

Large-scale VAR modeling based on functional
MRI data by applying a Bayes shrinkage
estimation and a modified local FDR procedure

by

Sung-Ho Kim and Namgil Lee

BK21 Research Report

11 - 04

December 20, 2011

DEPARTMENT OF MATHEMATICAL SCIENCES

The second stage of
BK21
Fostering A World Class Talent

KAIST

한국과학기술원
Korea Advanced Institute of Science and Technology

Large-scale VAR modeling based on functional MRI data by applying a Bayes shrinkage estimation and a modified local FDR procedure

Sung-Ho Kim

KAIST, Daejeon, South Korea

Namgil Lee

KAIST, Daejeon, South Korea

Summary.

Discovering effective connectivity between brain regions gained a lot of attention recently. A vector autoregressive (VAR) model is a simple and flexible approach for exploratory structural modeling where the involvement of a large number of brain regions is crucial to avoid confounding. In this paper we propose improved procedures for inferring large-scale causal networks of brain regions using functional MRI (fMRI) data. First, an empirical Bayesian procedure for shrinkage estimation of the model coefficients is recruited to prevent overfitting in the face of the high dimensionality of VAR models. Second, a modified local false discovery rate (FDR) procedure is suggested to identify nonzero coefficients by adaptively determining null distributions with respect to skewed distributions of partial correlations between regions. The effectiveness of the proposed procedures is demonstrated through simulated data sets in comparison with other methods known in literature. Through multitask fMRI data sets concerning Brodmann areas in both hemispheres, meaningful directed networks of the brain regions are obtained and, through some measures from graph theory, we discover (1) the adaptation of the networks of brain regions to different cognitive tasks in terms of changes in the level of involvement of each hemisphere, (2) the importance of memory regions in the medial part of the brain for efficient visual information processing, and (3) the adaptive changes of the networks to different cognitive demands in terms of changes in solution strategies.

Keywords: Brodmann area; local false discovery rate; Granger causality; graph theory; high-dimensional data

1. INTRODUCTION

Modeling communications among a set of brain regions which are involved in various cognitive tasks is one of the most prevalent researches in neuroscience. Whereas it is well established that the brain is segregated functionally in some depth, it is relatively less understood how the functional areas of the brain are interconnected and integrated. The fMRI is a non-invasive neuroimaging technique to measure time courses of hemodynamic responses in each voxel of the brain. Statistical models for the brain connectivity using fMRI data have arisen in recent years and are mainly categorized into two groups: the dynamic causal modelling (DCM)(Friston, Harrison and Penny, 2003) and the Granger causal modelling/mapping

Address for correspondence: Sung-Ho Kim, Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, South Korea
Email: sung-ho.kim@kaist.edu

(GCM)(Goebel et al., 2003; Valdes-Sosa, 2004). The comparisons and interpretations of the two approaches are well discussed in Friston (2009), Roebroeck, Formisano and Goebel (2011) and Friston (2011).

A typical example of GCM is a vector autoregressive (VAR) model, which is a simple linear stochastic model without exogenous input variables. When it is used to model fMRI time series data, nonzero coefficients of a VAR model represent Granger causality between brain regions. It became popular for analysis and prediction of economic and financial time series by Sims (1980), and it has been widely used in econometrics (Doan, Litterman and Sims, 1984), bioinformatics (Opgen-Rhein and Strimmer, 2007) and neuroscience (Harrison, Penny and Friston, 2003).

The strength of VAR models in brain connectivity analysis lies in its flexibility and structure exploration capability (Roebroeck, Formisano and Goebel, 2011). For the exploratory analysis, it is important to include as many of the brain regions relevant to the task as possible while avoiding overfitting. To cope with the increased demand to deal with high-dimensional VAR models concerning a large number of brain regions for a relatively short length of the time series, Valdes-Sosa et al. (2005) introduced a two-stage process to estimate sparse VAR models. The process makes a combined use of (1) penalized regression and (2) pruning of unlikely connections by means of the local false discovery rate (FDR) developed by Efron (2003, 2004). Through simulation experiments, Valdes-Sosa et al. (2005) showed that the penalized regression methods including ridge, LASSO and more, resulted in similarly good detection efficiency by presenting areas under ROC curves of the methods.

