Published in final edited form as:

Neuroimage. 2014 November 1; 101: 531–546. doi:10.1016/j.neuroimage.2014.06.052.

Evaluating Dynamic Bivariate Correlations in Resting-state fMRI: A comparison study and a new approach

Martin A. Lindquist^{a,*}, Yuting Xu^a, Mary Beth Nebel^b, and Brain S. Caffo^a

^aDepartment of Biostatistics, Johns Hopkins University, USA

^bKennedy Krieger Institute, Johns Hopkins University, USA

Abstract

To date, most functional Magnetic Resonance Imaging (fMRI) studies have assumed that the functional connectivity (FC) between time series from distinct brain regions is constant across time. However, recently, there has been increased interest in quantifying possible dynamic changes in FC during fMRI experiments, as it is thought this may provide insight into the fundamental workings of brain networks. In this work we focus on the specific problem of estimating the dynamic behavior of pair-wise correlations between time courses extracted from two different regions of the brain. We critique the commonly used sliding-windows technique, and discuss some alternative methods used to model volatility in the finance literature that could also prove useful in the neuroimaging setting. In particular, we focus on the Dynamic Conditional Correlation (DCC) model, which provides a model-based approach towards estimating dynamic correlations. We investigate the properties of several techniques in a series of simulation studies and find that DCC achieves the best overall balance between sensitivity and specificity in detecting dynamic changes in correlations. We also investigate its scalability beyond the bivariate case to demonstrate its utility for studying dynamic correlations between more than two brain regions. Finally, we illustrate its performance in an application to test-retest resting state fMRI data.

Keywords

Functional connectivity; fMRI; Dynamics; Resting state; Dynamic conditional correlations

1 Introduction

To date, most functional Magnetic Resonance Imaging (fMRI) studies have implicitly assumed that the functional connectivity (FC) between time series from distinct brain regions is constant across time. Recently, there has been an increased interest in attempting to quantify the dynamic changes in FC during the course of an fMRI experiment (e.g., Allen et al. (2012); Chang and Glover (2010); Handwerker et al. (2012); Hutchison et al. (2013); Jones et al. (2012); Kiviniemi et al. (2011); Cribben et al. (2012)); particularly during resting state. Changes in both the strength and directionality of functional connections have been

*Corresponding author: 615 N. Wolfe Street, E3634; Baltimore, MD 21205, mlindqui@jhsp.edu.

observed to vary across experimental runs (Hutchison et al., 2013). It is thought that this temporally varying information may be used to provide insight into the fundamental properties of brain networks. For example, studies have found correlation between changes in resting-state FC and simultaneous recorded electrophysiological data, as well as behavioral data (Allen et al., 2012; Chang et al., 2013; Tagliazucchi et al., 2012; Thompson et al., 2012), indicating a possible neuronal origin to the observed variation.

Though it is of increasing importance, interpreting temporal fluctuations in FC can be difficult due to low signal-to-noise ratio, physiological artifacts and variation in BOLD signal mean and variance over time (Hutchison et al., 2013). For these reasons, it is often difficult to determine whether observed fluctuations in FC should be attributed to neuronal activity or whether they are simply due to random noise, and thus significant research is still needed in the area. In particular, there remains uncertainty regarding the appropriate analysis strategy to use and how to properly interpret results.

While numerous metrics for evaluating functional connectivity exist, in this work we focus exclusively on pair-wise correlations between time courses from two regions of the brain. The most common approach towards studying these types of dynamic correlations has been the so-called sliding-window approach (Allen et al., 2012; Chang and Glover, 2010; Handwerker et al., 2012). Here, a window of fixed length w is selected, and data within that window are used to compute the correlation coefficient. The window is thereafter moved step-wise across time, providing a time-varying measure of the correlation between brain regions.

Though extremely simple to implement, this approach has a number of potential shortcomings, including the use of arbitrary window lengths, an inability to deal with abrupt changes and the fact that it gives equal weight to all observations that lie less than w time points in the past, but no weight to older observations (Engle, 2002; Lebo and Box-Steffensmeier, 2008). In this work we evaluate the properties of the sliding-window technique, as well as introduce several alternative model-based approaches that we believe may ultimately prove more useful for analyzing dynamic correlations in fMRI.

The problem of studying time-varying variances and correlations between multivariate time series, while relatively recent in the neuroimaging literature, has been extensively studied in the finance literature during the past few decades (see, for example, Bauwens et al. (2006) for a nice overview). Financial time series often exhibit time-varying conditional standard deviations (typically referred to as volatility) and correlations. Here volatility is seen as a measure of the risk of certain assets, while correlations between time series play an important role in asset allocation, risk management and portfolio selection (Tsay, 2006).

In the finance literature, simple methods such as sliding-windows (often referred to as “rolling-windows”) and related techniques have been widely used to model dynamic correlations. However, it is generally accepted that time series models provide superior results (Hansen and Lunde, 2005) compared to rolling-window methods. In particular, generalized autoregressive conditional heteroscedastic (GARCH) models (Bollerslev, 1986) have been shown to efficiently model both dynamic variances and correlations. These

techniques model covariance in an analogous manner as fMRI noise is modeled in standard GLM analysis using more conventional time series analysis techniques such as autoregressive (AR) and autoregressive moving-average (ARMA) models (Purdon et al., 2001; Lindquist, 2008). Recently, Engle (2002) introduced a particular variant of the multivariate GARCH model, denoted the dynamic conditional correlation (DCC) model, which has been shown to be particularly effective for estimating both time-varying variances and correlations (Lebo and Box-Steffensmeier, 2008). In this approach all the parameters are effectively estimated through quasi-maximum likelihood methods and require no *ad hoc* parameter settings. The model uses a sequential estimation scheme and a parsimonious parameterization that allows it to estimate potentially large covariance structures. Hence, the method promises to be scalable to situations where one wants to study dynamic correlations between more than two brain regions.

It is our belief that lessons from finance may help inform future studies of dynamic correlation in the neuroimaging context as well. To this end, in this work we introduce multivariate volatility models to the neuroimaging literature. Further, we investigate the properties of these and other commonly used techniques in a series of simulation studies. Our results indicate that techniques such as DCC show great promise in the analysis of dynamic FC. We conclude by applying the method to test-retest resting state fMRI data.

2 Methods

In this section we begin by setting up the problem and continue by introducing several methods for estimating dynamic correlations, which we seek to compare in our simulation studies. These include both sliding-window type methods and model-based multivariate volatility methods.

