{y} = [y_1, \cdots, y_m]'$ denote two random, real valued vectors, where $'$ denotes transpose. The covariance between $\mathbf{x}$ and $\mathbf{y}$ is

$$\sigma_{\mathbf{x},\mathbf{y}} = E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{y} - E\{\mathbf{y}\})'\} = \begin{bmatrix} \sigma_{x_1,y_1} & \cdots & \sigma_{x_1,y_m} \\ \vdots & \ddots & \vdots \\ \sigma_{x_n,y_1} & \cdots & \sigma_{x_n,y_m} \end{bmatrix}. \tag{2.1}$$

The variance of a scalar variate is defined as $\sigma^2_{x_i} = \sigma_{x_i,x_i}$.

Timeseries values are designated by lower case symbols with subscript t, and the uppercase counterpart denotes its frequency representation derived using the Fourier transform. For a timeseries x_t, denote the past p signal values of x at time t, from x_{t-1} to x_{t-p}, by $\mathbf{x}^{(p)}_{t-1} = [x_{t-1}, \cdots, x_{t-p}]$.

Temporally filtered signals and parameters derived from filtered signals are designated using the superscript (f). Variates standardised to zero mean and unit variance are indicated by superscript (s). The spectral counterparts to standardised signal variates, not standardised themselves, are also denoted by (s).

Where applicable, Greek symbols denote statistical measures while sample estimates are referred to by name, for example $\rho_{x,y}$ and $\mathrm{corr}\,(x_t, y_t)$ for correlation and sample correlation, respectively. Correlation refers to instantaneous (lag-0) correlation, unless a specific lag is prepended to the term. Estimates of model parameters are identified by a circumflex, $\hat{}$, prediction of a variate by a tilde, $\tilde{}$, approximate equivalence in distribution is denoted $\overset{a}{=}$, and $\overset{a}{\sim}$ indicates asymptotic equivalence with sample size. Finally, to aid readability, let $\sum_k \equiv \sum_{k=0}^{T-1}$.

2.2.2 The linear Gaussian model class and measures of dependence

The linear Gaussian model (LGM) class defines a linear relationship at each time instant, t, between a scalar, x_t, and a multivariate signal of dimension m and/or a vector of lagged

time series measurements, $\mathbf{z}_t$, hereafter referred to as the 'predictor'. The signal is assumed measured in additive Gaussian white noise, e_t, so that the model is given by

$$x_t = \mathbf{b}'\mathbf{z}_t + e_t, \tag{2.2}$$

where $\mathbf{b} = [b_1, \cdots, b_m]'$ is a vector of coefficients.

If the predictor is univariate (i.e. $m = 1$), the strength of instantaneous linear association between x_t and z_t can be determined using correlation, defined as (Whittaker, 2009)

$$\rho_{x,z} \triangleq \frac{\sigma_{x,z}}{\sigma_x \sigma_z}, \tag{2.3}$$

where σ_x and σ_z denote the standard deviation of x and z, respectively; recall that $\sigma_{x,z}$ denotes covariance between x and z.

For a multivariate predictor, correlation generalises to multiple correlation, with the coefficient of determination - i.e. multiple correlation squared - defined as (Hamilton, 1994),

$$\mathrm{R}^2_{x,\mathbf{z}} \triangleq \frac{\sigma_{\tilde{x},x}}{\sigma^2_x}, \tag{2.4}$$

where $\tilde{x}$ denotes the prediction of x so that, from Eq. (2.2), $\tilde{x} = \mathbf{b}'\mathbf{z}_t$.

Partial variance is a measure of the variance remaining in x_t after regressing out the influence of predictor $\mathbf{z}_t$ (Whittaker, 2009),

$$\sigma^2_{x|\mathbf{z}} \triangleq \sigma^2_{x-\tilde{x}} = \sigma^2_e. \tag{2.5}$$

Partial covariance is defined as the covariance between two stochastic signals after partialling out the influence of a common predictor. Consider a second process, $y_t = \mathbf{c}'\mathbf{z}_t + u_t$, where $\mathbf{c}$ denotes a vector of coefficients and u_t designates additive white Gaussian noise. If $\tilde{y}_t$ represents the prediction of y_t by $\mathbf{z}_t$, so that $\tilde{y}_t = \mathbf{c}'\mathbf{z}_t$, then (Whittaker, 2009)

$$\begin{aligned}\sigma_{x,y|\mathbf{z}} &\triangleq \sigma_{x-\tilde{x},y-\tilde{y}} \\ &= \sigma_{x-\mathbf{b}'\mathbf{z},y-\mathbf{c}'\mathbf{z}}.\end{aligned}$$

Partial correlation is the normalisation of partial covariance onto the range $[-1, 1]$,

$$\rho_{x,y|\mathbf{z}} \triangleq \frac{\sigma_{x,y|\mathbf{z}}}{\sigma_{x|\mathbf{z}}\sigma_{y|\mathbf{z}}}. \tag{2.6}$$

Zero partial correlation implies conditional independence between x_t and y_t, given knowledge of $\mathbf{z}_t$.

It should be noted that, for the LGM class, sample estimates of the model parameter

derived from maximum likelihood estimation (MLE) are equivalent to values acquired using least squares estimation (LSE) (Lupton, 1993) and, for the model described in Eq. (2.2), determined by (Whittaker, 2009)

$$\hat{\mathbf{b}} = (\mathrm{cov}\,(\mathbf{z}_t))^{-1}\mathrm{cov}\,(\mathbf{z}_t, x_t)\,. \tag{2.7}$$

For the case of univariate z_t, this leads to the MLE/LSE sample estimate of correlation (Whittaker, 2009),

$$\begin{aligned}\mathrm{corr}\,(z_t, x_t) &= \frac{\mathrm{cov}\,(z_t, x_t)}{\mathrm{std}\,(z_t)\,\mathrm{std}\,(x_t)} \\ &= \frac{\sum_{t=1}^{T}(z_t - \hat{\mu}_z)(x_t - \hat{\mu}_x)}{\sqrt{\sum_{t=1}^{T}(z_t - \hat{\mu}_z)^2 \sum_{t=1}^{T}(x_t - \hat{\mu}_x)^2}},\end{aligned} \tag{2.8}$$

where $\hat{\mu}_z$ and $\hat{\mu}_x$ denote the sample means of z_t and x_t, respectively.

Sample correlation can be equivalently expressed using standardised variates by a simple rearrangement of Eq. (2.8) (Whittaker, 2009),

$$\begin{aligned}\mathrm{corr}\,(z_t, x_t) &= \frac{1}{T}\sum_{t=1}^{T}\frac{(z_t - \hat{\mu}_z)}{\mathrm{std}\,(z_t)}\frac{(x_t - \hat{\mu}_x)}{\mathrm{std}\,(x_t)} \\ &= \frac{1}{T}\sum_{t=1}^{T} z_t^{(s)} x_t^{(s)},\end{aligned} \tag{2.9}$$

where $z_t^{(s)}$ and $x_t^{(s)}$ designate the standard normal counterparts of z_t and x_t, respectively, such that $z_t^{(s)} = \frac{z_t - \hat{\mu}_z}{\mathrm{std}(z_t)}$ and $x_t^{(s)} = \frac{x_t - \hat{\mu}_x}{\mathrm{std}(x_t)}$.

Correlation is used to refer to instantaneous, or lag-zero, correlation. To estimate time-lagged linear dependence between between signals, the cross-correlation function is used with lag specified by τ (Kumar, 2009),

$$\gamma_{z,x}(\tau) = \frac{1}{T}\sum_{t=1}^{T} z_t^* x_{t+\tau}, \tag{2.10}$$

where z_t^* indicates the complex conjugate of z_t. The Fourier transform of cross-correlation satisfies (Kumar, 2009)

$$\Gamma_{z,x}(k) = \frac{1}{T} Z_k^* X_k, \qquad k = 1, \cdots, T, \tag{2.11}$$

where Z_k and X_k denote the Fourier transforms of the vectors $\mathbf{z} = [z_1, \cdots, z_T]'$ and $\mathbf{x} = [x_1, \cdots, x_T]'$, evaluated at frequency index k, respectively. Consequently, application of the

inverse Fourier transform requires that

$$\gamma_{z,x}(\tau) = \frac{1}{T^2} \sum_{k=0}^{T-1} Z_k^* X_k e^{2\pi i \frac{k}{T} \tau}. \tag{2.12}$$

It is pertinent to note that the cross-correlation between standardised variates at lag zero is equivalent to correlation,

$$\rho_{z,x} \equiv \gamma_{z^{(s)}x^{(s)}}(0). \tag{2.13}$$

Finally, convolution between z_t and x_t is denoted by $z_t \star x_t$, and is defined as

$$z_t \star x_t \triangleq \frac{1}{t} \sum_{j=1}^{t} x_j z_{t-j}. \tag{2.14}$$

2.2.3 Statistical testing of correlation-based measures

Assuming Gaussian signals and independent samples, the variance of sample correlation is given by (Fisher, 1915)

$$\sigma^2_{\mathrm{corr}(z_t,x_t)} = \frac{(1-\rho^2_{z,x})^2}{T}. \tag{2.15}$$

Thus sample variance of correlation depends on the true correlation, rendering it difficult to significance test correlation estimates for which the true correlation is unknown. Fisher (1915) overcame this problem by proposing a variance stabilising transformation[1], known as Fisher's z-transformation,

$$\mathrm{r}\left(z_t, x_t\right) \triangleq \frac{1}{2} \ln \frac{1 + \mathrm{corr}\left(z_t, x_t\right)}{1 - \mathrm{corr}\left(z_t, x_t\right)}. \tag{2.16}$$

The transformation to $\mathrm{r}\left(\mathrm{corr}\left(z_t, x_t\right)\right)$ is advantageous as the variance of $\mathrm{r}\left(\mathrm{corr}\left(z_t, x_t\right)\right)$ is independent of true correlation and asymptotically normal with sample size. The significance of a correlation estimate can then be tested using (Fisher, 1915)

$$\mathrm{r}\left(\mathrm{corr}\left(z_t, x_t\right)\right) \sim \mathcal{N}\left(\mathrm{r}\left(\rho_{z,x}\right), \frac{1}{T-2}\right), \tag{2.17}$$

where $\mathrm{r}\left(\rho_{z,x}\right)$ is Fisher's z-transformation of the true correlation.

Under an assumption of zero correlation, an alternative method available for evaluating the significance of a correlation estimate is to transform $\mathrm{corr}\left(z_t, x_t\right)$ into a Student's t-variate

[1] In the case of sample data characterised by a variance dependent on the unknown true mean of the data, a variance stabilising transformation can be applied to render the variance independent of the mean.

(Kumar, 2009),

$$\mathrm{t}\left(z_t, x_t\right) = \mathrm{corr}\left(z_t, x_t\right) \sqrt{\frac{T-2}{1-\mathrm{corr}\left(z_t, x_t\right)^2}} \sim \boldsymbol{t}\left(T-2\right). \tag{2.18}$$

The analytic test variates detailed in Eqs. (2.17)-(2.18) assume that the constituent signals are drawn from a bivariate normal distribution with independent samples (Härdle and Simar, 2007, p.67). The distribution of sample correlation and its test variates are not robust to violation of the Gaussian assumption for small sample sizes (Kowalski, 1972).

The distribution of a sample estimate of partial correlation between x_t and y_t, after removing the influence of $\mathbf{z}_t$, Eq. (2.6), is given by (Fisher, 1924)

$$\mathrm{corr}\left(x_t, y_t|\mathbf{z}_t\right) \sim \mathcal{N}\left(\rho_{x_t,y_t|\mathbf{z}_t}, \frac{\left(1-\rho^2_{x_t,y_t|\mathbf{z}_t}\right)^2}{M}\right), \tag{2.19}$$

where $\rho_{x_t,y_t|\mathbf{z}_t}$ is the true partial correlation, T represents the total number of samples, m denotes the dimension of the predictor, $\mathbf{z}_t$, and M denotes the degrees of freedom, $M = T - m - 1$, contributing to the calculation. Similarly to correlation, the variance of sample partial correlation depends on the true partial correlation. Fisher (1924) showed that Fisher's z-transformation can be employed to remove this dependence,

$$\mathrm{r}\left(\mathrm{corr}\left(x_t, y_t|\mathbf{z}_t\right)\right) = \frac{1}{2}\ln\frac{1+\mathrm{corr}\left(x_t, y_t|\mathbf{z_t}\right)}{1-\mathrm{corr}\left(x_t, y_t|\mathbf{z}_t\right)}. \tag{2.20}$$

and demonstrated that the distribution of $\mathrm{r}\left(x_t, y_t|\mathbf{z}_t\right)$ is asymptotically Gaussian,

$$\mathrm{r}\left(\mathrm{corr}\left(x_t, y_t|\mathbf{z}_t\right)\right) \overset{a}{\sim} \mathcal{N}\left(\mathrm{r}\left(\rho_{x_t,y_t|\mathbf{z}_t}\right), \frac{1}{M}\right), \tag{2.21}$$

where $\mathrm{r}\left(\rho_{x_t,y_t|\mathbf{z}_t}\right)$ denotes Fisher's z-transformation of the true partial correlation, $\rho_{x_t,y_t|\mathbf{z}_t}$.

To measure linear influence from a multivariate predictor, $\mathbf{z}_t$, to a scalar target, x_t, the multiple correlation coefficient, Eq. (2.4), or partial variance, Eq. (2.5), can be employed. In both cases, the significance tests were designed to evaluate the incremental efficacy obtained from adding predictors. Therefore, nested models are employed to determine if the change in linear dependence between the two models - the models before and after the additional predictors are added - is significant. In the case of evaluating linear dependence between $\mathbf{z}_t$ and x_t under a null hypothesis of zero dependence, the efficacy of $\mathbf{z}_t$ is compared with having no predictor at all.

For the partial variance measure, nested models are compared by evaluating the change in the amount of signal variance remaining after regressing out predictors. The variance remaining for the model with no predictor at all is simply the variance of the signal itself.

Therefore, significance testing linear dependence between $\mathbf{z}_t$ and x_t under a null hypothesis of zero dependence, requires the partial variance test (Fox, 1947),

$$\begin{aligned} \mathrm{q}\left(\mathrm{var}\left(x_t - 0\right), \mathrm{var}\left(x_t - \tilde{x}_t\right)\right) &= \mathrm{q}\left(\mathrm{var}\left(x_t\right), \mathrm{var}\left(e_t\right)\right) \\ &= \frac{\mathrm{var}\left(x_t\right) - \mathrm{var}\left(e_t\right)}{\mathrm{var}\left(e_t\right)} \frac{M}{m} \\ \mathrm{q}\left(\mathrm{var}\left(x_t\right), \mathrm{var}\left(x_t - \tilde{x}_t\right)\right) &\sim F\left(m, M\right). \end{aligned} \tag{2.22}$$

Employing the same nested models concept to test the significance of a multiple correlation estimate requires the null hypothesis model with no predictor to have zero multiple correlation. In this case, the significance test for multiple correlation becomes (Fox, 1947),

$$\begin{aligned} \mathrm{h}\left(\mathrm{mcorr}\left(x_t, \mathbf{z}_t\right), \mathrm{mcorr}\left(x_t, 0\right)\right) &= \mathrm{h}\left(\mathrm{mcorr}\left(x_t, \mathbf{z}_t\right), 0\right) \\ &= \frac{\mathrm{mcorr}\left(x_t, \mathbf{z_t}\right)^2}{1 - \mathrm{mcorr}\left(x_t, \mathbf{z_t}\right)^2} \frac{M}{m} \\ \mathrm{h}\left(\mathrm{mcorr}\left(x_t, \mathbf{z}_t\right)\right) &\sim F\left(m, M\right). \end{aligned} \tag{2.23}$$

2.2.4 Autoregressive models

Gaussian noise-driven autoregressive (AR) models form a subclass of the linear Gaussian model (LGM) class in which the predictor contains lagged components of the signal itself, in a self (auto) regression (McQuarrie and Tsai, 1998, p.89). Model order, p, determines the longest lag present in the predictor. vector autoregressive (VAR) models are a multivariate generalisation of AR models (McQuarrie and Tsai, 1998, p.199). The VAR(p) model for timeseries x_t and y_t is defined by

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{a} & \mathbf{c} \\ \mathbf{b} & \mathbf{d} \end{bmatrix}' \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{y}_{t-1}^{(p)} \end{bmatrix} + \begin{bmatrix} e_t \\ u_t \end{bmatrix}, \qquad \begin{bmatrix} e_t \\ u_t \end{bmatrix} \sim \mathcal{N}\left(0, \mathbf{\Sigma}\right), \tag{2.24}$$

where $\mathbf{a} = [a_1, \ldots, a_p]'$, $\mathbf{b} = [b_1, \ldots, b_p]'$, $\mathbf{c} = [c_1, \ldots, c_p]'$, and $\mathbf{d} = [d_1, \ldots, d_p]'$ are vectors of autoregressive coefficients and e_t and u_t are the noise processes driving the signals x_t and y_t, respectively. Assume that there are T samples. The noise covariance matrix, $\mathbf{\Sigma}$, is determined as

$$\mathbf{\Sigma} \triangleq \begin{bmatrix} \varepsilon & \delta \\ \delta & \eta \end{bmatrix}. \tag{2.25}$$

AR models are the trivial VAR case for which $\delta = 0$ and the autoregressive coefficients $\mathbf{b} = \mathbf{c} = \mathbf{0}$, effectively decoupling the two timeseries into distinct univariate AR models.

Components of an AR model will be identified with subscript A. VAR models will be identified by subscript V.

2.2.5 Model order selection

Applying AR models to connectivity analysis requires determining model order, defined by the maximum lag included in the regression. The quality of a model's fit is assessed via prediction error, which is summarised by the variance of the estimated residual timeseries. Therefore, it is required to find the model order that minimises estimated residual variance. However, residual variance is a monotonic decreasing function of model order and hence increasing order can never result in a worse model fit (Rezek and Roberts, 1997). Therefore, since parsimony is desirable, model order is penalised. The most common approach to order selection is founded on trading model complexity (determined by model order) and prediction error, the exact details of which differ for the various order selection criteria. Parsimony is important in fMRI connectivity, where timeseries are often short, to maximise samples for parameter estimation and to avoid overfitting (overestimating model order) which models the noise and renders a model less applicable to other datasets generated from the same model specification (Goldenshluger and Zeeviy, 2001). A brief summary of the most common order selection criteria used in fMRI connectivity is provided here, but is by no means an exhaustive review.

The final prediction error (FPE) criterion (Akaike, 1969) was derived from the notion of a training set; a training set of size $T-(p+1)$ is used to estimate the final prediction error for a sample of size $T+p+1$ (Rezek and Roberts, 1997). FPE is susceptible to overestimating model order (Goldenshluger and Zeeviy, 2001). Akaike (1973) later proposed a generalised criterion derived from information theoretic concepts, which he termed Akaike's information criterion (AIC). The objective of AIC is to select the model that minimises, in a maximum likelihood sense, the expectation of the Kullback-Liebler distance between a model and the data (Akaike, 1973). FPE and AIC are asymptotically equivalent with sample size (Rezek and Roberts, 1997). AIC was shown to be problematic in that its variance does not tend to zero with increasing sample size (Rezek and Roberts, 1997); the minimum description length (MDL) was introduced to resolve this issue (Rissanen, 2007), penalising model order less severely than AIC. The Bayesian information criterion (BIC) maximises the posterior probability distribution of the model, assuming that the data can be described by a distribution in the exponential family (Schwarz, 1978). BIC imposes a more severe penalty on model complexity than AIC, and its variance tends to zero with increasing sample size (Shittu and Asemota, 2009). MDL and BIC are equivalent for the LGM class (Fitzgibbon et al., 2004). Finally, Hannan and Quinn (1979) derived an order selection criterion (HQC) based on AIC, but with consistency, meaning that with probability one, HQC will select the correct model as sample size tends to infinity.

For each criterion, the objective is to identify the most parsimonious model by minimising the order selection equation as a function of model order, p. Assume the criteria are being evaluated for an AR(p) model with residual variance $\hat{\varepsilon}_A$, and let T denote the number of

image volumes in the timeseries so that the degrees of freedom remaining for parameter estimation is $N = T - p$. The equations for several order selection criteria are then as follows:

$$\text{Akaike information criterion: } AIC = N\ln\left(\hat{\varepsilon}_A\right) + 2p, \tag{2.26}$$

$$\begin{cases}\text{Bayesian information criterion:}\\ \text{Minimum description length:}\end{cases} \quad BIC = MDL = N\ln\left(\hat{\varepsilon}_A\right) + p\ln\left(N\right), \tag{2.27}$$

$$\text{Hannan Quinn information criterion : } HQC = N\ln\left(\hat{\varepsilon}_A\right) + 2p\ln\left(\ln\left(N\right)\right). \tag{2.28}$$

$$\text{Final prediction error: } FPE = \frac{N\hat{\varepsilon}_A\left(N+p+1\right)}{N-p-1}, \tag{2.29}$$

$$\text{Akaike information criterion corrected: } AICc = \ln\left(\hat{\varepsilon}_A\right) + \frac{N+p}{N-p-2}, \tag{2.30}$$

Another order selection criterion that is used less frequently in the fMRI literature is the partial autocorrelation function (PACF). To calculate the order of an AR(p) model for timeseries x_t, the PACF is calculated for a range of lags. The PACF at lag k is defined as the autocorrelation between x_t and x_{t-k} that remains unaccounted for by the autocorrelation contained in lags 1 through to $k-1$. Theoretically, the PACF of an AR(p) model should be zero for all lags greater than p, though in practice such coefficients are independent and identically distributed (iid) Gaussian variates, distributed around a mean of zero with standard deviation $1/\sqrt{N}$ (Fuller, 1996). Each PACF coefficient is significance tested and model order is determined by the maximum lag of the set of all PACF coefficients that exceed the stated threshold.

The necessity of a threshold to distinguish significant PACF values implies that AR timeseries with actual coefficients less than the threshold cannot be modelled as they cannot be distinguished from noise.

2.2.5.1 Stability

The z-transform can be applied to a discrete time signal to determine its complex frequency domain representation. The transform facilitates analysis of the frequency characteristics of a model. If $X\left(\lambda\right)$ denotes the value of timeseries x_t at complex frequency λ then the z-transform is defined as (Ifeachor and Jervis, 1993, p.106),

$$X\left(\lambda\right) = \sum_{i=-\infty}^{\infty} x_i \lambda^{-i}. \tag{2.31}$$

Applied to the AR model for x_t, Eq. (2.24), with $\delta = 0$ and $\mathbf{c} = \mathbf{b} = \mathbf{0}$, the z-transform can be interpreted as a time-shift operator such that the power of λ^{-i} indicates the lag of

a sample in the sum (Shin and Hammond, 2008, p.149),

$$X(\lambda) = X(\lambda) \sum_{i=1}^{p} a_i \lambda^{-i} + E(\lambda), \tag{2.32}$$

where $E(\lambda)$ denotes the z-plane frequency representation of e_t.

In the z-domain, the response of an AR process can be characterised by its transfer function, which is a rational function of system output to system input. Rearrangement of Eq. (2.32) allows the transfer function of the AR model to be written as,

$$\begin{aligned} \mathrm{H}(\lambda) &= \frac{E(\lambda)}{X(\lambda)} \\ &= \frac{1}{1 - \sum_{i=1}^{p} a_i \lambda^{-i}}. \end{aligned} \tag{2.33}$$

Stability of an autoregressive model is governed via the poles of its transfer function, which are identified by the roots of the transfer function's denominator. This equation is often referred to as the characteristic equation. Expanding the denominator of the transfer function, the characteristic equation can be written,

$$\lambda^p - a_1 \lambda^{p-1} - \cdots - a_p = 0. \tag{2.34}$$

The values of λ that satisfy this equation are the poles of the AR(p) system. All poles must lie within the unit circle for the system to be stable, so that (Hamilton, 1994)

$$|\lambda_i| \leq 1, \quad \forall i = 1, \cdots, p \tag{2.35}$$

is a necessary and sufficient condition for stability (Prandoni and Vetterli, 2008, p.152).

2.2.5.2 Granger causality measures

Granger causality employs the notion of temporal precedence to formally define causal dependence in the context of timeseries (Granger, 1969); a timeseries, x_t, causally influences another timeseries, y_t, if knowledge of the past of x_t reduces uncertainty of y_t, where uncertainty is quantified by the proportion of signal variance that remains unaccounted for. Geweke (1982) extended the concepts for autoregressive models by decomposing dependence into instantaneous and directed dependence. Geweke's measure of Granger causality (GGC) was then defined as a sum of the two reciprocal directed influence measures plus an instantaneous influence measure, as will be defined below.

The total linear dependence between x_t and y_t of the bivariate VAR model defined in

Eqs. (2.24)-(2.25) is measured by directed Granger causality (DGC) from y_t to x_t, denoted $F_{y\to x}$, directed Granger causality from x_t to y_t, denoted $F_{x\to y}$, and instantaneous Granger causality (IGC) between x_t and y_t, designated by $F_{x\cdot y}$; summing these three measures gives the total linear dependence between two timeseries (Geweke, 1982),

$$F_{x,y} \triangleq F_{x\to y} + F_{y\to x} + F_{x\cdot y}, \tag{2.36}$$

where

$$\textbf{DGC:} \begin{cases} F_{y\to x} \triangleq \ln \frac{\mathrm{var}(e_{t,A})}{\mathrm{var}(e_{t,V})}, \\ F_{x\to y} \triangleq \ln \frac{\mathrm{var}(u_{t,A})}{\mathrm{var}(u_{t,V})}, \end{cases} \tag{2.37}$$

$$\textbf{IGC:} \quad F_{x\cdot y} \triangleq \ln \frac{\mathrm{var}\left(e_{t,V}\right)\mathrm{var}\left(u_{t,V}\right)}{\left|\mathrm{var}\left(\begin{bmatrix} e_{t,V} \\ u_{t,V} \end{bmatrix}\right)\right|}. \tag{2.38}$$

In an extension to multivariate models, Geweke (1984) introduced conditional instantaneous Granger causality (CIGC) and conditional directed Granger causality (CDGC), measures of dependence in which the influence between x_t and y_t is estimated given knowledge of a third timeseries, z_t, to provide robustness in the presence of mediated influence. Denote by V_{xz} the bivariate VAR model containing x and z, and by V_{xyz} the multivariate VAR model of x, y, and z, then

$$\textbf{CIGC:}\ F_{x\cdot y|z} = \ln \frac{\mathrm{var}\left(e_{t,V_{xyz}}\right)\mathrm{var}\left(u_{t,V_{xyz}}\right)}{\left|\mathrm{var}\left(\begin{bmatrix} e_{t,V_{xyz}} \\ u_{t,V_{xyz}} \end{bmatrix}\right)\right|}, \tag{2.39}$$

$$\textbf{CDGC:}\ F_{y\to x|z} = \ln \frac{\mathrm{var}\left(e_{t,V_{xy}}\right)}{\mathrm{var}\left(e_{t,V_{xyz}}\right)}. \tag{2.40}$$

Partial directed Granger causality (PDGC), introduced by Guo et al. (2008) and similar to conditional Granger causality, removes any additional covariance between the source and target variables to address concern regarding potentially unmodelled inputs that are common to all regions, such as cardiac noise. If w_t is the residual of z_t after regressing onto the VAR predictor, then

$$\textbf{PDGC:}\ F_{y\to x|\cdot z} \triangleq \ln \frac{\mathrm{var}\left(e_{t,V_{xz}}|w_{t,V_{xz}}\right)}{\mathrm{var}\left(e_{t,V_{xyz}}|w_{t,V_{xyz}}\right)}. \tag{2.41}$$

While the notation of '$|\cdot$' to denote PDGC is somewhat cumbersome, it is also intuitive in the sense that '$|z$' denotes regressing onto z's past, while '$\cdot z$' refers to the subsequent removal of any remaining instantaneous covariance between the residuals of x_t and z_t.

2.2.5.3 Statistical inference testing of Granger causality measures

The IGC and DGC dependence measures described in Chapter 2.2.5.2 each form test statistics when scaled by the degrees of freedom that contributed to the estimation, given by $N = T - 2p - 1$.

Under the null hypothesis, $H_0 : F_{x \cdot y} = 0$, the IGC measure, Eq. (2.38), is asymptotically χ^2 distributed (Geweke, 1982). Let $\mathrm{S}_{x \cdot y} = NF_{x \cdot y}$ denote the test variate for IGC. Then

$$\mathrm{S}_{x \cdot y} \overset{a}{\sim} \chi^2 (1) . \tag{2.42}$$

Recall that $\overset{a}{\sim}$ denotes asymptotic equivalence with sample size.

The null hypothesis for directed influence, measured via DGC, Eq. (2.37), is expressed as $H_0 : F_{y \to x} = 0$. The DGC inference test has an asymptotic $\chi^2(p)$ distribution (Geweke, 1982),

$$\mathrm{S}_{y \to x} \overset{a}{\sim} \chi^2 (p) , \tag{2.43}$$

where $\mathrm{S}_{y \to x} = NF_{y \to x}$ denotes the test variate for DGC.

Chapter 3

Signal Characteristics

3.1 Introduction

FMRI connectivity analyses typically estimate linear dependencies in the brain in the context of a LGM. Correlation-based data-driven methods implicitly employ a LGM (Härdle and Simar, 2007, p,.132) so that the predominance of correlation-based methods in the category (Friston, 1994; Wang and Xia, 2007; Sun et al., 2004; Marrelec et al., 2006) renders the majority of data-driven methods implicitly founded on a LGM. Members of the model-driven category that measure linear dependence between signals explicitly utilise an instance of the LGM class (McIntosh and Gonzalez-Lima, 1994; Gates et al., 2009; Kim et al., 2007). Additionally, AR models, a subclass of the LGM class, are utilised by model-driven methods that consider autocorrelation in the signals, such as the VAR method introduced by Harrison et al. (2003), and Granger causality techniques (Goebel et al., 2003; Roebroeck et al., 2005; Liao et al., 2010; Deshpande et al., 2009).

The use of a standard LGM is coupled with several implicit assumptions which must be validated to guarantee integrity of results. Signals are required to be wide-sense stationary, and the noise process is assumed iid Gaussian (Fox, 1947, p.113). Wide-sense stationarity requires the timeseries mean and variance to remain constant over time, while the iid Gaussianity assumption requires an absence of autocorrelation in the noise timeseries so that it is comprised of serially independent samples drawn from a Gaussian distribution. In the absence of model assumption conformance, a measure must be examined to ensure it is robust to the violation.

The importance of assumption validity suggests the need for a detailed appraisal of the experimental BOLD dataset employed in this thesis. Therefore, in this chapter we introduce the dataset and discuss the assumptions of a LGM with a review of their consideration in the fMRI literature, and an appraisal of the experimental dataset employed throughout this thesis. The key contributions of this chapter are now outlined.

1. LGM model assumptions: Contrary to the LGM Gaussianity assumption for noise, BOLD FMRI noise is understood to be Rician distribution. A Rician distribution is well approximated by a Gaussian distribution for high SNR environments. Validity of the Gaussianity assumption for LGM models is considered with a review of the literature and an appraisal of an experimental resting state BOLD dataset. LGM models also assume temporal sample independence and wide-sense stationarity of signals. The validity of these assumptions is also examined, including a review of their consideration in the fMRI literature and an assessment of the experimental dataset.

2. AR model stability: AR models incorporating linear autoregression form a subclass of the LGM class that requires additional assumptions. The self regression contained in AR models can render them unstable, and hence appraising an AR model for stability is crucial in ensuring it is a meaningful model. This chapter contains a discussion of the AR test for

stability and its use in fMRI connectivity literature.

3. AR order selection: Model order selection is a significant issue for AR models; identifying the most parsimonious model is important in order to capture all signal dynamics, while retaining as many degrees of freedom as possible for parameter estimation (Goldenshluger and Zeeviy, 2001). Overfitting also compromises performance of a model because the additional degrees of freedom model the particular noise process in the timeseries sample, which is not generalisable to later instances of the process (de Waele and Broersen, 2003). This chapter contains a review of order selection criteria employed in the fMRI connectivity literature, and compares the performance of each using an experimental resting state dataset.

4. AR model applicability: An AR residual timeseries is required to be stationary, Gaussian and white. Therefore, if voxel timeseries are indeed generated from an AR process, the residual should conform to these assumptions with correct model order. Using a resting state BOLD dataset the applicability of AR models to fMRI resting state data analysis is evaluated in this chapter by examining residuals for assumption conformance across a range of model orders. The performance of AR models is further scrutinised by calculating the proportion of voxel signal variance explained by the models. Additionally, the performance of VAR models, using both brain and non-brain voxels as predictors, is examined to ascertain whether it is signal of interest or noise that is being explained by a VAR model.

This chapter is organised as follows. The dataset utilised for validating theoretical results contained in the thesis is introduced. Section 3.2 describes acquisition of the resting state BOLD dataset. This dataset is examined for conformance to LGM assumptions, key point 1 above, in Section 3.3, and the applicability of AR models, key points 2–4 above, in Section 3.4. As distinct to other chapters in this thesis, each of the key points is explored without division into explicit methods and results sections.

3.2 Experimental data acquisition

Three healthy controls were scanned on a 3T Siemens Tim TRIO MRI scanner using two different procedures: 1) resting state BOLD echo planar imaging and 2) a motor task performance BOLD EPI. For both procedures we collected 219 volumes for each participant (repetition time = 1600 ms; echo time = 20 ms; flip angle = 90; 24 trans-axial slices, matrix = 64×64, acquisition voxel size=$3.125 \times 3.125 \times 5.5$ mm^3)[1]. The motor task was a block design containing alternating periods of 30s each. During the active period the subjects were instructed to press either the left or right button according to an arrow displayed on the screen that changed randomly every second. During the baseline period, subjects were instructed to focus on a crosshair at the screen's centre point. Slice timing correction and

[1]Data courtesy of the Florey Neuroscience Institutes, Parkville.

motion correction were performed using FSL FEAT (Jenkinson et al., 2002) and data were detrended using MATLAB to mitigate the effects of coil heating.

The left hemisphere primary motor cortex (LMC) and the right hemisphere primary motor cortex (RMC) were identified via a conventional general linear model (GLM) activation analysis (Friston et al., 1994) applied to the motor task data. LMC and RMC timeseries were formed for each subject by averaging the five most activated voxels in each region-of-interest. Brain and non-brain timeseries were formed by averaging five random voxels from within and outside the brain, respectively.

3.3 Linear Gaussian Model Assumptions

In this section, the assumptions for utilising LGMs are examined. Previous discussions of each assumption in the fMRI literature are reviewed, followed by an appraisal of the validity of the assumption for the resting state BOLD dataset employed in this thesis (Section 3.2).