On the other hand, Opgen-Rhein and Strimmer (2007) suggested to use a nonparametric shrinkage estimation method of Schäfer and Strimmer (2005b) for estimation of VAR model coefficients with an application to modeling high-dimensional biological data. They showed by simulation that the suggested shrinkage estimation is superior to ridge and LASSO regressions in terms of positive predictive value when the length of time series data is shorter than the dimensionality. However, they noticed that for large sample size the nonparametric shrinkage estimator appears to be prone to overshrinking, which leads to an increase of false positives.

In this paper, we propose improvements to the two-stage processes of Valdes-Sosa et al. (2005) and Opgen-Rhein and Strimmer (2007). First, for the penalized regression, we notice that one of the reasons of the overshrinking of the nonparametric shrinkage estimator is due to the independence assumption of data in the derivation in Schäfer and Strimmer (2005b), which is obviously not true for time series data. We propose an improved shrinkage estimation procedure based on an empirical Bayesian approach which correctly incorporates correlations in the data.

Second, for the local FDR, it is often the case that empirical distributions of test statistics obtained from fMRI data are skewed and have shifted modes. Schwartzman et al. (2009) illustrate such examples which arise from fMRI and DTI data when selecting significant activations of voxels. Schwartzman et al. (2009) and Efron (2004) also built some models for simulations in which the shift of the mode of empirical distributions is caused by a shift of theoretical null distributions. However, in this paper we show it is actually hard to tell whether the shift has been caused by the shift of null distributions or the mixture of null and alternative distributions in case that VAR models are considered in the selection of nonzero coefficients through simulations. Accordingly we propose a procedure to estimate null distributions assuming that the null distributions are not shifted. We also provide a theoretical justification of the proposed procedure together with experimental evaluations.

On the other hand, we concern the fMRI experiment where each of the total 27 male subjects undergoes psychometric tests consisting of 5 types of problems. The 5 types of problems are supposed to assess different aspects of spatial ability of individuals and may have different difficulty levels. The spatial ability is considered to consist of various factors involved in a process of generating, retaining, retrieving, and transforming well-structured visual images (Lohman, 1994). Such factors include spatial visualization, spatial orientation, and spatial relations. Several psychometric tests have been developed to distinguish such factors. For instance, the tests such as a form board test, a hole punching test, and a surface development test were suggested to measure the factor of spatial visualization, and the tests such as card rotations and cube comparisons were suggested to measure the factor of spatial relations (Ekstrom et al., 1976; French, Ekstrom and Price, 1963).

It is well known that individuals differ in the way they solve spatial tasks: different people use different strategies in solving the same test items. Glück et al. (2002) also have shown that training or practice can lead to changes in strategy use in a spatial task. In general, strategies for solving test items in a type of spatial ability tests are classified into two categories: holistic and analytic strategies (Glück and Fitting, 2003). These strategies are not mutually exclusive categories, but rather they are the poles of a continuum. Glück and Fitting (2003) also reported that people use analytic strategies more with increasing task difficulty. Most people can solve easy tasks by holistic strategies, whereas with increased complexity of spatial tasks, strategies become more analytic.

After all, by investigating the inferred large-scale causal networks of brain regions obtained from the fMRI data, we can discover in part the dynamics underlying the cognitive process of visual information processing. Since the structures of the networks may closely related to the strategies that the subjects used for the test, we can compare the networks of different problem types to discover common neural substrates in the spatial tasks and examine changes in strategies over the different problem types. Specifically, by using graph theoretic measures such as degree and betweenness centrality, we discover (1) the adaptation of the networks of brain regions to different cognitive tasks, (2) the importance of memory regions in the medial part of the brain for efficient visual information processing, and (3) the adaptive changes of the networks to different cognitive demands in terms of changes in solution strategies.

This paper is organized as follows. In Section 2 we describe the fMRI data and the proposed methods for inferring large-scale causal networks. In Section 3 we present results of simulation and analysis on the fMRI data. In Section 4 we discuss the implications along with concluding remarks.





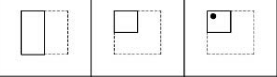





2. METHODS

2.1. Data Sets

For fMRI experiment, 27 young male college students were asked to participate in the study. They are between 20-23 years old. The fMRI data were acquired from a 3.0T ISOL FORTE scanner (ISOL Technology, Gyeonggi, Korea). A total of 177 scans of whole-brain images were acquired using a T2*-weighted single-shot echo-planar imaging (EPI) sequence (repetition time (TR) = 3,000 msec, echo time (TE) = 35 msec, number of slices = 36, slice thickness = 3 mm, matrix size = 64×64 , field of view = $220 \text{ mm} \times 220 \text{ mm}$).