2.1 Problem Set-up

Let us assume that we are interested in studying the relationship between two time series $y_{1,t}$ and $y_{2,t}$, measured over two separate regions of interest (ROIs) in the brain, at equally spaced time points $t = 0, \dots, T$. Let $\mathbf{y}_t = (y_{1,t}, y_{2,t})^T$ be a vector containing the values of both time series at time t and assume that

$$\mathbf{y}_t = \mu_t + \mathbf{e}_t, \quad (1)$$

where $\mu_t = (\mu_{1,t}, \mu_{2,t})^T$ is the conditional mean of $\mathbf{y}_t$ using all information in the time series observed up to time t , denoted $E_{t-1}(\mathbf{y}_t)$. The noise term $\mathbf{e}_t$ has mean zero and its conditional covariance matrix at time t can be expressed as

$$\Sigma_t = \begin{pmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{pmatrix}. \quad (2)$$

Here the diagonal terms represent the conditional variance of $y_{i,t}$ using all information in the time course observed up to time t , written $\sigma_{i,t}^2 = E_{t-1}((y_{i,t} - \mu_{i,t})^2)$ for $i = 1, 2$. The square

root of this entity, σ_{it} , is typically referred to as the volatility of the time series in the financial literature. The off-diagonal term is $\sigma_{12,t} = \sigma_{1,t}\sigma_{2,t}\rho_t$ where

$$\rho_t = \frac{E_{t-1}((y_{1,t} - \mu_{1,t})(y_{2,t} - \mu_{2,t}))}{\sqrt{E_{t-1}((y_{1,t} - \mu_{1,t})^2)E_{t-1}((y_{2,t} - \mu_{2,t})^2)}} \quad (3)$$

represents the conditional correlation coefficient. Under this definition the conditional correlation at time t relies on information that is observed up to time $t - 1$. Importantly, this quantity is guaranteed to lie in the interval $[-1, 1]$ for all possible realizations of these random variables, as well as, for any linear combination.

Throughout we assume without loss of generality that $\mu_t = \mathbf{0}$ and $\mathbf{y}_t = \mathbf{e}_t$. Under this assumption, Eq. 3 simplifies as follows:

$$\rho_t = \frac{E_{t-1}(y_{1,t}y_{2,t})}{\sqrt{E_{t-1}(y_{1,t}^2)E_{t-1}(y_{2,t}^2)}}. \quad (4)$$

The conditional covariance matrix defined in Eq. 2 can alternatively be written in matrix form as

$$\Sigma_t = D_t R_t D_t \quad (5)$$

where D_t is a diagonal matrix consisting of the conditional standard deviations of the time series, i.e. $D_t = \text{diag}\{\sigma_{1,t}, \sigma_{2,t}\}$ and R_t is the correlation matrix,

$$R_t = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}. \quad (6)$$

Throughout this paper, our primary concern will be in developing methods for estimating the components of the conditional covariance matrix, with the particular goal of estimating $\sigma_{1,t}^2$, $\sigma_{2,t}^2$ and ρ_t . We begin by discussing sliding-windows techniques and move on to discuss multivariate volatility models.

2.2 Sliding-Window Techniques

Perhaps the simplest approach towards estimating the elements of the covariance matrix is to use the sliding-window technique. In particular, sliding-window correlations have received substantial interest in the recent neuroimaging literature (Allen et al., 2012; Chang and Glover, 2010; Handwerker et al., 2012). Here, a time window of fixed length w is selected, and data points within that window are used to calculate the correlation coefficients. The window is thereafter moved across time and a new correlation coefficient is computed for each time point.

In recent work, Chang and Glover (2010) define the sliding-window correlation at time t as follows:

$$\hat{\rho}_t = \frac{\sum_{s=t}^{t+w-1} y_{1,s} y_{2,s}}{\sqrt{(\sum_{s=t}^{t+w-1} y_{1,s}^2) (\sum_{s=t}^{t+w-1} y_{2,s}^2)}}. \quad (7)$$

According to this definition the correlation at time t is based on w future measurements of the time courses. Though it should be noted that in resting state experiments, which is the focus of their work, the exact timing of the dynamic correlation is not meaningful. We still prefer to define the window using only past values as it provides a more suitable estimate of the conditional correlation defined in Eq. 4.

For this reason, we define the sliding-window so that it gives equal weight to all observations that lie less than w time points in the past and zero weight to all older observations. Hence, the general form of the estimate is given by

$$\hat{\rho}_t = \frac{\sum_{s=t-w-1}^{t-1} (y_{1,s} - \hat{\mu}_{1,s}) (y_{2,s} - \hat{\mu}_{2,s})}{\sqrt{(\sum_{s=t-w-1}^{t-1} (y_{1,s} - \hat{\mu}_{1,s})^2) (\sum_{s=t-w-1}^{t-1} (y_{2,s} - \hat{\mu}_{2,s})^2)}}, \quad (8)$$

where $\hat{\mu}_{i,t}$, $i = 1, 2$ represents the estimated time-varying mean. Assuming $\mu = 0$, we can instead use

$$\hat{\rho}_t = \frac{\sum_{s=t-w-1}^{t-1} y_{1,s} y_{2,s}}{\sqrt{(\sum_{s=t-w-1}^{t-1} y_{1,s}^2) (\sum_{s=t-w-1}^{t-1} y_{2,s}^2)}}. \quad (9)$$

Though not commonly performed in the neuroimaging literature, it is a simple extension to use the sliding-window technique to estimate the conditional variance. We define, for the case when $\mu = 0$, the sliding-window variance at time t as follows:

$$\hat{\sigma}_{i,t}^2 = \frac{1}{w-1} \sum_{s=t-w-1}^{t-1} y_{i,s}^2. \quad (10)$$

While the sliding-window technique allows for a simple approach for exploring changes in connectivity, it has a number of obvious shortcomings. First, it gives equal weight to all observations less than w time points in the past and 0 weight to all others. Hence, the removal of a highly influential outlying data point will cause a sudden change in the dynamic correlation that may be mistaken for an important aspect of brain connectivity. To circumvent this problem, Allen et al. (2012) suggested the use of a tapered sliding-window. Here, the sliding-window (width = 22 TRs) is convolved with a Gaussian kernel ($\sigma = 3$ TRs). This allows points to gradually enter and exit from the window as it moves across time. It should be noted that they define t to be the middle of the subsequent window, thus giving equal weight to future and past values.

Second, the window length is typically chosen in an arbitrary manner. While methods exist for automatically selecting window lengths (see for example Ombao and Van Bellegem (2008) which proposes a method for automatically selecting the optimal window width for estimating local coherence), these methods have yet to find wide usage in the field. While it is beneficial to have a short window to better detect transient changes in connectivity, a large window is often necessary to allow for robust estimation of the correlation coefficient. Fig. 1 illustrates the effect of window length on the estimated conditional correlation fit to null data (i.e. two uncorrelated time courses). Note, that when using short windows, correlations are susceptible to large variations. In fact, a 95% confidence interval for null data will lie roughly between $\pm 2/\sqrt{w}$, which for a window length of 15 observations would roughly lie between $[-0.5, 0.5]$. Hence, using these settings random fluctuations can mistakenly be seen as meaningful time-varying correlations. The width of the 95% confidence interval is halved if we increase the window length four-fold to 60 observations. However, this comes at the cost of reduced sensitivity to minor changes in correlation.