As the name implies, linear Gaussian models assume that error terms are characterised by a Gaussian distribution. Although LGM models require an assumption of Gaussianity, fMRI noise is understood to be Rician distributed. This is because voxel intensity values are derived from the magnitude of complex Gaussian components (Bernstein et al., 1989). Nonetheless, the Rician distribution is well approximated by a Gaussian distribution for voxels with sufficient SNR; Gudbjartsson and Patz (1995) showed that, in the context of structural MRI data, the Gaussian assumption for Rician voxel intensity values is reasonable even for fairly small SNR, requiring the ratio of signal mean to noise standard deviation to be greater than three.

Many researchers have found support for the assumption of Gaussianity. Wink and Roerdink (2006) assessed the validity of a Gaussian assumption for fMRI data and found that it is acceptable, and further demonstrated that the difference between two Rician distributions can be well approximated by a Gaussian distribution. This is highly relevant for BOLD data in which only comparisons between timeseries is meaningful. Petersson et al. (1999) found the Gaussian approximation valid on the basis of Gaussian spatial smoothing and justified by the central limit theorem. Conversely, Hanson and Bly (2001) argued that BOLD noise is Gamma distributed, motivating the result by interpreting voxels as a source or sink for exponential processes with similar decay rates. They highlight the danger in incorrectly assuming Gaussian distributed noise, namely that the heavier tail of the Gamma distribution significance tested as a Gaussian distribution results in a higher false positive rate. However, Chen et al. (2003a) asserted that the result in Hanson and Bly (2001) stemmed the spatial distribution of noise since spatial and temporal samples were pooled, and further showed that a Gaussian approximation was valid for the temporal distribution of BOLD noise. The distribution of voxel intensity values of resting state data is discussed in more detail in

Chapter 4.

Gaussianity of voxel timeseries for the experimental resting state dataset employed in this thesis (Section 3.2) was assessed using the Anderson-Darling test ($p = 0.01$) (Anderson and Darling, 1952). The Anderson-Darling test attempts to transform the sample data to a uniform distribution using the hypothesised distribution, in this case Gaussian. A distance metric is then employed to evaluate the uniformity of transformed data. The Shapiro-Wilk (Shapiro and Wilk, 1965) and Kolmogorov-Smirnov (Smirnov, 1948) tests were also applied with very similar results so only the Anderson-Darling test results are presented here.

Each subject had between 88% and 92% of brain voxels considered normal according to the Anderson-Darling test for normality. For non-brain voxels, the percentage for normality reduced to between 50% and 60% for all subjects. These results are in accordance with the Gaussianity approximation for Rician distributed voxels acquired with sufficiently high SNR. As expected, for all voxels the three parameter Student's t distribution fit as well or better than a Gaussian distribution, where distribution fit was evaluated in a log-likelihood sense. An example brain voxel is shown in Fig. 3.1 with both a Gaussian and a three parameter Student's t distribution optimally fit to the voxel's intensity samples using maximum likelihood estimation (MLE). The brain voxel was considered normal using the Anderson-Darling test for Gaussianity but the Student's t distribution was optimal. The distribution of voxel intensity is discussed further in Chapter 4.

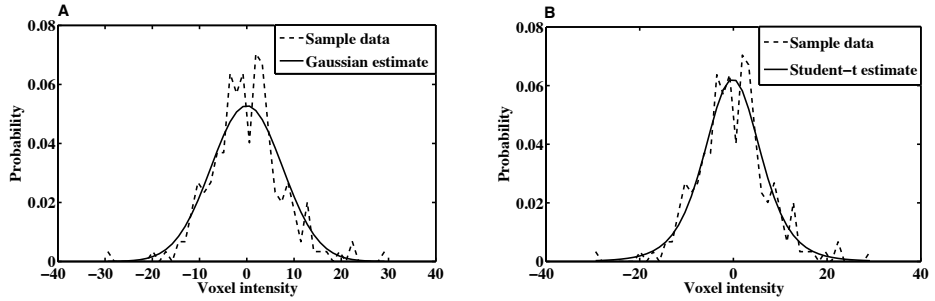

Figure 3.1: MLE (solid line) fit to an example voxel's sample intensity values (dashed line) for: **A:** a Gaussian distribution, **B:** a three parameter Student's t distribution (LHS), the Student's t distribution is optimal in a log-likelihood sense.

An important assumption of the LGM class is wide-sense stationarity, which requires the first and second moments of a timeseries to remain constant over time (Hamilton, 1994). Granger and Newbold (1974) demonstrated the importance of a stationarity assumption by showing that its violation causes incongruities between the actual and expected distributions of the null hypothesis test variate, in some cases leading to spurious regression. Moreover, non-stationarity can distort connectivity estimates by inadvertently applying an implicit time-varying weight to samples (Cohen et al., 2003). Random noise occurring at samples

with a large weight can appear connected whilst signal of interest occurring at low-weight samples can be obscured by noise. Therefore, if the non-stationarity remains uncorrected, connectivity results may not be reflective of the underlying neuronal dependence.

Non-stationarity is a known issue for fMRI data owing to both non-stationary noise processes (Diedrichsen and Shadmehr, 2005), such as subject movement, and non-stationary signal power, a repercussion of inhomogeneous RF amplification (Tanase et al., 2011). We tested voxels for stationarity using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Bhargava, 1986), which is a unit root test with a null hypothesis of stationarity.

90% of non-brain voxels were found to be stationary, while the proportion of stationary brain voxels was consistently between 50% and 60%, and not significantly tissue specific (data not shown). Stationarity in the context of fMRI connectivity is considered further in Chapter 4.

A problematic assumption of the LGM class for BOLD data is the assumption of sample independence or, equivalently, an absence of autocorrelation in the noise process. Bullmore et al. (1996a) first identified autocorrelation in the residuals of an fMRI activation study. It is now accepted that fMRI noise can contain significant autocorrelation due to underlying physiological fluctuations (Purdon and Weisskoff, 1998) and that, since SNR varies with acquisition sampling rate, the extent of autocorrelation attributable to physiological fluctuations varies also. Several methods have been proposed to address the problem of autocorrelation. Precolouring imposes a known autocorrelation function on the noise via temporal smoothing of voxel timeseries (Worsley and Friston, 1995), while prewhitening transforms the data to remove any autocorrelation in the error terms (Bullmore et al., 1996a). Prewhitening is optimal in terms of variance of parameter and noise estimates, but on the unlikely premise that the autocorrelation structure is known (Bullmore et al., 2001; Woolrich et al., 2001). Lack of understanding of the autocorrelation profile leads to bias in statistical inference of model parameters (Marchini and Smith, 2003). Sample variance of autocorrelation estimates is high so Gautama and Hulle (2004) proposed averaging across all voxels within the brain to increase degrees of freedom of the estimate and, therefore, reduce sample variance. However, to complicate matters spatial variability of autocorrelation exists that can not be accounted for by tissue variability (Bullmore et al., 1996a). If spatial variability remains unaddressed averaging across brain voxels will underestimate or overestimate and autocorrelation for specific voxels (Gautama and Hulle, 2004). Consequently, it is preferable to estimate noise autocorrelation for voxels individually (Locascio et al., 1997).

Voxel timeseries from the experimental dataset were examined for sample independence using the Breusch-Godfrey test, which also works in the presence of lagged dependent variables (Breusch, 1979). Three lags ($p = 1, 2, 3$) were included in a voxel's test for independence.

Non-brain voxels were found to be predominantly white (greater than 90% for all subjects), and brain voxels consisted of many non-white voxels, close to 50% for all subjects. In

this case, some tissue specificity was exhibited, with cerebral spinal fluid (CSF) and grey matter (GM) both exhibiting higher rates of non-white voxels (approximately 60% for all subjects) while the white matter (WM) revealed lower rates of sample dependence, being approximately 20% for all subjects. The impact of autocorrelation on the distribution of sample correlation is detailed in Chapter 5. Voxel timeseries that were both non-stationary and non-white did exhibit tissue specificity, predominantly in the CSF (Fig. 3.2).

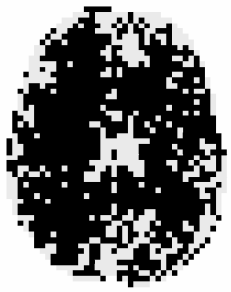

Figure 3.2: Example slice (z=15) from subject 1 showing voxels that are both non-white and non-stationary (red).

3.4 Autoregressive model assumptions

We examine the applicability of AR models to resting state BOLD data and investigate issues including model stability and order selection. A brief discussion of how AR models are employed in fMRI connectivity literature is first provided.

The demonstrated presence of autocorrelation in BOLD signals prompted a focus on autoregressive models in the fMRI connectivity literature. Harrison et al. (2003) employed a single multivariate VAR model with autoregressive coefficients interpreted as connection strength. Where there were multiple lags depicting directed influence between regions, a table of connection strengths was provided. Coefficients whose standard deviation about the mean crossed zero were discarded as insignificant. Valdes-Sosa et al. (2005) employed a sparse VAR model to avoid the necessity of *a priori* knowledge for region selection. A model order of one was assumed and an iterative process employed, during which coefficients were successively forced to zero according to their size.

Granger causality employs AR models to identify temporal precedence relationships from lagged dependence. Geweke's Granger causality (GGC) measures utilise the concept of Granger causality but decompose dependence into instantaneous and directed influence

components. GGC measures have been applied extensively in fMRI connectivity (Goebel et al., 2003; Roebroeck et al., 2005; Abler et al., 2005; Chen et al., 2006; Zhou et al., 2009a), with a concurrent stream of frequency domain GGC implementations (Chen et al., 2006; Zhou et al., 2009a; Demirci et al., 2009; Deshpande et al., 2009; Sato et al., 2009, 2006a). Implicit in the application of Granger causality methodology is the assumption that voxel timeseries are described well by AR models.

3.4.1 Applicability of AR models to fMRI connectivity

The applicability of AR models to fMRI resting state data analysis was evaluated by examining the conformance of LGM assumptions with differing model order. An AR residual timeseries is required to be stationary, Gaussian and white. Therefore, if voxel timeseries are indeed generated from an AR process, the residual should be seen to conform to these assumptions with sufficient order. There is a pronounced lack of inclusion of the AR(0) model in order selection within fMRI connectivity literature (Harrison et al., 2003; Goebel et al., 2003; Deshpande et al., 2009; Cecchi et al., 2003); if LGM assumptions are met for an AR(0) residual then an AR model is unnecessary.

We examined assumption conformance for model orders between zero and six. For each model order we employed MLE to determine regression coefficients for the univariate AR model (Eq. (2.24)). Conformance to AR model assumptions was examined using: the Anderson-Darling test for Gaussianity (Anderson and Darling, 1952), the Breusch-Godfrey test for independence (Breusch, 1979), and the KPSS test for stationarity (Bhargava, 1986). In each case the significance level, $\alpha = 0.01$. The results are shown in Fig. 3.3.

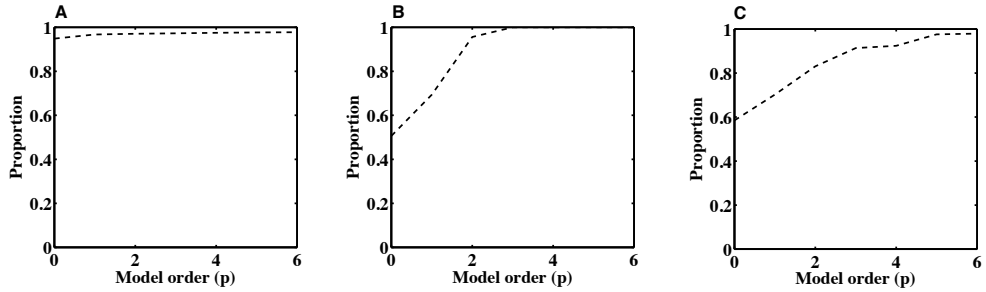

Figure 3.3: Proportion of brain voxels with AR(p) residual timeseries satisfying LGM assumptions for model order, $p \in [0, 6]$. **A:** Test of Gaussianity using the Anderson-Darling test. **B:** Test for sample independence using the Breusch-Godfrey test. **C:** Test for stationarity using the KPSS test.

Increasing model order increases the number of brain voxels with residual timeseries that conform to LGM assumptions (Fig. 3.3). An AR(6) model culminates in almost 100% of

brain voxel residuals being white, Gaussian and stationary, compared with approximately 50% of voxel timeseries satisfying all three assumptions. This result suggests that AR models are valid for modelling fMRI data. Indeed, an example voxel timeseries (Fig. 3.4A) shows that the fitted AR(3) signal captures the dynamics of the voxel timeseries well. However, closer inspection reveals that despite the improved conformance to model assumptions, the residuals of the AR(3) model were almost identical to the AR(0) model, indicating poor model performance. An example is shown in Fig. 3.4B, in which the voxel timeseries is non-white and non-stationary, but after fitting an AR model (of order 3 in this case), the residual becomes white and stationary. However, the signal component of the AR model has not necessarily captured all of the signal dynamics. This may be indicative of an inappropriate model type or perhaps of the small percentage of signal fluctuation attributable to signal of interest, often cited as less that 10% (Kruger et al., 2001). This example demonstrates that it cannot be assumed that a single model type is applicable to all brain voxels.

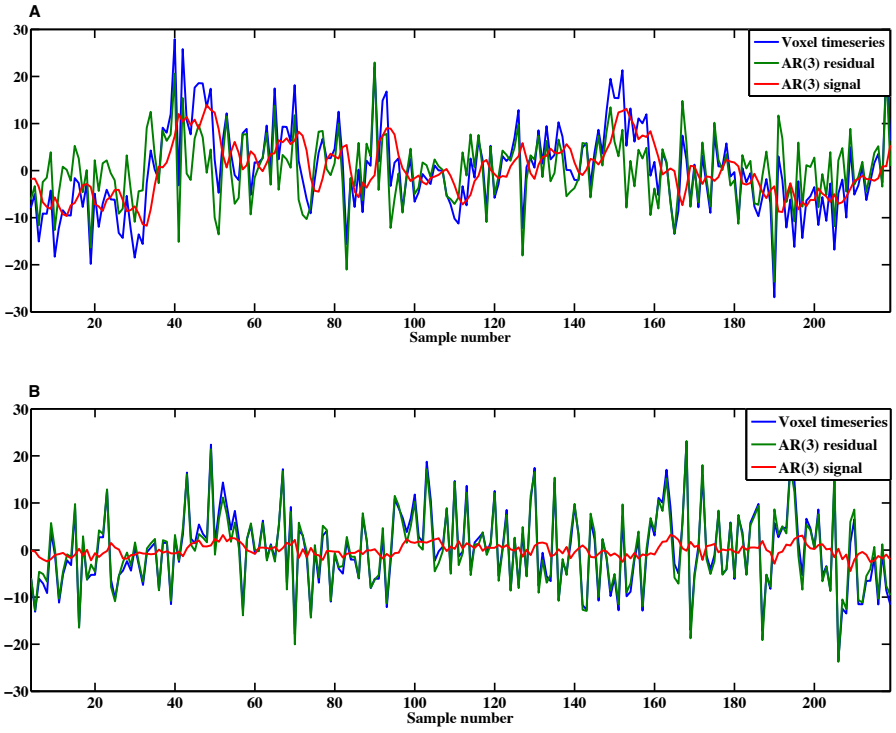

Figure 3.4: Example voxel timeseries and AR residuals. Each voxel timeseries (blue) is considered non-white and non-stationary prior to fitting an AR(3) model, while each residual (green) is considered white and stationary after fitting an AR(3) signal (red). Note that the voxel timeseries have been truncated to account for the initiation period of the AR(3) model. **A:** A good fit; the fitted AR signal follows the dynamics of the voxel timeseries. **B:** A poor fit; the fitted AR signal has not captured voxel timeseries dynamics despite the residual passing iid Gaussianity tests.

It is well known that increasing model order will always result in reduced residual variance since the additional degrees of freedom are used to model the noise process (Rezek and Roberts, 1997). However, for a true AR process, there will be a sharp reduction in incremental signal variance explained with increasing order once the true lag is exceeded (Rezek and Roberts, 1997). This concept is tested on experimental BOLD data by fitting an AR model to the LMC timeseries with model orders between one and ten (Fig. 3.5A). VAR models were also fit to the LMC timeseries to model the impact of the RMC, brain voxels, and non-brain voxels on residual variance across a range of model orders (Fig. 3.5A) (see Section 3.2 for details on how these timeseries were generated). An example of the residual timeseries for each model is shown in Fig. 3.5B.

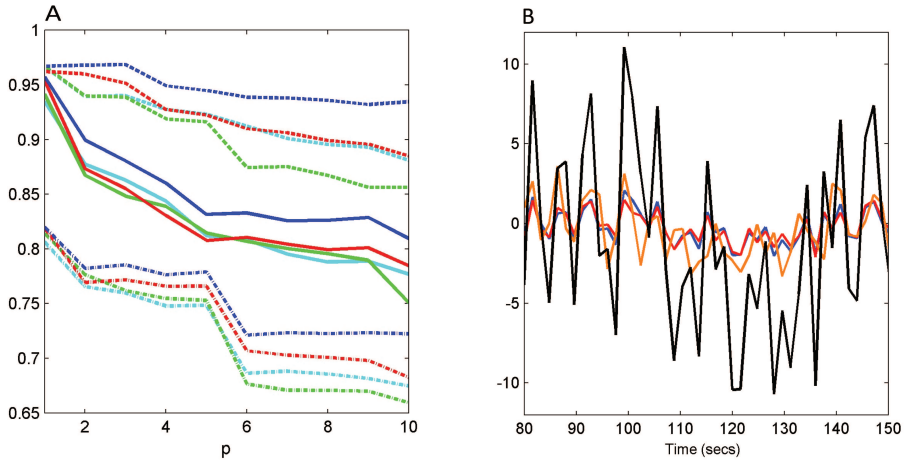

Figure 3.5: AR and VAR model fits to the LMC of three subjects. **A:** Residual variance of LMC shown as a proportion of original signal variance, and plotted as a function of model order, for model(timeseries): AR(LMC) (blue), VAR(LMC,RMC) (red), VAR(LMC,brain) (cyan), and VAR(LMC,non-brain) (green), for subjects: 1 (solid), 2 (dashed) and 3 (dashed-dotted). **B:** Example segment of LMC timeseries for subject 2 (black), with estimated signals from models: AR(1) fit (blue), VAR(1) fit with RMC (red), VAR(5) fit with RMC (orange).

Residual variance decreased with increasing order when modelling the LMC timeseries with both randomly selected voxels and voxels drawn from the contralateral motor cortex (Fig. 3.5A). While contralateral timeseries decreased residual variance in VAR models, we have demonstrated that randomly selected timeseries from both within and outside the brain produce the same effect; non-brain signals diminished residual variance comparably to all brain signals. A salient observation is the impoverished description that AR models pro-

vide of resting state BOLD signals (Fig. 3.5B). These results suggest that the VAR models are modelling the noise in the LMC timeseries rather than signal of interest. It is possible that components of the noise are common to the timeseries extracted from LMC, RMC, brain, and non-brain, causing the comparable performance of all regions, although the poor fits suggest this is not a primary factor. It appears more likely that all VAR models perform comparably because all model the LMC timeseries poorly.

3.4.2 Order estimation criteria

Use of AR models introduces the complication of identifying model order, which requires ascertaining the maximum lag present in the regressor. Identifying the most parsimonious model is important to optimise model performance; underfitting results in unmodelled signal dynamics whilst overfitting compromises predictive performance by modelling the noise process, which is not generalisable (de Waele and Broersen, 2003). There are many methods available for order selection, most of which require minimising an expression for model error and impose a penalty for complexity (Section 2.2.5). The BIC (Schwarz, 1978) imposes a harsher penalty than AIC (Shittu and Asemota, 2009), whilst AIC is known to overfit because of being derived asymptotic with length but applied to finite samples (de Waele and Broersen, 2003). HQC (Hannan and Quinn, 1979) lies between AIC and BIC in terms of probability of overestimating and underestimating model order (Asghar and Abid, 2007).

In the fMRI connectivity literature, Chang and Glover (2010) and Roebroeck et al. (2005) employed BIC for order estimation, while Baraldi et al. (2007) used AIC. (Roebroeck et al., 2005) estimated model order by finding the minimum of AIC, BIC, and HQC, and Tagaris et al. (1997) and Locascio et al. (1997). employed the PACF is sometimes used and it is also common in the fMRI literature to simply assume an AR model order, typically AR(1) (Bullmore et al., 1996a; Goebel et al., 2003; Abler et al., 2005).

AIC, BIC, and HQC, defined in Section 2.2.5 were applied to the experimental resting state dataset (Section 3.2) to determine how different criteria estimate model order, and to establish how dependent the resulting model is on the choice of criteria. For model orders between zero and ten, MLE (Eq. (2.7)) was employed to estimate the regression coefficient vector, a_A, for the univariate AR model (Eq. (2.24)). For each voxel timeseries, the order selection criteria were minimised across model order according to Eqs. (2.26)-(2.28).

The estimated AR model order was found to vary significantly between both subject and criteria. Using the notation [subject 1, subject 2, subject 3], for the LMC timeseries AIC evaluated model order to be $p = [5, 1, 1]$, respectively. BIC estimated LMC model order for the three subjects to be $[4, 1, 1]$. For the RMC, AIC determined model order for the subjects to be $p = [1, 5, 5]$, while BIC estimated model order as $[1, 1, 2]$. Evidently, there is a lack of consistency across both subject and selection criteria.

The estimates of model order for three slices of subject 1 are shown for AIC, BIC, and HQC (Fig. 3.6). All three measures indicate that a single model order cannot be assumed valid across all brain voxels, and certainly a model order of 1 is insufficient. Further to this, all three order selection criteria exhibit tissue specificity, particularly BIC. AIC estimates model order comparatively higher than HQC and BIC, which is in agreement with its known tendency to overfit the data (de Waele and Broersen, 2003). HQC appears less conservative than BIC, while less prone to overfitting than AIC. BIC estimates are the most conservative, which is expected since it's penalty term is most severe (Shittu and Asemota, 2009). Although FPE was examined, these results have not been incorporated since they were identical to AIC; this is credible since AIC and FPE are asymptotically equivalent (Liew, 2004).

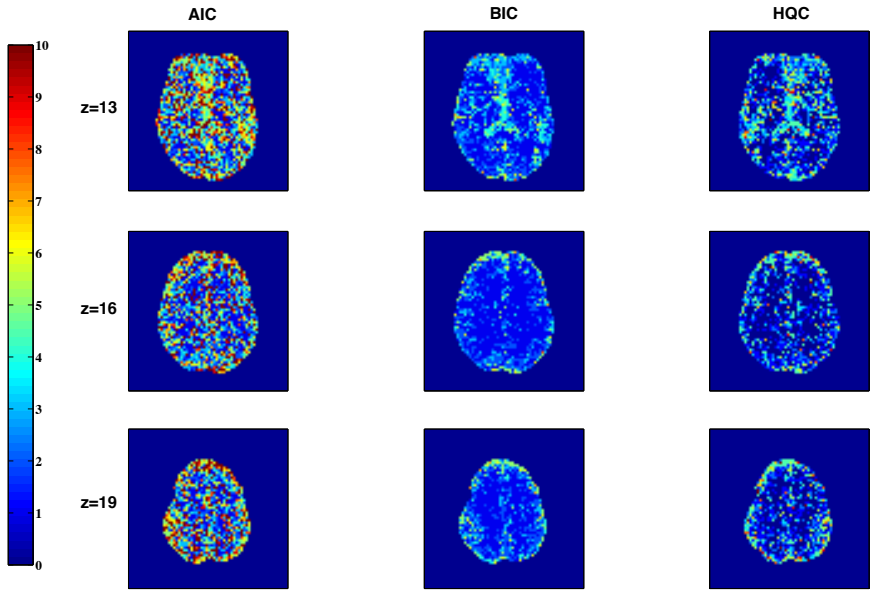

Figure 3.6: Order selection criteria applied to resting state data brain voxels of subject 1. For each model order between zero and ten MLE was used to estimate the AR coefficient vector and calculate residual variance. The residual variance that minimised each order selection criteria was used to identify model order. **AIC**: Akaike information criterion, **BIC:** Bayesian information criterion, **HQC:** Hannan-Quinn criterion.

An AR(1) model is frequently assumed in fMRI connectivity literature (Bullmore et al., 1996a; Goebel et al., 2003; Abler et al., 2005). We calculated the AR(1) coefficient for all brain voxels across subjects (Fig. 3.7). The estimated AR(1) coefficient displays tissue specificity, with high values in grey matter and lower values in white matter. These demarcations are reminiscent of the low frequency artifact demonstrated by Lund et al. (2006), and attributed to hardware error. If so, this has ramifications for GGC measures that utilise

lagged regressors to identify causal relationships.

3.4.3 AR model stability

An AR model can be described by its characteristic equation, the roots of which determine model stability (see Section 2.2.5.1). The roots, or poles, dictate the behaviour of an AR model because they designate the dominant frequencies and relative strength of each frequency. Poles with absolute value greater than 1 result in an unstable model with signal magnitude that will diverge to infinity over time. In the context of fMRI, stability of an AR model is crucial for a meaningful biological model that prevents the voxel intensity, and hence haemodynamic response, from growing without limitation. Accordingly, Valdes-Sosa et al. (2005) rejected sparse VAR models with poles greater than one to ensure stability.

We examined use of the temporal coefficient size to determine significance of lagged samples in an AR regression (Harrison et al., 2003), by comparing the poles of two AR systems that differed only in the inclusion of a small coefficient. The impact of removing an apparently insignificant coefficient on stability was examined by assessing the poles of the following two AR systems:

$$\text{AR system 1:} \quad x_t = 1.1x_{t-1} + 0.15x_{t-2} + e_t, \tag{3.1}$$

$$\text{AR system 2:} \quad y_t = 1.1y_{t-1} + 0y_{t-2} + u_t, \tag{3.2}$$

where the model for x_t denotes an AR system with a coefficient at lag 2, while the model describing y_t represents the AR system with the coefficient set to zero. Stability results are shown in Fig. 3.8. The first model, containing a small coefficient for the lag 2 sample, is stable, since the poles are inside the unit circle. After forcing the small coefficient of the second lag to zero, the poles of the model indicate that it is no longer stable, and signal magnitude will diverge to infinity. In this example, a seemingly insignificant temporal coefficient is seen to be critical to stability of the model.

3.5 Conclusions

The validity of linear connectivity methods is contingent on the timeseries conforming to often implicit and untested assumptions of the linear Gaussian model class. We have assessed an experimental fMRI resting state dataset for conformance to the LGM assumptions. While the Gaussianity assumption was found to be valid, the stationarity and whiteness assumptions were more problematic. These assumptions are considered in detail in Chapters 4 and 5.

Granger causality methodologies are increasingly employed in fMRI connectivity analyses.

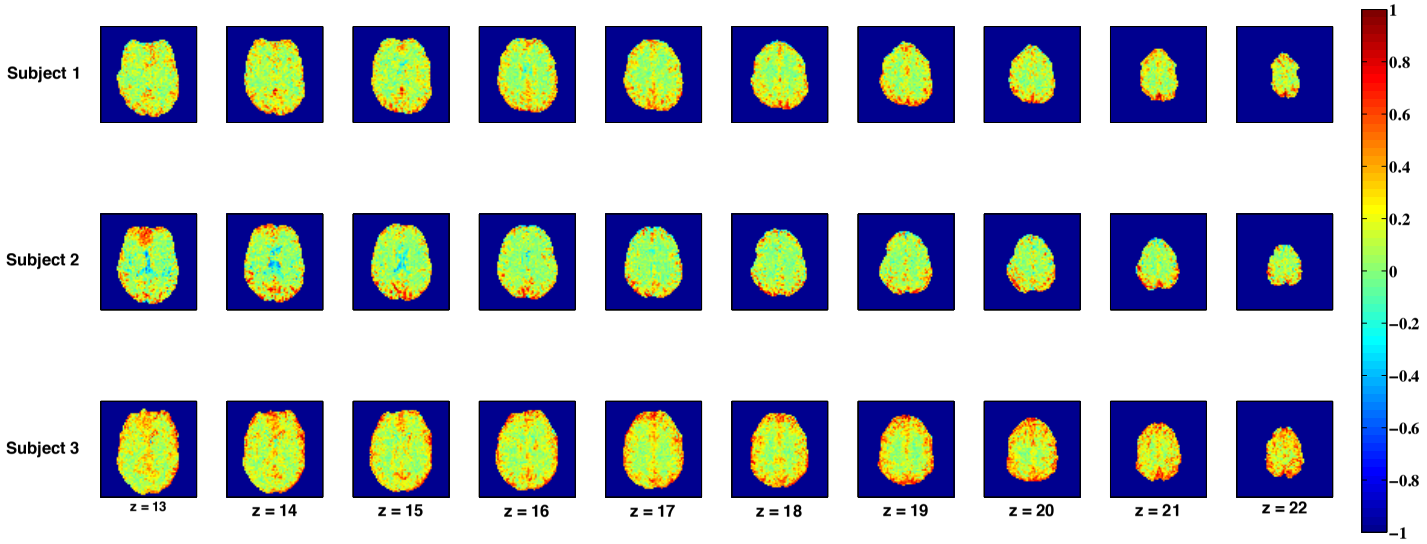

Figure 3.7: AR(1) coefficient values calculated using MLE for all brain voxels of subjects 1, 2, and 3 (slices z=13 to z=22 shown). Note the tissue specificity of coefficient values.

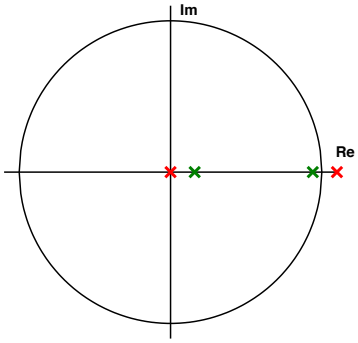

Figure 3.8: Poles of two different AR models differing in the inclusion of one small coefficient on the second lagged sample, and demonstrating the necessity for establishing AR model stability via poles; $x_t = 1.1x_{t-1} - 0.15x_{t-2} + e_t$ (green), $y_t = 1.1y_{t-1} + 0y_{t-2} + u_t$ (red).

The validity of Granger causality in fMRI relies on the assumption that timeseries are well-described by stationary and stochastic AR models of determinable order. While a VAR model that includes the contralateral timeseries decreases the residual variance in VAR models, we have demonstrated that randomly selected timeseries from within and outside the brain produce the same effect. AR models provide impoverished descriptions of resting state BOLD signals, and are not robust across order selection procedures or model-training/model-validation timeseries decompositions. Granger causality methodology must, therefore, be applied to experimental fMRI with caution.

Chapter 4

Non-stationarity in fMRI Connectivity

4.1 Introduction

FMRI connectivity analysis utilises BOLD data to determine how spatially remote regions cooperate to fulfill a task, or connections between regions observed when simply at rest. BOLD data is notoriously noisy, being impacted by both equipment related noise, such as coil interference (Bandettini et al., 1998) and scanner drift (Bianciardi et al., 2009), and subject noise, such as physiological interference (Biswal et al., 1995) or subject movement (Barry et al., 2010). Both equipment and subject noise can be irregular and therefore non-stationary. Scanner equipment can generate non-stationary signal variance from inhomogeneous RF amplification (Tanase et al., 2011), coil resistance and inductive losses (Gudbjartsson and Patz, 1995), disparities in scanner performance (Weisskoff, 1996), and signal attenuation from diffusion and dissemination amongst surrounding voxels (Stejskal, 1965). Delay between slice acquisitions can render the introduced non-stationarity slice dependent, which is significant for connectivity in which comparison between voxels is key, in contrast to an activation analysis in which voxels are analysed independently (Kimberg, 2008).

FMRI connectivity analysis frequently employs a linear Gaussian model (LGM) to estimate linear dependence between a seed timeseries and all brain voxels (Marrelec et al., 2005; Goebel et al., 2003; Wang and Xia, 2007; Tana et al., 2008). A core assumption of the LGM class is weak stationarity of the constituent timeseries, requiring the mean and variance of each signal to be constant over time. Violation of this assumption has implications for integrity of connectivity results since the distribution of the test statistic is altered (Granger and Newbold, 1974). Linear dependence estimates may also be skewed as time-varying variance is analogous to applying a (non-constant) weight to each sample; samples associated with higher variance receive a greater weighting and are, therefore, more significant in the correlation estimate (Cohen et al., 2003). Random noise occurring at samples characterised by large variance can skew results and even induce spurious connectivity (Granger and Newbold, 1974). Therefore, if the non-stationarity remains uncorrected, connectivity results may not be reflective of the underlying linear dependence between voxels.

Non-stationarity of BOLD timeseries has been examined in an effort to resolve the loss of integrity associated with a stationarity assumption for non-stationary data. Lund et al. (2006) modelled non-stationarity of noise via nuisance variable regression in an fMRI activation analysis. This restricts removal of non-stationary processes to those explicitly modelled in the regressors. Diedrichsen and Shadmehr (2005) addressed time-varying volume variance in a fMRI activation study by employing a linear model with Gaussian noise characterised by time dependent variance. The non-stationary process was assumed to apply globally, affecting all voxels identically. This choice was motivated empirically by observing correlation between voxel-wise time dependent variances and global scaling parameters. Sample variance was estimated for each image and employed as a weighting in a least squares es-

timate of regression coefficients. Restricted maximum likelihood (Bartlett, 1937) was used to determine regression coefficients. Luo and Puthusserypady (2007) consider the noise model proposed by Diedrichsen and Shadmehr (2005), employing a Bayesian framework to estimate the time-varying weights.

The necessity for modelling non-stationarity is typically argued with visual evidence (Calhoun et al., 1999; Lund et al., 2006; Diedrichsen and Shadmehr, 2005) or is assumed (Oikonomou et al., 2009; Long et al., 2005), so that there is a notable absence of statistical evidence demonstrating non-stationarity (Park et al., 2010). Furthermore, modelling of non-stationary processes has typically been limited to fMRI activation data; there have been few investigations into non-stationarity of resting state signals. Typical modelling of non-stationarity has focused on addressing non-stationary noise processes (Calhoun et al., 1999; Lund et al., 2006; Diedrichsen and Shadmehr, 2005) or non-stationary dependence measures (Long et al., 2005; Chang and Glover, 2010; Niazy et al., 2011; Park et al., 2010); there has been less attention paid to non-stationary signal variance.