2.3 Multivariate Volatility Models

In this section we discuss two multivariate volatility models commonly used in the finance literature. The first, denoted the exponential weighted moving average (EWMA) approach, shares some similarities to sliding-windows. However, it provides solutions to some of the more obvious shortcomings of that approach. The second, the DCC model, provides a parametric approach towards estimating dynamic correlations, much like auto-regressive (AR) and auto-regressive moving averages (ARMA) models provide a parametric approach towards modeling fMRI noise (Purdon et al., 2001).

2.3.1 EWMA—The first multivariate volatility model we introduce provides an alternative to the tapered sliding-window. The EWMA approach applies declining weights to past observations in the time series based on a parameter λ , and is based upon the following recursion

$$\Sigma_t = (1-\lambda)e_{t-1}e'_{t-1} + \lambda\Sigma_{t-1}, \quad (11)$$

where Σ_t is the conditional covariance matrix. This approach places the most weight on recent observations, and for each step away from t values become gradually down-weighted by a factor λ , before eventually being removed from further computations.

Decomposing the covariance matrix in Eq. 11, we can express the conditional variances and covariance as follows:

$$\begin{aligned} \sigma_{i,t}^2 &= (1-\lambda)e_{i,t-1}e'_{i,t-1} + \lambda\sigma_{i,t-1}^2 \\ &= (1-\lambda)y_{i,t-1}y'_{i,t-1} + \lambda\sigma_{i,t-1}^2 \end{aligned} \quad (12)$$

and

$$\sigma_{12,t} = (1-\lambda)e_{1,t-1}e_{2,t-1}' + \lambda\sigma_{12,t-1}. \quad (13)$$

The parameter λ must take values between 0 and 1, and it determines how responsive the estimate of the covariance matrix is to the most recent time points. A small value of λ gives high weight to recent time points, while a large value produces estimates that respond more gradually to new information. The value determines how many data points are included in the calculation and serves the equivalent purpose to the window size used in the sliding-window approach. Another important property of the approach is that as long as the recursion is initialized with a positive-definite matrix, it will remain so throughout the sequence.

Often the value of λ is set arbitrarily, with 0.94 being a popular value in the finance literature (Sheppard, 2012). However, if one assumes $\mathbf{y}_t$ to be bivariate Gaussian, it is straightforward to estimate λ through maximum likelihood estimation. Here, the log-likelihood function that should be maximized can be written:

$$\log\{L(\lambda)\} \propto -\frac{1}{2}\sum_{t=1}^T |\Sigma_t| - \frac{1}{2}\sum_{t=1}^T \mathbf{e}_t' \Sigma_t^{-1} \mathbf{e}_t. \quad (14)$$

This function can be maximized using any standard search algorithm and the value that λ takes at the optimum recorded.

Fig. 2 illustrates the effect of λ on the estimated conditional correlation fit to null data (i.e. two uncorrelated time courses). Note, that as λ gets smaller, less of the past data is used in the calculation of the correlation. Hence, if one uses small values of λ one is susceptible to similar problems that arise when using short window lengths when computing sliding-window correlations. Hence, it is beneficial to estimate this important parameter directly from the data.

Finally, it should be noted that a related EWMA approach has previously been used to detect dynamic changes in the mean BOLD response (Lindquist et al., 2007). There the goal was to detect so-called “change points” representing state-changes where the mean signal changes values, thus allowing one to model slowly varying processes with uncertain onset times and durations of underlying psychological activity. In contrast, this work considers dynamic correlations, which are perhaps more relevant for resting state analyses.

2.3.2 DCC—The second multivariate volatility model we introduce is the DCC model (Engle, 2002), which is an approach towards estimating the conditional variances and correlations that has become increasingly popular in the finance literature during the past decade. However, before introducing DCC, we must first discuss GARCH processes (Engle, 1982; Bollerslev, 1986), which are often used to model volatility in univariate time series. They provide flexible models for the variance in much the same manner that commonly used time series models, such as ARMA and AR, provide models for the mean. GARCH models express the conditional variance of a single time series at time t as a linear

combination of past values of the conditional variance and of the squared process itself. To illustrate, let us assume that we are observing a univariate process

$$y_t = \sigma_t \varepsilon_t \quad (15)$$

where ε_t is a $N(0, 1)$ random variable and σ_t represents the time-varying variance term we seek to model. In a GARCH(1,1) process the conditional variance is expressed as

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (16)$$

where $\omega > 0$, $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$. Here the term α controls the impact of past values of the time series on the variance and β controls the impact of past values of the conditional variance on its present value.

It is interesting to note that if we set $\omega = 0$, $\alpha = 1 - \lambda$ and $\beta = \lambda$, the GARCH(1,1) model expressed in Eq. 16 can be written:

$$\sigma_t^2 = (1 - \lambda) y_{t-1}^2 + \lambda \sigma_{t-1}^2 \quad (17)$$

which is equivalent to the EWMA model for the variance described in Eq. 12. Hence, though quite different in appearance, the GARCH(1,1) model provides a generalization of the EWMA model.

While GARCH(1,1) processes are the most widely used in practice, there exist a more general class of GARCH(p,q) models. Here the p most recent observations of y_t^2 and the q most recent estimates of σ_t^2 are used in the updated estimate of σ_t^2 . The model takes the form:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (18)$$

with $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \geq 0$, $\omega > 0$ and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$. While we include the general formulation for completeness, we note that models with large values of p and q may not be appropriate for fMRI data sampled at 2s time intervals, though with decreasing TR values they may become increasingly important. The parameters of the GARCH model can be estimated using maximum likelihood, and most statistical software packages have built-in functions for fitting GARCH models.

While many multivariate GARCH models exist that can be used to estimate dynamic correlations, it has been shown that the DCC model out-performs the rest (Engle, 2002). To illustrate the DCC approach, again assume $\mathbf{y}_t = \varepsilon_t$ is a bivariate mean zero time series with conditional covariance matrix Σ_t . The first order form of DCC can be expressed as follows:

$$\sigma_{i,t}^2 = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 \text{ for } i=1, 2 \quad (19)$$

$$\mathbf{D}_t = \text{diag}\{\sigma_{1,t}, \sigma_{2,t}\} \quad (20)$$

$$\varepsilon_t = \mathbf{D}_t^{-1} \mathbf{e}_t \quad (21)$$

$$\mathbf{Q}_t = (1 - \theta_1 - \theta_2) \bar{\mathbf{Q}} + \theta_1 \varepsilon_{t-1} \varepsilon_{t-1}' + \theta_2 \mathbf{Q}_{t-1} \quad (22)$$

$$\mathbf{R}_t = \text{diag}\{\mathbf{Q}_t\}^{-1/2} \mathbf{Q}_t \text{diag}\{\mathbf{Q}_t\}^{-1/2} \quad (23)$$

$$\Sigma_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \quad (24)$$

The DCC algorithm basically consists of two steps. In the first step (Eqs. 19–21), univariate GARCH(1,1) models are fit (Eq. 19) to each of the two univariate time series that make up $\mathbf{y}_t$, and used to compute standardized residuals (Eq. 21). In the second step (Eqs. 22–24), an EWMA-type method is applied to the standardized residuals to compute a non-normalized version of the time-varying correlation matrix $\mathbf{R}_t$ (Eq. 22). Here $\bar{\mathbf{Q}}$ represents the unconditional covariance matrix of ε_t and (θ_1, θ_2) are non-negative scalars satisfying $0 < \theta_1 + \theta_2 < 1$. Eq. 23 is simply a rescaling step to ensure a proper correlation matrix is created, while Eq. 24 computes the time-varying covariance matrix. The model parameters $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \theta_1, \theta_2)$ can be estimated using a two-stage approach. In the first stage, time-varying variances are estimated for each time series. In the second stage, the standardized residuals are used to estimate the dynamic correlations. This two-stage approach has been shown (Engle and Sheppard, 2001; Engle, 2002) to provide estimates that are consistent and asymptotically normal with a variance that can be computed using the generalized method of moments approach. has been shown (Engle and Sheppard, 2001; Engle, 2002) to provide estimates that are consistent and asymptotically normal with a variance that can be computed using the generalized method of moments approach.