The distribution of resting state BOLD noise is typically assumed Gaussian in fMRI connectivity analyses (Friston, 1994; Wang and Xia, 2007; Sun et al., 2004; Marrelec et al., 2006). While the fMRI literature contains experimental evidence for this assumption (Wink and Roerdink (2006); Petersson et al. (1999)), there has also been evidence to the contrary: Hanson and Bly (2001) argued that BOLD noise is Gamma distributed, while Chen et al. (2003a) showed that the Gaussianity assumption is valid only when voxel timeseries are normalised by voxel standard deviation, which they proposed reduces non-Gaussianity introduced by head movement.

We propose a model of the non-stationarity in voxel intensities introduced by fluctuations in slice-dependent signal power. Slice variance of resting state data is empirically shown to be non-stationary and optimally characterised by an inverse Gamma distribution. The distribution of intensity values for voxels with non-stationary variance is analytically determined to be a generalisation of a Student's t distribution. The impact of non-stationary signal variance on connectivity is determined via analytic derivation of correlation between non-stationary signals. We establish that the expected value of correlation between non-stationary signals is altered, and propose a correction that is analytically shown to restore both signal stationarity and correlation results, thereby mitigating the effect of spurious correlation for non-stationary timeseries. Theoretical results are corroborated with empirical evidence that demonstrates the utility of our correction in restoring stationarity and integrity of connectivity results.

This chapter is organised as follows. General background detailing notation and definitions of variance and correlation are be reviewed in Section 2.2.2. Theoretical results are detailed in Section 4.2, incorporating a derivation for the distribution of voxel intensity for signals with time-varying variance (Section 4.2.2), a description of the non-stationary signal

model, and the consequent expression for correlation between non-stationary timeseries with proposed correction (Section 4.2.3). Empirical and experimental methods are described in Section 4.3, followed by a description of results in Section 4.4 and discussion in Section 4.5.

4.2 Theoretical results

In the acquisition of BOLD data, two-dimensional slices are typically obtained serially, on a millisecond time scale. This can result in changes in the equipment and subject environment, ultimately modifying signal characteristics between slices, causing non-stationarity of voxel timeseries (Calhoun et al., 1999). FMRI analyses typically assume stationarity of voxel timeseries and therefore it is desirable to determine a correction to restore stationarity. Voxels within the same slice are acquired simultaneously and hence signal power within a slice should be constant. We propose a model of non-stationarity for resting state BOLD data in which each two-dimensional slice is characterised by a distinct non-stationary variance process. The distribution of voxel intensity is analytically derived for this model and the impact of the non-stationarity on correlation is determined. Crucially, a correction is proposed that is demonstrated to restore stationarity and correct correlation values so they reflect the true linear dependence between voxel timeseries.

4.2.1 A non-stationary model of resting state timeseries

We begin by proposing a model for the measured non-stationary signal in a voxel x, situated in slice n. The signal model decomposes the voxel intensity into stationary and non-stationary components,

$$x_{n,t} = x^{(s)}_{n,t}\sigma_{n,t}, \quad t = 1, \ldots, T, \tag{4.1}$$

where $x^{(s)}_{n,t}$ denotes the stationary component at time t, and $\sigma_{n,t}$ is a time-varying multiplicative variance process distinct to slice n, distributed according to an inverse gamma distribution (see Section 4.4),

$$\sigma^2_{n,t} \sim IG\left(\alpha_n, \frac{1}{\beta_n}\right). \tag{4.2}$$

Here α_n and β_n are slice specific shape and scale parameters, respectively, and IG denotes the inverse gamma distribution.

The resting state stationary component is assumed to be Gaussian distributed (Wink and Roerdink, 2006),

$$x^{(s)}_{n,t} \sim \mathcal{N}\left(0, \sigma^2_x\right), \tag{4.3}$$

4.2.2 Distribution of voxel intensity

We analytically derive the distribution of measured non-stationary voxel intensity in resting state fMRI given the model proposed in Section 4.2.1.

As $x_{n,t}^{(s)}$ and $\sigma_{n,t}$ are independent random variables,

$$x_{n,t} \sim \mathcal{N}\left(0, \sigma_x^2 \sigma_{n,t}^2\right). \tag{4.4}$$

We now determine the distribution of non-stationary voxel intensity when $\sigma_{n,t}$ is unknown. Let $\alpha = \alpha_n$ and $\beta = \beta_n$ for readability, and introduce a precision parameter, $\tau_t = \frac{1}{\sigma_{n,t}^2}$, observing that Eq. (4.2) requires τ_t to be gamma distributed according to, $\tau_t \sim Ga(\alpha, \beta)$ (Hamada et al., 2008, p.392), so that

$$p_{Ga}(\tau_t\,;\,\alpha,\beta) = x^{\alpha-1}e^{-\beta x}\frac{\beta^\alpha}{\Gamma(\alpha)}. \tag{4.5}$$

Here Γ is the gamma function.

Let $z_t = \frac{x_{n,t}}{\sigma_x}$. Then Eq. (4.4) becomes

$$z_t \sim \mathcal{N}\left(0, \frac{1}{\tau_t}\right), \tag{4.6}$$

with conditional probability density function

$$p(z_t \mid \tau_t) = \sqrt{\frac{\tau_t}{2\pi}} e^{-\frac{\tau_t}{2}(z_t)^2}. \tag{4.7}$$

The unconditional distribution of z_t is determined by integration over τ_t, as follows:

$$\begin{aligned} p(z_t) &= \int_0^\infty p(z_t \mid \tau_t)\, p(\tau_t)\, d\tau \\ &= \int_0^\infty \frac{\tau_t^{\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{\tau_t}{2}z_t^2} \frac{\beta^\alpha}{\Gamma(\alpha)} \tau_t^{\alpha-1} e^{-\beta\tau_t} d\tau \\ &= \int_0^\infty \tau_t^{\alpha-\frac{1}{2}} e^{-\tau_t\left(\beta+\frac{z_t^2}{2}\right)} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\sqrt{2\pi}} d\tau. \end{aligned} \tag{4.8}$$

Substituting $a = \alpha + \frac{1}{2}$ and $b = \beta + \frac{z_t^2}{2}$, gives

$$\begin{aligned} p(z_t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(a)}{b^a} \frac{1}{\sqrt{2\pi}} \int_0^\infty \tau_t^{a-1} e^{-\tau_t b} \frac{b^a}{\Gamma(a)} d\tau \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(a)}{b^a} \frac{1}{\sqrt{2\pi}}. \end{aligned} \tag{4.9}$$

Therefore,

$$\begin{aligned} p\left(z_t\right) &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\left(\beta+\frac{z_t^2}{2}\right)^{\alpha+\frac{1}{2}}} \frac{1}{\sqrt{2\pi}} \\ &= \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)} \left(\frac{1}{2\pi\beta}\right)^{\frac{1}{2}} \left(\frac{\beta+\left(\frac{z_t}{\sqrt{2}}\right)^2}{\beta}\right)^{-\left(\alpha+\frac{1}{2}\right)} . \end{aligned} \tag{4.10}$$

Substituting $x_{n,t} = z_t\sigma_x$ back into Eq. (4.10), orchestrated via a change of variables (Papoulis and Pillai, 2002), requires

$$\begin{aligned} p\left(x_{n,t}\right) &= p\left(z_t\right)\left|\frac{dz_t}{dx_{n,t}}\right| \\ &= \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)} \left(\frac{1}{2\pi\beta}\right)^{\frac{1}{2}} \left(\frac{\beta+\left(\frac{\frac{z_t}{\sigma_x}}{\sqrt{2}}\right)^2}{\beta}\right)^{-\left(\alpha+\frac{1}{2}\right)} \frac{1}{\sigma_x} \\ &= \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)} \left(\frac{1}{2\pi\beta\sigma_x^2}\right)^{\frac{1}{2}} \left(1+\frac{z_t^2}{2\sigma_x^2\beta}\right)^{-\left(\alpha+\frac{1}{2}\right)} . \end{aligned} \tag{4.11}$$

The three parameter Student's t distribution is (Härdle and Simar, 2007, p.129)

$$p_t\left(z\,;\,\mu,\nu,\lambda\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\lambda}{\pi\nu}\right)^{\frac{1}{2}} \left[1+\frac{\lambda\left(z-\mu\right)^2}{\nu}\right]^{-\frac{\nu+1}{2}} , \tag{4.12}$$

Therefore the voxel intensity distribution, Eq. (4.11), is an instance of the three parameter Student's-t distribution where $\mu = 0$, $\nu = 2\alpha$ and $\frac{\lambda}{\nu} = \frac{1}{2\beta\sigma_x^2}$. Consequently, the voxel intensity of sampled non-stationary resting state BOLD data described by the model in Section 4.2.1 is characterised by a generalised Student-t distribution.

4.2.3 Impact of non-stationarity on fMRI connectivity

Many fMRI connectivity methods, such as correlation-based measures (Friston et al., 1994; Wang and Xia, 2007), Granger causality methods (Goebel et al., 2003; Guo et al., 2008; Chen et al., 2006), and structural equation modelling (McIntosh and Gonzalez-Lima, 1994), assume stationarity of voxel timeseries. We now propose a correction to apply to voxel timeseries that restores stationarity.

Let $x^{(c)}_{n,t}$ denote a corrected voxel timeseries, given by

$$x^{(c)}_{n,t} = \frac{x_{n,t}}{\hat{\sigma}_{n,t}} \tag{4.13}$$

where $\hat{\sigma}^2_{n,t}$ is the sample variance of voxel intensities within slice n at time t. It is assumed that this sample variance will include both the time-varying variance of interest, $\sigma^2_{n,t}$, and a scale-factor, σ_n, such that

$$\begin{aligned} x^{(c)}_{n,t} &= \frac{x^{(s)}_{n,t}\sigma_{n,t}}{\sigma_{n,t}\sigma_n} \\ &= \frac{x^{(s)}_{n,t}}{\sigma_n}. \end{aligned} \tag{4.14}$$

Observe that the corrected voxel timeseries is now stationary, but with a variance that differs by a constant factor from that of the variance of the stationary component, $x^{(s)}_{n,t}$, Eq. (4.3). Consequently,

$$x^{(c)}_{n,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_x}{\sigma^2_n}\right). \tag{4.15}$$

Correlation-based measures assume stationarity of timeseries and are known to be susceptible to spurious significance if this assumption is violated (Granger and Newbold, 1974). We now derive an expression for correlation between non-stationary timeseries, and demonstrate that the corrected correlation achieves the value equivalent to the correlation of the underlying stationary component.

Sample correlation requires an estimate of the variance of each voxel timeseries. From the model for non-stationarity, the variance of a voxel timeseries can be determined as

$$\begin{aligned} \sigma^2_{x_{n,t}} &= E\left\{{x_{n,t}}^2\right\} - E^2\left\{x_{n,t}\right\} \\ &= E\left\{x^{(s)2}_{n,t}\sigma^2_{n,t}\right\} - E^2\left\{x^{(s)}_{n,t}\sigma_{n,t}\right\} \\ &= E\left\{x^{(s)2}_{n,t}\right\} E\left\{\sigma^2_{n,t}\right\} - E^2\left\{x^{(s)}_{n,t}\right\} E^2\left\{\sigma_{n,t}\right\} \\ &= \mu_{x^{(s)2}_{n,t}}\mu_{\sigma^2_{n,t}}, \end{aligned} \tag{4.16}$$

where the last line results from the zero mean of the stationary voxel timeseries, Eq. (4.3).

Let $\rho_{x^{(s)}_n, y^{(s)}_m}$ represent the true correlation between stationary voxel signals $x_{n,t}$ and $y_{m,t}$, and let ρ_{x_n, y_m} denote the correlation between the weighted counterparts. Let corr ( ) denote sample correlation so that sample correlation between measured non-stationary signals is

expressed as

$$\begin{aligned}\operatorname{corr}\left(x_{n,t}, y_{m,t}\right) &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_{n,t}^{(s)}\sigma_{n,t}}{\sqrt{\mu_{x_{n,t}^{(s)2}}\mu_{\sigma_{n,t}^2}}}\right)\left(\frac{y_{m,t}^{(s)}\sigma_{n,t}}{\sqrt{\mu_{y_{m,t}^{(s)2}}\mu_{\sigma_{m,t}^2}}}\right) \\ &= \frac{\mu_{\sigma_{n,t}}\mu_{\sigma_{m,t}}}{\sqrt{\mu_{\sigma_{n,t}^2}\mu_{\sigma_{m,t}^2}}}\operatorname{corr}\left(x_{n,t}^{(s)}, y_{m,t}^{(s)}\right), \end{aligned} \tag{4.17}$$

where the last line results from independence between voxel intensity values and the time-varying weights.

Denote the weighting of sample correlation between non-stationary variates by ω so that

$$\begin{aligned}\omega &= \frac{\mu_{\sigma_{n,t}}\mu_{\sigma_{m,t}}}{\sqrt{\mu_{\sigma_{n,t}^2}\mu_{\sigma_{m,t}^2}}} \\ &= \frac{\sum\sigma_{n,t}\sum\sigma_{m,t}}{\sqrt{T\sum\sigma_{n,t}^2 T\sum\sigma_{m,t}^2}}. \end{aligned} \tag{4.18}$$

Jensen's inequality (Jensen, 1906), applied with unitary weights, provides the result that

$$\begin{aligned}\frac{\sum\sigma_{n,t}}{T^2} &\leq \frac{\sum\sigma_{n,t}^2}{T}, \\ \frac{\sum\sigma_{m,t}}{T^2} &\leq \frac{\sum\sigma_{m,t}^2}{T} \end{aligned} \tag{4.19}$$

which implies

$$\begin{aligned}\left(\sum_{t=1}^{T}\sigma_{n,t}\right)^2 &\leq T\sum\sigma_{n,t}^2, \\ \left(\sum_{t=1}^{T}\sigma_{m,t}\right)^2 &\leq T\sum\sigma_{m,t}^2, \end{aligned} \tag{4.20}$$

so that

$$\begin{aligned}\sum_{t=1}^{T}\sigma_{n,t} &\leq \left(T\sum\sigma_{n,t}^2\right)^{\frac{1}{2}}, \\ \sum_{t=1}^{T}\sigma_{m,t} &\leq \left(T\sum\sigma_{m,t}^2\right)^{\frac{1}{2}}. \end{aligned} \tag{4.21}$$

Substituting this result into Eq. (4.18) gives

$$\omega \leq 1. \tag{4.22}$$

Consequently, time-varying weights necessarily reduce sample correlation between non-

stationary signals.

Correlation between corrected timeseries is given by

$$\begin{aligned}
\operatorname{corr}\left(x_{n,t}^{(c)}, y_{m,t}^{(c)}\right) &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_{n,t}^{(c)} - \mu_{x_{n,t}^{(c)}}}{\sigma_{x_{n,t}^{(c)}}}\right)\left(\frac{y_{m,t}^{(c)} - \mu_{y_{m,t}^{(c)}}}{\sigma_{y_{m,t}^{(c)}}}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\frac{x_{n,t}}{\sigma_{n,t}}}{\frac{\sigma_x\sigma_{n,t}}{\sigma_n\sigma_{n,t}}}\right)\left(\frac{\frac{y_{m,t}}{\sigma_{m,t}}}{\frac{\sigma_y\sigma_{m,t}}{\sigma_m\sigma_{m,t}}}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_{n,t}^{(s)}}{\frac{\sigma_x}{\sigma_n}}\right)\left(\frac{y_{m,t}^{(s)}}{\frac{\sigma_y}{\sigma_m}}\right) \\
&= \operatorname{corr}\left(x_n^{(s)}, y_m^{(s)}\right).
\end{aligned} \tag{4.23}$$

This demonstrates that correlation between corrected voxel timeseries is equivalent to the correlation between the stationary component of the voxel timeseries; the impact of the non-stationary variance component has been removed and the underlying linear dependence has been recovered, as required.

4.3 Methods

4.3.1 Simulated data

MATLAB was used to generate simulation datasets of 2000 timeseries pairs. Denote the constituent timeseries in each pair by $x_{n,t}^{(s)}$ and $y_{m,t}^{(s)}$. Each pair was generated from a bivariate normal distribution characterised by: timeseries length, $T \in \{200, 300, 400, 500\}$; true correlation, $\rho \in [-1, 1]$; mean values of $x_{n,t}^{(s)}$ and $y_{m,t}^{(s)}$, denoted μ_x and μ_y, respectively, drawn from the range $[-100, 100]$; and the variance of $x_{n,t}^{(s)}$ and $y_{m,t}^{(s)}$, denoted σ_x and σ_y, respectively, both randomly selected from the range $(0, 100]$. For each dataset, a single time-varying variance process was generated for $x_{n,t}^{(s)}$ timeseries and another for $y_{m,t}^{(s)}$ timeseries to emulate voxels within a slice. Non-stationary variance samples were randomly drawn from an inverse gamma distribution, with the two parameters randomly selected from the range $[1, 10]$. Each timeseries in the dataset was then weighted by the square root of the appropriate time-varying variance process, according to whether it was designated $x_{n,t}^{(s)}$ or $y_{m,t}^{(s)}$ within the pair.

Time-varying slice variance was estimated by calculating sample variance across all $x_{n,t}^{(s)}$ signals, and across all $y_{m,t}^{(s)}$ signals, in the dataset. The precision correction, Eq. (4.14), was applied to generate timeseries corrected for non-stationarity.

Correlation was calculated between each timeseries pair before being weighted by the non-stationary variance process, after being weighted, and after the correction was applied.

4.3.2 Experimental data

Experimental data acquisition was described in Section 3.2. The left motor cortex (LMC) was identified for each subject using the activation dataset acquired during the motor task. Average timeseries were created for the LMC of each subject by averaging across voxels in the region of interest (ROI).

For each subject, the time-varying slice variance was estimated by calculating the sample variance of voxel intensity in each slice at each time point. The best-fit distribution for the variance process of each slice was determined by calculating the MLEs for distribution parameters and identifying the MLE distribution amongst the set of Weibull, Gaussian, Gamma, Inverse Gamma, Student's t, Exponential, Log-normal, Laplace, Rayleigh, Poisson, and Cauchy distributions, that attained the minimum negative log-likelihood. The Wilcoxon signed-rank test was used to determine if slice variance processes were drawn from distributions with equal mean (Wilcoxon, 1945). This choice was motivated by the non-normal distribution of slice variance. Slice variance was also tested for stationarity, for which the augmented Dickey Fuller test was employed (Said and Dickey, 1984). In both cases, the results were significance tested with $\alpha = 0.01$.

Corrected voxel timeseries were formed by weighting voxel samples according to time-varying variance, as per Eq. (4.14). Seed-voxel correlation maps were generated with a LMC seed for datasets with and without applying the non-stationarity correction. Correlation estimates were significance tested using Fisher's z-transformation, Eqs. (2.16)-(2.17), and the resulting maps thresholded ($\alpha = 0.01$ Bonferroni corrected).

4.4 Results

4.4.1 Results for the non-stationary model of resting state timeseries

For each subject and for all slices within each subject's dataset, the inverse gamma distribution was the distribution that best characterised the temporal slice variance. The optimal inverse gamma distributions are shown for all slices in a single subject (Fig. 4.1). Despite significant variation in the MLE parameters, as demonstrated in Fig. 4.2, in each case the distribution of the sample variance process is seen to be accurately characterised by an inverse gamma distribution.

The probability distribution of Gaussian intensity values with inverse gamma non-stationary variance was analytically derived to be a generalisation of a Student's t distribution, Eq. (4.11). Experimental intensity values corroborate the theoretical result; for each subject, the temporal distribution of both slice intensity and individual voxel intensity was deter-

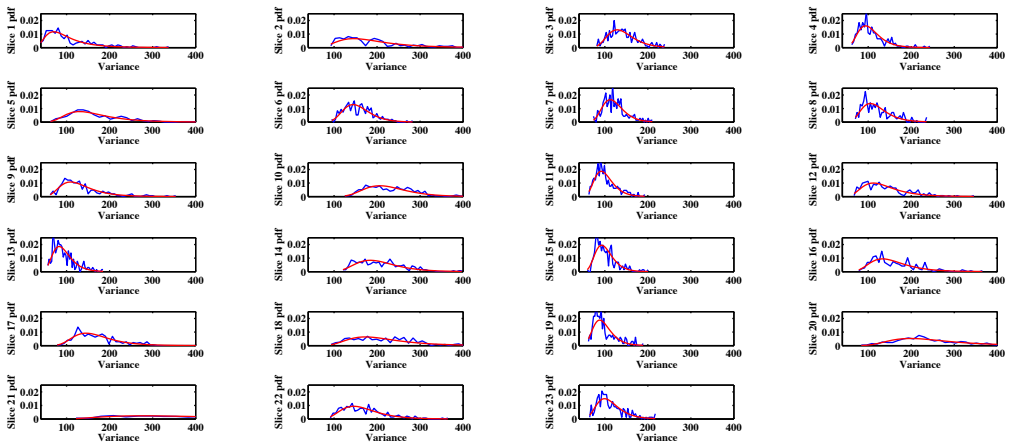

Figure 4.1: Distribution of slice sample variance. Slice sample variance values (blue) and MLE inverse gamma distribution fit to slice variance (red), across slices 1-24.

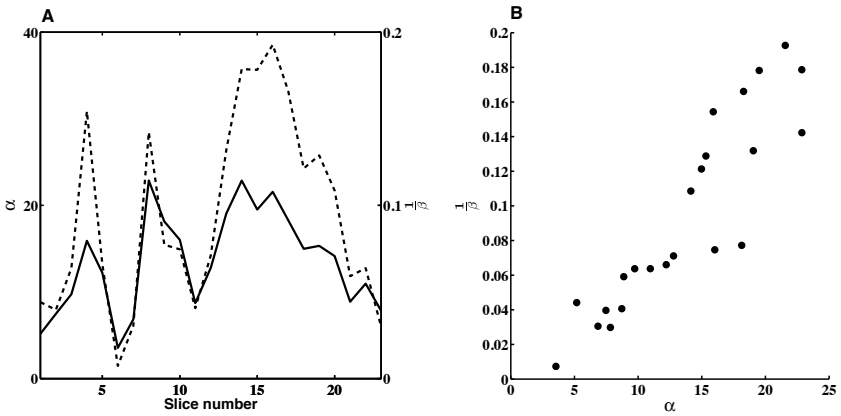

Figure 4.2: Maximum likelihood parameters for slice sample variance values fit optimally to the inverse gamma distribution, $IG(\alpha, \frac{1}{\beta})$. **A:** Plot of α (solid) and $\frac{1}{\beta}$ (dashed) across slice number. Note that each parameter has a separate y-axis labelling because of the size disparity, so that α is labelled on the left side of the figure whilst $\frac{1}{\beta}$ is labelled on the right side of the figure. **B:** Scatter plot of $\frac{1}{\beta}$ against α.

mined to be best characterised by a three parameter Student's t distribution (Fig. 4.3). The optimal three parameter Student's t distributions for the slice and voxel are similar, although the voxel distribution has significantly reduced degrees of freedom. Since a Student's t distribution is asymptotically Gaussian with degrees of freedom our empirical results support the Gaussianity assumption inherent in a LGM.

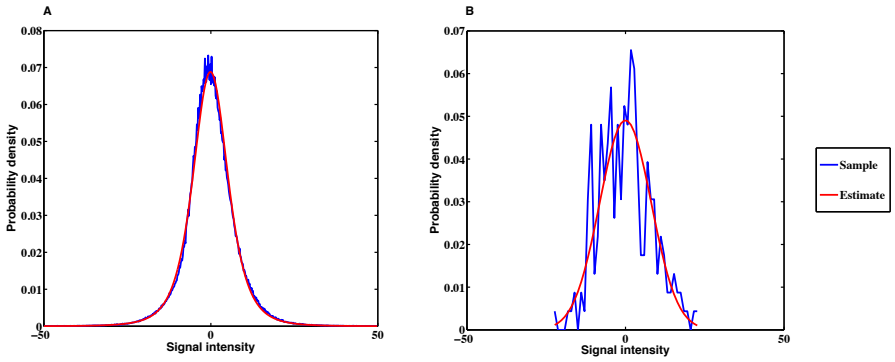

Figure 4.3: Three parameter Student's t distribution (red) fit to experimental data values (blue) for a single subject, drawn from **A:** all voxels within a slice z=16 and across time, **B:** a single voxel timeseries within z=16 (x=32, y=32).

Our theoretical results in Section 4.2.1 were derived using a model of slice-dependent, time-varying variance. This model differs to that of Diedrichsen and Shadmehr (2005), in which a global time-varying variance was assumed so that all voxels within a 3-d image were associated with a single non-stationary weighting. Our model of slice dependent, non-stationary variance was validated using the experimental dataset. Fig. 4.4 shows the temporal variance processes for two exemplar slices, demonstrating both non-stationarity and the dissimilarity of the time-varying process for each slice. A more formal examination of dissimilarity was performed using the Wilcoxin rank-sum test to test for equality of slice variance distributions between all slice pairs. Fig. 4.5 shows the result of the paired tests in matrix form; blue indicates that the null hypothesis of identical means is supported while red indicates rejection of the null hypothesis, suggesting that sample variance values for the slice pair are drawn from different distributions. The result clearly establishes that slice variance distributions are dissimilar in a majority of cases, supporting our model of slice dependent variance.

Stationarity of slices was evaluated using the augmented Dickey-Fuller (ADF) stationarity test applied to each slice. For all subjects, more than 70% of the slice variance processes were found to be non-stationary.

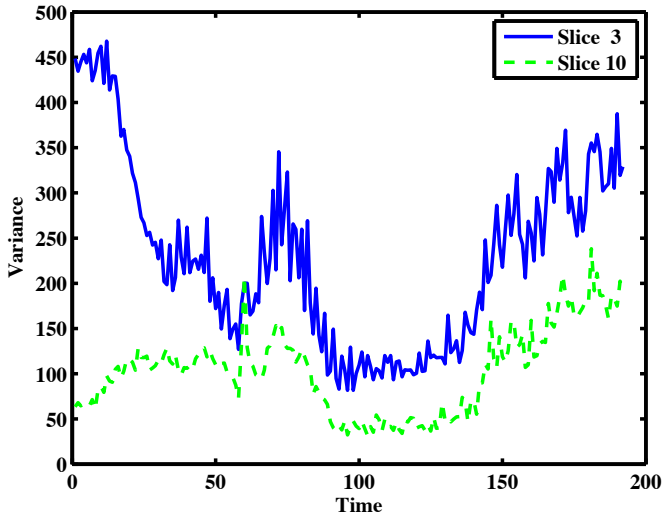

Figure 4.4: Example slice variance processes from the resting state dataset of subject 3, for slice z=3 (blue) and slice z=10 (red)

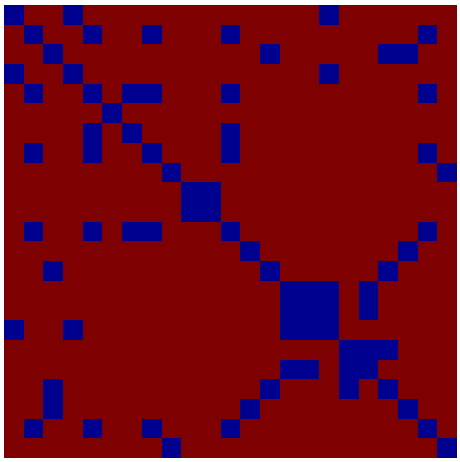

Figure 4.5: Test of slice dependence for time-varying variance. The Wilcoxon rank-sum test is applied to each slice pair to ascertain equivalence of slice variance distributions. Blue indicates that the null hypothesis of equivalence is supported whilst red indicates rejection of the null hypothesis ($\alpha = 0.01$).

4.4.2 Results showing the impact of non-stationarity on fMRI connectivity

Non-stationary variance is problematic when methods that assume stationarity are employed to estimate brain connectivity. The theoretical results detailed in Section 4.2.1 model non-stationarity derived from temporal changes in signal power. The impact of non-stationarity on correlation was analytically established in Section 4.2.3, in which it was shown that a non-stationary weighting alters the expected value of sample correlation, which is critical for fMRI connectivity results. In this section, the theoretical results are corroborated with empirical evidence from both simulated and experimental data.

4.4.2.1 Empirical data

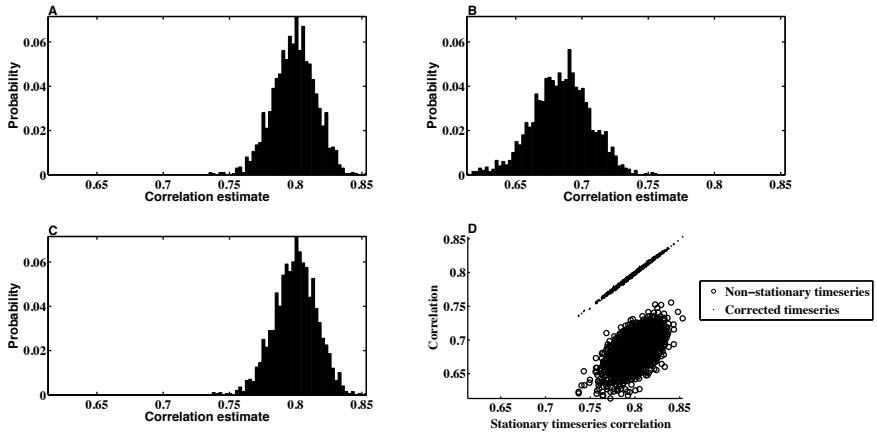

Figure 4.6: **A:** Histogram of correlation estimates between stationary timeseries pairs with true correlation, $\rho = 0.8$. **B:** Histogram of correlation estimates acquired from non-stationary timeseries pairs, with variance for each timeseries within a pair randomly derived from $IG(2, 2)$ or $IG(3, 2)$. **C:** Histogram of correlation between timeseries pairs corrected for non-stationarity. **D:** Scatter plots between original correlation estimates drawn from stationary timeseries against correlation estimates of weighted, non-stationary timeseries (dots) and correlation estimates derived from timeseries corrected to restore stationarity (circles).

Empirical datasets were used to examine the impact of non-stationary variance on sample correlation, and the utility of the voxel timeseries correction proposed in Eq. (4.14). A histogram of correlation estimates derived from the underlying stationary timeseries (Fig. 4.6A) shows an expected correlation that matches the true correlation of 0.8. Correlation estimates acquired from timeseries pairs weighted by a non-stationary inverse gamma process (Fig. 4.6B) have a clearly transformed expected value, decreased to 0.68 in precise agreement with the analytic expression derived in Eq. (4.17). Furthermore, the variance of the corre-

lation estimates for the weighted timeseries is increased, thereby reducing confidence in the estimates and inducing spurious correlation. The histogram of correlation estimates between timeseries corrected for non-stationarity, Eq. (4.14), (Fig. 4.6C), demonstrates the efficacy of our correction in restoring stationarity to the distribution of sample correlation, thereby reestablishing the expected value of sample correlation to the true correlation (Fig. 4.6C). Fig. 4.6D contains a scatterplot between sample correlation values calculated from the stationary timeseries, against both sample correlation of the non-stationary timeseries and those from the corrected timeseries. While the histograms are indicative of changes in the distribution of sample correlation, the scatterplots depict changes in individual correlation estimates as a consequence of the time-varying weighting and subsequent correction. Correlation estimates of weighted timeseries differ by up to 0.25 from the correlation contained in the original, stationary timeseries, with a standard deviation of 0.15. Conversely, the scatterplot between stationary correlation estimates and those derived from timeseries corrected to restore stationarity are almost indistinguishable.

4.4.2.2 Experimental data

The timeseries correction, Eq. (4.14), was applied to the experimental resting state dataset introduced in Section 4.3.2 to examine its ability to restore stationarity. The weighting of correlation anticipated by the theory as a consequence of the non-stationary variance, Eq. (4.18), was found to be in close agreement with experimental values. For example, for voxels co-located with the LMC, the anticipated weighting of correlation was 0.97 of the true underling correlation for subject 1, while experimentally the change was 0.93. Furthermore, the scatterplot between experimental correlation estimates before and after correction (Fig. 4.7) is reminiscent of the scatterplot generated using empirical data (Fig. 4.6D). Experimental correlation estimates are not acquired from the same true correlation, as is the case for the empirical data, so there is a larger range in correlation values for the experimental data. Correlation values estimated from timeseries corrected for stationarity deviated from uncorrected sample correlation values by up to 0.25, with an standard deviation of 0.05.

The utility of the non-stationarity correction is demonstrated in Fig. 4.8, in which seed-voxel correlation maps with a LMC seed is shown prior, and subsequent, to the correction for non-stationary slice variance, Eq. (4.14), for two different subjects. The z-maps for both subjects show a visible difference in connectivity to the LMC as a result of correcting for non-stationarity. Uncorrected z-maps contain a higher number of outliers than z-maps acquired from corrected timeseries, suggesting that uncorrected maps contain some connectivity artificially induced by non-stationarity. Connectivity maps generated from corrected datasets show bilateral connectivity in the primary motor cortex, in closer agreement to expectations for resting state connectivity with the motor cortex (Biswal et al., 1995).

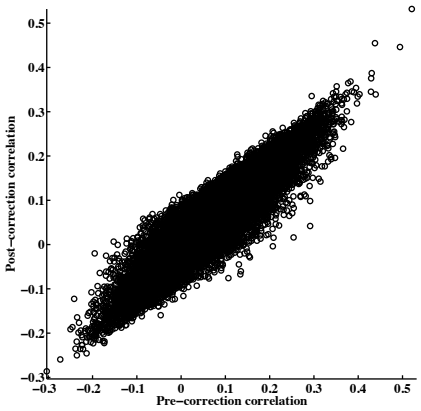

Figure 4.7: Scatterplot of correlation estimates between the LMC and brain voxels (slice 19), before correcting for non-stationarity against correlation estimates acquired after applying the correction to restore stationarity.

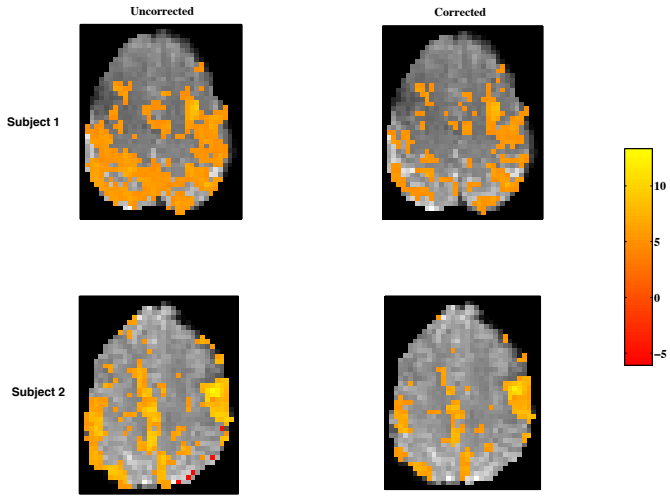

Figure 4.8: Maps of significant LMC-seed voxel correlation for uncorrected timeseries, and timeseries corrected for non-stationary signal power, tested using Fisher's z-transformation. For each subject, the number of outliers for corrected datasets is reduced. Spatial smoothing can cause clusters of outliers, which are visibly reduced after correction, creating significance maps in closer conformance with expected results.