Fig. 3 illustrates the estimated conditional correlation fit to null data (i.e. two uncorrelated time courses) using DCC. Clearly the method is able to closely follow the true value and appears to be a good candidate for use with fMRI data.

Due to the parametric nature of the model, confidence bands can be created for the dynamic correlation term by using the first two moments of the DCC parameters to simulate their empirical distribution under the assumption of joint normality. Monte-Carlo methods can then be used to repeatedly draw from the estimated distribution of the coefficients, refit the model and create confidence intervals for the dynamic correlations. Details of the estimation and inference procedure are outlined in the Appendix.

2.4 Simulations

To evaluate the performance of the different methods for estimating time-varying correlations, we performed a series of simulation studies. For the first three, we generated

random data $y_t = (y_{1,t}, y_{2,t})^T$ for each time point $t = 1, \dots, T$ using a mean-zero bivariate normal distribution. In each case the covariance matrix of the distribution was set to

$$\Sigma_t = \begin{pmatrix} 2 & p(t) \\ p(t) & 3 \end{pmatrix}$$

where the covariance term, $p(t)$, was allowed to vary across time for $t = 1, \dots, T$. This allowed us to control the dynamic relationship between the two time courses $y_{1,t}$ and $y_{2,t}$ throughout the time series. For each simulation, the value of $p(t)$ and T were set as follows:

1. $p(t) = 0$ for all values of $t = 1, \dots, T$. This represents the case where the time series are uncorrelated across the entire time course, i.e. null data. The number of time points T varied between 150, 300, 600 and 1000.
2. $p(t) = \sin(t/\)$ for $t = 1, \dots, T$, $\ = 1024/(2^k)$ and $k = 1, \dots, 4$. This represents a slowly varying periodic change in correlation. In this simulation we fix $T = 600$.
3. $p(t)$ is equal to a Gaussian kernel with mean 250 and standard deviation $15*k$ for $t = 1, \dots, T$ and $k = 1, \dots, 4$. Hence, $p(t)$ is non-zero in an interval that lies approximately within ± 3 standard deviations of 250. This represents a more transient state change from a state of no correlation to an enhanced state of positive correlation. In this simulation we fix $T = 600$.

For each choice of $p(t)$ and number of time points T , the simulations were repeated 1000 times. In each simulation, we fit the sliding-window technique using varying window lengths of 15, 30, 60 and 120 time points, in order to investigate the method's sensitivity to this parameter choice. We also fit the tapered sliding-window using a window length of 22 time points convolved with a Gaussian kernel with standard deviation of 3 time points, the EWMA approach (using the MLE of λ) and the first-order DCC model to the data. In total we evaluated seven different approaches, counting the four variants of the sliding-window approach.

In Simulation 1 we computed the maximum estimated correlation across time for each method and value of T . The goal was to investigate each method's potential for overstating the correlation found in the null data, with small values preferable to large. In addition, to evaluate the approach for computing confidence bounds for DCC we also computed 99% confidence intervals and evaluated their coverage of 0 across time. This was not performed for the other techniques, as methods for constructing confidence intervals are not readily available for them.

For each repetition of Simulations 1–3 we measured the mean square error (MSE) between the estimated correlation and the true value (i.e., $p(t)/\sqrt{6}$) to quantify the different methods' abilities to effectively track dynamic changes in connectivity. In addition, for each repetition we also computed what we refer to as the "oracle sliding window" (OSW) for comparison with the DCC approach. The OSW is defined as the sliding window whose value of w minimizes the MSE, i.e. that best fits the *true* underlying dynamic correlation.

The OSW was computed by searching across all window lengths ranging from 30 to 120 in order to find the optimum. It is important to note that the OSW is an idealized statistic that can only be estimated if the underlying truth is known, and thus it cannot be used on experimental data. However, in the context of a simulation study it allows us to compute how well the sliding window technique can theoretically do if we were to always choose the optimal window length in every given situation, and thus provides us with a “gold standard” for the approach.

Finally, in Simulation 4 we investigate the scalability of the DCC algorithm. Here we set the covariance matrix Σ_t to be an $N \times N$ matrix consisting of three equally sized block diagonal elements. Each of these consisted of an $N/3 \times N/3$ matrix with the value 2 in the diagonals and $p_i(t)$ in the off-diagonal elements, where $i = 1, 2, 3$ represents the three submatrices. The values of $p_i(t)$ were chosen as follows: $p_1(t) = 0$, $p_2(t) = \sin(t/512)$ and $p_3(t) = 0.5(1 + \sin(t/16))$ for $t = 1, \dots, T$, where $T = 600$. For a set of values for N ranging from 6 to 102 (in steps of 6), we generated random data y_t for each time point $t = 1, \dots, T$ using a mean-zero multivariate normal distribution with covariance matrix Σ_t . For each repetition we assessed both the performance of the DCC model based upon the MSE and its computation time.

2.5 Experimental Data

We applied the DCC approach to the “Multimodal MRI Reproducibility Resource” Landman et al. (2011), colloquially known as the Kirby 21 dataset, which is publicly available through the Neuroimaging Informatics Tools and Resources Clearinghouse (www.nitrc.org). The Kirby 21 dataset consists of test-retest MRI scans from 21 healthy adult volunteers with no history of neurological conditions (11 male and 10 female, aged 31.76 ± 9.47 years) that were collected using a variety of modern MRI modalities including structural T1-weighted scans and resting state fMRI scans.

Each resting state scan was 7-min long and was acquired using a single-shot, partially parallel (SENSE) gradient-recalled echo planar sequence with an ascending slice order (TR/TE, 2000/30 ms; FA, 75; SENSE acceleration factor of 2; 3-mm axial slices with a 1-mm slice gap) and an 8-channel head coil. Participants were instructed to relax and fixate on a cross-hair while remaining as still as possible. The two resting state scans were separated by a short break during which the participant exited the scanner.