4.5 Discussion

Connectivity analyses of resting state fMRI data typically require stationarity of timeseries. However, this assumption is violated by non-stationarity derived from noise sources such as head movement, or by variable signal power originating from sources such as inhomogeneous RF amplification (Tanase et al., 2011) and signal attenuation (Stejskal, 1965). Violation of the stationarity assumption impacts connectivity results by implicitly applying a weighting to timeseries elements. Random noise occurring at a sample characterised by high variance may contribute more to a correlation estimate than a sample containing signal of interest but located in a region of low signal variance. Consequently, correcting for non-stationarity is not only desirable but necessary (Cohen et al., 2003).

In the context of fMRI activation data, Diedrichsen and Shadmehr (2005) addressed the non-stationarity problem by estimating a global weighting for each image volume to model non-stationary subject movement, while Lund et al. (2006) added regressors to model non-stationary physiological subject noise. Non-stationarity of resting state data has received considerably less attention, with a shift in focus to establishing time-varying linear dependence (Long et al., 2005; Chang and Glover, 2010; Niazy et al., 2011; Park et al., 2010).

We propose a model for non-stationarity motivated by time-varying signal power during acquisition. Since voxels within a slice are acquired simultaneously in 2-d EPI acquisitions, we examined slice variance, describing temporal changes in the variance of voxel intensities within a slice. We established empirically that slice variance is non-stationary and can be well modelled by an inverse gamma distribution, in agreement with Hansen et al. (2003). Importantly, each slice variance process was found to be particular to that slice, contrary to the global assumption made by Diedrichsen and Shadmehr (2005). Incorrectly assuming an identical non-stationary process for all slices adversely effects connectivity results because signal of interest may be ignored based on an increase in noise variance in the overall image volume.

We have proposed an analytic expression for the distribution of normally distributed voxel intensity with non-stationary variance described by an inverse gamma distribution. The resulting distribution of voxel intensity was shown to be a generalisation of a Student's t distribution, which was further validated using an experimental resting state dataset.

Having modelled the non-stationary distribution of fMRI voxel intensities, we subsequently considered the impact of non-stationary slice variance in the context of connectivity analyses, deriving an analytic expression for correlation between non-stationary signals. Sample correlation for the time-varying processes was established as a function of the time-varying variance. Empirical results corroborated the theoretical results and further demonstrated a discernible change in the variance of sample correlation between non-stationary timeseries.

A precision correction to restore stationarity of signals was proposed, which produced cor-

rected stationary signals that differed from the original stationary signals by a constant scale factor, an inconsequential effect for correlation, since correlation is insensitive to scale. The efficacy of our correction was validated using simulation data; after applying our proposed correction the mean value of sample correlation coincided with the true correlation between the underlying stationary signals. Experimental data showed reduced outlier connectivity and the number of significant connections reduced in accordance with the smaller variance of stationary signals. Furthermore, the change in mean correlation agreed with that anticipated by the analytically derived expression for correlation between non-stationary signals.

4.6 Conclusion

Linear fMRI connectivity methods typically require stationarity of timeseries. We have demonstrated that resting state fMRI data do not meet this requirement but rather have time-varying slice-dependent variance that is well modelled by an inverse gamma distribution. The resulting voxel intensity distribution under this model is a generalisation of a Student's t distribution. The impact of non-stationary signal power on connectivity was established by analytically deriving an expression for correlation between non-stationary timeseries. Subsequently, a correction was proposed and validated using both empirical and experimental datasets.

Chapter 5

Filtering Induces Correlation in fMRI Resting State Data

5.1 Introduction

BOLD fMRI data can be employed in a fMRI connectivity analyses to gain insight into the functionality of neural networks in the brain. The BOLD signal is compromised by a myriad of noise sources including equipment-related noise, such as coil interference (Bandettini et al., 1998) and scanner drift (Bianciardi et al., 2009), and subject-initiated noise, such as physiological interference (Biswal et al., 1995) and subject movement (Barry et al., 2010). For BOLD signals obtained at 3T, the signal of interest comprises less than 10% of signal fluctuation, and is instead dominated by physiological noise (Kruger et al., 2001) and scanner drift (Bianciardi et al., 2009). This composition of the BOLD signal strongly motivates the use of preprocessing methods to suppress noise that, in the context of fMRI connectivity analyses, attempt to minimise artificial correlation emanating from noise processes such as synchronised cardiac and respiration (Birn et al., 2006). Temporal filtering is one such preprocessing method, aimed at removing noise concentrated at frequency bandwidths that do not coincide with signals of interest. Physiological noise derived from respiratory and cardiac function is concentrated at comparatively high frequencies (> 0.1Hz) (Cordes et al., 2001; Lowe et al., 1998; Thomas et al., 2002), while scanner drift is localised to frequencies below 0.01Hz (Bianciardi et al., 2009). Functional correlations typically subtend frequencies between 0.01-0.1Hz (Biswal et al., 1995; Cordes et al., 2001; Demirci et al., 2009; Salvador et al., 2005a). Filter specification and implementation employed in fMRI connectivity literature varies considerably, the most common filter specification being 0.01-0.08Hz (Auer, 2008; Cauda et al., 2011; Uddin et al., 2009; Zhong et al., 2009). For example, Lowe et al. (1998) and Hampson et al. (2002) apply a low-pass finite impulse response (FIR) filter to remove all frequencies above 0.08Hz. Liao et al. (2010) preprocess using a bandpass filter at 0.005-0.17Hz by forcing the appropriate frequency coefficients to zero, while Luca et al. (2006) employ a high-pass filter at 0.004Hz, implemented using a straight line fitting approach.

FMRI connectivity analyses frequently use seed-voxel correlation, in which sample correlation is calculated between a seed signal and all brain voxels (Friston, 1994; Biswal et al., 1995). The significance of a correlation estimate is typically determined via Fisher's z-transformation (Fisher, 1915) or Student's t-test (Anderson, 2003), generating a map of z-scores or t-scores, respectively. The distribution of sample correlation and, therefore, the distribution of associated significance test variates is conventionally derived using an assumption that voxel timeseries are comprised of temporally independent samples. The use of filtering in fMRI connectivity analyses violates the independence assumption by introducing autocorrelation to the timeseries and, as a consequence, statistical inference results may be invalidated. Moreover, these heterogeneous approaches to temporal filtering contain widely varying degrees of departure from the independence assumption. This has ramifications for the validity of inter-experimental comparison of results.

The variance of sample correlation proposed by Fisher (1915) is dependent on timeseries length, which dictates the degrees of freedom of the correlation estimate. By introducing autocorrelation, temporal filtering reduces the effective degrees of freedom (Kruggel et al., 2002). Friston et al. (1994) considered the problem of sample dependence in the context of fMRI activation analyses, demonstrating that the effective degrees of freedom contributing to sample correlation depends on the extent of autocorrelation in the constituent signals. Their proposed correction to z-score estimates, valid for uncorrelated signals, is calculated from the spectral densities of the constituent signals, necessitating an estimate for each signal pair. Bullmore et al. (2000) addressed the issue of sample dependence in correlation estimates by modelling autocorrelation as a first order autoregressive process and extracting the residual of the autoregression. The ratio of the residual variance to the original variance in the timeseries was then used to weight timeseries length to obtain an estimate of the effective degrees of freedom. The efficacy of this approach depends heavily on the model of autocorrelation in the timeseries.

In this chapter we derive an analytic expression for the distribution of correlation estimates acquired from temporally filtered data, from which the corrected degrees of freedom contributing to the estimate is determined. The distribution of sample correlation is expressed as a function of the complex frequency response of the filter. Crucially, the result is valid for any true correlation and arbitrary filter specifications. Furthermore, we present corrections to the significance tests of correlation, including Fisher's z-transformation and Student's t-test. We demonstrate that, in the absence of our correction, the distributions of the correlation test variates derived from filtered data do not coincide with expected distributions predicated on the independence assumption. As a consequence, intended confidence intervals are inadvertently rendered obsolete, with increased false positive rates and artificially induced correlation. The utility of our corrections for mitigating this effect and restoring confidence intervals is demonstrated using both empirical and experimental data. Our proposed corrections therefore support noise suppression via filtering, while ameliorating the problem of artificially induced connectivity. Importantly, restoring the integrity of confidence intervals supports meaningful inter-experimental comparison of results.

In this chapter analytic results are discussed in the context of resting state connectivity. However, as the corrections are valid for arbitrary true correlation, they are also applicable to task-activation connectivity analyses, which is the subject of current work. The results in this chapter also have implications for group level connectivity analyses given that the variance of group correlation estimates has contributions from both within-subject and between-subject variance (Penny et al., 2004c).

This chapter is organised as follows. In Section 2.2, the necessary theoretical background was was introduced. Section 5.2 contains our theoretical results, including the distribution of sample correlation between filtered signals, and distributions for z-score and t-score test variates derived from filtered signals. Theoretical results are verified both empirically and

experimentally, the methods of which are described in Section 5.3, with results presented in Section 5.4. Section 5.5 presents a discussion of these results.

5.2 Theoretical Results

We derive an analytic expression for the distribution of sample correlation between filtered signals, and show that the variance of sample correlation is a function of the complex frequency response of the filter. We subsequently derive corrected expressions for statistical inference tests of sample correlation, including both Fisher's z-transformation and the Student's t-test. In so doing, we demonstrate that temporal filtering induces artificial correlation by reducing the effective degrees of freedom of the estimate.

5.2.1 Filtering induces correlation

Temporal filtering employed as a preprocessing step prior to fMRI connectivity analysis typically applies an identical filter to all voxels; consequently, the distribution of sample correlation is first derived under the assumption that an identical filter is applied to each timeseries. It is assumed that the timeseries are white, Gaussian signals containing arbitrary contemporaneous correlation.

Initially suppose that x_t and y_t are identically filtered using a filter described by the complex frequency response f_k, for all frequencies k. Our first result is that the distribution of sample correlation between the filtered signals, $x_t^{(f)}$ and $y_t^{(f)}$, is

$$\operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right) \overset{a}{\sim} \mathcal{N}\left(\rho_{x,y}, \frac{\left(1-\rho_{x,y}^2\right)^2}{\kappa}\right), \tag{5.1}$$

where the corrected degrees of freedom between filtered signals, κ, is

$$\kappa \triangleq \frac{\left(\sum_k f_k^2\right)^2}{\sum_k f_k^4}. \tag{5.2}$$

An upper bound on the corrected degrees of freedom is

$$\kappa \leq T. \tag{5.3}$$

The upper bound for the corrected degrees of freedom is obtained if and only if power is constant across frequencies, implying that the signal is white and, therefore, samples are independent. Eq. (5.3) supports the intuitive interpretation of κ as a corrected degrees of

freedom. Thus, temporal filtering of timeseries may reduce the degrees of freedom contributing to a correlation estimate.

Proof of Eqs. (5.1)-(5.3). A necessary first step toward deriving the sample correlation between filtered signals is to determine expressions for the spectral constituents of both the unfiltered and filtered standardised timeseries.

Timeseries x_t and y_t can be written as the Fourier transform of iid Gaussian frequency components,

$$\begin{aligned} x_t &= \frac{1}{T}\sum_k X_k e^{2\pi i \frac{k}{T} t}, \\ y_t &= \frac{1}{T}\sum_k Y_k e^{2\pi i \frac{k}{T} t}, \end{aligned} \tag{5.4}$$

where each frequency component is distributed as (Fuller, 1996)

$$\begin{aligned} X_k &\sim \mathcal{N}\left(0, T\sigma_x^2\right), \\ Y_k &\sim \mathcal{N}\left(0, T\sigma_y^2\right). \end{aligned} \tag{5.5}$$

The temporal variates, Eq. (5.4), can be transformed to standard normal variates by

$$\begin{aligned} x_t^{(s)} &\triangleq \frac{x_t}{\sigma_x} \sim \mathcal{N}(0,1) \\ y_t^{(s)} &\triangleq \frac{y_t}{\sigma_y} \sim \mathcal{N}(0,1) \end{aligned} \tag{5.6}$$

so that, from Eq. (5.5),

$$\begin{aligned} X_k^{(s)} &\triangleq \frac{X_k}{\sigma_x} \sim \mathcal{N}(0,T) \\ Y_k^{(s)} &\triangleq \frac{Y_k}{\sigma_y} \sim \mathcal{N}(0,T). \end{aligned} \tag{5.7}$$

If a filter, described by complex frequency response f_k, at each frequency k, is applied to both x_t and y_t, the resulting filtered signals are

$$\begin{aligned} x_t^{(f)} &\triangleq \frac{1}{T}\sum_k X_k f_k e^{2\pi i \frac{k}{T} t} \\ y_t^{(f)} &\triangleq \frac{1}{T}\sum_k Y_k f_k e^{2\pi i \frac{k}{T} t}. \end{aligned} \tag{5.8}$$

Note that to preserve realness of the temporal signals, the imaginary frequency components of the filter are assumed to be conjugate symmetric around 0.

Employing the distribution of the spectral variates, Eq. (5.5), variance of the filtered signal sample at time t, $x_t^{(f)}$, can be expressed as

$$\sigma^2_{x^{(f)}} = \frac{1}{T^2}\sum_k \left(\sigma^2_{X_k^{(f)} e^{2\pi i \frac{k}{T} t}}\right), \tag{5.9}$$

where $\sigma^2_{X_k^{(f)} e^{2\pi i \frac{k}{T} t}}$ denotes the variance of the spectral variate at frequency k, weighted according to the inverse Fourier transform. Letting $\mathcal{R}$ and $\mathcal{I}$ denote real and imaginary components, respectively, Eq. (5.9) can be rearranged so that

$$\begin{aligned}
\sigma^2_{x^{(f)}} &= \frac{1}{T^2}\sum_k \left(\sigma^2_{\mathcal{R}\{X_k^{(f)}\}} + \sigma^2_{\mathcal{I}\{X_k^{(f)}\}}\right) \\
&= \frac{1}{T^2}\sum_k \left({f_k}^* f_k \sigma^2_{\mathcal{R}\{X_k\}} + {f_k}^* f_k \sigma^2_{\mathcal{I}\{X_k\}}\right) \\
&= \frac{1}{T^2}\sum_k \left(f_k^2 \sigma^2_{\mathcal{R}\{X_k\}} + f_k^2 \sigma^2_{\mathcal{I}\{X_k\}}\right),
\end{aligned} \tag{5.10}$$

since the conjugate symmetry of the signals requires the imaginary components to sum to zero. Furthermore, real and imaginary components of the spectral variates have equal variance (Millioz and Martin, 2010) so that

$$\begin{aligned}
\sigma^2_{x^{(f)}} &= \frac{1}{T^2}\sum_k \left(\frac{f_k^2}{2}\sigma^2_{X_k} + \frac{f_k^2}{2}\sigma^2_{X_k}\right) \\
&= \frac{\sigma_x^2}{T}\sum_k f_k^2,
\end{aligned} \tag{5.11}$$

Similarly, for $y_t^{(f)}$,

$$\sigma^2_{y^{(f)}} = \frac{\sigma_y^2}{T}\sum_k f_k^2. \tag{5.12}$$

Therefore, standardising the filtered timeseries results in

$$\begin{aligned}
x_t^{(f,s)} &= \frac{x_t^{(f)}}{\sigma_x \sqrt{\frac{\sum_k f_k^2}{T}}} \\
y_t^{(f,s)} &= \frac{y_t^{(f)}}{\sigma_y \sqrt{\frac{\sum_k f_k^2}{T}}},
\end{aligned} \tag{5.13}$$

which gives rise to the complementary spectral variates

$$\begin{aligned} X_k^{(f,s)} &= \frac{X_k f_k}{\sigma_x \sqrt{\frac{\sum_k f_k^2}{T}}} \\ Y_k^{(f,s)} &= \frac{Y_k f_k}{\sigma_y \sqrt{\frac{\sum_k f_k^2}{T}}}. \end{aligned} \tag{5.14}$$

For unfiltered bivariate normal random variables sample correlation, $\text{corr}\,(x_t, y_t)$, is calculated as

$$\begin{aligned} \text{corr}\,(x_t, y_t) &= \frac{\text{cov}\,(x_t, y_t)}{\text{std}\,(x_t)\,\text{std}\,(y_t)} \\ &= \frac{\sum_{t=1}^{T}(x_t - \hat{\mu}_{x_t})(y_t - \hat{\mu}_{y_t})}{\sqrt{\sum_{t=1}^{T}(x_t - \hat{\mu}_{x_t})^2 \sum_{t=1}^{T}(y_t - \hat{\mu}_{y_t})^2}}, \end{aligned} \tag{5.15}$$

where $\hat{\mu}_{x_t}$ and $\hat{\mu}_{y_t}$ denote the sample means of x_t and y_t, respectively. Sample correlation can be equivalently expressed using standardised variates by a simple rearrangement of Eq. (5.15) (Whittaker, 2009),

$$\begin{aligned} \text{corr}\,(x_t, y_t) &= \frac{1}{T}\sum_{t=1}^{T}\frac{(x_t - \hat{\mu}_{x_t})}{\text{std}\,(x_t)}\frac{(y_t - \hat{\mu}_{y_t})}{\text{std}\,(y_t)} \\ &= \frac{1}{T}\sum_{t=1}^{T} x_t^{(s)} y_t^{(s)}, \end{aligned} \tag{5.16}$$

where $x_t^{(s)}$ and $y_t^{(s)}$ denote the standard normal counterparts of x_t and y_t, respectively, such that $x_t^{(s)} = \frac{(x_t - \hat{\mu}_{x_t})}{\text{std}(x_t)}$, and $y_t^{(s)} = \frac{(y_t - \hat{\mu}_{y_t})}{\text{std}(y_t)}$. Correlation is equivalent to cross-correlation, Eq. (2.10), between standardised variates at lag zero,

$$\rho_{x,y} = \gamma_{x^{(s)}y^{(s)}}(0). \tag{5.17}$$

Employing the convolution theorem, the Fourier transform of cross-correlation satisfies (Kumar, 2009)

$$\Gamma_{x,y}(k) = \frac{1}{T} X_k^* Y_k, \qquad k = 1, \cdots, T, \tag{5.18}$$

where X_k and Y_k denote the values for frequency k, of the Fourier transforms of the vectors $\mathbf{x} = [x_1, \cdots, x_T]'$ and $\mathbf{y} = [y_1, \cdots, y_T]'$, respectively. Consequently, application of the

inverse Fourier transform requires that

$$\gamma_{x,y}(\tau) = \frac{1}{T^2} \sum_{k=0}^{T-1} X_k^* Y_k e^{2\pi i \frac{k}{T} \tau}. \tag{5.19}$$

To derive the distribution of sample correlation between filtered variates express sample correlation as lag-0 cross-correlation (Eqs. (5.17)-(5.19)) between Eq. (5.14), which gives

$$\begin{aligned}
\operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right) &= \frac{1}{T^2} \sum_k X_k^{(f,s)^*} Y_k^{(f,s)} e^{2\pi i \frac{k}{T} 0} \\
&= \frac{1}{T^2} \sum_k \left(\frac{(X_k f_k)^*}{\sigma_x \sqrt{\frac{\sum_k f_k^2}{T}}} \frac{Y_k f_k}{\sigma_y \sqrt{\frac{\sum_k f_k^2}{T}}} \right) \\
&= \frac{1}{T \sum_k f_k^2} \sum_k \left(\frac{X_k^*}{\sigma_x} \frac{Y_k}{\sigma_y} {f_k}^* f_k \right) \\
&= \frac{1}{T \sum_k f_k^2} \sum_k \left(\frac{X_k^*}{\sigma_x} \frac{Y_k}{\sigma_y} f_k^2 \right),
\end{aligned} \tag{5.20}$$

where the last line follows from the conjugate symmetry of the filtered signals.

Although the product terms, $X_k^* Y_k$, Eq. (5.20), have a Laplacian distribution (see Appendix 5.A), the central limit theorem dictates that the sum tends to a Gaussian distribution. To describe the distribution of sample correlation, we therefore only require the mean and variance as sufficient statistics of the Gaussian distribution.

Application of Eq. (5.7) to Eq. (5.20), in conjunction with the iid attribute of the unfiltered spectral components, yields the expected value of $\operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right)$:

$$\begin{aligned}
\mathrm{E}\left\{\operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right\} &= \frac{1}{T \sum_k f_k^2} \mathrm{E}\left\{X_k^{(s)} Y_k^{(s)}\right\} \sum_k T f_k^2 \\
&= \frac{\rho_{x,y}}{T \sum_k f_k^2} \sum_k T f_k^2 \\
&= \frac{\rho_{x,y}}{\sum_k f_k^2} \sum_k f_k^2 \\
&= \rho_{x,y}.
\end{aligned} \tag{5.21}$$

To determine the variance of sample correlation, rewriting Eq. (5.20) using the standardised spectral components of the serially independent signals, Eq. (5.7), and employing the

independence of spectral variates obtains

$$\sigma^2_{\mathrm{corr}\left(x_t^{(f)},y_t^{(f)}\right)} = \frac{1}{T^2\left(\sum_k f_k^2\right)^2}\sum_k \sigma^2_{X_k^{(s)*}Y_k^{(s)}} f_k^4. \tag{5.22}$$

It remains to determine $\sigma^2_{X_k^{(s)}Y_k^{(s)}}$. Considering the standardised form of sample correlation, Eq. (5.16), we can write,

$$\sigma^2_{\mathrm{corr}(x_t,y_t)} = \frac{1}{T^2}\sum_{t=1}^{T}\sigma^2_{x^{(s)}y^{(s)}}, \tag{5.23}$$

since the unfiltered signals are comprised of independent temporal samples. Rearranging Eq. (5.23), and employing the known variance of sample correlation between unfiltered signals, Eq. (2.15), gives

$$\sigma^2_{x^{(s)}y^{(s)}} = \left(1-\rho^2_{x,y}\right)^2, \tag{5.24}$$

which implies that

$$\sigma^2_{X_k^{(s)}Y_k^{(s)}} = T^2\left(1-\rho^2_{x,y}\right)^2. \tag{5.25}$$

Substitution of Eq. (5.25) in Eq. (5.22) yields the sample variance of correlation between filtered signals,

$$\begin{aligned}\sigma^2_{\mathrm{corr}\left(x_t^{(f)},y_t^{(f)}\right)} &= \frac{1}{T^2\left(\sum_k f_k^2\right)^2}\sum_k T^2\left(1-\rho^2_{x,y}\right)^2 f_k^4 \\ &= \frac{\sum_k f_k^4}{\left(\sum_k f_k^2\right)^2}\left(1-\rho^2_{x,y}\right)^2,\end{aligned} \tag{5.26}$$

and the proof for Eqs. (5.1)-(5.2) is complete.

It now remains to prove Eq. (5.3). Recall that the corrected degrees of freedom, Eq. (5.2), is given by

$$\kappa \triangleq \frac{\left(\sum_k f_k^2\right)^2}{\sum_k \left(f_k^2\right)^2}.$$

Denote the power at frequency k by $p_k = f_k^2$, so that

$$\kappa = \frac{\left(\sum_k p_k\right)^2}{\sum_k p_k^2}. \tag{5.27}$$

Application of Jensen's inequality with identically unitary weights for all frequencies requires that (Jensen, 1906)

$$\left(\frac{\sum_k p_k}{T}\right)^2 \leq \frac{\sum_k p_k^2}{T}. \tag{5.28}$$

Noting that p_k is necessarily positive, rearranging gives

$$\frac{\left(\sum_k p_k\right)^2}{\sum_k p_k^2} \leq T. \tag{5.29}$$

Substituting f_k^2 for p_k completes the proof of Eq. (5.3).

□

The result in Eq. (5.1) generalises easily to the case in which signals are filtered multiple times. The concept of multiple filters can also be employed to model sample dependence existing prior to temporal filtering. Where there are a total of n filters applied to each signal, and $f_{i,k}$ denotes the complex response of filter i at frequency k, the net complex frequency response of the filters can be determined by

$$f_k = \prod_{i=1}^{n} f_{i,k}. \tag{5.30}$$

The corrected degrees of freedom, Eq. (5.1), can then be determined as usual.

Proof. The proof follows directly from the definition of a temporally filtered signal, Eq. (5.8).

□

The distribution of sample correlation, Eq. (5.1), was derived assuming that signals were filtered using a common filter or common multiple filters. This assumption can be relaxed and the distribution of sample correlation determined for the case where signals are filtered using different filters. Let signals x_t and y_t be temporally filtered by filters specified by complex frequency responses $f_{x,k}$ and $f_{y,k}$, respectively, for all frequencies, k. Our result is

that the distribution of sample correlation between the filtered signals, $x_t^{(f_x)}$ and $y_t^{(f_y)}$, is

$$\operatorname{corr}\left(x_t^{(f_x)}, y_t^{(f_y)}\right) \overset{a}{\sim} \mathcal{N}\left(\rho_{x,y}\frac{\sum_k f_{x,k}f_{y,k}}{\sqrt{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}}, \left(1-\rho_{x,y}^2\right)^2 \frac{\sum_k f_{x,k}^2 f_{y,k}^2}{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}\right). \tag{5.31}$$

Note that filtering the timeseries with disparate filters modifies the expected sample correlation from the original true correlation, $\rho_{x,y}$.

Proof. Note that this derivation follows the same procedure as the derivation of Eq. (5.1) and is included for completeness.

Temporal filtering of white, Gaussian timeseries x_t and y_t by disparate filters respectively defined by complex frequency responses, $f_{x,k}$ and $f_{y,k}$, for all frequencies k, generates filtered signals $x_t^{(f_x)}$ and $y_t^{(f_y)}$, respectively. The filtered signals can be expressed as

$$\begin{aligned} x_t^{(f_x)} &= \frac{1}{T}\sum_k X_k f_{x,k} e^{2\pi i \frac{k}{T} t}, \\ y_t^{(f_y)} &= \frac{1}{T}\sum_k Y_k f_{y,k} e^{2\pi i \frac{k}{T} t}. \end{aligned} \tag{5.32}$$

Variance of the filtered signals can be determined by application of Eq. (5.11) to be

$$\begin{aligned} \sigma^2_{x^{(f_x)}} &= \frac{\sigma_x^2}{T}\sum_k f_{x,k}^2 \\ \sigma^2_{y^{(f_y)}} &= \frac{\sigma_y^2}{T}\sum_k f_{y,k}^2. \end{aligned} \tag{5.33}$$

Therefore, standardising the filtered time series, $x_t^{(f_x)}$ and $y_t^{(f_y)}$, requires

$$x_t^{(f_x s)} = \frac{x_t^{(f_x)}}{\sigma_x \sqrt{\frac{\sum_k f_{x,k}^2}{T}}} \tag{5.34}$$

$$y_t^{(f_y s)} = \frac{y_t^{(f_y)}}{\sigma_y \sqrt{\frac{\sum_k f_{y,k}^2}{T}}}, \tag{5.35}$$

which generates the complementary spectral variates

$$X_k^{(f_x s)} = \frac{X_k f_{x,k}}{\sigma_x \sqrt{\frac{\sum_k f_{x,k}^2}{T}}} \qquad (5.36)$$
$$Y_k^{(f_y s)} = \frac{Y_k f_{y,k}}{\sigma_y \sqrt{\frac{\sum_k f_{y,k}^2}{T}}}.$$

Expressing sample correlation as a special case of cross-correlation, Eq. (5.17), and deriving cross-correlation from its spectral components, Eq. (5.19), results in

$$\begin{aligned}
\mathrm{corr}\left(x_t^{(f_x)}, y_t^{(f_y)}\right) &= \frac{1}{T^2}\sum_{k=0}^{T-1} X_k^{(f_x s)^*} Y_k^{(f_y s)} \\
&= \frac{1}{T^2}\sum_k \left(\frac{{X_k}^* {f_{x,k}}^*}{\sigma_x \sqrt{\frac{\sum_k f_{x,k}^2}{T}}} \frac{Y_k f_{y,k}}{\sigma_y \sqrt{\frac{\sum_k f_{y,k}^2}{T}}} \right) \\
&= \frac{1}{T\sqrt{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}} \sum_k \left(\frac{X_k^*}{\sigma_x}\frac{Y_k}{\sigma_y} {f_{x,k}}^* f_{y,k} \right) \\
&= \frac{1}{T\sqrt{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}} \sum_k \left(\frac{X_k}{\sigma_x}\frac{Y_k}{\sigma_y} f_{x,k} f_{y,k} \right), \qquad (5.37)
\end{aligned}$$

where the last line results because the filtered signals are assumed to be real.

Application of the central limit theorem requires that the sum tends to a Gaussian distribution. Consequently, mean and variance are sufficient to describe the distribution of sample correlation between timeseries filtered with disparate filters.

The expected value of $\mathrm{corr}\left(x_t^{(f_x)}, y_t^{(f_y)}\right)$ can be derived from Eq. (5.37) by observing that the standardised spectral components of the unfiltered timeseries, Eq. (5.7), are iid;

$$\begin{aligned}
\mathrm{E}\left\{\mathrm{corr}\left(x_t^{(f_x)}, y_t^{(f_y)}\right)\right\} &= \frac{1}{T\sqrt{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}} \mathrm{E}\left\{X_k^{(s)} Y_k^{(s)}\right\} \sum_k \left(f_{x,k} f_{y,k}\right) \\
&= \frac{\rho_{x,y} T}{T\sqrt{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}} \sum_k \left(f_{x,k} f_{y,k}\right) \\
&= \rho_{x,y} \frac{\sum_k f_{x,k} f_{y,k}}{\sqrt{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}}. \qquad (5.38)
\end{aligned}$$

To determine the variance of sample correlation between disparately filtered signals rewrite Eq. (5.37) using the standardised spectral components of unfiltered signals $X_k^{(s)}$ and $Y_k^{(s)}$, Eq. (5.7), as

$$\sigma^2_{\mathrm{corr}\left(x_t^{(f_x)}, y_t^{(f)}\right)} = \frac{1}{T^2\sqrt{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2}} \sum_k \sigma^2_{X_k^{(s)*} Y_k^{(s)}} f_{x,k}^2 f_{y,k}^2. \tag{5.39}$$

Application of Eq. (5.25) then gives

$$\begin{aligned}\sigma^2_{\mathrm{corr}\left(x_t^{(f_x)}, y_t^{(f_y)}{}_{[y]}\right)} &= \frac{1}{T^2 \sum_k f_{x,k}^2 \sum_k f_{y,k}^2} \sum_k T^2 \left(1-\rho_{x,y}^2\right)^2 f_{x,k}^2 f_{y,k}^2 \\ &= \left(1-\rho_{x,y}^2\right)^2 \frac{\sum_k f_{x,k}^2 f_{y,k}^2}{\sum_k f_{x,k}^2 \sum_k f_{y,k}^2},\end{aligned} \tag{5.40}$$

and the proof is complete.

□

5.2.2 Correcting for induced correlation

Statistical inference testing of correlation estimates is crucial to identify significant functional connections. The distribution of a significant test variate for correlation depends on the distribution of sample correlation. The result in Eq. (5.1) can be utilised to correct statistical testing of correlation estimates between filtered signals. We derive corrected distributions for Fisher's z-transformation, $\mathrm{r}\left(\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right)$ and the Student's t-test variate, $\mathrm{t}\left(\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right)$.

Again assuming that x_t and y_t are identically filtered, using a filter described by the complex frequency response f_k, at frequency k, our result is that the distribution of Fisher's z-transformation of $\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)$ is

$$\mathrm{r}\left(\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right) \sim \mathcal{N}\left(\mathrm{r}\left(\rho_{x,y}\right), \frac{1}{\kappa}\right). \tag{5.41}$$

Thus, for identical filters the mean is unaffected but, as a consequence of the upper bound on κ, Eq. (5.3) temporal filtering of white signals necessarily increases the variance of sample correlation. The upper bound is obtained only in the case of white signals.

For an ideal filter defined by a low cut-off frequency, f_l, and a high cut-off frequency, f_h, we demonstrate that the distribution of Fisher's z-transformation of sample correlation reduces

to

$$\mathrm{r}\left(\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right) \overset{a}{\sim} \mathcal{N}\left(\mathrm{r}\left(\rho_{x,y}\right), \frac{1}{2T_s(f_h - f_l)T}\right), \tag{5.42}$$

where T_s denotes the sampling time for acquisition. The corrected degrees of freedom is, therefore,

$$\kappa = 2T_s(f_h - f_l)T. \tag{5.43}$$

Recall that for a null hypothesis of zero true correlation, sample estimates can be tested for significance using a Student's t-test. We show that sample correlation between filtered signals can be transformed to a Student's t-test variate by

$$\mathrm{t}\left(\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right) = \mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right) \sqrt{\frac{\kappa}{\left(1 - \mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)^2\right)}} \tag{5.44}$$

with distribution

$$\mathrm{t}\left(\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right) \sim \boldsymbol{t}\left(\frac{\left(\sum\limits_k |f_k|^2 \star |f_k|^2\right)^2}{2\sum\limits_k (|f_k|^2 \star |f_k|^2)^2}\right). \tag{5.45}$$

Proof of Eqs. (5.41)-(5.45). The variance of $\mathrm{r}\left(\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right)$ can be derived via the delta method (Oehlert, 1992), which states that if $\mathrm{corr} \sim \mathcal{N}\left(\rho, \sigma^2_{\mathrm{corr}}\right)$ then for some function $g(\mathrm{corr})$ with derivative defined at ρ, a first order approximation of its distribution is given by

$$g(\mathrm{corr}) \overset{D}{\to} \mathcal{N}\left(g(\rho), \sigma^2_{\mathrm{corr}} g'(\rho)^2\right), \tag{5.46}$$

where $\overset{D}{\to}$ denotes convergence in distribution.