Image processing was performed using SPM8 and custom MATLAB scripts. Images were registered to the first functional volume and normalized to MNI space using unified segmentation/normalization (SPM8). Functional data were adjusted for slice-time acquisition, as well as participant motion and were transformed to MNI space. Nuisance covariates from white matter and CSF were estimated using CompCor (Behzadi et al., 2007) and regressed from the data along with the absolute and differential motion realignment estimates, and linear trends. Data were then spatially smoothed (6mm kernel). Data from one participant was excluded from further analysis due to a misalignment of the first and second resting state scans.

Following Chang and Glover (Chang and Glover, 2010), a region in the posterior cingulate (3mm-radius sphere, centered at $x = 6$, $y = 58$, and $z = 28$) was selected as the primary ROI

for the default-mode network. In addition, the five foci identified in Chang and Glover as showing the greatest variation in dynamic correlation with the posterior cingulate cortex (PCC) were selected. Average time courses were extracted (3mm radius spheres) from these regions which included the right inferior parietal cortex (ROI1; $x = 34$, $y = 58$, and $z = 44$), the right inferior frontal operculum/BA44 (ROI2; $x = 50$, $y = 18$, and $z = 32$), the right inferior temporal cortex (ROI3; $x = 58$, $y = 38$, and $z = 16$), the right inferior orbitofrontal cortex (ROI4; $x = 50$, $y = 26$, and $z = 8$), and the anterior cingulate cortex (ACC)/BA24 (ROI5; $x = 2$, $y = 26$, and $z = 24$).

To evaluate whether the regions exhibited similar temporal activity, the pair-wise correlation coefficients between the raw time series of the 5 ROIs were computed for each scan. As shown in Fig. 4, ROIs 1, 2 and 3 appear highly correlated with one another, and distinct from ROIs 4 and 5. This is consistent between both scans. However, contrary to Chang and Glover (Chang and Glover, 2010), the latter two regions do not appear highly correlated in our data set. One possible reason for this discrepancy is that we used the same, relatively small, sized seeds as Chang and Glover, and it is possible the placement of these seeds do not align perfectly with the regions of interest in our data set. However, since our goal was to get sample time courses to illustrate DCC, and not necessarily replicate the previous study we proceed using these time courses with the caveat that the results obtained using ROIs 1–3 may be more reflective of past work.

The DCC approach was used to estimate the dynamic correlation between the PCC and each of the five ROIs. The range of variability was recorded, as were 99% confidence intervals for the correlation. The latter was used to determine whether the correlations varied significantly over time.

3 Results

3.1 Simulations

Simulation 1—For the first simulation, the two time courses were designed to be uncorrelated throughout the entire time course. In this setting we would expect the estimated dynamic correlation to take values close to zero across the entire range of the time course. To illustrate the potential for the different methods to overestimate the correlation, we compute the maximum correlation obtained across time for each method discussed using simulated data of varying length ($T = 150, 300, 600, 1000$). Fig. 5 shows results for the sliding-window technique using window lengths of 15, 30, 60 and 120 time points. Clearly for shorter windows this approach consistently gives maximum correlations in the range 0.5 – 0.6, and at times even approaches 0.9, based purely on random noise. Hence, using this approach one would expect to get large correlations purely by chance. This result is further confounded by the fact that the correlations will vary smoothly due to the manner in which computations are performed (e.g., see Fig. 1), thus providing dynamic time courses that give the illusion of providing important signal, while in reality simply responding to natural variation due to noise. This problem is alleviated by increasing the window length. However, on the flipside this will reduce the method's sensitivity to observe actual correlations differing from 0.

Fig. 6 shows similar results obtained using the EWMA and DCC models. For EWMA the maximum correlation lies somewhere between those found using sliding-windows of length 30 and 60. For the DCC approach the maximum correlation observed across time is significantly lower than those reported using the sliding-window technique. It tends to report maximum correlations in the range of 0 – 0.2, rarely giving rise to maximum correlations higher than 0.5, and therefore noise will be far less likely to be erroneously interpreted as important signal.

Fig. 7 shows the proportion of times the 99% confidence intervals computed using the DCC approach did not cover 0 as a function of time for $T = 600$. Results for other values of T (not shown here) were similar. Clearly, all values lie well below the nominal value of 0.01, and the Monte Carlo procedure appears to have appropriate coverage.

The left panel of Fig. 8 shows an example of the true correlation plotted together with the estimated correlations obtained using EWMA, DCC, sliding-windows and tapered sliding-windows for a single repetition of the simulation. The right panel shows boxplots of the mean square error (MSE) for all 1000 repetitions of the simulation when $T = 600$. The MSE is computed by calculating the mean of the squared difference between the estimated dynamic correlation and the true correlation (0). Clearly the DCC model performs very well compared with the other models, as the MSE for the DCC model is an order of magnitude smaller than most of the other methods (i.e. EWMA, tapered sliding windows, sliding windows with lengths 15 – 60). In general, as illustrated in the left panel, the estimated correlation tends to remain close to zero across the entire span of the time series, while the other methods tend to oscillate in the range between -0.4 and 0.4 . Only the sliding window with length 120 is comparable in performance. This isn't surprising as the truth is a static correlation and including as many points as possible in the computation will improve the results.

Finally, we compared the DCC with the oracle sliding-window, computed using the value of w that minimizes the MSE. Fig. 9A shows a boxplot for the percent difference in MSE between DCC and OSW, with negative values indicating lower MSE values for DCC. Clearly, DCC outperforms the OSW in almost all replications, with an average improvement of 0.8%. Fig. 9D shows a boxplot of the optimal values of w in the repeated realizations. The most common value is 120, indicating that one should use as much of the data as possible. Since, there is no actual time-varying correlation between time courses, the difference between OSW and DCC will decrease as w approaches T .

Simulation 2—The left panels of Fig. 10 show examples of the true correlations used in this simulation, which are slowly varying periodic functions of differing frequencies. Also shown in the plot are the estimated dynamic correlations obtained using EWMA, DCC, sliding windows and tapered sliding windows for a single repetition of the simulation. The right panels show boxplots of the MSE for all 1000 repetitions of the simulation. For the more slowly varying correlations, DCC and the sliding-windows techniques with longer windows (60 and 120) perform best. However, as the frequencies of the true correlations increase the performance of the sliding-window techniques worsen and DCC comes to perform best.

This can also be seen in Fig. 9B where the OSW outperforms the DCC in a majority of repetitions for the more slowly varying correlations, while the opposite results hold as the frequency increases. Interestingly, for these later frequencies the variation of the optimal window length is quite substantial (see Fig. 9E), illustrating the inherent difficulties involved in determining an appropriate window length in a real-world situation. These results show that the OSW and DCC perform roughly equivalent to one another. This is particularly remarkable as the OSW uses the true value to calibrate itself. This indicates that even if one were able to always choose the optimal window length in the sliding-window analysis, its performance would not necessarily improve upon that of DCC which is entirely data driven.