Let

$$g(\mathrm{corr}) \triangleq \mathrm{r}\left(\mathrm{corr}\right) = \frac{1}{2}\ln\frac{1 + \mathrm{corr}}{1 - \mathrm{corr}} \tag{5.47}$$

so that

$$g'(\mathrm{corr}) = \frac{1}{1 - \mathrm{corr}^2}. \tag{5.48}$$

Applying Eq. (5.1) to Eq. (5.46) gives

$$\begin{aligned}\sigma^2_{\mathrm{r}\left(\mathrm{corr}\left(x_t^{(f)},y_t^{(f)}\right)\right)} &= \sigma^2_{\mathrm{corr}\left(x_t^{(f)},y_t^{(f)}\right)}\left(g'(\rho_{x^{(f)},y^{(f)}})\right)^2 \\ &= \frac{\left(1-\rho^2_{x^{(f)},y^{(f)}}\right)^2}{\kappa}\frac{1}{\left(1-\rho_{x^{(f)},y^{(f)}}\right)^2} \\ &= \frac{1}{\kappa}, \end{aligned} \tag{5.49}$$

proving the variance component. From Eq. (5.1), the mean of $\mathrm{corr}\left(x_t^{(f)},y_t^{(f)}\right)$ is simply $\rho_{x,y}$; transforming the mean using $g(\rho_{x,y})$, the result for Eq. (5.41) follows.

To prove Eqs. (5.42)-(5.43), define an ideal filter by a low cut-off frequency, f_l, and a high cut-off frequency, f_h. For a timeseries of length T denote the interval between discrete frequencies by (Atlas, 2009, p.54)

$$\Delta f = \frac{1}{T}, \tag{5.50}$$

so that $f_l = \Delta f k_l$ and $f_h = \Delta f k_h$, where k_l and k_h denote indices for the low and high cut-off frequencies, respectively.

Since the filter is ideal

$$f_k = \begin{cases} 0, & \text{for } k < f_l, f_h < k \\ 1, & \text{for } f_l \leq k \leq f_h. \end{cases} \tag{5.51}$$

so that the expression for corrected degrees of freedom between filtered signals, Eq. (5.2), reduces to

$$\begin{aligned}\kappa &= \frac{\left(2\left(k_h-k_l\right)\right)^2}{2\left(k_h-k_l\right)} \\ &= 2\left(k_h-k_l\right) \\ &= \frac{2\left(f_h-f_l\right)}{\Delta f}. \end{aligned} \tag{5.52}$$

The sampling interval can be expressed as (Westbrook et al., 2005, pp.77-81)

$$T_s = \frac{1}{T\Delta f}. \tag{5.53}$$

Substituting this into Eq. (5.52) gives

$$\kappa = 2T_s T\left(f_h - f_l\right), \tag{5.54}$$

and the proof of Eqs. (5.42)-(5.43) is complete.

It now remains to prove Eqs. (5.44)-(5.45). As the t-test is only valid when testing the null case of zero correlation, it suffices for this proof to consider only the case in which $x_t^{(f)}$ and $y_t^{(f)}$ are independent.

To aid readability in deriving the transformation of sample correlation to a Student's t statistic, define

$$w_t = x_t^{(f,s)} y_t^{(f,s)} \tag{5.55}$$

with sample mean

$$\begin{aligned} \hat{\mu}_w &= \frac{1}{T}\sum_{t=1}^{T} w_t \\ &= \frac{1}{T}\sum_{t=1}^{T} x_t^{(f,s)} y_t^{(f,s)}, \end{aligned} \tag{5.56}$$

which is equivalent to sample correlation between filtered signals, $\text{corr}\left(x_t^{(f)}, y_t^{(f)}\right)$, Eq. (5.16).

The Student's t-test statistic is acquired from the ratio of a standard normal variate to an independent χ variate. The t-test for sample correlation obtains these variables from descriptive statistics of w_t.

The standard normal variate is derived by standardising the sample mean, $\hat{\mu}_w$. From Eq. (5.1), μ_w (i.e. $\text{corr}\left(x_t^{(f)}, y_t^{(f)}\right)$) has an asymptotically normal distribution with mean zero, and can be standardised by,

$$\begin{aligned} \hat{\mu}_w^{(s)} &= \frac{\hat{\mu}_w}{\sigma_{\mu_w}} \\ &= \hat{\mu}_w \sqrt{\kappa}. \end{aligned} \tag{5.57}$$

To construct the χ variate, consider the sample variance of w_t,

$$\begin{aligned} \text{var}\,(w) &= \frac{1}{T-1}\sum_{t=1}^{T}\left(w_t^2 - \hat{\mu}_w^2\right) \\ &= \frac{T}{T-1}\left(1 - \hat{\mu}_w^2\right), \end{aligned} \tag{5.58}$$

where the last line is a consequence of Eq. (5.55), which describes w_t as the product of independent, standard normal variates, so that $E\{w\} = 0$ and $\sigma_w^2 = 1$. Application of

Parseval's theorem (Kammler, 2000) under the null hypothesis requires that

$$\begin{aligned}\sum_{t=1}^{T} w_t^2 &= \frac{1}{T}\sum_k |W_k|^2 \\ \mathrm{var}\,(w) &= \frac{1}{T}\mathrm{var}\,(W)\,,\end{aligned} \tag{5.59}$$

which permits the use of the sample variance of W_k to determine the signal variance for w_t. Since w_t is derived from the product of temporal variables, $x_t^{(f,s)}$ and $y_t^{(f,s)}$, Eq. (5.55), the convolution theorem dictates that W_k is determined by the convolution of the frequency domain counterparts, $X_k^{(f,s)}$ and $Y_k^{(f,s)}$, Eq. (5.14), so that

$$\begin{aligned}W_k &= X_k^{(f,s)} \star Y_k^{(f,s)} \\ &= \frac{T}{T\sigma_x\sigma_y \sum_k f_k^2}\left(f_k X_k \star f_k Y_k\right) \\ &= \frac{1}{\sum_k f_k^2}\sum_{j=1}^{N} f_j f_{k-j} X_k^{(s)} Y_{k-j}^{(s)}.\end{aligned} \tag{5.60}$$

Although each term in the sum has a Laplacian distribution (see Appendix 5.A), application of the central limit theorem results in an asymptotic normal distribution for W_k, with zero mean and variance,

$$\begin{aligned}\mathrm{var}\,(W) &= \frac{1}{\sum_k f_k^2}\sum_{j=1}^{N} |f_j|^2 |f_{k-j}|^2 \sigma^2_{X_j^{(s)}} \sigma^2_{Y_{k-j}^{(s)}} \\ &= \frac{T^2}{\left(\sum_k f_k^2\right)^2} |f_k|^2 \star |f_k|^2.\end{aligned} \tag{5.61}$$

Consequently, the distribution of the variance of W_k and hence w_t (Eq. (5.59)) can be considered a weighted sum of χ^2 variates, which can be approximated by a χ^2 distribution with effective degrees of freedom determined by the Welch-Satterthwaite's equation

(Satterthwaite, 1946; Welch, 1947),

$$\begin{aligned} d &= \frac{\left(\sum\limits_k |W_k|^2\right)^2}{\sum\limits_k |W_k|^4} \\ &= \frac{\left(\frac{T^2}{\left(\sum\limits_k f_k^2\right)^2} \sum\limits_k |f_k|^2 \star |f_k|^2\right)^2}{2\sum\limits_k \frac{T^4}{\left(\sum\limits_k f_k^2\right)^4} (|f_k|^2 \star |f_k|^2)^2} \\ &= \frac{\left(\sum\limits_k |f_k|^2 \star |f_k|^2\right)^2}{2\sum\limits_k \left(|f_k|^2 \star |f_k|^2\right)^2}, \end{aligned} \tag{5.62}$$

where the factor of $\frac{1}{2}$ is a consequence of d being an even function.

The standard normal and χ variates from which the Student's t-test statistic is constructed are required to be independent. Given that the sample mean and sample variance of a Gaussian variable are independent[1] (Whittaker, 2009), a ratio of the sample mean, $\hat{\mu}_w$, and sample variance, $\mathrm{var}\,(w)$, both scaled by σ_w, satisfies the independence requirement. To this end define

$$\begin{aligned} v &\triangleq (T-1)\frac{\mathrm{var}\,(w)}{\sigma_w^2} \\ &= T\left(1-\hat{\mu}_w^2\right), \end{aligned} \tag{5.63}$$

since $\sigma_w^2 = 1$.

The Student's t variate can now be constructed from the ratio of Eqs. (5.57)-(5.63),

$$\frac{\hat{\mu}_w^{(s)}}{\sqrt{v}} \sim \boldsymbol{t}\,(d)\,. \tag{5.64}$$

Substituting $\mathrm{corr}\left(x_t^{(f)}, y_t^{(f)}\right)$ for $\hat{\mu}_w$, and employing the expression derived for the degrees

[1] The sample mean and sample variance of a Gaussian, $z_t \sim \mathcal{N}\left(\mu_z, \sigma_z^2\right)$, are independent since $E\{\mu_z(z_t - \mu_z)\} = 0$.

of freedom Eq. (5.62) results in

$$t\left(\operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right) = \operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\sqrt{\frac{\kappa}{1-\operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right)^2}}$$

$$t\left(\operatorname{corr}\left(x_t^{(f)}, y_t^{(f)}\right)\right) \sim \boldsymbol{t}\left(\frac{\left(\sum_k |f_k|^2 \star |f_k|^2\right)^2}{2\sum_k \left(|f_k|^2 \star |f_k|^2\right)^2}\right). \tag{5.65}$$

This completes the proof for Eqs. (5.44)-(5.45). □

5.3 Methods

5.3.1 Simulated data

MATLAB was used to generate simulation datasets, each containing 2000 pairs of timeseries from a bivariate normal distribution, characterised by timeseries length $T \in [500, 10000]$, variance $\sigma^2 \in [1, 100]$, and correlation $\rho \in (-1, 1)$. Correlation induced between signals was independent of frequency and a sampling frequency of $T_s = 1$ second was used.

A set of filters, containing both realistic and unrealistic filter designs, was constructed to test the effect of different power distributions amongst signal frequencies on κ. FIR filters were included in the filter set as realistic filters with linear phase response. AR filters were incorporated as fMRI signals are known to contain autocorrelation. Step and ramp filters were impractical filter designs that were included to examine the generality of results. The complex response vector of each filter was scaled by $s \in \{\frac{1}{2}, 1, 2, 4, 8\}$ to determine the impact of modifying signal power and, in the case of multiple filters being applied, the net effect of multiple filters with differing scale factors.

FIR filters were parameterised by low cut-off frequency $f_l \in [0, 0.5)$, high cut-off frequency $f_h \in (0, 0.5]$, such that $f_l < f_h$ and scale of amplitude response, s. AR filters were generated using a random lag order $p \in \{0, 10\}$, and a randomised coefficient vector $\mathbf{b}$, tested for stability. Step filters were constructed by setting the frequency response between f_l and f_h to three different amplitudes, $\frac{s}{3}, \frac{s}{2}, s$, at equal divisions between f_l and f_h. The amplitude response of ramp filters increased linearly between f_l and f_h from 0 to s. An example of each filter type is shown in Fig. 5.1. For each filter specification, a frequency response vector was generated, $\mathbf{f} = [f_0, \cdots, f_T]'$, from which the corrected degrees of freedom, κ, was calculated.

To determine the impact of filtering on both white and non-white signals, and to verify the result for multiple filters, Eq. (5.30), random combinations of up to three filters were applied to datasets. Where a single filter was employed, it was applied to each timeseries

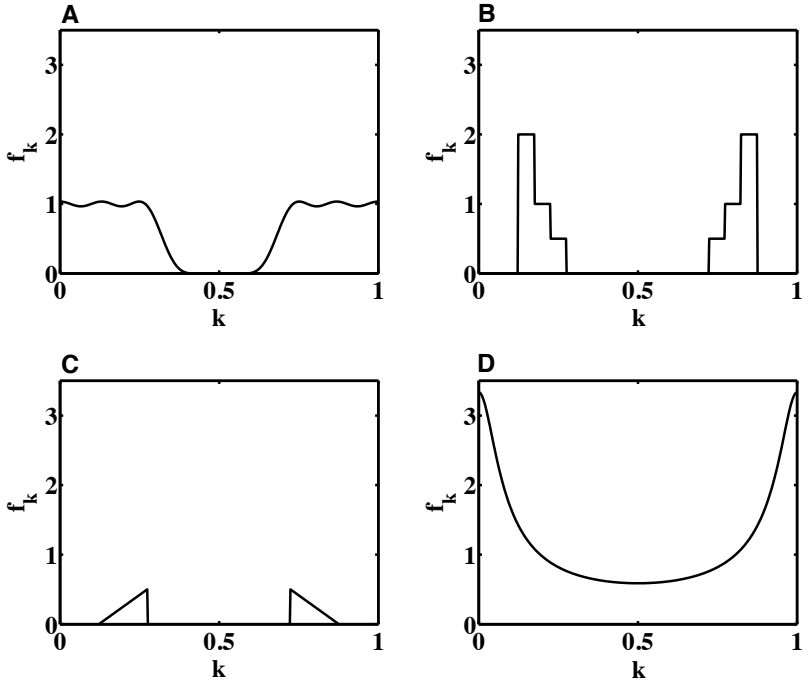

Figure 5.1: Amplitude response of example filters from each filter class. **A:** FIR filter with cut-off frequencies $f_l = 0$, $f_h = 0.25$ and scaling parameter $s = 1$. **B:** Step filter with cut-off frequencies $f_l = 0.125$, $f_h = 0.275$ and scale parameter $s = 2$. **C:** Ramp filter with cut-off frequencies $f_l = 0.125$, $f_h = 0.275$ and scale parameter $s = 0.5$. **D:** AR(1) filter with coefficient of 0.9.

after inducing correlation between signal pairs. For the case of two filters, one was applied prior to correlation being induced between timeseries pairs, and the other was applied after correlation induction. Finally, in the case of three filters, one was applied prior to correlation being induced between timeseries and the other two were applied post correlation induction. The analytic expression for the corrected variance of sample correlation, Eq. (5.1), was evaluated using the net complex response of the multiple filters according to Eq. (5.30), where necessary. Sample correlation was calculated for each signal pair in the dataset and the variance across the sample correlation estimates determined. The empirical and analytic estimates for sample correlation were plotted against a sum of the filter's normalised power spectral density.

Corrected distributions for the sample correlation test statistics were verified by generating empirical distributions of the test statistics with and without the filtering correction. For each simulation dataset sample correlation between timeseries pairs was recorded both before and after filtering with random combinations of up to three filters. Fisher's z-transformation was applied and the resulting z-scores standardised to enable comparison. The z-scores generated from unfiltered data were standardised using the distribution predicated on independent samples, Eq. (2.17), while z-scores generated from filtered data were standardised using the uncorrected distribution and the corrected distribution that accounts for filtering, Eq. (5.41). A histogram was generated for each case. The same process was employed to generate empirical distributions for the t-test using the uncorrected test statistic, Eq. (2.18), and the test variate adjusted for filtering, Eq. (5.44), using the corrected distribution, Eq. (5.45).

5.3.2 Experimental data

The resting state BOLD dataset introduced in Section 3.2 was employed. All images were motion corrected and smoothed using an isotropic Gaussian 6x6x6mm^3 kernel, applied using MATLAB . A conventional general linear model (GLM) activation analysis (Friston et al., 1994) was applied to the motor task data to identify the LMC. An average hemodynamic timeseries was created for the LMC of each subject by averaging across voxels in the region of interest (ROI).

For each subject, seed-voxel correlation maps were calculated using the averaged LMC as the seed. A low-pass FIR filter with cut-off, $f_h = 0.1$Hz, was then applied, and the correlation maps re-calculated. For the unfiltered data, voxels significantly connected to the LMC were identified by thresholding the z-scores based on the standard distribution derived for independent samples, Eq. (2.17), with $\alpha < 0.01$ (Bonferroni corrected). Z-scores derived from filtered data were thresholded using both the uncorrected distribution, and the distribution corrected for filtering, Eq. (5.41), also with $\alpha < 0.01$ (Bonferroni corrected).

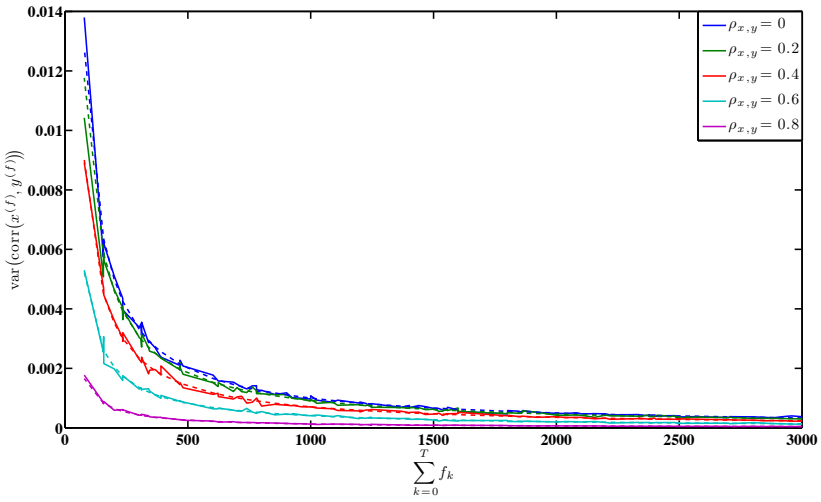

Figure 5.2: Variance of sample correlation estimated from simulations (solid lines), and predicted using the filter's complex response (dashed lines), for various true correlation values.

5.4 Results

The theoretical results presented in Section 5.2 established an analytic expressions for the distribution of sample correlation between filtered timeseries based on the complex frequency responses of the filters. Analytic expressions were also provided for the test variates of sample correlation of filtered timeseries, using both Fisher's z-transformation and Student's t-test. Empirical results validating the corrections are now presented, with experimental results demonstrating the significance of the corrections in connectivity analyses.

5.4.1 Empirical results

The variance of sample correlation was recorded after applying each filter type to each simulation dataset. Additionally, for each simulation, the variance of sample correlation predicted by the analytic correction proposed in Eq. (5.1) was recorded. The simulation results (Fig. 5.2) demonstrate that the sample variance of correlation between filtered signals is accurately predicted by the analytic expression in Eq. (5.1), derived from the filter's complex amplitude response. Furthermore, the correction is clearly valid for all true correlation values. The figure also indicates that the variance of sample correlation becomes arbitrarily large as the remaining filter power becomes infinitesimally small.

Histograms of correlation test variates derived from both unfiltered and filtered Gaussian noise were generated to empirically compare the distributions of samples. Fig. 5.3 shows an example result generated using a true correlation of $\rho_{x,y} = 0.3$. All true correlation

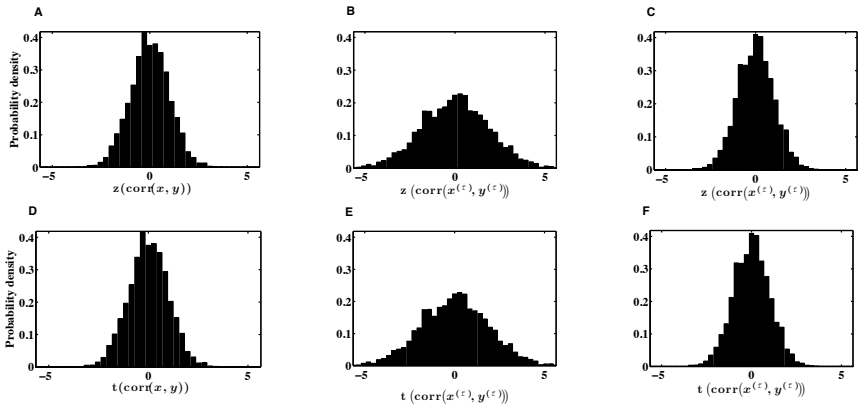

Figure 5.3: Empirical distributions of correlation test variates derived from Fisher's z-transformation, $\mathrm{z}\left(\mathrm{corr}\left(x_t, y_t\right)\right)$, and Student's t-test, $\mathrm{t}\left(\mathrm{corr}\left(x_t, y_t\right)\right)$, sampled from both unfiltered and filtered simulation data with; true correlation, $\rho_{x,y} = 0.3$; $T = 1000$; $\sigma^2 = 1$. A low-pass FIR filter with a cut-off frequency of 0.08Hz was used for filtering. **A:** Fisher's z-scores for unfiltered data. **B:** Uncorrected z-scores for filtered data. **C:** Corrected z-scores for filtered data. **D:** t-scores derived from the t-test for unfiltered data. **E:** Uncorrected t-scores for filtered data. **F:** Corrected t-scores for filtered data.

values generated histograms centred at 0 (data not shown) since z-scores and t-scores were standardised, and the expected value of sample correlation is unchanged when filtering signals with a common filter, Eq. (5.1). Standardised and corrected test variates derived from filtered timeseries have the same distribution as standardised test variates acquired from unfiltered timeseries, and conform with the standard normal distribution, indicating that the samples were standardised using the true distributions of the test variates. The uncorrected empirical distribution derived from filtered datasets show clearly the impact of temporal filtering; the measures are characterised by visibly increased variance and no longer conform to the standard normal distribution from which they were drawn, so that thresholding to identify significant sample correlation values no longer yields meaningful results. An intended z-score confidence interval of 95% is reduced to 60% if uncorrected and predicated on the violated assumption of sample independence. Conversely, thresholding the corrected measures derived from filtered data successfully restores the intended confidence interval to 95%.

5.4.2 Experimental results

To examine the impact of our correction in the context of fMRI connectivity seed-voxel correlation maps were computed for two subjects using both unfiltered and filtered resting state BOLD data (Fig. 5.4). Thresholded maps generated from unfiltered data and significance tested using z-scores (Fig. 5.4A,G) and t-scores (Fig. 5.4D,J) show marked variation both across inference tests and between subjects. Connectivity maps generated from filtered data

and significance tested using uncorrected measures show an inordinate proportion of voxels connected to the LMC (Fig. 5.4B,E,H,K), and exhibit substantial differences between subjects but improved consistency across significance test. Connectivity maps generated from filtered data and significance tested using our corrected expressions contain a feasible proportion of significant voxels connected to the LMC seed, with connectivity in conformance with the literature (Fig. 5.4C,F,I,L) (Vollmar et al., 2011; Xiong et al., 1999; Park et al., 2011), and exhibit improved consistency in connectivity both across the inference tests and between subjects.

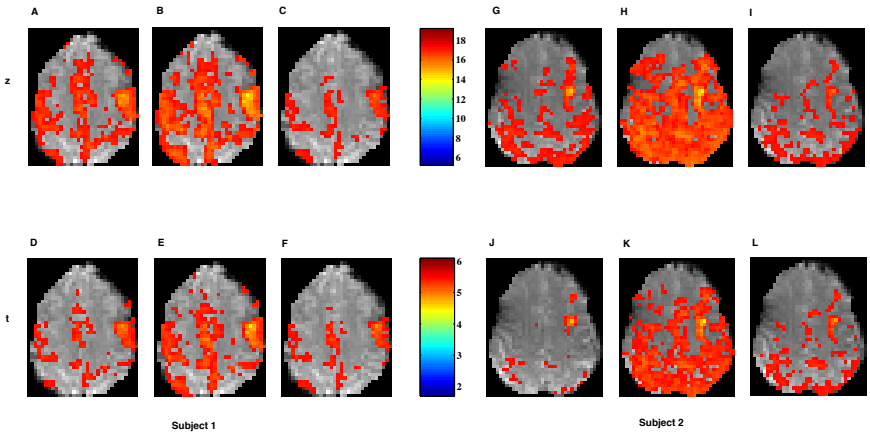

Figure 5.4: Seed-voxel correlation maps derived from sample correlation values for unfiltered, and filtered, data, significance tested using Fisher's z-transformation, $\mathrm{r}\left(\mathrm{corr}\left(x_t, y_t\right)\right)$, and the t-test, $\mathrm{t}\left(\mathrm{corr}\left(x_t, y_t\right)\right)$, thresholded at $p < 0.05$ (corrected). An FIR filter with a cut-off frequency of 0.1Hz was used for filtering. Subject 1: panels A-F. Subject 2: panels G-L. **A, G:** Fisher's z-scores for unfiltered data. **B, H:** Uncorrected z-scores for filtered data. **C, I:** Corrected z-scores for filtered data. **D, J:** t-scores derived from the t-test applied to unfiltered data. **E, K:** Uncorrected t-scores for filtered data. **F, L:** Corrected t-scores for filtered data.

5.5 Discussion

BOLD data is notoriously noisy, promoting the emergence of preprocessing techniques in fMRI connectivity analyses. Temporal filtering is a preprocessing technique that aims to reduce power at frequencies containing predominantly noise, thereby improving SNR. However, a secondary effect of temporal filtering, crucial in the context of fMRI connectivity, is its impact on the distribution of sample correlation. The expression for the variance of sample correlation, proposed by Fisher (1915) is founded on an assumption of sample independence and, necessarily, an absence of autocorrelation in the constituent signals. The

process of filtering violates this assumption by introducing autocorrelation to the signals. We have demonstrated that temporal filtering reduces the degrees of freedom contributing to correlation estimates, thus increasing the associated variance. We have addressed this issue by deriving an analytic expression for the post filtering distribution of sample correlation as a function of the complex frequency response of the filter.

As the distribution of sample correlation governs the distribution of its test variates, z-scores and t-scores are also impacted by temporal filtering. We have derived corrected distributions for both z-score and t-score estimates as a function of the filter frequency response, thereby restoring integrity to statistical inference results acquired from filtered datasets. If the filter-induced change in distribution is not accounted for, the consequent increase in variance will inflate the false positive rate, artificially inducing correlation.

As the extent of filtering and hence departure from sample independence increases, the expression for the variance of sample correlation proposed by Fisher (1915) becomes increasingly inaccurate. BOLD data are characterised by a narrow frequency band containing signal of interest that renders fMRI correlation estimates vulnerable to significant deviation from the expected distribution. Furthermore, the absence of a filtering standard, in conjunction with diverse frequency spectra induced by differing sample rates, impedes comparison of results across different experimental configurations. Use of our corrections for significance testing ensures confidence intervals are correct and enables comparison of results across heterogeneous experimental conditions.

The importance of our corrections is exemplified in application to experimental resting state data, described in Section 5.4. Inconsistency of the correlation maps between the unfiltered test statistics suggests that noise is corrupting results, advocating the need for temporal filtering to suppress noise. However, while the uncorrected test statistics acquired from filtered data show improved consistency across the test types, the improbable proportion of significant voxels is not reflective of expected results, but instead suggests severely compromised confidence intervals (Biswal et al., 1995; Xiong et al., 1999). The correlation maps derived from the corrected test statistics show improved consistency across both inference tests and subjects, and conform to established expectations (Wu et al., 2011; Vollmar et al., 2011).

The composition of BOLD signals is known to be dominated by noise, components of which are isolated in specific frequency bands. Consequently, BOLD signals may not exhibit white frequency spectra prior to temporal filtering. In this case the frequency spectra can be considered a naturally occurring filter. Eq. (5.30) addresses this eventuality by stating that, in the case of multiple filters, the distribution of sample correlation is determined by the net filter gain. This remark is verified empirically in Section 5.4, which further demonstrates that the correction remains valid if a filter is applied prior to correlation being induced. The implications of this result are two-fold. First, filters can be employed to model temporal

structure in the original, unprocessed data, allowing for relaxation of the independence assumption for BOLD signals. Secondly, the corrections are valid both when decimated frequencies contain only noise, and when filtering extracts signal of interest.

Friston et al. (1994) recognised the significance of autocorrelation in calculation of a corrected degrees of freedom for sample correlation. However, they estimate the variance of sample lag-0 correlation across multiple instances using the variance of cross-correlation across multiple lags. This is only valid for signals containing zero correlation at all lags. Indeed, Fisher (1915) showed that the variance of sample correlation should diminish with increasing strength of correlation, while the variance of the cross-correlation function actually increases with true correlation. This difference renders the correction proposed by Friston et al. (1994) incompatible with Fisher's z-transformation, which is applied to sample correlation to generate a test variate independent of true correlation.

The complex responses of a finite impulse response filter and an infinite impulse response filter can be approximated by an ideal filter response, described by its low and high cut-off frequencies. In this case, the corrected degrees of freedom can be determined simply from the filter cut-off frequencies, and the variance of Fisher's z-score reduces to the simple expression described in Eq. (5.42). This result is particularly useful in the context of fMRI connectivity, where the use of such filters is prevalent.

The impact of temporal filtering on the distribution of sample correlation estimates has implications for group analyses also, given that total variance is a function of within-subject variance (Penny et al., 2004c). An in depth examination of the impact on group analyses is beyond the scope of this thesis and is the subject of current work.

5.6 Conclusion

In this chapter we have demonstrated that the fMRI preprocessing step of temporal filtering increases the sample variance of correlation, thereby inflating the false positive rate and artificially inducing connectivity. We have derived an analytic expression for the variance of sample correlation as a function of the filter frequency response. Furthermore, we have established corrected expressions for the z-score and t-score test variates of sample correlation, restoring confidence intervals and integrity of results. We advocate employing the proposed correction to ameliorate the problem of artificially induced connectivity while retaining the improved noise suppression acquired from temporal filtering.

5.A Product of complex Gaussian variates

The product of two Gaussian random variables results in a product-normal variate. Assuming we take the product of $x \sim p_{\mathcal{N}}\left(x\,;\,0,\sigma_x^2\right)$ and $y \sim p_{\mathcal{N}}\left(y\,;\,0,\sigma_y^2\right)\big)$, the resulting product normal variate, $xy \sim p_{\mathcal{PN}}\left(xy\,;\,\sigma_x\sigma_y\right)$, has the characteristic function $\frac{1}{(1+\sigma_x^2\sigma_y^2)^{\frac{1}{2}}}$ (Pearson et al., 1929). The sum of two identical product normal variates has the characteristic function

$$\psi = \frac{1}{(1+\sigma_x^2\sigma_y^2t^2)^{\frac{1}{2}}}\frac{1}{(1+\sigma_x^2\sigma_y^2t^2)^{\frac{1}{2}}} = \frac{1}{(1+\sigma_x^2\sigma_y^2t^2)} \sim p_{\mathcal{L}apl}\left(xy\,;\,0,\sigma_x\sigma_y\right). \tag{5.66}$$

Thus, the sum of two identical product normal variables generates a Laplacian variable.

Chapter 6

The Equivalence of Linear Gaussian Connectivity Techniques

6.1 Introduction

FMRI connectivity analysis techniques have traditionally been categorised into data-driven methods and model-based methods, termed functional and effective connectivity, respectively, by Friston et al. (1993). Data-driven connectivity techniques most commonly extract information from whole-brain correlation or covariance maps (Friston, 1994; Rogers et al., 2007) and therefore do not incorporate a notion of directional influence, purportedly since correlation is a scalar measure of total linear connectivity between two voxels (Friston et al., 1993; Sato et al., 2006b; Rogers et al., 2007); correlative connectivity measures are described as "ambiguous with respect to underlying directed interactions" (Roebroeck et al., 2005, pg. 230). In contrast, model-based methods, such as structural equation modelling (SEM) (McIntosh and Gonzalez-Lima, 1994; Gonzalez-Lima and McIntosh, 1994) and dynamic causal modelling (DCM) (Friston et al., 2003b), employ some form of *a priori* knowledge to examine selected networks in the brain, based on a parametric model of the relationships between network nodes (Friston, 1994).

Marrelec et al. (2005) examined the relationship between data-driven and model-based connectivity techniques and demonstrated that, while unmediated correlation fails to capture the complexity of interactions modelled in SEM, conditional correlation that mediates relationships between network nodes is sufficient to express the relationships that arise in a linear Gaussian SEM analysis. Their work provided the underpinnings of a unified view of connectivity analysis techniques that transcends the functional/effective dichotomy. However, it did not consider measures describing lagged influence, such as Granger causality.

A concurrent stream of connectivity analysis techniques arose in the medical imaging literature based on the concepts of Granger causality. Granger (1969) defined the axiom of causality, that cause must precede effect, in the context of time series by asserting that a time series y causally influences x if one is better able to predict x knowing the past of y. Granger causality, also termed Granger-Wiener causality with appropriate deference to the seminal work of Wiener (1956), refers to the quantification of causality in stochastic time series through analysis of the variance of the residual processes. Geweke (1982) extended Granger's application of causality to autoregressive models by the introduction of both a directed and an instantaneous Granger causality measure, which together form a measure of total linear dependence between two time series.

The application of Geweke's Granger causality (Geweke's Granger causality (GGC)) measures to fMRI data seemingly enlarged the pool of fMRI connectivity analysis techniques, "promising further insights" (Zhou et al., 2009a, pg. 2198) into neural connectivity by providing methods that were simultaneously functional, taking a whole-brain approach, and effective, by their dependence on a parametric model (Goebel et al., 2003; Roebroeck et al., 2005; Abler et al., 2005). While linear Granger causality measures are known to be limited in their ability to infer causality from fMRI timeseries (Smith et al., 2011), their uptake in

fMRI analyses has been rapid. The breadth of GGC techniques to-date spans conditional GGC measures (Chen et al., 2006; Zhou et al., 2009a), partial GGC measures (Guo et al., 2008), frequency domain implementations (Chen et al., 2006; Zhou et al., 2009a; Demirci et al., 2009; Deshpande et al., 2009; Sato et al., 2009), and Bayesian (Valdes-Sosa, 2004), time-varying (Sato et al., 2006a) and nonlinear (Li et al., 2010) methods.

While similarities between GGC and correlation-based connectivity methods have been acknowledged, e.g. Rogers et al. (2010, pg. 1062) note that "partial Granger causality is based on the concept of partial covariance...", the precise nature of the relationship between the two has remained oblique. The primary objective of this chapter is to ascertain whether connectivity identified by GGC measures can be equally assessed using lagged or partial correlation measures. We investigate the relationship between GGC and correlation-based connectivity methods through an examination of how GGC measures can be exactly expressed as remappings of correlation-based quantities. Furthermore, we show that the statistical inference tests applied to GGC and correlation-based connectivity are equivalent. Our aim was to develop a clear understanding of the theoretical basis of linear connectivity analysis methods for human brain mapping researchers, in particular, to establish the place of Granger causality as a surrogate for correlation-based techniques.

The chapter is organised as follows. The main contributions of the chapter are described in Section 6.2; that GGC measures can be expressed as nonlinear functions of correlation-based quantities, and that the statistical inference tests of GGC measures are equivalent to those applied to their correlation-based counterparts. The theoretical results are then verified empirically (Section 6.3 and Section 6.4). The chapter then concludes with a discussion (Section 6.5) regarding implications of the equivalence mappings between GGC measures and correlation-based measures for fMRI connectivity.

6.2 Theoretical Results

We investigate dependencies between correlation-based measures and GGC measures in the following sections, and present the primary contribution of this chapter; that all GGC measures are remappings of traditional correlation-based connectivity measures. We subsequently show that GGC inference tests provide the same information as inference tests of their counterpart correlation-based measures.

6.2.1 GGC Measures as Functions of Correlation

The following five statements of equivalence demonstrate that all GGC measures are

Equivalence 1. *Instantaneous Granger causality is a remapping of partial correlation:*

$$F_{x\cdot y} = -\ln\left(1 - \operatorname{corr}\left(x_t, y_t|\mathbf{x}_{t-1}^{(p)}, \mathbf{y}_{t-1}^{(p)}\right)^2\right).$$

Proof. Rearranging instantaneous Granger causality (IGC), Eq. (2.38), gives

$$\begin{aligned} F_{x\cdot y} &= \ln\frac{\operatorname{var}\left(e_{t,V}\right)\operatorname{var}\left(u_{t,V}\right)}{\operatorname{var}\left(e_{t,V}\right)\operatorname{var}\left(u_{t,V}\right) - \operatorname{cov}\left(e_{t,V}, u_{t,V}\right)^2} \\ &= -\ln\left(1 - \frac{\operatorname{cov}\left(e_{t,V}, u_{t,V}\right)^2}{\operatorname{var}\left(e_{t,V}\right)\operatorname{var}\left(u_{t,V}\right)}\right). \end{aligned} \tag{6.1}$$

Using the definition of partial variance, Eq. (2.5), and employing partial correlation, Eq. (2.6), the equivalence follows. □

It follows trivially from the first statement of equivalence that for order-zero VAR models, instantaneous Granger causality is a remapping of correlation,

$$F_{x\cdot y} = -\ln(1 - \operatorname{corr}\left(x_t, y_t\right)^2).$$

Equivalence 2. *Directed Granger causality is a remapping of partial variance and the multiple correlation coefficient:*

$$\begin{aligned} F_{y\to x} &= \ln\frac{\operatorname{var}\left(x_t|\mathbf{x}_{t-1}^{(p)}\right)}{\operatorname{var}\left(x_t|\mathbf{x}_{t-1}^{(p)}, \mathbf{y}_{t-1}^{(p)}\right)} \\ &= \ln\frac{1 - \operatorname{mcorr}\left(x_t, \mathbf{x}_{t-1}^{(p)}\right)^2}{1 - \operatorname{mcorr}\left(x_t, \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{y}_{t-1}^{(p)} \end{bmatrix}\right)^2}. \end{aligned}$$

Proof. The equivalence of directed Granger causality (DGC) to a log ratio of partial variances is an immediate consequence of the definition of partial variance in a LGM (Davidson and MacKinnon, 2004; Greene, 2008). For completeness, using the definition of partial variance, Eq. (2.5),

$$F_{y\to x} = \ln\frac{\operatorname{var}\left(x_t|\mathbf{x}_{t-1}^{(p)}\right)}{\operatorname{var}\left(x_t|\mathbf{x}_{t-1}^{(p)}, \mathbf{y}_{t-1}^{(p)}\right)}, \tag{6.2}$$

It remains to show that DGC can be expressed as a remapping of the multiple correlation coefficient. Recall that the prediction of x_t is designated by $\tilde{x}_t$. Consider first the AR model, Eq. (2.24), in which $\mathbf{x}_{t-1}^{(p)}$ is used to predict x_t, and coefficient vectors $\mathbf{b}_A$ and $\mathbf{c}_A$ are $\mathbf{0}$.

From the properties of covariance (Whittaker, 2009),

$$\begin{aligned}\mathrm{var}\left(\tilde{x}_{t,A}\right) &= \mathrm{cov}\left(\tilde{x}_{t,A}, \tilde{x}_{t,A}\right) \\ &= \mathrm{cov}\left(\tilde{x}_{t,A}, x_t - e_{t,A}\right) \\ &= \mathrm{cov}\left(\tilde{x}_{t,A}, x_t\right) - \mathrm{cov}\left(\tilde{x}_{t,A}, e_{t,A}\right) \\ &= \mathrm{cov}\left(\tilde{x}_{t,A}, x_t\right), \end{aligned} \tag{6.3}$$

where the last line results from the orthogonality of the predictor and residual timeseries. Applying this relationship to the definition of multiple correlation, Eq. (2.4), gives the sample estimate of multiple correlation,

$$\mathrm{mcorr}\left(x_t, \mathbf{x}_{t-1}^{(p)}\right)^2 = \frac{\mathrm{var}\left(\tilde{x}_{t,A}\right)}{\mathrm{var}\left(x_t\right)}. \tag{6.4}$$

Since the signal variance can be decomposed into orthogonal components (Hamilton, 1994)

$$\mathrm{var}\left(x_t\right) = \mathrm{var}\left(\tilde{x}_{t,A}\right) + \mathrm{var}\left(e_{t,A}\right), \tag{6.5}$$

multiple correlation, Eq. (2.4), can be expressed in terms of residual variance (Whittaker, 2009),

$$\mathrm{mcorr}\left(x_t, \mathbf{x}_{t-1}^{(p)}\right)^2 = 1 - \frac{\mathrm{var}\left(e_{t,A}\right)}{\mathrm{var}\left(x_t\right)}. \tag{6.6}$$

Similarly, multiple correlation for the VAR model can be expressed as

$$\mathrm{mcorr}\left(x_t, \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{y}_{t-1}^{(p)} \end{bmatrix}\right)^2 = 1 - \frac{\mathrm{var}\left(e_{t,V}\right)}{\mathrm{var}\left(x_t\right)}. \tag{6.7}$$

Finally, employing the definition of partial variance, Eq. (2.5), gives the result. □

The next statements follow in the same manner as equivalences, and detail equivalence relationships between conditional Granger causality measures, and correlation-based measures. 1 and 2.

Equivalence 3. *Conditional instantaneous Granger causality is a remapping of partial correlation:*

$$F_{x\cdot y|z} = -\ln\left(1 - \mathrm{corr}\left(x_t, y_t | \mathbf{x}_{t-1}^{(p)}, \mathbf{y}_{t-1}^{(p)}, \mathbf{z}_{t-1}^{(p)}\right)^2\right).$$

Proof. The derivation follows directly from Equivalence 1 and application of the Frisch-Waugh-Lovell theorem (Lovell, 2008) which states that regression can be applied sequentially, so that the influence of $\mathbf{z}_{t-1}^{(p)}$ can be removed, and then Equivalence 1 applied. □

Equivalence 4. *Conditional directed Granger causality is a remapping of partial variance and multiple correlation:*

$$\begin{aligned} F_{y\to x|z} &= \ln \frac{\operatorname{var}\left(x_t|\mathbf{x}_{t-1}^{(p)}, \mathbf{z}_{t-1}^{(p)}\right)}{\operatorname{var}\left(x_t|\mathbf{x}_{t-1}^{(p)}, \mathbf{y}_{t-1}^{(p)}, \mathbf{z}_{t-1}^{(p)}\right)} \\ &= \ln \frac{1-\operatorname{mcorr}\left(x_t, \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{z}_{t-1}^{(p)} \end{bmatrix}\right)^2}{1-\operatorname{mcorr}\left(x_t, \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{y}_{t-1}^{(p)} \\ \mathbf{z}_{t-1}^{(p)} \end{bmatrix}\right)^2}. \end{aligned}$$

Proof. The derivation follows directly from Equivalence 2 and application of the Frisch-Waugh-Lovell theorem (Lovell, 2008). □

Partial directed Granger causality was introduced to cater for unmodelled inputs that are common to all regions, by removing any remaining covariance between residuals after regressing out lagged influence.

Equivalence 5. *Partial directed Granger causality is a remapping of directed and instantaneous measures:*

$$F_{y\to x|\cdot z} = F_{y\to x|z} + F_{x\cdot z|y} - F_{x\cdot z}.$$

Proof. The result follows from direct substitution of the relevant expressions, as follows:

$$\begin{aligned} F_{y\to x|\cdot z} =& \ln \frac{\operatorname{var}\left(e_{t,V_{xz}}|w_{t,V_{xz}}\right)}{\operatorname{var}\left(e_{t,V_{xyz}}|w_{t,V_{xyz}}\right)} \\ =& \ln \frac{\operatorname{var}\left(e_{t,V_{xz}}\right) - \operatorname{cov}\left(e_{t,V_{xz}}, w_{t,V_{xz}}\right)\left(\operatorname{var}\left(w_{t,V_{xz}}\right)\right)^{-1}\operatorname{cov}\left(w_{t,V_{xz}}, e_{t,V_{xz}}\right)}{\operatorname{var}\left(e_{t,V_{xyz}}\right) - \operatorname{cov}\left(e_{t,V_{xyz}}, w_{t,V_{xyz}}\right)\left(\operatorname{var}\left(w_{t,V_{xyz}}\right)\right)^{-1}\operatorname{cov}\left(w_{t,V_{xyz}}, e_{t,V_{xyz}}\right)} \\ =& \ln \frac{\operatorname{var}\left(e_{t,V_{xz}}\right)\left(1-\operatorname{corr}\left(e_{t,V_{xz}}, w_{t,V_{xz}}\right)^2\right)}{\operatorname{var}\left(e_{t,V_{xyz}}\right)\left(1-\operatorname{corr}\left(e_{t,V_{xyz}}, w_{t,V_{xyz}}\right)^2\right)} \\ =& \ln \frac{\operatorname{var}\left(x_t|\mathbf{x}_{t-p}, \mathbf{z}_{t-p}\right)\left(1-\operatorname{corr}\left(x_t, z_t|\mathbf{x}_{t-p}, \mathbf{z}_{t-p}\right)^2\right)}{\operatorname{var}\left(x_t|\mathbf{x}_{t-p}, \mathbf{y}_{t-p}, \mathbf{z}_{t-p}\right)\left(1-\operatorname{corr}\left(x_t, z_t|\mathbf{x}_{t-p}, \mathbf{y}_{t-p}, \mathbf{z}_{t-p}\right)^2\right)} \\ =& \ln \frac{\operatorname{var}\left(x_t|\mathbf{x}_{t-p}, \mathbf{z}_{t-p}\right)}{\operatorname{var}\left(x_t|\mathbf{x}_{t-p}, \mathbf{y}_{t-p}, \mathbf{z}_{t-p}\right)} - \ln\left(1-\operatorname{corr}\left(x_t, z_t|\mathbf{x}_{t-p}, \mathbf{y}_{t-p}, \mathbf{z}_{t-p}\right)^2\right) \\ &+ \ln\left(1-\operatorname{corr}\left(x_t, z_t|\mathbf{x}_{t-p}, \mathbf{z}_{t-p}\right)^2\right) \\ =& F_{y\to x|z} + F_{x\cdot z|y} - F_{x\cdot z} \end{aligned} \tag{6.8}$$

□

6.2.2 The Equivalence Between GGC and Correlation-based Statistical Inference Tests

Each measure of linear dependence described in Appendix 2.2.2, and each Granger causality measure, forms a test statistic that is distributed exactly, asymptotically, or approximately, according to a given distribution. Comparison of the distributions yields insight into any source of difference between the application of Granger-based and variance-based measures.

The statistical inference test for zero partial correlation, Eq. (2.19), is compared to the inference test for IGC, Eq. (2.42).

Equivalence 6. *The statistical inference tests of zero partial correlation and zero instantaneous Granger causality are asymptotically equivalent.*

Proof. To aid readability, throughout this derivation let $\rho = \rho_{x_t,y_t|\mathbf{x}_{t-1}^{(p)},\mathbf{y}_{t-1}^{(p)}}$, and $F = F_{x_t \cdot y_t}$, so that Equivalence 1 becomes $F = -\ln(1-\rho^2)$.

In Chapter 6.2.1, IGC was shown to be a remapping of partial correlation, Eq. (1). Taylor's theorem (Pedrick, 1994) can be employed to approximate this function using a fourth order Taylor expansion,

$$\begin{aligned} F \approx F(\rho_0) + \left.\frac{dF}{d\rho}\right|_{\rho=\rho_0}(\rho-\rho) + \left.\frac{d^2F}{d\rho^2}\right|_{\rho=\rho_0}\frac{(\rho-\rho)^2}{2} + \left.\frac{d^3F}{d\rho^3}\right|_{\rho=\rho_0}\frac{(\rho-\rho)^3}{3!} \\ + \left.\frac{d^4F}{d\rho^4}\right|_{\rho=\rho_0}\frac{(\rho-\rho)^4}{4!}, \end{aligned} \tag{6.9}$$

where ρ_0 is the centre point of the Taylor expansion. The derivative functions can easily be derived as

$$\begin{aligned} \frac{dF}{d\rho} &= -\frac{2\rho}{\rho^2-1} \\ \frac{d^2F}{d\rho^2} &= \frac{4\rho^2}{(\rho^2-1)^2} + \frac{2}{1-\rho^2} \\ \frac{d^3F}{d\rho^3} &= \frac{12\rho}{(\rho^2-1)^2} + \frac{16\rho^3}{(\rho^2-1)^3} \\ \frac{d^4F}{d\rho^4} &= \frac{12\rho}{(\rho^2-1)^2} - \frac{96\rho^2}{(\rho^2-1)^3} + \frac{96\rho^4}{(\rho^2-1)^4}. \end{aligned}$$

Under a null hypothesis of $H_0 : \rho = 0$, the expected value for sample partial correlation is 0, and hence it is intuitive to centre the Taylor expansion at 0, so that Eq. (6.9) becomes

$$F \approx \rho^2. \tag{6.10}$$

Following the same process to approximate Fisher's z-transformation from partial correlation

results in

$$\mathrm{r} \approx \rho + \frac{\rho^3}{3} + \frac{\rho^4}{6}. \tag{6.11}$$

Under the null hypothesis of zero partial correlation, sample partial correlation values are expected to be small, and hence this approximation can be further approximated by

$$\mathrm{r} \approx \rho. \tag{6.12}$$

This result is expected since under the null hypothesis the distribution of sample correlation, Eq. (2.19), matches the null hypothesis distribution of Fisher's z-transformation, Eq. (2.21). Applying this result to Eq. (6.10) immediately gives

$$F \approx \mathrm{r}^2. \tag{6.13}$$

Let

$$\mathrm{S} = \phi(\mathrm{r}) \approx N\mathrm{r}^2, \tag{6.14}$$

which has derivative

$$\phi'(\mathrm{r}) = \frac{d\mathrm{S}}{d\mathrm{r}} = 2N\mathrm{r}. \tag{6.15}$$

Under the null hypothesis the distribution of the partial correlation test variate is given by

$$\mathrm{r} \sim \mathcal{N}\left(0, \frac{1}{N}\right). \tag{6.16}$$

Expressing the distribution of the IGC test statistic using a change of variables from the partial correlation test variate requires that (Papoulis and Pillai, 2002)

$$p_{\chi^2}(\mathrm{S}) = 2p_{\mathcal{N}}\left(\phi^{-1}(\mathrm{S})\right)\left|\frac{1}{\phi'(\phi^{-1}(\mathrm{S}))}\right|, \tag{6.17}$$

where the factor of 2 is a consequence of the non-injective mapping from r to F. Expanding gives,

$$\frac{1}{\sqrt{2\pi}}(\mathrm{S})^{-\frac{1}{2}}e^{\frac{\mathrm{S}}{2}} = \sqrt{\frac{N}{2\pi}}e^{-\frac{\mathrm{S}N}{N2}}\left|\frac{1}{2\sqrt{\mathrm{S}N}}\right|, \tag{6.18}$$

and the proof is complete.

□

The significance of directed influence from y to x is assessed by comparing the bivariate

VAR model, in which x_t is a function of $[x_{t-1}^{(p)}, y_{t-1}^{(p)}]$, with the AR model, in which x_t is a function of only $x_{t-1}^{(p)}$. The reduced parameter set of the AR model renders it nested within the VAR model, which is a key consideration for statistical inference testing. Note that N continues to denote degrees of freedom such that $N = T - 2p - 1$.

Using the multiple correlation coefficient, the test for zero directed influence is expressed via a null hypothesis, $H_0 : \mathrm{R}^2_{x_t, \mathbf{x}_{t-1}^{(p)}, \mathbf{y}_{t-1}^{(p)}} = \mathrm{R}^2_{x_t, \mathbf{x}_{t-1}^{(p)}}$. For nested models with Gaussian noise, the resulting test statistic has an F distribution, Eq. (2.22), where, for clarity, $\mathrm{mcorr}_A^2 = \mathrm{mcorr}\left(x_t, \mathbf{x}_{t-1}^{(p)}\right)^2$ and $\mathrm{mcorr}_V^2 = \mathrm{mcorr}\left(x_t, \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{y}_{t-1}^{(p)} \end{bmatrix}\right)^2$.

$$\mathrm{h}\left(\mathrm{mcorr}_V^2, \mathrm{mcorr}_A^2\right) = \frac{\frac{\mathrm{mcorr}_V^2 - \mathrm{mcorr}_A^2}{p}}{\frac{1 - \mathrm{mcorr}_V^2}{N}} \tag{6.19}$$

For partial variance, the null hypothesis for zero directed influence is $H_0 : \sigma_{e_V}^2 = \sigma_{e_A}^2$. In the case of nested models with Gaussian noise, the test for partial variance has an F distribution, Eq. (2.23),

$$\mathrm{q}\left(\mathrm{var}\left(e_{t,A}\right), \mathrm{var}\left(e_{t,V}\right)\right) = \frac{\frac{\mathrm{var}(e_{t,A}) - \mathrm{var}(e_{t,V})}{p}}{\frac{\mathrm{var}(e_{t,V})}{N}} \sim F\left(p, N\right). \tag{6.20}$$

The null hypothesis for directed influence measured via DGC, Eq. (2.43), is compared to the distribution of multiple correlation and partial variance estimates under the null hypothesis.

Equivalence 7. *The statistical inference tests of multiple correlation, partial variance, and directed Granger causality are asymptotically equivalent for linear Gaussian models.*

Proof. Considering first the multiple correlation and partial variance, a proof of the equivalence of their statistical tests can be found in Greene (2008, p.90), and is provided here for completeness. Rearrangement of the test statistic for multiple correlation, Eq. (2.23), results in

$$\mathrm{h}\left(\mathrm{mcorr}_V^2, \mathrm{mcorr}_A^2\right) = \frac{\frac{\mathrm{mcorr}(V)^2 - \mathrm{mcorr}(A)^2}{p}}{\frac{1 - \mathrm{mcorr}(V)^2}{N}} \tag{6.21}$$

$$= \frac{N}{p}\left(\frac{\mathrm{var}\left(e_{t,A}\right)}{\mathrm{var}\left(e_{t,A}\right)} - 1\right), \tag{6.22}$$

and, for partial variance, Eq. (2.22),

$$\mathrm{q}\left(\mathrm{var}\left(e_{t,A}\right),\mathrm{var}\left(e_{t,V}\right)\right) = \frac{\frac{\mathrm{var}(e_{t,A})-\mathrm{var}(e_{t,V})}{p}}{\frac{\mathrm{var}(e_{t,V})}{N}} \tag{6.23}$$

$$= \frac{N}{p}\left(\frac{\mathrm{var}\left(e_{t,A}\right)}{\mathrm{var}\left(e_{t,V}\right)} - 1\right) \tag{6.24}$$

$$= \mathrm{h}\left(\mathrm{mcorr}_V^2, \mathrm{mcorr}_A^2\right), \tag{6.25}$$

which demonstrates the equivalence of the test statistics for multiple correlation and partial variance, and hence the equivalence of the inference tests themselves.

Equivalence between testing for zero directed influence using partial variance and DGC is examined by comparison of the distributions in Eq. (2.22) and Eq. (2.43), orchestrated via a change of variables (Papoulis and Pillai, 2002). For simplicity denote $\mathrm{q}\left(\mathrm{var}\left(e_{t,A}\right),\mathrm{var}\left(e_{t,V}\right)\right)$ by q. Exploiting the relationship in Equivalence 2,

$$L \triangleq \theta(\mathrm{S}_{y\to x}) = \frac{Ne^{F_{y\to x}} - N}{p}, \tag{6.26}$$

from which the derivative is obtained,

$$\frac{dq}{dN_{y\to x}} = \frac{e^{F_{y\to x}}}{p}. \tag{6.27}$$

If $p_F\left(q\right)$ denotes the pdf of q and $p_{\chi^2}\left(\mathrm{S}_{y\to x}\right)$ the pdf of $\mathrm{S}_{y\to x}$, then applying a change of variables gives the result,

$$p_{\chi^2}\left(\mathrm{S}_{y\to x}\right) \overset{a}{=} p_F\left(\theta\left(\mathrm{S}_{y\to x}\right)\right)\left|\frac{dq}{d\mathrm{S}_{y\to x}}\right|. \tag{6.28}$$

The asymptotic equivalence of the tests for DGC, Eq. (2.43), and partial variance (and hence multiple correlation) is demonstrated by showing that the probability distributions of the test statistics are asymptotically equivalent. Let $N = T - 2p - 1$ designate the degrees of freedom, denote $\mathrm{q}\left(\mathrm{var}\left(e_{t,A}\right),\mathrm{var}\left(e_{t,V}\right)\right)$ by q, let $F\left(p,N\right)$ designate the pdf of q, and $\chi^2\left(p\right)$ be the pdf of $\mathrm{S}_{y\to x}$. Introduce

$$k = \frac{\mathrm{var}\left(e_{t,A}\right)}{\mathrm{var}\left(e_{t,V}\right)}, \tag{6.29}$$

which results in,

$$\begin{aligned} \mathrm{S}_{y\to x} &= N\ln(k) \sim \chi^2(\mathrm{S}_{y\to x}\,;\,p), \\ q &= (k-1)\frac{N}{p} \sim F\left(p,N\right). \end{aligned} \tag{6.30}$$

Rearranging and applying Equivalence 2 gives

$$\mathrm{S}_{y\to x} = \theta^{-1}(q) = N\ln(\frac{qp}{N}+1) \tag{6.31}$$

from which the derivative is obtained,

$$\frac{d\mathrm{S}_{y\to x}}{dq} = \frac{Np}{qp+N}. \tag{6.32}$$

Using a change of variables we want to show that

$$\begin{aligned} p_F(q) &\overset{a}{=} p_{\chi^2}\left(g^{-1}(q)\right)\left|\frac{d\mathrm{S}_{x\to y}}{dq}\right|, \\ \frac{\sqrt{\frac{(qp)^p N^N}{(qp+N)^{p+N}}}}{qB\left(\frac{p}{2},\frac{N}{2}\right)} &\overset{a}{=} \frac{\left(N\ln\left(\frac{qp}{N}+1\right)\right)^{\frac{p}{2}-1}}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)} e^{-\frac{N\ln(qp)}{N}+1}\left|\frac{Np}{qp+N}\right|, \end{aligned} \tag{6.33}$$

where B is the beta function.

The $\ln\left(\frac{qp}{N}+1\right) = \ln\left(\frac{\mathrm{var}(e_{t,A})-\mathrm{var}(e_{t,V})}{\mathrm{var}(e_{t,A})}+1\right)$ term can be simplified to $\frac{qp}{N}$ by assuming that $\mathrm{var}\left(e_{t,V}\right) \gg \left(\mathrm{var}\left(e_{t,A}\right) - \mathrm{var}\left(e_{t,V}\right)\right)$, which gives

$$\begin{aligned} \frac{\sqrt{\frac{(qp)^p N^N}{(qp+N)^{p+N}}}}{qB\left(\frac{p}{2},\frac{N}{2}\right)} &\approx \frac{p^{\frac{p}{2}} N^{\frac{N}{2}+1} q^{\frac{p}{2}-1}}{2^{\frac{p}{2}}\Gamma\frac{p}{2}\left(qp+N\right)^{\frac{N}{2}+1}} \\ \Rightarrow \left(\frac{N}{qp+N}\right)^{\frac{p}{2}} &\overset{a}{=} \frac{N}{qp+N}, \end{aligned} \tag{6.34}$$

proving Eq. (6.33), which can then be rearranged to give

$$p_{\chi^2}\left(\mathrm{S}_{y\to x}\right) \overset{a}{=} p_F\left(g\left(\mathrm{S}_{y\to x}\right)\right)\left|\frac{dq}{d\mathrm{S}_{x\to y}}\right|. \tag{6.35}$$

The result proves asymptotic equivalence of DGC and partial variance with length, and equivalence when model order, p, is 2.

□

6.3 Methods

6.3.1 Probability distribution generation

Probability distributions were generated according to the known tests for zero partial correlation, Eq. (2.21), and zero IGC, Eq. (2.42). Equivalence 6 can be expressed as,

$$p_{\chi^2}\left(\mathrm{S}_{x\cdot y}\right) \approx 2p_{\mathcal{N}}\left(\mathrm{r}\right)\left|\frac{d\mathrm{r}}{d\mathrm{S}_{x\cdot y}}\right|, \tag{6.36}$$

where the factor of 2 is a consequence of the non-injective mapping from r to $\mathrm{S}_{x\cdot y}$. To incorporate the effect of time series length on the $\chi^2\left(1\right)$ distribution of IGC, the range of $F_{x\cdot y}$ was made constant and scaled by various values of N. The distribution for IGC was then derived from partial correlation, using the relationship described in Eq. (6.36). The differences between known and derived distributions of IGC, as a function of time series length, were then calculated. The same process was applied to Fisher's z-transform of partial correlation.

The seventh statement of equivalence was similarly validated numerically using MATLAB . Probability distributions were generated for the known $\chi^2\left(p\right)$ distribution of the DGC test statistic, for $p \in \{1, \cdots, 4\}$. The effect of time series length on the $\chi^2\left(p\right)$ distribution of $\mathrm{S}_{y\to x}$ was modelled by fixing the range of $F_{y\to x}$, and scaling the values by various values of N. The distribution of the DGC test was then derived from partial variance, as described in Eq. (6.28). The differences between the known and derived distributions for DGC were then calculated and the results displayed graphically in Fig. 6.2.

6.3.2 Synthetic data generation and analysis

MATLAB was used to generate simulation data for bivariate VAR models. 5000 datasets were generated for each unique model configuration based on model order, p, noise covariance, δ, time series length, T, and directed influence regression coefficient vector, $\mathbf{b}$, such that $p \in \{1, 2, 3, 4\}, \delta \in \{0, 0.2, 0.4, 0.6, 0.8\}, T \in \{50, 100, 200\}, \mathbf{b} \in \{\mathbf{0}, \text{random}\}$. Non-zero regression coefficients were randomly generated and stability of the resulting model validated before use.

Receiver operating characteristic (ROC) curves were generated for each statistical inference test to compare the sensitivity and specificity of influence measures across various bivariate VAR model configurations. For instantaneous influence, a null hypotheses of zero instantaneous influence was defined by $H_0 : \delta = 0$, while an alternative hypothesis was one of $H_1 : \delta = 0.2$, $H_1 : \delta = 0.4$, or $H_1 : \delta = 0.6$. ROC curves were generated for IGC, tested for significance using a χ^2 test, Eq. (2.42), and for correlation and partial correlation, both tested for significance using Fisher's z-transformation, Eq. (2.21). For directed influence

measures, a null hypothesis of zero directed influence was defined by $H_0 : \mathbf{b} = \mathbf{0}$, while the alternative hypothesis was defined by $H_1 : \mathbf{b} \neq \mathbf{0}$. ROC curves were generated for multiple correlation, tested as an F-statistic, Eq. (2.23), partial variance, also tested as an F-statistic, Eq. (2.22), and DGC, tested as a χ^2 variate, Eq. (2.43).

For each model configuration, the associated datasets were used to generate empirical null distributions for each measure of dependence using three different resampling procedures: randomisation (e.g. Cordes et al. (2000); Deshpande et al. (2008)), permutation (e.g. Kayser et al. (2009)), and a boostrapping surrogate analysis (Theiler et al., 1990) (e.g. Roebroeck et al. (2005); Abler et al. (2005); Guo et al. (2008)) which we will term bootstrapping. For each resampling method, the influence contained in a dataset was eliminated in a process applicable to that method. For randomisation, the phase of each time series was randomised to retain the power spectrum while destroying temporal information. For permutation, we exchanged the time point labels by transposing the first and second halves of one of the time series. For bootstrapping, the maximum likelihood parameters were determined and a new dataset generated using the MLE model parameters, but with the parameters controlling influence modified to satisfy the null hypothesis. The influence measures were then calculated for this modified dataset satisfying the null hypothesis, and the results from 5000 datasets amalgamated to create an empirical null distribution for each influence measure. The significance of an influence measure calculated for a real dataset was determined by the percentage of values in the empirical null distribution that were further from the null hypothesis than the given value.

For every influence measure, both instantaneous and directed, three ROC curves were generated using the empirical null distributions resulting from the three different resampling techniques: randomisation, permutation, and bootstrapping.

6.4 Results

The seven statements of equivalence presented in the Methods section demonstrate that Geweke's measures of linear dependence evaluate the same aspects of a LGM as partial correlation and partial variance, for instantaneous and directed influence, respectively. Numerical results that validate the equivalence of the statistical distributions are presented, together, with the results of the application of GGC and correlation-based measures to the synthetic data.

6.4.1 Validation of the equivalence of the statistical inference tests

The asymptotic equivalence between the tests of zero partial correlation and zero IGC was demonstrated as follows. The probability distributions of $\mathrm{S}_{x \cdot y}$, for various time series

lengths, are shown in Fig. 6.1A. The normal distribution is plotted for mean 0, and variance a function of time series length in Fig. 6.1B, depicting the distribution of Fisher's z-transformation of partial correlation. The distribution of the test for IGC, derived from partial correlation using Eq. (6.36), is contained in Fig. 6.1C, while the distribution of the test for partial correlation, derived from IGC is displayed in Fig. 6.1D. Finally, Figs. 6.1E and 6.1F show the error between the actual and derived distributions for the tests for IGC and partial correlation, respectively. Errors on the order of 10^{-4} and 10^{-3}, diminishing with increasing N, clearly demonstrate that the tests for zero partial correlation and IGC are asymptotically equivalent with time series length.

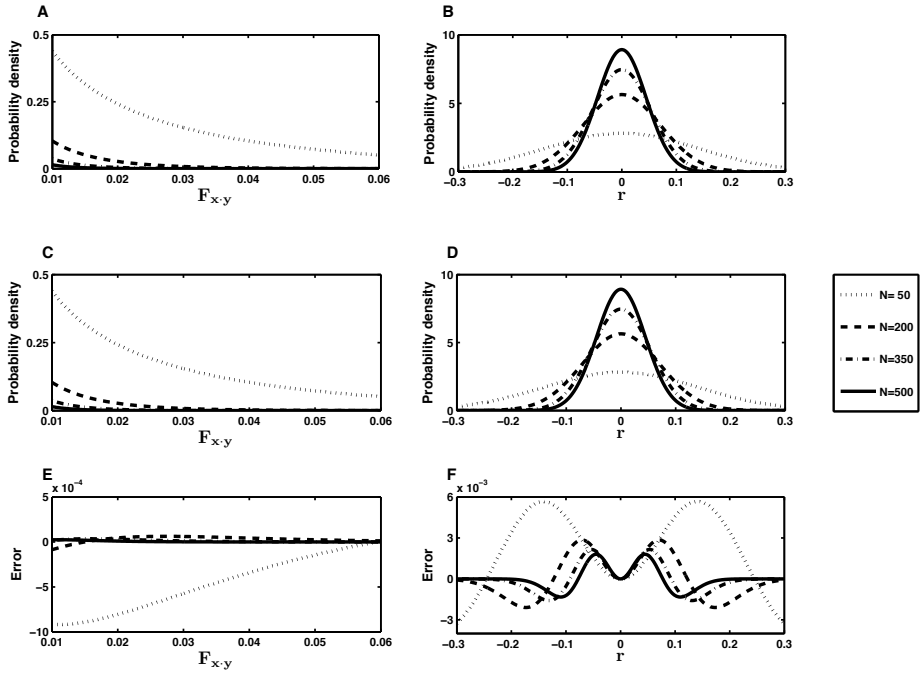

Figure 6.1: Equivalence of instantaneous influence test distributions. Let $r = \mathrm{r}\left(\mathrm{corr}\left(x_t, y_t | \mathbf{x}_{t-1}^{(p)}, \mathbf{y}_{t-1}^{(p)}\right)\right)$ **A:** Pdf of $S_{x\cdot y} \sim \chi^2(1)$. **B:** Pdf of $r \sim \mathcal{N}(0, \frac{1}{N})$. **C:** Pdf of $S_{x\cdot y} \sim p_{\mathcal{N}}(0, \frac{1}{N}) \left|\frac{dr}{dS_{x\cdot y}}\right|$. **D:** Pdf of $r \sim p_{\chi^2(1)} \left|\frac{dS_{x\cdot y}}{dt}\right|$ **E:** Error in estimating the distribution of IGC from r - Fisher's z-transformation of partial correlation. **F:** Error in estimating the distribution of r from IGC.

Asymptotic equivalence between tests of multiple correlation, partial covariance and DGC (the seventh equivalence) was verified as follows. The $\chi^2(p)$ probability distribution of the DGC test is displayed in Fig. 6.2A, as a function of model order, p. The distribution of the DGC test derived from partial variance is illustrated in Fig. 6.2B, while the difference between the known and derived distributions for DGC is shown as an error in Fig. 6.2C. The small error values, diminishing with time series length, support the theoretical findings.

Furthermore, for order $p = 2$, the distributions are identical for all values of N, while further increases in p increase the magnitude of error. This finding was predicted by the theory, as detailed fully in the derivation of Equivalence 7.

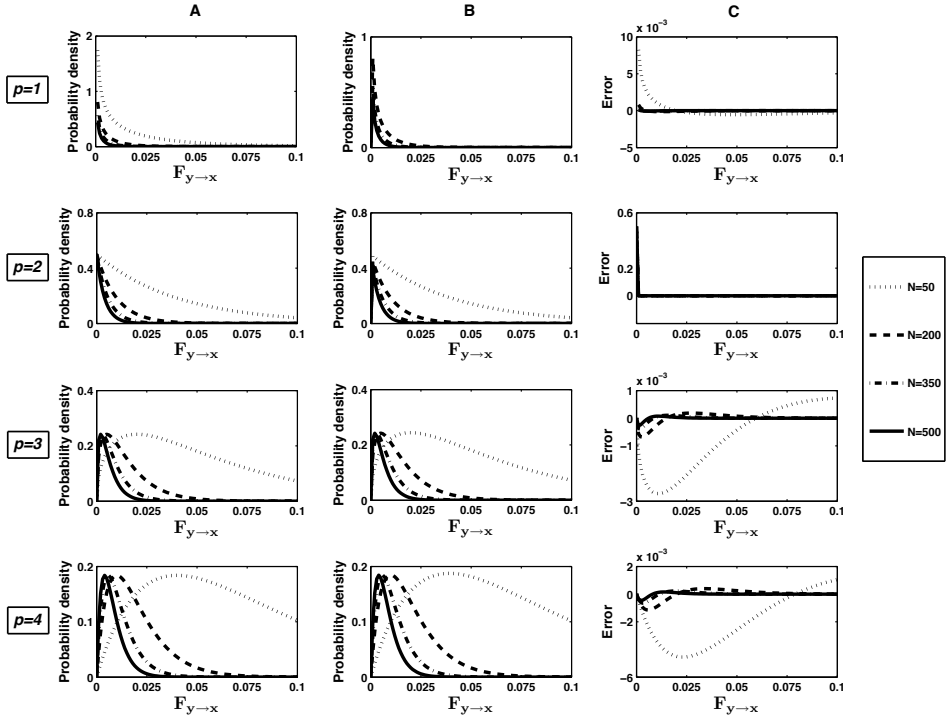

Figure 6.2: Equivalence of directed influence test distributions. **A:** pdf of $S_{y\to x} \sim \chi^2(p)$. **B:** pdf of $S_{y\to x} \sim p_F(p, N) \left|\frac{dq}{dS_{y\to x}}\right|$. **C:** Error in estimating the distribution of DGC from q. Each row gives the probability distributions for a different model order, $p \in \{1, 2, 3, 4\}$.