Simulation 3—The left panels of Fig. 11 show examples of the true correlation, in this case meant to represent transient state-changes depicted by Gaussian curves of varying width (15, 30, 45, 60), together with the estimated correlations obtained using EWMA, DCC, sliding-windows and tapered sliding-windows for a single repetition of the simulation. The right panels show boxplots of the MSE computed solely over intervals with non-zero correlation (defined as intervals within three standard deviations of the mean located at time point 250) for all 1000 repetitions of the simulation. The MSE was computed in this manner in order to properly evaluate the methods ability to pick up the transient changes, and not be unduly influenced by their ability to detect static null correlation, which was evaluated in Simulation 1. For three out of the four simulations, the DCC model outperforms the other approaches. However, for short state changes (see top row) it performs slightly worse than the EWMA model. This illustrates that the benefits of the DCC model may be less apparent when attempting to estimate rapid state changes, and alternative approaches may be preferable. As the width of the Gaussian increases, the performance of DCC and the sliding-window techniques with longer window lengths (60 and 120) improve, with DCC performing best.

A similar story can be seen in Fig. 9C where the performance of the DCC becomes marginally better than the OSW for wider Gaussian kernels. However, as illustrated in Fig. 9F the process of determining the optimal window length needed to attain the performance of the OSW promises to be difficult as its value is shown to vary substantially across replications of simulations based upon the same underlying true correlation.

Simulation 4—The top panels of Fig. 12 show the mean and variance of the MSE for each component of the matrix Σ_t for the case when there were $N = 90$ time courses included in the analysis. The results, which are representative of other values of N , are consistent with those seen in the bivariate case. The bottom panel of Fig. 12 shows the computation time for DCC as a function of the number of nodes include in the analysis. The results indicate a near linear increase in complexity. On a MacBook Pro (2.4 GHz Intel Core i7) the bivariate case had an average computation time of 1.7s, while for the case with 100 nodes it increased to 63.3s.

3.2 Experimental Data

Here we illustrate the application of DCC to the Kirby 21 data set. Fig. 13 shows the ranges of the estimated dynamic correlations between the PCC and ROIs 1–5. The ranges vary significantly both across regions and subjects. Interestingly, there is also a large amount of variation across scans for the same region and subject.

We further computed 99% confidence intervals for the correlation and determined how often the intervals did not cover the static correlation coefficient between the PCC and the region in question. The idea is that correlations that consistently cover the static value are less likely to truly vary across time. Fig. 14 show heat maps of the number of time points for each region, subject and scan. Again, there appears to be a large amount of variability in the values across regions, subjects and scans.

Fig 15 displays data from three illustrative subjects. The first subject demonstrates a small range in connectivity and there are only a few points where the static correlation lies outside of the confidence bounds. The second illustrative subject shows a large range, but there are still only a few points outside of the intervals. Finally, in the third example, the subject has a large range and many points outside of the interval. The estimated correlation is extremely noisy and it leads to the suspicion that the large amount of variation may be driven by noise. In addition, somewhat troubling, the behavior of these correlations are not consistent across repeated scans for the same subjects and regions.

4 Discussion

In recent years there has been a growing interest in estimating potential dynamic correlations between two brain regions, as it is thought to provide important information about the properties of brain networks. In this paper we study the behavior of several variants of the commonly used sliding-windows technique and introduce two multivariate volatility models often used in the finance literature to the neuroimaging community.

In general, the paper illustrates that sliding-window techniques may not be particularly useful for tracking dynamic correlations. For short window lengths, our simulations clearly show that even completely random noise will give rise to dynamic profiles that appear to show compelling dynamic changes in correlation across time. This problem is less pronounced as the window length increases, however, this comes at the cost of lowered sensitivity to short-term changes in dynamic correlation. Another issue with sliding-window techniques is that their window length is typically chosen in an *ad hoc* manner, when it would be preferable if it were directly estimated from the data. This is presumably done primarily to reduce computational burden, as methods for automatic window size selection are available in the time series literature. For example, Ombao and Van Belle (2008) suggest a simple algorithm for coherence estimation, that is a specific application of a more general approach developed by Lepskii (1990).

Several of the shortcomings of the sliding-window technique are addressed by using multivariate volatility models that are typically used to study the relationship between the volatilities and co-volatilities of several financial markets. We recommend two such models

not currently in use in the neuroimaging community, namely the EWMA and DCC models. Both have found extensive use in the finance literature for modeling time-varying variances and correlations and are generally thought to be preferable to sliding-window type approaches (Bauwens et al., 2006). While the sliding window estimator is non-parametric, both the DCC and EWMA models are parametric. These types of approaches tend to be more powerful as long as the model is reasonably accurate. However, with the usual caveat that if not accurate the more flexible non-parametric methods may be preferable.

EWMA has previously been applied to neuroimaging data to model unexpected changes in the mean BOLD signal across time (Lindquist and Wager, 2005; Lindquist et al., 2007; Wager et al., 2009). In this work the focus is instead on estimating the dynamic nature of the variances of the BOLD signal and the correlations between different regions. EWMA shares some similarities with sliding-window techniques in that the correlation is computed using only data contained within a certain interval of time. However, in EWMA the data points are proportional to their distance from the point t , with weights decaying exponentially (hence the name) as points move further away from the time point of interest. Thus, points are removed from the moving windows in a gradual manner more akin to the tapered sliding-windows approach. The weighting is determined by a parameter λ which can be estimated using maximum likelihood methods under the assumption that the time courses follow a bivariate normal distribution. Hence, EWMA effectively allows the window length to be computed in a data-driven manner tailored to the data at hand.

DCC has to the best of our knowledge never been applied to fMRI data. It shares certain similarities with time series models commonly used to describe fMRI noise, such as the autoregressive (AR) and autoregressive moving-average (ARMA) models (Purdon et al., 2001). DCC allows for a two-stage estimator. In the first stage, time-varying variances are estimated for each time series. In the second stage, the standardized residuals are used to estimate the dynamic correlations. This two-stage approach has been shown (Engle and Sheppard, 2001; Engle, 2002) to provide estimates that are consistent and asymptotically normal with a variance that can be computed using the generalized method of moments approach. This provides us with a framework for performing inference (e.g., constructing confidence intervals) on both the parameters of the model, as well as, the time-varying correlations themselves.

Our simulations show that the performance of DCC is roughly equivalent to that of the “oracle sliding window”, which is allowed to tailor itself to the true underlying correlation. This is remarkable as the OSW is an idealized statistic that can’t be computed on experimental data and simply acts as a theoretical gold standard for how well the sliding window technique would perform if it always choose the optimal window length. In short, the results indicate that even if we could always choose the optimal window length we would still not improve upon the results of DCC using the sliding window approach.

Though the focus of this paper is on bivariate correlations, one of the primary benefits of DCC is its scalability. In general when studying covariances in the N -variate case, a total of $N(N+1)/2$ variances and covariances need to be estimated, some of which may be time-varying. In addition, the resulting covariance matrix must be positive definite at each point

in time. DCC circumvents these problems by using the above mentioned two stage approach. In the first stage each time series is analyzed separately. Hence, moving from the bivariate to N-variate case necessitates computing an additional $N - 2$ univariate GARCH models. However, the complexity of each individual GARCH model is unaffected. In the second stage, there are only two parameters to estimate regardless of the dimension. Hence, this stage scales nicely as well. This near linear increase in complexity is illustrated in Fig. 12. On the flipside a potential shortcoming of this approach is that the dynamic behavior of the correlation matrix is only governed by two parameters, which places certain limitations on its time evolution.