6.4.2 Application of GGC and correlation-based measures

To understand how the GGC and correlation-based connectivity measures perform when applied to sample data, we examined the statistical inference tests available for each measure using simulated data.

For all VAR models the ROC curves for the three resampling techniques, randomisation, permutation, and bootstrapping, showed negligible difference across all inference measures, to the extent that it was not possible to distinguish the individual empirical ROC curves of any measure. We have, therefore, only included the results for a single resampling method, randomisation, to improve figure clarity.

ROC curves illustrating the performance of instantaneous influence measures are shown in Fig. 6.3. The effects of both noise covariance, δ (Fig. 6.3A), and time series length,

T (Fig. 6.3B), on the performance of each influence measure was examined using VAR(0) datasets. Performance was found to be strongly and positively correlated with both noise covariance and sample size. Importantly, a single ROC curve is visible for each distinct model examined, clearly demonstrating the negligible difference between the instantaneous tests and showing an identical dependence on noise covariance and time series length. Furthermore, the close agreement between analytic and empirical tests supports the accuracy of the analytic distributions, while the agreement between the empirical results of each measure reinforces Equivalence 1.

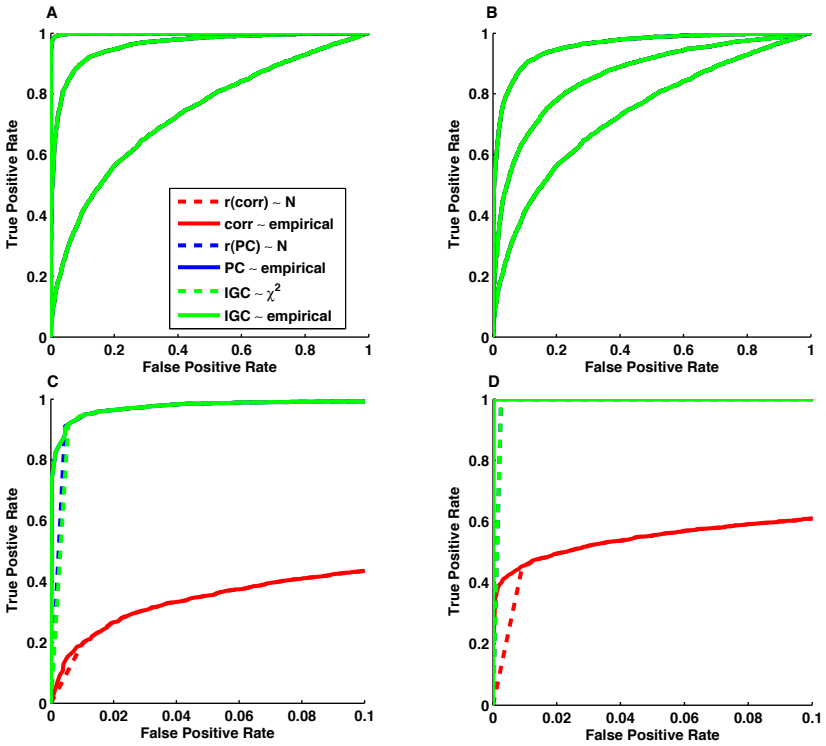

Figure 6.3: Sensitivity and specificity characteristics of instantaneous influence for bivariate VAR models. Comparisons of IGC (green), partial correlation (PC,red) and correlation (corr,blue). The legend shown in A is applicable to all figures. Where a single ROC plot is visible for a model, all inference tests performed with negligible difference. Note the enlargement of axis in C and D to enable distinction between analytic and empirical inference tests. **A:** ROC curves generated from VAR(0) datasets with T = 50, and three difference noise covariance values, such that; $H_0 : \delta = 0$: inner curve, $H_1 : \delta = 0.2$; middle curve, $H_1 : \delta = 0.4$; outer curve, $H_1 : \delta = 0.6$. **B:** ROC curves generated from VAR(0) datasets with $H_0 : \delta = 0, H_1 : \delta = 0.2$, and time series length: inner curve, T = 50; middle curve, T = 100; outer curve, T = 200. **C:** ROC curves generated from VAR(4) data, with T = 50, and $H_0 : \delta = 0, H_1 : \delta = 0.6$. Note that IGC and partial correlation are indistinguishable. **D:** ROC curves generated from VAR(4) data, with T = 200, and $H_0 : \delta = 0, H_1 : \delta = 0.6$. Again, IGC and partial correlation are indistinguishable.

A slight performance difference between empirical and analytic tests was observed for large noise covariance values, diminishing with sample size, suggesting an asymptotic equivalence between empirical and analytic tests. This effect is illustrated in Figs. 6.3C and 6.3D, which use an enlarged false positive axis for ROC curves generated from VAR(2) datasets with 50 and 200 samples, respectively. The strong performance of partial correlation and IGC in the presence of lagged influence demonstrates how effectively directed influence was regressed out when model order was to be assumed known. Conversely, (lag-0) correlation does not model directed influence so its performance was, predictably, severely impacted for non-zero model orders (Figs. 6.3C and 6.3D).

The ROC curves generated from directed influence tests are shown in Fig. 6.4. For each model, we tested for zero directed influence using the χ^2 DGC test, the F distributed multiple correlation test, and the F distributed test for partial variance. For each distinct model, a single ROC curve is visible for all inference tests, both analytic and empirical (Figs. 6.4A and 6.4B), and all measures showed identical dependence on sample size. The agreement between the analytic measures supports equivalence of the directed influence tests (Equivalence 7), while that of the empirical distributions substantiates the deterministic mapping between directed influence measures (Equivalence 2). Furthermore, the negligible difference between analytic and empirical tests highlights the correctness of the analytic tests.

Zoomed-in axes are shown for ROC curves generated from VAR(4) datasets to highlight the marginal difference between the probability density functions of $\mathrm{S}_{y\to x}$ and the test variates for both partial variance and multiple correlation (Fig. 6.4C). The increased error between the probability density functions generated with $N = 50$ (Fig. 6.4C) and $N = 200$ (Fig. 6.4D) supports the theoretical finding of asymptotic equivalence with time series length, Eq. (6.35). The ROC curves of directed influence tests did not change with noise covariance, which is expected because the measures are functions of lagged influence, expressed in the regressors of an LGM.

The ROC curves for both instantaneous and directed influence measures confirmed the theoretical results. The ROC curves generated for conditional measures (both instantaneous and directed) also corroborated the theory, but as the only change for conditional measures is the addition of regression term(s) for the third variate, the ROC curves exhibit the same behaviour as those for the non-conditional measures, and have thus been excluded.

6.5 Discussion

Linear connectivity analyses based on the concepts of Granger causality lie between the data-driven method class and model-based method class. Data-driven methods are comprised of correlation-based techniques including partial correlation and partial variance which mea-

sure the degree of linear association between variates (Marrelec et al., 2006; Wang and Xia, 2007; Tana et al., 2008). The model-based methods are attributed with the ability to reveal directional dependencies (Goebel et al., 2003; Abler et al., 2005; Eichler, 2005). In this chapter we have demonstrated that Geweke's measures of Granger causality (Geweke, 1982) are remappings of correlation-based measures, and that conditional and partial Granger causality measures are similarly remappings of correlation-based quantities. Furthermore, we have provided evidence that correlation-based and Granger causality statistical inference tests are, in both theory and practice, equivalent. These results call into question the utility of Granger causality in connectivity analyses, beyond that elucidated with traditional correlation-based metrics.

It should be noted that while the use of correlation-based measures imply an instance of the LGM class, application of GGC measures are not restricted by this implicit model assumption. For a non-Gaussian noise process comprised of independent and identically distributed samples, the GGC test variates can still be described by an asymptotic chi-square distribution (Geweke, 1982). Thus for highly non-Gaussian noise that may be present in experimental BOLD data, it is anticipated that the GGC inference tests would remain more robust than the correlation-based equivalents. Theoretical analysis of robustness to model/test mismatch, and analysis of the necessarily ill-posed multivariate models that extend from the bivariate models considered herein, are beyond the scope of this paper and the subject of ongoing research.

A corollary of the equivalence results demonstrated in this chapter is the ubiquity of partial correlation and partial variance in linear connectivity methods. Partial correlation and partial variance can describe both directed and instantaneous influence, optionally in the presence of exogenous inputs for a LGM of arbitrary dimensionality and order, limited only by the fMRI time series length. In the time domain, correlation is a trivial case of partial correlation, and the equivalence between multiple correlation (Wang and Xia, 2007) and partial variance is well known (Whittaker, 2009; Greene, 2008) Frequency domain linear connectivity methods fall under the same umbrella, as Marrelec et al. (2005) have shown that correlation, coherence (Sun et al., 2004), and mutual information (Salvador et al., 2005b) extract the same information from a LGM. Marrelec et al. (2006) introduced spatial partial correlation as a fMRI connectivity method, applying it to models incorporating only instantaneous influence which were thus unable to discern directionality. They observed that spatial partial correlation has both data-driven attributes, as it is correlation-based, and model-based attributes, such as the requirement to perform node selection. Marrelec et al. (2007, pg. 1188) later concluded that spatial partial correlation 'provides a pertinent first step to bridge the gap that has endured between functional and effective connectivity'. Our results augment this statement, by establishing equivalence between GGC analysis procedures and their spatio-temporal correlation-based counterparts.

SEM, the widely used model-based linear connectivity technique, employs a goodness-of-fit

test to guide node selection (McIntosh and Gonzalez-Lima, 1994) that can alternatively be approached using a likelihood ratio test (Kline, 2005; Schumacker and Lomax, 2004). In the case of Gaussian noise, the likelihood ratio is equivalent to the DGC test (Geweke, 1982) and is, therefore, asymptotically equivalent to the partial variance test by Equivalence 7, further bridging the gap between classes of connectivity methods. SEM differs from the linear connectivity measures discussed thus far, however, in its use of a scalar regression coefficient to denote influence from a source to a target region. In order to maintain a scalar description of the influence from the source to the target, the inclusion of lagged influence is precluded as it requires multiple regression terms for the source. SEM is therefore not applicable to all members of the LGM class. Furthermore, regression coefficients should not be used to designate the strength of linear association because they are unit dependent (Eichler, 2005). Harrison et al. (2003) similarly used scalar regression coefficients to designate connection strength. Their application to VAR(p) models, however, incorporated multiple influence types which resulted in multiple connection coefficients for each source. The graphical tabulation of the multiple estimated coefficients is both cumbersome to summarise for large models and, we believe, potentially misleading; the coefficients of a VAR(p) model form characteristic polynomials, the roots (poles) of which determine system behaviour, not the absolute weights of individual coefficients. This critique is, however, beyond the scope of the current chapter .

Conditional and/or partial connectivity methods, whether presented as Granger causality or correlation-based, and similarly SEM, apply to a single multivariate model that necessitates a node selection process due to the small time series length relative to the number of voxels. In contrast, seed correlation and pairwise Granger causality methods apply many bivariate models in the search for connectivity networks, thereby eliminating the need for node selection, but also rescinding the capacity to remove mediated influence. It is this *a priori* classification of network model type, in conjunction with the choice to model instantaneous and/or directed influence via selection of model order, that differentiates linear connectivity methods, rather than the distinguishing feature being a methodological difference in metric or statistical inference test.

6.6 Conclusion

The equivalence of Geweke's measures of Granger causality and correlation-based connectivity measures and their corresponding statistical inference tests has been delineated in this chapter . Geweke's measures of Granger causality provide a framework in which to describe linear connectivity, but offer no further insight beyond correlation-based techniques. We have clarified distinctions, both perceived and real, between linear connectivity methods in their application to instantiations of the linear Gaussian model class, and we have highlighted the centrality of correlation-based measures in all linear connectivity methods.

While these theoretical results may be unsurprising to the reader with advanced statistical knowledge, they highlight the importance of a clear understanding of the theoretical basis of connectivity analysis methods for human brain mapping researchers. Our conclusion is that assignment of a linear Gaussian connectivity analysis method into one or more of the data-driven, model-based and Granger causality classes is ultimately a function of node selection and model configuration, rather than a distinction rooted in methodological difference.

6.A Parameterising Instantaneous Influence

The source of the covariance, δ, between the noise time series of the VAR model defined in Eq. (2.24) is presented here. The residuals that are used to estimate the noise processes contain any of the original time series not partialled out in the regression. The covariance between the residuals can be shown to contain instantaneous covariance between model variates x_t and y_t where it exists.

Assume that the two signals, x_t and y_t, co-vary instantaneously, so that x_t is a function of $[y_t, \mathbf{y}_{t-1}^{(p)}]$, and y_t depends on $[x_t, \mathbf{x}_{t-1}^{(p)}]$. If $[x_t, \mathbf{x}_{t-1}^{(p)}, y_t, \mathbf{y}_{t-1}^{(p)}]$ constitutes the predictor set, then instantaneous influence is partialled out in the regression. Consequently, the noise processes will have a diagonal covariance matrix, with all covariance terms set to 0. This model is defined as

$$\begin{bmatrix} 1 & -b_0 \\ -c_0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \mathbf{a} & \mathbf{c} \\ \mathbf{b} & \mathbf{d} \end{bmatrix}' \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{y}_{t-1}^{(p)} \end{bmatrix} + \begin{bmatrix} e_t \\ u_t, \end{bmatrix}, \quad \begin{bmatrix} e_t \\ u_t \end{bmatrix} \sim \mathcal{N}(0, \mathbf{\Sigma}), \quad (6.37)$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \varepsilon & 0 \\ 0 & \eta \end{bmatrix}. \quad (6.38)$$

Premultiplying both sides of Eq. (6.37) by the instantaneous linear dependence matrix results in

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & -b_0 \\ -c_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{a} & \mathbf{c} \\ \mathbf{b} & \mathbf{d} \end{bmatrix}' \begin{bmatrix} \mathbf{x}_{t-1}^{(p)} \\ \mathbf{y}_{t-1}^{(p)} \end{bmatrix} + \begin{bmatrix} 1 & -b_0 \\ -c_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} e_t \\ u_t \end{bmatrix}. \quad (6.39)$$

The noise covariance matrix is now a function of the instantaneous covariance between x_t and y_t,

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -b_0 \\ -c_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon & 0 \\ 0 & \eta \end{bmatrix} \left(\begin{bmatrix} 1 & -b_0 \\ -c_0 & 1 \end{bmatrix}^{-1} \right)' . \tag{6.40}$$

Eq. (6.40) demonstrates that, for a VAR model with instantaneous terms excluded from the regression, the instantaneous covariance between x_t and y_t is reparameterised in the covariance matrix of the residual.

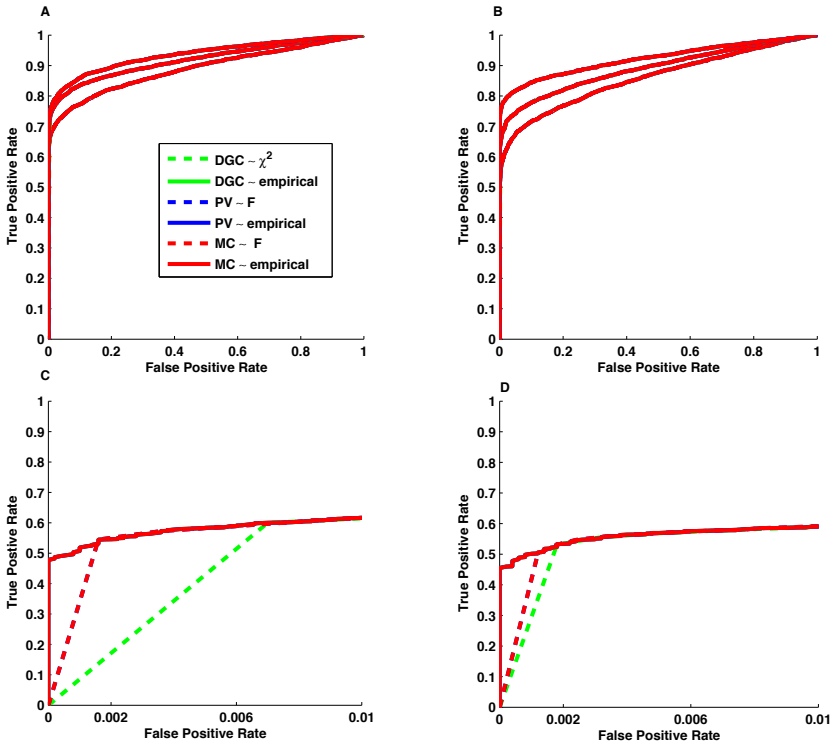

Figure 6.4: Sensitivity and specificity characteristics of directed influence for bivariate VAR models. Comparison of multiple correlation (MC,red), partial variance (PV,blue), and DGC (green) tests. Where a single ROC plot is visible for a model, all inference tests performed with negligible difference. The axes on the bottom panels are enlarged to highlight ROC curve differences. The legend displayed in panel A is applicable to all panels. **A:** ROC curves generated from bivariate VAR(2) datasets with $T = 100$, $H_0 : \mathbf{b} = \mathbf{0}, H_1 : \mathbf{b} \neq \mathbf{0}$, and three different model orders: inner curve, $p = 2$; middle curve, $p = 1$; outer curve, $p = 4$. **B:** ROC curves generated from VAR(2) datasets with $H_0 : \mathbf{b} = \mathbf{0}, H_1 : \mathbf{b} \neq \mathbf{0}$, and across three different time series length: inner curve, $T = 50$; middle curve, $T = 100$; outer curve, $T = 200$. **C:** ROC curves generated from VAR(4) data, with $T = 50$, and $H_0 : \mathbf{b} = \mathbf{0}, H_1 : \mathbf{b} \neq \mathbf{0}$. Note that DGC and partial variance are indistinguishable. **D:** ROC curves generated from VAR(4) data, with $T = 200$, and $H_0 : \mathbf{b} = \mathbf{0}, H_1 : \mathbf{b} \neq \mathbf{0}$. Again, DGC and partial variance are indistinguishable.

Chapter 7

Conclusion

Connectivity methods for analysis of fMRI data have traditionally been categorised into data-driven and model-driven methods. Distinction between the two classes of methods is cited as being founded on the use, or lack thereof, of a data model (Friston, 1994; Rogers et al., 2007). This lack of data model in data-driven connectivity methods supposedly results in an inability to identify directional influence and a reduced ability to draw conclusions about the connectivity network (Friston et al., 1993; Sato et al., 2006b; Rogers et al., 2007). Conversely, use of a data model by model-driven methods is accredited with the ability to distinguish causal influence and draw stronger conclusions regarding the connectivity network (McIntosh and Gonzalez-Lima, 1994; Gonzalez-Lima and McIntosh, 1994; Friston et al., 2003b; Rogers et al., 2007). This thesis has scrutinised the validity of the data-driven/model-driven dichotomy by clarifying methodological dependencies between linear techniques drawn from both modalities.

LGMs comprise the majority of data models employed in fMRI connectivity analyses. Data-driven connectivity methods are primarily correlation-based, and hence make an implicit LGM assumption before (typically) applying many bivariate LGM models to extract whole-brain connectivity maps of linear dependence between timeseries pairs. It is customary for model-driven connectivity methods to explicitly employ a single multivariate LGM, specifying constituent regions using *a priori* knowledge. Whether engaged explicitly or implicitly, LGM models impose several data assumptions, conformance to which is required to guarantee integrity of connectivity results. Required assumptions include Gaussianity of LGM residuals, stationarity of residuals, sample independence, and linearity of dependence.

In contraventions of the LGM assumption, BOLD data is known to contain Rician distributed residuals (Gudbjartsson and Patz, 1995), non-stationary noise and signal power (Diedrichsen and Shadmehr, 2005), and temporal sample dependence (Bullmore et al., 2000). Furthermore, preprocessing methods employed to address the prevalence of noise in BOLD signals can exacerbate the extent of the violation of assumptions. As a subclass of the LGM class, autoregressive models introduce an additional requirement of stability to prevent divergent models. The material contained in this thesis has rectified known infringements of the LGM assumptions in resting state fMRI data to mitigate the potential for artificially induced connectivity.

7.1 Thesis contributions

The significant contributions of this thesis to fMRI connectivity research are summarised as follows.

Distribution of non-stationary timeseries: A model for non-stationarity was formulated to address the acknowledged problem of variable signal power in BOLD data acquisition, which was shown to be slice specific in resting state data. Slice variances were

empirically demonstrated to be have dissimilar distributions and be accurately characterised by an inverse gamma distribution. The temporal distribution of non-stationary voxel timeseries was analytically derived to be a generalised Student's t distribution. Experimental data supported the model, with both slice and individual voxels being well characterised described by a Student's t distribution. This result provides support for the Gaussianity assumption inherent in the use of a LGM.

Correlation derived from non-stationary timeseries: An analytic expression for correlation between non-stationary timeseries was derived and shown to differ from the correlation between stationary signals. A proposed correction was both theoretically and empirically demonstrated to recover stationarity and restore the expected value of correlation between corrected signals to the underlying linear dependence.

Distribution of sample correlation derived from temporally filtered timeseries: Temporal filtering as a preprocessing step in fMRI analyses was shown to increase the sample variance of correlation, thereby inflating the false positive rate and artificially inducing connectivity. An analytic expression for the variance of sample correlation was derived as a function of the complex filter frequency response. The result is valid for both multiple filters and different filters applied to constituent timeseries. Consequently, the finding can be utilised to correct correlation estimates for the sample dependence inherent in fMRI timeseries prior to temporal filtering.

Correcting statistical tests of correlation between filtered timeseries: Corrected expressions for statistical tests of sample correlation between filtered timeseries:, including Fisher's z-transformation and Student's t-test, were derived. The corrections were shown to restore confidence intervals and ameliorate artificially induced correlation resulting from temporal filtering. Our corrections enable the use of temporal filtering for noise suppression whilst maintaining the integrity of connectivity results.

Equivalence of linear Granger causality and correlation-based connectivity methods: Geweke's measures of Granger causality connectivity, which describe instantaneous and directed linear influence between sets of voxel time series, were demonstrated to be remappings of correlation-based metrics. Statistical inference tests for Granger causality measures were theoretically proven to be equivalent to inference tests of correlation-based metrics.

Reclassification of linear Gaussian connectivity methods: The equivalence of linear Granger causality and correlation-based connectivity methods has highlighted the centrality of partial correlation and partial variance in linear connectivity analyses, bridging the gap between functional and effective connectivity techniques. Consequently, assignment of a linear Gaussian connectivity method into either the functional or effective connectivity class was established to be a function of node selection and preprocessing, rather than a distinction derived from methodological difference.

7.2 Future research

The work contained in this thesis presents many avenues for extension and further research. Some possible future directions are discussed here.

The modelling of non-stationary signal power proposed in Chapter 4, while advantageous, is limited in its equal weighting of signal and noise non-stationarity. Noise processes in fMRI contain known non-stationarities, often assumed global, such as the noise introduced by head movement, which would weight the noise process but not impact the signal. Consequently, it would be beneficial to introduce a model incorporating both a signal power non-stationary parameter and a distinct non-stationary noise parameter.

FMRI datasets are acquired in the spatial frequency domain, and hence it may be beneficial to model non-stationary signal power in the spatial frequency domain. Further to this, deriving an analytic expression for the variance of sample correlation acquired from non-stationary timeseries, rather than only the means as in Chapter 4, would provide a more complete description of the impact of non-stationarity on correlation.

In Chapter 4 non-stationary signal power was assumed to be identical for all voxels within a slice. However, visual inspection of the data suggests that there may be region-dependent non-stationary effects. It would be interesting to examine the same signal power model with region-specific non-stationary weighting and determine if there is increased efficacy, where efficacy could be defined as the change in the number of non-stationary voxel timeseries after applying the non-stationarity correction. A limitation to this approach would be in the number of samples contributing to the non-stationary variance estimates, both in terms of overall sample size and the disparate sample sizes for different regions, which may complicate rigorous statistical testing.

The distribution of sample correlation derived for temporally filtered timeseries in Chapter 5 is for single subject analysis only. Generalisation of the result to group connectivity analyses would make it more widely applicable. Potentially, correcting the distributions of sample correlation estimates at the single subject level is sufficient, allowing corrected estimates to be combined at the group level as usual. However, this conjecture requires theoretical and experimental investigation to support it. The distribution of sample correlation would also be more widely applicable to fMRI connectivity analyses if methodology was developed to apply it address known autocorrelations in the raw fMRI timeseries. A valid profile of temporal autocorrelation in raw BOLD timeseries is required in order to model sample dependence with a temporal filter.

The impact of temporal filtering on correlation suggests that there is likely to be similar impact on autoregressive coefficients (since both are derived from LGM coefficients). Indeed, our preliminary investigations suggest that AR(1) coefficients are heavily impacted by the degree of temporal filtering. Research has begun into the expression of autoregressive

coefficients as a function of the complex preprocessing filter response, the results for which will have important implications for directed influence in Granger causality.

7.3 Conclusion

This thesis has examined the validity of the data-driven/model-driven dichotomy present in fMRI connectivity analyses by clarifying methodological dependencies between linear techniques drawn from both modalities. In particular, Granger causality measures which qualify instantaneous and directed influence are demonstrated to be remappings of correlation-based metrics. Consequently, rather than a distinction rooted in methodological difference, the dichotomy between functional and effective connectivity methods is ultimately a function of model configuration realised in the selection of nodes, the choice to model instantaneous and/or directed influence, and the choice to employ many bivariate models or a single multivariate model.

The LGM class is at the root of linear fMRI connectivity, being implicitly employed by correlation-based data-driven methods and explicitly utilised by linear model-driven techniques. The work in this thesis has addressed violated LGM assumptions extant in both raw and preprocessed BOLD data. A model for the distribution of voxel intensity in the presence of non-stationary signal power was proposed and validated using resting state experimental BOLD data. Additionally, the impact of non-stationary signal power on correlation was deduced and the expected value of sample correlation shown to differ to the true correlation from which the signals were generated. A correction was proposed and shown to restore signal stationarity and rectify the change in the expected value of sample correlation.

Work in this thesis has also addressed the problematic assumption of sample independence in fMRI. Temporal filtering exacerbates the problem by filtering out components of the frequency spectra, necessarily introducing autocorrelation. To enable filtering for noise suppression while maintaining the integrity of connectivity results, the distribution of sample correlation was derived as a function of the complex frequency response of the temporal filter. Furthermore, corrections were proposed for Fisher's z-test and Student's t-test to re-establish valid statistical inference testing of correlation estimates.

It is hoped that the theoretical results in this thesis will impact on the every day use of functional connectivity methods and thereby improve the integrity of functional brain imaging research.

Bibliography

Abdi, H., 2007. Bonferroni and Šidák corrections for multiple comparisons. Encyclopedia of Measurement and Statistics.

Abler, B., Roebroeck, A., Goebel, R., Hose, A., Schfnfeldt-Lecuona, C., Hole, G., Walter, H., 2005. Investigating directed influences between activated brain areas in a motor-response task using fMRI. Magnetic Resonance Imaging 24, 181–185.

Akaike, H., 1969. Fitting autoregressive models for prediction. Annals of the Institute of Statistical Mathematics 21, 243–247.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B., Caski, F. (Eds.), Second International Symposium on Information Theory. Vol. 27. p. 267.

Amaro, E., Barker, G., 2006. Study design in fMRI: Basic principles. Brain and Cognition 60 (3), 220–232.

Anderson, T., 2003. An introduction to multivariate statistical analysis, 3rd Edition. Wiley.

Anderson, T., Darling, D., 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Statistics 23, 193–212.

Asghar, Z., Abid, I., 2007. Performance of lag length selection criteria in three different situations.

Atlas, S., 2009. Magnetic resonance imaging of the brain and spine, 4th Edition. Lippincott Williams and Wilkins.

Auer, D., 2008. Spontaneous low-frequency blood oxygenation level-dependent fluctuations and functional connectivity analysis of the "resting" brain. Magnetic Resonance Imaging 26, 1055–1064.

Bandettini, P., 2003. Functional MRI, 2nd Edition. Vol. 9. Elsevier Science, pp. 271–298.

Bandettini, P., 2006. Functional magnetic resonance imaging. The MIT Press, pp. 1–61.

Bandettini, P., 2009a. Neural correlates of thinking, 1st Edition. Springer.

Bandettini, P., 2009b. What's new in neuroimaging methods. Annals of the New York Academy of Sciences 1156, 260–293.

Bandettini, P., Jesmanowicz, A., Kylen, J., Birn, R., Hyde, J., 1998. Functional MRI of brain activation induced by scanner acoustic noise. Magnetic Resonance in Medicine 39, 410–416.

Bannister, P., Brady, J., Jenkinson, M., 2007. Integrating temporal information with a non-rigid method of motion correction for function magnetic resonance images. Image and Vision Computing 25 (3), 311–320.

Baraldi, P., Manginelli, A., Maieron, M., Liberati, D., Porro, C., 2007. An ARX model-based approach to trial by trial identification of fMRI-BOLD responses. NeuroImage 37, 189–201.

Barry, R., Williams, J., Klassen, L., Gallivan, J., Culham, J., Menon, R., 2010. Evaluation of preprocessing steps to compensate for magnetic field distortions due to body movements in bold fMRI. Magnetic Resonance Imaging 28 (2), 235–244.

Bartlett, M., 1937. Properties of sufficiency and statistical tests. In: Proceedings of the Royal Society of London. pp. 268–282.

Beckmann, DeLuca, Devlin, Smith, 2005. Investigation into resting state connectivity using ICA. Philosophical Transactions of the Royal Society, 1001–1013.

Benjamini, Y., Hockberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57, 289–300.

Benter, P., Chou, C., 1987. Practical issues in structural modeling. Sociological Methods and Research 16 (1), 78–117.

Bernasconi, C., Konig, P., 1999. On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. Biological Cybernetics, 199–210.

Bernstein, M., Thomasson, D., Perman, W., 1989. Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. Med Phys 15 (5), 81381.

Beutel, J., Kundel, H., van Metter, R. (Eds.), 2000. Handbook of medical imaging: Physics and psychophysics. Vol. 1. SPIE - The International Society for Optical Engineering.

Bhargava, A., 1986. On the theory of testing for unit roots in observed time series. Review of Economic Studies 53, 369384.

Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S., de Zwart, J., Shmueli, K., Duyn, J., 2009. Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: A 7 T study. Magnetic Resonance Imaging 27 (8), 1019–1029.

Birn, R., Diamond, J., Smith, M., Bandettini, P., 2006. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. NeuroImage 31, 1536–1548.

Biswal, B., Ulmer, J., 1999. Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. Journal of Computer Assisted Tomography 23 (2), 265–271.

Biswal, B., Yetkin, F., Haughton, V., Hyde, J., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine 34, 537–541.

Bloch, F., 1946. Nuclear induction. The American Physical Society 70, 460–474.

Bodurka, J., Ye, F., Petridou, N., Murphy, K., Bandettini, P., 2007. Mapping the MRI voxel volume in which thermal noise matches physiological noise - implications for fMRI. NeuroImage 34, 542–549.

Bonferroni, C., 1935. Il calcolo delle assicurazioni su gruppi di teste. In: Studi in Onore del Professore Salvatore Ortu Carboni. Rome, pp. 13–60.

Boomsma, A., 1985. Nonconvergence, improper solutions, and starting values in lisrel maximum likelihood estimation. Psychometrika 50, 229–242.

Brett, M., Penny, W., Kiebel, S., 2003. An Introduction to Random Field Theory. Wellcome Trust Centre for Neuroimaging, pp. 1–23.

Breusch, T., 1979. Testing for autocorrelation in dynamic linear models. Australian Economic Papers 17, 334355.

Büchel, C., Friston, K., 1997a. Effective connectivity in neuroimaging. Wellcome Department of Cognitive Neurology, Ch. 6, pp. 1–18.

Büchel, C., Friston, K., 1997b. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cerebral Cortex Dec 7, 768–778.

Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., Sham, P., 1996a. Statistical methods of estimation and inference for functional MR image analysis. Magnetic Resonance in Medicine 35, 261–277.

Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., Sharma, T., 2000. How good is good enough in path analysis of fMRI data? NeuroImage 11, 289–301.

Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., Carpenter, T., Brammer, M., 2001. Colored noise and computational inference in neurophysiological fMRI time series analysis: Resampling methods in time and wavelet domains. Human Brain Mapping 12, 61–78.

Bullmore, E., Rabe-Hesketh, S., Morris, R., Williams, S., Gregory, L., Gray, J., Brammer, M., 1996b. Functional magnetic resonance image analysis of a large-scale neurocognitive network. NeuroImage 4, 16–33.

Buxton, R., Wong, E., Frank, L., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model. Magnetic Resonance in Medicine 39.

Calhoun, V., Adah, T., Pearlson, G., 1999. (Non)stationarity of temporal dynamics in fMRI. In: Annual Fall Meeting of the Biomedical Engineering Society. p. 1079.

Calhoun, V., Adali, T., Pearlson, G., Pekar, J., 2001. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Human Brain Mapping 13, 43–53.

Cauda, F., D'Agata, F., Sacco, K., Duca, S., Geminiani, G., Vercelli, A., 2011. Functional connectivity of the insula in the brain. NeuroImage 55, 8–23.

Cecchi, G., Garg, R., Rao, A., 2003. Inferring brain dynamics using Granger causality on fMRI data. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 604–607.

Chang, C., Glover, G., 2010. Time-frequency characteristic of resting state brain connectivity measured with fMRI. NeuroImage 50 (1), 81–98.