In a series of simulation studies, the DCC model was shown to outperform the other studied techniques. This included situations where there was no correlation between time courses and when there was slowly varying periodic fluctuations. DCC had some difficulties with very fast short-term state changes. This is illustrated in Simulation 3 and confirmed in other simulations with very rapid state changes not shown here. In these scenarios, DCC had a tendency to miss the change and instead erroneously estimate it as zero correlation. For these types of situations it is instead our recommendation to use a technique such as Dynamic Connectivity Regression (Cribben et al., 2012, 2013) which may be more appropriate.

DCC was used to study the bivariate dynamic connectivity between PCC, a node in the default-mode network, and 5 ROIS that were identified in the literature due to the fact that their dynamic correlations exhibited a high degree of variability across time. Using test-retest data, we observed a fair amount of variability in the behavior of these correlations across subjects, regions and scans. Though variability between regions and subjects in the same scan is understandable due to the task-free nature of the experiment, the variability across scans for the same subject and region is cause for greater concern. It calls into question the reproducibility of the dynamic correlations, and also our ability to effectively link it to behavioral data.

In conclusion, we believe DCC is an attractive option for effective estimation of dynamic changes in correlation in fMRI data. In contrast to commonly used sliding-windows techniques, we have shown that DCC is not easily confused by changes in correlation induced solely by random noise. In addition, all the parameters are effectively estimated through quasi-maximum likelihood methods and require no *ad hoc* parameter settings. Finally, the asymptotic theory of DCC provides a mechanism for statistical inference that is not readily available when using other techniques for estimating dynamic correlations.

Acknowledgments

This research was supported by NIH grant R01EB016061.

References

- Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*. 2012
- Bauwens L, Laurent S, Rombouts JV. Multivariate garch models: a survey. *Journal of applied econometrics*. 2006; 21(1):79–109.

- Behzadi Y, Restom K, Liao J, Liu TT. A component based noise correction method (compcor) for bold and perfusion based fmri. *Neuroimage*. 2007; 37(1):90–101. [PubMed: 17560126]
- Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*. 1986; 31(3):307–327.
- Chang C, Glover GH. Time–frequency dynamics of resting-state brain connectivity measured with fmri. *Neuroimage*. 2010; 50(1):81–98. [PubMed: 20006716]
- Chang C, Liu Z, Chen MC, Liu X, Duyn JH. Eeg correlates of time-varying bold functional connectivity. *Neuroimage*. 2013
- Cribben I, Haraldsdottir R, Atlas LY, Wager TD, Lindquist MA. Dynamic connectivity regression: Determining state-related changes in brain connectivity. *NeuroImage*. 2012; 61(4):907–920. [PubMed: 22484408]
- Cribben I, Wager TD, Lindquist MA. Detecting functional connectivity change points for single-subject fmri data. *Frontiers in computational neuroscience*. 2013;7. [PubMed: 23444270]
- Engle R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*. 2002; 20(3):339–350.
- Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*. 1982:987–1007.
- Engle, RF.; Sheppard, K. Technical report. National Bureau of Economic Research; 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate garch.
- Handwerker DA, Roopchansingh V, Gonzalez-Castillo J, Bandettini PA. Periodic changes in fmri connectivity. *Neuroimage*. 2012
- Hansen PR, Lunde A. A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*. 2005; 20(7):873–889.
- Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Penna SD, Duyn J, Glover G, Gonzalez-Castillo J, et al. Dynamic functional connectivity: Promises, issues, and interpretations. *NeuroImage*. 2013
- Jones DT, Vemuri P, Murphy MC, Gunter JL, Senjem ML, Machulda MM, Przybelski SA, Gregg BE, Kantarci K, Knopman DS, et al. Non-stationarity in the “resting brain’s” modular architecture. *PloS one*. 2012; 7(6):e39731. [PubMed: 22761880]
- Kiviniemi V, Vire T, Remes J, Elseoud AA, Starck T, Tervonen O, Nikkinen J. A sliding time-window ica reveals spatial variability of the default mode network in time. *Brain connectivity*. 2011; 1(4): 339–347. [PubMed: 22432423]
- Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JA, Bogovic JA, Hua J, Chen M, Jarso S, et al. Multi-parametric neuroimaging reproducibility: a 3-t resource study. *Neuroimage*. 2011; 54(4):2854–2866. [PubMed: 21094686]
- Lebo MJ, Box-Steffensmeier JM. Dynamic conditional correlations in political science. *American Journal of Political Science*. 2008; 52(3):688–704.
- Lepskii O. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability & Its Applications*. 1990; 35(3):454–466.
- Lindquist, M.; Wager, TD. Applied Data Analytic Techniques for “Turning Points Research. 2005. Application of change-point theory to modeling state-related activity in fmri.
- Lindquist MA. The statistical analysis of fmri data. *Statistical Science*. 2008; 23(4):439–464.
- Lindquist MA, Waugh C, Wager TD. Modeling state-related fmri activity using change-point theory. *Neuroimage*. 2007; 35(3):1125–1141. [PubMed: 17360198]
- Ombao H, Van Bellegem S. Evolutionary coherence of nonstationary signals. *Signal Processing, IEEE Transactions on*. 2008; 56(6):2259–2266.
- Purdon PL, Solo V, Weisskoff RM, Brown EN. Locally regularized spatiotemporal modeling and model comparison for functional mri. *NeuroImage*. 2001; 14(4):912–923. [PubMed: 11554810]
- Sheppard K. Forecasting high dimensional covariance matrices. *Handbook of Volatility Models and Their Applications*. 2012:103–125.

- Tagliazucchi E, Von Wegner F, Morzelewski A, Brodbeck V, Laufs H. Dynamic bold functional connectivity in humans and its electrophysiological correlates. *Frontiers in human neuroscience*. 2012;6. [PubMed: 22375109]
- Thompson GJ, Magnuson ME, Merritt MD, Schwarb H, Pan W-J, McKinley A, Tripp LD, Schumacher EH, Keilholz SD. Short-time windows of correlation between large-scale functional brain networks predict vigilance intraindividually and interindividually. *Human brain mapping*. 2012
- Tsay RS. Multivariate volatility models. *Lecture Notes-Monograph Series*. 2006:210–222.
- Wager TD, Waugh CE, Lindquist M, Noll DC, Fredrickson BL, Taylor SF. Brain mediators of cardiovascular responses to social threat: part i: Reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *Neuroimage*. 2009; 47(3):821–835. [PubMed: 19465137]

Appendix

Let $\gamma = (\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2)$ represent the parameters of the two GARCH models and $\phi = (\theta_1, \theta_2)$ the parameters of the correlation process. It can be shown (Engle and Sheppard, 2001; Engle, 2002) that the log-likelihood function for the DCC model can be expressed as the sum of a volatility term (corresponding to the individual GARCH models for each of the two time series) and a correlation term. The log-likelihood can be written:

$$\mathcal{L}(\gamma, \phi) = \mathcal{L}_V(\gamma) + \mathcal{L}_C(\gamma, \phi) \quad (25)$$

where

$$\mathcal{L}_V(\gamma) = -\frac{1}{2} \sum_{t=1}^T (2 \log(2\pi) + \log|D_t|^2 + r_t^T D_t^{-2} r_t) \quad (26)$$

$$\mathcal{L}_C(\gamma, \phi) = -\frac{1}{2} \sum_{t=1}^T \log|R_t| + \varepsilon_t^T R_t^{-1} \varepsilon_t - \varepsilon_t^T \varepsilon_t \quad (27)$$

The parameters (γ, ϕ) can be estimated using a two-stage approach. First, the parameters of the volatility model are estimated by computing:

$$\hat{\gamma} = \arg\max_{\gamma} \mathcal{L}_V(\gamma). \quad (28)$$

Thereafter, the parameters of the correlation model can be estimated by plugging $\hat{\gamma}$ into $\mathcal{L}_C(\gamma, \phi)$ and computing:

$$\hat{\phi} = \arg\max_{\phi} \mathcal{L}_C(\hat{\gamma}, \phi). \quad (29)$$

Both solutions can be obtained using standard maximization techniques.

Under certain very general conditions (Engle and Sheppard, 2001), the estimates obtained using this two-stage approach are consistent and asymptotically normal with variance $A^{-1}BA^{-1}$ where

$$A = \begin{pmatrix} \nabla_{\gamma\gamma} \log f_V(\gamma_0) & 0 \\ \nabla_{\gamma\phi} \log f_C(\gamma_0, \phi_0) & \nabla_{\phi\phi} \log f_C(\gamma_0, \phi_0) \end{pmatrix} \quad (30)$$

and

$$B = \text{var} \left[\sum_{t=1}^T \{ T^{-1/2} \nabla_{\gamma}^T \log f_V(r_t, \gamma_0), T^{-1/2} \nabla_{\phi}^T \log f_C(r_t, \gamma_0, \phi_0) \} \right]. \quad (31)$$

Here ∇_x represents a partial derivative with respect to a parameter x , and f_V and f_C are the likelihood functions for the volatility and correlation portion of the likelihood function, respectively.

These variances can be approximated using finite difference methods and used to approximate a distribution for the model parameters. Using Monte Carlo sampling, we can randomly generate a number of draws for these distributions, calculate the dynamic correlations for each set of parameters, and use this to construct confidence intervals.

Figure 1.

Example of the estimated correlation using the sliding windows approach with window lengths $w = 15, 30, 45, 60$. Here the time courses are independent of one another and, hence, there should be no correlation between time courses.

Figure 2.

Example of the estimated correlation using the EWMA approach with $\lambda = 0.7, 0.8, 0.9, 0.95$.

Here the time courses are independent of one another and there should be no correlation between time courses.

Figure 3.
Example of the estimated correlation using the DCC model. Here the time courses are independent of one another and there should be no correlation between time courses.

Figure 4.
The Fisher z-transformed correlation coefficients between the 5 ROIs for both scans. High pair-wise correlation was observed between ROIs 1–3. There also appears to be a high degree of between-scan reliability in the measures.

Figure 5.

Boxplots of the maximum correlation computed using the sliding-window approach with window length $w = 15, 30, 60, 120$ for time courses of length $T = 150, 300, 600, 1000$. In this simulation the time courses are generated to be independent and hence the correlation between time courses should ideally be 0.

Figure 6.

Boxplots of the maximum correlation computed using the EWMA and DCC models for time courses of length $T = 150, 300, 600, 1000$. In this simulation the time courses are generated to be independent and the estimated correlation between time courses should ideally be 0.

Figure 7.

The proportion of times the 99% confidence intervals computed using DCC did not cover 0 as a function of time for the case when $T = 600$. Results were similar for other values of T . Clearly, all values lie below the nominal value of 0.01.

Figure 8.

Results from Simulation 1 with $T = 600$. The left panel shows the results for a single iteration of the simulation. The truth is no correlation across time and fits for the sliding-window with lengths 15 and 120, the tapered sliding-window, EWMA and DCC are shown overlayed on the plot. The right panel shows a boxplot of the MSE between the estimated dynamic correlations and the truth for each of the seven methods over the 1000 realizations of the simulation.

Figure 9.

A–C. Comparison of the oracle sliding-window and the DCC model for each of the first three simulation studies. The boxplots represent the percent difference in MSE between the two methods based on 1000 realizations of the simulation. As the OSW requires the true correlation to be known it provides a theoretical gold standard for the sliding window approach. D–F. The boxplots show the window lengths that obtained the optimal MSE across the 1000 realizations of each simulation.

Figure 10.

Results of Simulation 2. The left panels show the results for a single iteration for each of the four parts of the simulation (one per row). The truth is a slowly varying periodic correlation across time and fits for EWMA, DCC, the tapered sliding-window, and sliding-windows using the worst and best fitting window length are shown in the plot. The right panel shows a boxplot of the MSE between the estimated dynamic correlations and the truth for each of the seven methods over the 1000 realizations of the simulation.

Figure 11.

Results of Simulation 3. The left panels show the results for a single iteration for each of the four parts of the simulation (one per row). The truth is the correlation is zero except in a range corresponding to a Gaussian curve centered at 250 with standard deviations 15, 30, 45, and 60. The fits for EWMA, DCC, the tapered sliding-window, and sliding-windows using the worst and best fitting window length are shown overlayed on the plot. The right panel shows a boxplot of the MSE between the estimated dynamic correlations and the truth for each of the seven methods over the 1000 realizations of the simulation. The MSE are computed solely over intervals with non-zero correlation, defined as intervals within three standard deviations of the mean located at time point 250.

Figure 12.

Results of Simulation 4. The top panels show the mean (left) and variance (right) of the MSE for each component of the matrix Σ_t when $N = 90$ time courses were included in the analysis. The bottom panel shows the computation time for DCC as a function of N , showing a near linear increase in complexity. The shaded area indicates one standard error around the mean.

Figure 13.

The range of dynamic correlation values between the PCC and ROIs 1–5. The vertical lines show the minimum and maximum correlation coefficients across time and the dot indicates the mean. Each subject has two measures, one corresponding to scan 1 (red) and the other to scan 2 (blue).

Figure 14.

The number of time points in which the 99% confidence intervals for the correlation did not cover the static correlation coefficient between the PCC and ROIs 1–5 for each subject, region and scan.

Figure 15.

The raw time series and estimated dynamic correlations between the PCC and ROI 1 for (A) subject 2 (scan 2); (B) subject 5 (scan 2); and (C) subject 6 (scan 2). The confidence intervals are provided as is the static correlation value (dotted line).