Chen, C., Tyler, C., Baseler, H., 2003a. Statistical properties of BOLD magnetic resonance activity in the human brain. NeuroImage 20, 1096–1109.

Chen, N., Dickey, C., Yoo, S., Guttmann, C., Panych, L., 2003b. Selection of voxel size and slice orientation for fMRI in the presence of susceptibility field gradients: application to imaging of the amygdala. NeuroImage 19 (3), 817–825.

Chen, Y., Bressler, S., Ding, M., 2006. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. Journal of Neuroscience Methods 150, 228–237.

Cohen, J., Cohen, P., West, S., Aiken, L., 2003. Applied multiple regression/correlation analysis for the behavioral sciences, 3rd Edition. Vol. 1. Lawrence Erlbaum Associates, Inc., Publishers.

Cordes, D., Haughton, V., Arfanakis, K., Carew, J., Turski, P., Moritz, C., Quigley, M., Meyerand, M., 2001. Frequencies contributing to functional connectivity in the cerebral cortex in 'resting-state' data. American Journal of Neuroradiology, 1326–1333.

Cordes, D., Haughton, V., Arfanakis, K., Wendt, G., Turski, P., Moritz, C., Quigley, M., Meyerand, M., 2000. Mapping functionally related regions of brain with functional connectivity MR imaging. American Journal of Neuroradiology 21, 1636–1644.

Costagli, M., Waggoner, R., Ueno, K., Tanaka, K., Cheng, K., 2009. Correction of 3D body motion in fMRI time series by independent estimation of rotational and translational effects in k-space. NeuoImage 45 (3), 749–757.

Dagli, S., Ingeholm, J., Haxby, J., 1999. Localization of cardiac-induced signal change in fMRI. NeuroImage 9, 407–415.

Davidson, R., MacKinnon, J., 2004. Econometric Theory and Methods. Oxford University Press.

de Marco, G., Devauchelle, B., Berquin, 2009. Brain functional modeling, what do we measure with fMRI data? Neuroscience Research 64, 12–19.

de Waele, S., Broersen, P., 2003. Order selection for vector autoregressive models. IEEE Transactions on Signal Processing 51, 427–433.

Demirci, O., Stevens, M., Andreasen, N., Michael, A., Liu, J., White, T., Pearlson, G., Clark, V., Calhoun, V., 2009. Investigation of relationships between fMRI brain networks in the spectral domain using ICA and Granger causality reveals distinct differences between schizophrenia patients and healthy controls. NeuroImage 46, 419–431.

Deshpande, G., Hu, X., Lacey, S., Stilla, R., Sathian, K., 2010. Object familiarity modulates effective connectivity during haptic shape perception. NeuroImage 49, 1991–2000.

Deshpande, G., Hua, X., Stillab, R., Sathian, K., 2008. Effective connectivity during haptic perception: A study using Granger causality analysis of functional magnetic resonance imaging data. NeuroImage 40, 1807–1814.

Deshpande, G., LaConte, S., James, G., Peltier, S., Hu, X., 2009. Multivariate Granger causality analysis of fMRI data. Human Brain Mapping 30, 1361–1373.

Diedrichsen, J., Shadmehr, R., 2005. Detecting and adjusting for artifacts in fMRI time series data. NeuroImage 27, 624–634.

Efron, B., 2010. Large-scale inference: empirical Bayes methods for estimation, testing, and prediction. Cambridge University Press.

Eichler, M., 2005. A graphical approach for evaluating effective connectivity in neural systems. Philosophical Transactions of the Royal Society B 360, 953–967.

Fisher, R., 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10, 507–521.

Fisher, R., 1924. The distribution of the partial correlation coefficient. Metron 3, 329–332.

Fitzgibbon, L., Dowe, D., Vahid, F., 2004. Minimum message length autoregressive model order selection. In: Intelligent Sensing and Information Processing. pp. 439–444.

Fox, J., 1947. Applied regression analysis, linear models, and related methods. SAGE publications.

Friston, K., 1994. Functional and effective connectivity in neuroimaging: A synthesis. Human Brain Mapping 2, 56–78.

Friston, K., 2009. Causal modelling and brain connectivity in functional magnetic resonance imaging. Public Library of Science Biology, 220–225.

Friston, K., Büchel, C., 2007. Functional connectivity: Eigenimages and multivariate analyses. In: Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W. (Eds.), Statistical Parametric Mapping. Academic Press, Ch. Functional Connectivity: Eigenimages and multivariate analyses, pp. 492–507.

Friston, K., Büchel, C., Fink, G., Morris, J., Rolls, E., Dolan, R., 1997. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6, 218–229.

Friston, K., Frith, C., Liddle, P., Frackowiak, 1993. Functional connectivity: The principal-component analysis of large data sets. Journal of Cerebral Blood Flow & Metabolism 13, 5–14.

Friston, K., Glaser, D., Mechelli, A., Turner, R., C.Price (Eds.), 2003a. Hemodynamic Modeling. Academic Press, pp. 178–192.

Friston, K., Harrison, L., Penny, W., 2003b. Dynamic causal modelling. NeuroImage 19, 1273–1302.

Friston, K., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Human Brain Mapping 1, 153–171.

Friston, K., Williams, S., Howard, R., Frackowiak, R., Turner, R., 1996. Movement-related effect inf fMRI time-series. Magnetic resonance in medicine 35 (3), 346–355.

Fuller, W., 1996. Introduction to statistical time series, 2nd Edition. Wiley.

Gaglianese, A., Costagli, M., Bernardi, G., Sani, L., Ricciardi, E., Pietrini, P., 2011. Conditional Granger causality analysis of fMRI data shows a direct connection from LGN to hMT+ bypassing V1. In: Proceedings of the International Society for Magnetic Resonance in Medicine. p. 3554.

Gates, K., Molenaar, P., Hillary, F., Ram, N., Rovine, F., 2009. Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. NeuroImage 50 (3), 1118–1125.

Gautama, T., Hulle, M. V., 2004. Optimal spatial regularisation of autocorrelation estimates in fMRI analysis. NeuroImage 23 (3), 1203–1216.

Genovese, C., Lazar, N., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 4, 870–878.

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association 77, 304–313.

Geweke, J., 1984. Measures of conditional linear dependence and feedback between time series. Journal of the American Statistical Association 79, 907–915.

Glover, G., Krueger, G., 2002. Optimum voxel size in bold fMRI. In: Proceedings of the International Society for Magnetic Resonance in Medicine. p. 240.

Goebel, R., Roebroeck, A., Kim, D., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modelling and Granger causality mapping. Magnetic Resonance Imaging 21, 1251–1261.

Goldenshluger, A., Zeeviy, A., 2001. Non-asymptotic bounds for autoregressive time series modeling. Annals of Statistics 29 (2), 417–444.

Goltz, F., 1881. Discussion on the localization of function in the cortex cerebri. In: Transactions of the 7th International Medical Congress. pp. 218–228.

Goncalves, M., Hall, D., Johnsrude, I., Haggard, M., 2001. Can meaningful effective connectivities be obtained between auditory cortical regions? NeuroImage 14, 1353–1360.

Gonzalez-Lima, F., McIntosh, A., 1994. Neural network interactions related to auditory learning analyzed with structural equation modeling. Human Brain Mapping 2, 23–44.

Granger, C., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Granger, C., Hatanaka, M., 1964. Spectral analysis of economic time series. Princeton University Press, Princeton.

Granger, C., Newbold, P., 1974. Spurious regressions in econometrics. Journal of Econometrics 2, 111–112.

Greene, W., 2008. Econometric Analysis, 6th Edition. Pearson/Prentice Hall.

Greicius, M., Krasnow, B., Reiss, A., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proceedings of the National Academy of Sciences USA 100 (1), 253–258.

Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI data. Magnetic Resonance in Medicine 34, 910–914.

Guo, S., Seth, A., Kendrick, K., Zhou, C., Feng, J., 2008. Partial Granger causality - eliminating exogenous inputs and latent variables. Journal of Neuroscience Methods 172, 79–93.

Hajnal, J., Myers, R., Oatridge, A., Schwieso, J., Young, I., Bydder, G., 1994. Artifacts due to stimulus correlated motion in functional imaging of the brain. Magnetic Resonance in Medicine 31, 283–291.

Hamada, M., Wilson, A., Reese, C., Martz, H., 2008. Bayesian reliability. Springer.

Hamilton, J., 1994. Time Series Analysis, 1st Edition. Princeton University Press.

Hampson, M., Peterson, B., Skudlarski, P., Gatenby, J., Gore, J., 2002. Detection of functional connectivity using temporal correlations in MR images. Human Brain Mapping 15, 247–262.

Hannan, E., Quinn, B., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society, B 41, 190–195.

Hansen, L., Purushotham, A., Kim, S., 2003. Testing competing hypothesis about single trial fMRI, 1st Edition. Massachusetts Institute of Technology.

Hanson, S., Bly, B., 2001. The distribution of BOLD susceptibility effects in the brain is non-Gaussian. NeuroReport 12 (9), 1971–1977.

Härdle, W., Simar, L., 2007. Applied multivariate statistical analysis, 2nd Edition. Springer.

Harrison, L., Penny, W., Friston, K., 2003. Multivariate autoregressive modeling of fMRI time series. NeuroImage 19, 1477–1491.

Hashemi, R., Bradley, W., Lisanti, C., 2010. MRI: The basics, 3rd Edition. Lippincott Williams and Wilkins.

Henson, R., Friston, K., 2007. Statistical parametric mapping: The analysis of functional brain images. Elsevier, pp. 178–192.

Hinrichs, H., Heinze, H., Schoenfeld, M., 2006. Causal visual interactions as revealed by an information theoretic measure and fMRI. NeuroImage 31 (3), 1051–1060.

Horwitz, B., 2005. Integrating neuroscientific data across spatiotemporal scales. Comptes Rendus Biologies 328 (2), 109–118.

Horwitz, B., Grady, C., Haxby, J., Ungerleider, L., Schapiro, M., Mishkin, M., Rapoport, S., 1992. Functional associations among human posterior extrastriate brain regions during object and spatial vision. Journal of Cognitive Neuroscience 4, 311–32.

Howseman, A., Josephs, O., Rees, G., Friston, K., 1997. Special issues in functional magnetic resonance imaging. Tech. rep., The Wellcome Department of Cognitive Neurology, Institute of Neurology, London.

Hoyle, R., 1995. Structural equation modeling: Concepts, issues, and applications. Sage Publications.

Hu, L., Bentler, P., Kano, Y., 1992. Can test statistics in covariance structure analysis be trusted? Psychological Bulletin 112, 351–362.

Husain, F., Tagamets, M., Fromm, S., Braun, A., Horwitz, B., 2004. Relating neuronal dynamics for auditory object processing to neuroimaging activity: A computational modeling and an fMRI study. NeuroImage 21, . 21, 1701–1720.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: Algorithms and applications. Neural Networks 13, 411–430.

Ifeachor, E., Jervis, B., 1993. Digital system processing: A practical approach. Addison-Wesley.

Jenkinson, M., Bannister, P., Brady, J., , Smith, S., 2002. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17 (2), 825–841.

Jensen, E., Martinez, P., Litvan, H., Vereecke, H., Rodriguez, B., Struys, M., 2007. Recent advances in AEP/EEG indices for estimating hypnotic depth during general anesthesia. In: Akay, M. (Ed.), Handbook of Neural Engineering. John Wiley & Sons, Inc., pp. 535–553.

Jensen, J., 1906. Sur les fonctions convexes et les ingalits entre les valeurs moyennes. Acta Mathematica 30 (1), 175–193.

Jezzard, P., 1999. Sources of distortion in functional MRI data. Human Brain Mapping 8, 80–85.

Johnson, A., Proctor, R., 2004. Attention: Theory and practice, 1st Edition. Sage Publications Ltd.

Junning, L., Wang, Z., McKeown, M., 2008. Controlling the false discovery rate in modeling brain functional connectivity. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP). pp. 2105–2108.

Jreskog, K., Srbom, D., 1996. LISREL 8 User's Reference Guide. Scientific Software International, Chicago, 8th Edition.

Kammler, D., 2000. A First Course in Fourier Analysis, 22nd Edition. Prentice-Hall.

Kayser, A., Sun, F., D'Esposito, M., 2009. A comparison of Granger causality and coherence in fMRI-based analysis of the motor system. Human Brain Mapping 30, 3475–94.

Kenny, D., Judd, C., 1984. Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin 96, 201–210.

Kiebel, S., Kloppel, S., Weiskopf, N., Friston, K., 2007. Dynamic causal modeling: a generative model of slice timing in fMRI. NeuroImage 34 (4), 1487–1496.

Kim, J., Zhu, W., Chang, L., Bentler, P., Ernst, T., 2007. Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. Human Brain Mapping 28, 85–93.

Kim, S., Richter, W., Ugurbil, K., 1997. Limitations of temporal resolution in functional MRI. Magnetic Resonance in Medicine 37, 631–636.

Kimberg, D., 2008. A functional connectivity primer. Center for Functional Neuroimaging University of Pennsylvania.

Kline, R., 2005. Principles and practice of structural equation modeling, 2nd Edition. The Guilford Press.

Kowalski, C., 1972. On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. The Journal of the Royal Statistical Society, Series C, 1–12.

Kriegeskorte, N., Bodurka, J., Bandettini, P., 2008. Artifactual time-course correlations in echo-planar fMRI with implications for studies of brain function. International Journal of Imaging Systems and Technology 18 (5-6), 345–349.

Kruger, G., Kastrup, A., Glover, G., 2001. Neuroimaging at 1.5 T and 3.0 T: Comparison of oxygenation-sensitive magnetic resonance imaging. Magnetic Resonance in Medicine 45, 595–604.

Kruggel, F., Pelegrini-Issac, M., Benali, H., 2002. Estimating the effective degrees of freedom in univariate multiple regression analysis. Medical Image Analysis 6, 63–75.

Kumar, R., 2009. Signals and systems. PHI Learning Pvt. Ltd.

Lauterbur, P., 1973. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. Nature 242, 190–191.

Le, T., Hu, X., 1995. Potential pitfalls of principal component analysis in fMRI. In: Proceedings of the Magnetic Resonance in Medicine. p. 820.

Lee, L., Friston, K., Horwitz, B., 2006. Large-scale neural models and dynamic causal modelling. NeuroImage 30, 1243–1254.

Li, X., Marrelec, G., Hess, R., Benali, H., 2010. A nonlinear identification method to study effective connectivity in functional MRI. Medical Image Analysis 14, 30–38.

Liao, W., Mantini, D., Zhang, Z., Pan, Z., Ding, J., Gong, Q., Yang, Y., Chen, H., 2010. Evaluating the effective connectivity of resting state networks using conditional Granger causality. Biological Cybernetics 102, 57–69.

Liew, V., Apr 2004. On autoregressive order selection criteria. Computational Economics 0404001, The Economics Working Paper Archive.

Lindquist, M. (Ed.), June 2011. The network of network methods, and a case study of SEM. Organization of Human Brain Mapping, Quebec, Canada.

Locascio, J., Jennings, P., Moore, C., Corkin, S., 1997. Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. Human Brain Mapping 5, 168–193.

Londei, A., D'Ausilio, A., Basso, D., Belardinelli, M., 2006. A new method for detecting causality in fMRI data of cognitive processing. Cognitive Processing 7, 42–52.

Londei, A., DAusilio, A., Basso, D., Sestieri, C., Gratta, D., Romani, G., Belardinelli, M. O., 2007. Brain network for passive word listening as evaluated with ICA and Granger causality. Brain Research Bulletin 72, 284–292.

Long, C., Brown, E., Triantafyllou, C., Aharon, I., Wald, L., Solo, V., 2005. Nonstationary noise estimation in functional MRI. NeuroImage 28 (4), 890–903.

Lovell, M., 2008. A simple proof of the Frisch-Waugh-Lovell theorem. Journal of Economic Education 39, 88–91.

Lowe, M., Mock, B. J., Sorenson, J., 1998. Functional connectivity in single and multislice echoplanar imaging using resting state fluctuations. NeuroImage 7, 119–132.

Luca, M., Beckmann, C., Stefano, N., Matthews, P., Smith, S., 2006. FMRI resting state networks define distinct modes of long-distance interactions in the human brain. NeuroImage 29, 1359–1367.

Lund, T., Madsen, K., Sidaros, K., Luo, W., Nichols, T., 2006. Non-white noise in fMRI: Does modelling have an impact. NeuroImage 29, 54–66.

Luo, H., Puthusserypady, S., 2007. fMRI data anaysis with nonstationary noise models: a Bayesian approach. IEEE Transactions on Biomedical Engineering 54 (9), 1621–1630.

Lupton, R., 1993. Statistics in theory and practice, 1st Edition. Princeton University Press.

Ma, L., Wang, B., Chen, X., Xiong, J., 2007. Detecting functional connectivity in the resting brain: A comparison between ICA and CCA. Magnetic Resonance Imaging 25, 47–56.

Marchini, J., Smith, S., 2003. On bias in the estimation of autocorrelations for fMRI voxel time-series analysis. NeuroImage 18 (1), 83–90.

Marrelec, G., Daunizeau, J., Pelegrini-Issac, M., Doyon, J., Benali, H., September 2005. Conditional correlation as a measure of mediated interactivity in fMRI and MEG/EEG. IEEE Transactions on Signal Processing 53 (9), 3503–3516.

Marrelec, G., Horwitz, B., Kim, J., Pelegrini-Issac, M., Benali, H., Doyon, J., 2007. Using partial correlation to enhance structural equation modeling of functional MRI data. Magnetic Resonance Imaging 25 (8), 1181–1189.

Marrelec, G., Kim, J., Doyon, J., Horwitz, B., 2009. Large-scale neural model validation of partial correlation analysis for effective connectivity investigation in functional MRI. Human Brain Mapping 30 (3), 941–950.

Marrelec, G., Krainik, A., Duffau, H., Pelegrini-Issac, M., Lehricy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional MRI. NeuroImage 32 (1), 228–237.

McIntosh, A., Gonzalez-Lima, F., 1994. Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping 2, 2–22.

McIntosh, A., Grady, C., Ungerleider, L., Haxby, J., Rapoport, S., Horwitz, B., 1994a. Network analysis of cortical visual pathways mapped with PET. Journal of Neuroscience 14, 655–666.

McIntosh, A., Horwitz, B., Haxby, J., Ungerleider, L., Grady, C., 1994b. Functional cortical networks during short-term recognition memory for faces. Society for Neuroscience 20, 362.

McKeown, Sejinowski, B., 1998. Analysis of fMRI data by blind separation in to separate independent spatial components. Human Brain Mapping 6, 160–188.

McQuarrie, A., Tsai, C., 1998. Regression and time series model selection. World. Scientific Publishing Co., 455.

Millioz, F., Martin, N., 2010. Estimation of a white gaussian noise in the short time Fourier transform based on the spectral kurtosis of the minimal statistics: application to underwater noise. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. Dallas, Texas, United States, pp. 5638–5641.

Moeller, J., Strother, S., 1991. A regional covariance approach to the analysis of functional patterns in position emission tomographic data. Journal of Cerebral Blood Flow & Metabolism 11, A121–A135.

Moler, C., 2004. Numerical computing with MATLAB. Society for Industrial and Applied Mathematics.

Moon, Y., Rajagopalan, B., Lall, U., 1995. Estimation of mutual information using kernel density estimators. Physical Review E 52, 2318–2321.

Moonen, C., Bandettini, P. (Eds.), 2000. Functional MRI, 1st Edition. Springer-Verlag.

Mulert, C., Lemieux, L. (Eds.), 2010. EEG-fMRI: Physiological Basis, Technique and Applications. Springer.

Niazy, R., Xie, J., Miller, K., Beckmann, C., Smith, S., 2011. Spectral characteristics of resting state networks. Progress in Brain Research 193, 259–276.

Oehlert, G., 1992. A note on the delta method. The American Statistician 46 (1), 27–29.

Ogawa, S., Lee, T., Nayak, A., Glynn, P., 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magnetic Resonance in Medicine 14, 68–78.

Oikonomou, V., Tripoliti, E., Fotiadis, D., 2009. A sparse linear model for the analysis of fMRI data with non stationary noise. In: 4th International IEEE EMBS Conference on Neural Engineering. pp. 253–257.

Papoulis, A., Pillai, S., 2002. Probability, random variables and stochastic processes. McGraw Hill.

Park, C., Chang, W., Ohn, S., Kim, S., Bang, O., Pascual-Leone, A., Kim, Y., 2011. Longitudinal changes of resting-state functional connectivity during motor recovery after stroke. Stroke 42 (5), 1357–62.

Park, C., Lazar, N., Ahn, J., Sornborger, A., 2010. A multiscale analysis of the temporal characteristics of resting-state fMRI data. Journal of Neuroscience Methods 193 (2), 334–342.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2, 559–572.

Pearson, K., Jeffery, G., Elderton, E., 1929. On the distribution of the first product moment-coefficient, in samples drawn from an indefinitely large normal population. Biometrika 21, 164–201.

Pedrick, G., 1994. A first course in analysis. Springer-Verlag.

Penny, W., Roberts, S., 2002. Bayesian multivariate autoregressive models with structured priors. In: IEE Proceedings - Vision Image and Signal Processing. Vol. 149. pp. 33–41.

Penny, W., Stephan, K., Mechelli, A., Friston, K., 2004a. Comparing dynamic causal models. NeuroImage 22, 1157–1172.

Penny, W., Stephan, K., Mechelli, A., Friston, K., 2004b. Modelling functional integration: a comparison of structural equation and dynamic causal models. NeuroImage 23.

Penny, W. D., Holmes, A. J., Friston, K., 2004c. Random effects analysis, 2nd Edition. Academic Press.

Petersson, K., Nichols, T., Poline, J., Holmes, A., 1999. Statistical limitations in functional neuroimaging ii. signal detection and statistical inference. Philosophical Transactions of the Royal Society 354, 1261–1281.

Poldrack, R., Mumford, J., Nichols, T., 2011. Handbook of functional MRI data analysis, 1st Edition. Cambridge University Press.

Prandoni, P., Vetterli, M. (Eds.), 2008. Signal processing for communications. CRC Press.

Purdon, P., Weisskoff, R., 1998. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. Human Brain Mapping 6, 239–249.

Rahman, N., 1968. A course in theoretical statistics. Charles Griffin and Company.

Rezek, A., Roberts, S., 1997. Parametric model order estimation: a brief review. In: Model Based Digital Signal Processing Techniques in the Analysis of Biomedical Signals. pp. 1–6.

Rissanen, J., 2007. Information and complexity in statistical modeling, 1st Edition. Vol. 3. Springer Science+Business Media.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 230–242.

Rogers, B. P., Katwal, S. B., Morgan, V. L., Asplund, C. L., Gore, J. C., 2010. Functional MRI and multivariate autoregressive models. Magnetic Resonance Imaging 28, 1058–1065.

Rogers, B. P., Morgan, V. L., Newton, A., Gore, J. C., 2007. Assessing functional connectivity in the human brain by fMRI. Magnetic Resonance Imaging 25, 1347–1357.

Roy, Sherrington, 1890. On the regulation of the blood supply of the brain. Journal of Physiology 11, 85–108.

Russell, T., Zelaya, F., Bressan, R., Bandettini, P., 2002. Functional neuroimaging: an introduction to the technology, methodology, interpretation, and applications, 1st Edition. Dunitz Press, pp. 1–50.

Rykhlevskaia, E., Fabiani, M., Gratton, G., 2005. Lagged covariance structure models for studying functional connectivity in the brain. NeuroImage 30, 1203–1218.

Said, E., Dickey, D., 1984. Testing for unit roots in autoregressive moving average models of unknown order. Biometrika 71, 599–607.

Saito, Harashima, 1981. Tracking of information with-in multichannel record: Causal analysis in EEG. Vol. 35. North-Holland Biomedical Press, pp. 133–146.

Salvador, R., Martinez, A., Pomarol-Clotet, E., Sarro, S., Suckling, J., Bullmore, E., 2007. Frequency based mutual information measures between clusters of brain regions in functional magnetic resonance imaging. NeuroImage 35, 83–88.

Salvador, R., Suckling, J., Coleman, M., Pickard, J., Menon, D., Bullmore, E., 2005a. Neurophysiological architecture of functional magnetic resonance images of human brain. Cerebral Cortex 15 (9), 1332–1342.

Salvador, R., Suckling, J., Schwarzbauer, C., Bullmore, E., 2005b. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Philosophical Transactions of the Royal Society B 360, 937–946.

Sarty, G., 2006. Computing brain activity maps from fMRI time-series images. Cambridge.

Sato, J., Junior, E., Takahashi, D., Felix, M., Brammer, M., Morettin, P., 2006a. A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. NeuroImage 31, 187–196.

Sato, J., Takahashi, D., Arcuri, S., Sameshima, K., Morettin, P., Baccala, L., 2009. Frequency domain connectivity identification: An application of partial directed coherence in fMRI. Human Brain Mapping 30, 452–461.

Sato, J., Takahashi, D., Cardoso, E., Martin, M., Junior, E., Morettin, P., 2006b. Intervention models in functional connectivity identification applied to fMRI. International Journal of Biomedical Imaging 2006, 1–7.

Satterthwaite, F., 1946. An approximate distribution of estimates of variance components. Biometrics Bulletin 2, 110–114.

Schumacker, R., Lomax, R., 2004. A beginner's guide to structural equation modeling, 2nd Edition. Lawrence Erlbaum Associates.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Shapiro, S., Wilk, M., 1965. An analysis of variance test for normality (complete samples). Biometrika 52 (3), 591.

Shin, K., Hammond, J., 2008. Fundamentals of signal processing for sound and vibration engineers. Vol. 10. John Wiley and Sons Ltd.

Shittu, O., Asemota, M., 2009. Comparison of criteria for estimating the order of autoregressive process: A Monte Carlo approach. European Jounal of Scientific Research 30 (3), 409–416.

Smirnov, N., 1948. Tables for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistic 19, 279.

Smith, S., Miller, K., Salimi-Khorshidi, G., Webster, M., Beckmann, C., Nichols, T., Ramsey, J., Woolrich, M., 2011. Network modelling methods for fMRI. NeuroImage 54 (2), 875–891.

Stejskal, E., 1965. Use of spinechoes in a pulsed magnetic-field gradient to study anisotropic restricted diffusion and flow. Journal of Chemical Physics 43.

Storkey, A., Simonotto, E., Whalley, H., Lawrie, S., Murray, L., McGonigle, D., 2007. Learning structural equation models for fMRI. Advances in Neural Information Processing Systems 19, 1329–1336.

Strother, S., 2006. Evaluating fMRI preprocessing pipelines. Engineering in Medicine and Biology Magazine 25, 27–41.

Sun, F. T., Miller, L. M., D'Esposito, M., 2004. Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. NeuroImage 21, 647–658.

Tagaris, G., Richter, W., Kim, S., Georgopoulos, A., 1997. Box-Jenkins intervention analysis of functional magnetic resonance imaging data. Neuroscience Research 27, 289–294.

Tana, M., Bianchi, A., Vitali, P., Villani, F., Cerutti, S., August 2008. Exploring interregional brain interactivity in temporal lobe epilepsy using partial correlation analysis of fMRI data. In: 30th Annual International IEEE EMBS Conference. IEEE, Vancouver, British Columbia, Canada, pp. 4423–4426.

Tanaka, J., 1987. "how big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. Child Development 58 (1), 134–146.

Tanase, C., O'Hara, J., Davis, D., Boada, F., Buonocore, M. H., Carter, C. S., 2011. Identifying and separating the RF fluctuations from the measurement noise. In: Proceedings of the International Society of Magnetic Resonance in Medicine. Vol. 19. p. 3629.

Theiler, J., Galdrikian, B., Longtin, Eubank, S., Farmer, J., 1990. Using surrogate data to detect nonlinearity in time series. In: North Atlantic Treaty Organization (NATO) advanced research workshop on nonlinear modelling and forecasting. Santa Fe, United States, pp. 85–77.

Thomas, C., Harshman, R., Menon, R., 2002. Noise reduction in BOLD-based fMRI using component analysis. NeuroImage 17, 1521–1537.

Thulborn, K., Waterton, J., Matthews, P., Radda, G., 1982. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. Biochimica et Biophysica Acta 714 (2), 265–270.

Tong, Y., Lindsey, K., Frederick, B., 2011. Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI. Journal of Cerebral Blood Flow and Metabolism.

Uddin, L., Kelly, A., Biswal, B., Castellanos, F., Milham, M., 2009. Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. Human Brain Mapping 30, 625–637.

Valdes-Sosa, P., 2004. Spatio-temporal autoregressive models defined over brain manifolds. Neuroinformatics 2, 239–250.

Valdes-Sosa, P., Sanchez-Bornot, J., Lage-Castellanos, A., Vega-Hernandez, M., Bosch-Bayard, J., Melie-Garcia, L., Canales-Rodriguez, E., 2005. Estimating brain functional connectivity with sparse multivariate autoregression. Philosophical Transactions of the Royal Society B 360, 969–981.

van de Ven, V., Formisano, E., Prvulovic, D., Roeder, C., Linden, D., 2004. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Human Brain Mapping 22, 165–78.

Viswanatham, A., Freeman, R., 2007. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. Nature Neuroscience 10, 1308–1312.

Vollmar, C., OMuircheartaigh, J., Barker, G., Symms, M., Thompson, P., Kumari, V., Duncan, J., Janz, D., Richardson, M., Koepp, M., 2011. Motor system hyperconnectivity in juvenile myoclonic epilepsy: a cognitive functional magnetic resonance imaging study. Brain 134 (6), 1710–1719.

Waldorp, L., Christoffels, I., van de Ven, V., 2011. Effective connectivity of fMRI data using ancestral graph theory: dealing with missing regions. NeuroImage 54 (4), 2695–2705.

Wang, Y., Xia, J., 2007. Functional interactivity in fMRI using multiple seeds correlation analyses - novel methods and comparisons. In: Information Processing in Medical Imaging. 20th International Conference. pp. 147–59.

Wang, Y., Xia, J., 2009. Unified framework for robust estimation of brain networks from fMRI using temporal and spatial correlation analyses. IEEE Transactions on Medical Imaging 28, 1296–1307.

Weibull, A., Gustavsson, H., Mattsson, S., Svensson, J., 2008. Investigation of spatial resolution, partial volume effects and smoothing in functional MRI using artificial 3D time series. NeuroImage 41 (2), 346–353.

Weisskoff, R., 1996. Simple measurement of scanner stability for functional NMR imaging of activation in the brain. Magnetic Resonance in Medicine 36, 643–645.

Welch, B., 1947. The generalization of "Student's" problem when several different population variances are involved. Biometrika 34, 28–35.

Westbrook, C., 2010. MRI at a glance, 2nd Edition. Wiley-Blackwell.

Westbrook, C., Roth, C., Talbot, J., 2005. MRI in practice, 3rd Edition. Blackwell Publishing.

Whittaker, J., 2009. Graphical models in applied multivariate statistics, 3rd Edition. Wiley Publishing.

Wiener, 1956. The theory of prediction. In: Beckenbach, E. F. (Ed.), Modern mathematics for the engineer. Vol. 1. McGraw-Hill, pp. 165–190.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1 (6), 80–83.

Wink, A., Roerdink, J., 2006. BOLD noise assumptions in fMRI. International Journal of Biomedical Imaging 2006, 1–11.

Wise, R., Newling, B., Gates, A., Xing, D., Carpenter, T., Hall, L., 1996. Measurement of pulsatile flow using MRI and a bayesian technique of probability analysis. Magnetic Resonance Imaging 14 (2), 173– 185.

Woodward, P., 2001. MRI for technologists, 2nd Edition. McGraw Hill.

Woolrich, M., Ripley, B., Brady, M., Smith, S., 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. NeuroImage 14, 1370–1386.

Woolrich, M. W., Jenkinson, M., Brady, J. M., Smith, S. M., 2004. Fully bayesian spatio-temporal modeling of fMRI data. IEEE Transactions on Medical Imaging 23 (2), 213–231.

Worsley, K., Chen, J., Lerch, J., Evans, A., May 2005. Comparing functional connectivity via thresholding correlations and singular value decomposition. Philosophical Transactions of the Royal Society B 360, 913–920.

Worsley, K., Friston, K., 1995. Analysis of fMRI time-series revisted - again. NeuroImage 2, 173–181.

Wu, T., Long, X., Wang, L., Hallett, M., Zang, Y., Li, K., Chan, P., 2011. Functional connectivity of cortical motor areas in the resting state in parkinson's diseas. Human Brain Mapping 32 (9), 1443–1457.

Xiong, J., Parsons, L., Gao, J., Fox, P., 1999. Interregional connectivity to primary motor cortex revealed using MRI resting state images. Human Brain Mapping 8, 151–156.

Yang, K., Rajapakse, J. C., 2004. ICA gives higher-order functional connectivity of brain. Neural Information Processing - Letters and Review 2, 27–32.

Yue, Y., Loh, J., Lindquist, M., 2010. Adaptive spatial smoothing of fMRI images. Statistics and Its Interface 3, 3–13.

Yule, G., 1927. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. Philosophical Transactions of the Royal Society 226, 267–298.

Zarahn, E., Aguirre, G. K., DEsposito, M., 1997. Empirical analyses of bold fmri statistics. NeuroImage 5 (3), 179–197.

Zhang, J., Nassef, A., Mahfouf, M., Linkens, D., El-Samahy, E., Hockey, G., Nickel, P., Roberts, A., 2006. Modelling and analysis of HRV under physical and mental workloads. In: Feng, D., Zaytoon, J. (Eds.), Modelling and Control in Biomedical Systems 2006. Elsevier, pp. 189–193.

Zhong, Y., Wang, H., Lu, G., Zhang, Z., Jiao, Q., Liu, Y., 2009. Detecting functional connectivity in fMRI using PCA and regression analysis. Brain Topography 22, 134–44.

Zhou, Z., Chen, Y., Ding, M., Wright, P., Lu, Z., Liu, Y., 2009a. Analyzing brain networks with PCA and conditional Granger causality. Human Brain Mapping 30, 2197–2206.

Zhou, Z., Ding, M., Chen, Y., Wright, P., Lu, Z., Liu, Y., 2009b. Detecting directional influence in fMRI connectivity analysis using PCA based Granger causality. Brain Research 1289, 22–29.

Zhuang, J., LaConte, S., Peltier, S., Zhang, K., Hu, X., 2005. Connectivity exploration with structural equation modeling: An fMRI study of bimanual motor coordination. NeuroImage 25, 462–470.

Zhuang, J., Peltier, S., He, S., LaConte, S., Hu, X., 2008. Mapping the connectivity with structural equation modeling in an fMRI study of shape-from-motion task. NeuroImage 42, 799–806.