All subjects took psychometric tests during scanning. The tests consist of 42 problems, each of which falls into one of the five problem types: picture completion (12 problems),

Table 1. Examples of the figure frames of stimulus figures and test probes for the five problem types

problem type	1	2	3	4	5
stimulus figure					
test probe					

mental rotation (12 problems), surface development (6 problems), aperture passing (6 problems), and hole punching (6 problems). The problems of each problem type were partitioned into three sets and the total of 15 sets were presented to each subject in a given order. An instruction on how to solve the problems was presented for 6 seconds before the problems in each problem set were presented for 21 seconds. Each problem was displayed with two figure frames: a stimulus figure and a test probe. Table 1 shows examples of the figure frames of stimulus figures and test probes for the five problem types. If the test probe is recognized as a correct answer to the stimulus figure, the subject is supposed to press the left mouse button to answer 'Yes'; otherwise press the right mouse button to answer 'No'. The subject may not press mouse buttons if he or she isn't sure of the answer. The scores were obtained by the number of correct answers subtracted by the number of incorrect answers, and the difference were scaled linearly to lie between 0-100.

The fMRI data were preprocessed using the SPM software (SPM8, Wellcome Institute of Cognitive Neurology, London, UK). The preprocessing step included spatial realignment to the mean volume of a series of images and normalization to the MNI template brain. The data were further partitioned based on the Brodmann areas separated in the left and right hemispheres (Brodmann, 1909). The total 41 Brodmann areas in each hemisphere were selected for the brain connectivity analysis. Table 4 in Appendix B summarizes the anatomical labels and functions of the 41 Brodmann areas, which were produced mostly based on Wikipedia (2011). After all, the fMRI time series data within each of the Brodmann areas was spatially averaged to yield 82 time series.

2.2. VAR Models

A VAR model is a dynamic model that is useful for representing inter-relationship among a set of random variables in time series data. Let $\mathbf{y}_t = [y_{t1}, \dots, y_{td}]'$, $t = 1, \dots, T$, denote $(d \times 1)$ vectors of time series variables and T is the total number of observations, i.e., scans. A VAR model of order p is represented by

$$\mathbf{y}_t = \mathbf{c} + \sum_{k=1}^p A_k \mathbf{y}_{t-k} + \varepsilon_t, \quad (1)$$

where A_k , $k = 1, \dots, p$, are $(d \times d)$ coefficient matrices, \mathbf{c} is a $(d \times 1)$ vector, and ε_t is a $(d \times 1)$ noise vector process with mean zero and covariance matrix V , i.e., $E[\varepsilon_t] = \mathbf{0}$, and $E[\varepsilon_t \varepsilon_\tau'] = V$ if $t = \tau$ and 0 if $t \neq \tau$. Let $\mathbf{x}_t = [\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}]'$, $t = p + 1, \dots, T$, denote

$(dp \times 1)$ vectors of predictors. Then the VAR model (1) is re-expressed as

$$\mathbf{y}_t = \mathbf{c} + \Phi' \mathbf{x}_t + \varepsilon_t$$

where $\Phi = [A_1, \dots, A_p]'$, or, equivalently in terms of components as

$$y_{tj} = c_j + \sum_{i=1}^{dp} \phi_{ij} x_{ti} + \varepsilon_{tj} \quad (2)$$

for $j = 1, \dots, d$, where $\{\phi_{ij}\}$ are the model coefficients.

A VAR model represents a causal network whose model structure is determined by its nonzero coefficients. As indicated in (1), the causal relationship between any two variables is defined as follows: the i th variable $y_{.i}$ does not Granger-cause the j th variable $y_{.j}$ if the coefficient matrices satisfy $(A_1)_{ji} = \dots = (A_p)_{ji} = 0$ (Granger, 1969). In a graphical representation of a VAR model structure, each node corresponds to a variable and each directed edge corresponds to the Granger-causality between the connected variables, the arrow heading from a causal node to its effect node.

In the following subsections, we propose a two-step procedure for detecting nonzero coefficients from a given multivariate time series data: the empirical Bayesian shrinkage estimation procedure and the modified local FDR procedure.

2.3. Empirical Bayesian Shrinkage Estimation Procedure

The proposed shrinkage estimation procedure separates the model coefficients into variance parts and correlation parts, and estimation is made separately on each of the parts. We assume that the vector process \mathbf{y}_t is covariance-stationary, that is, its first and second order moments are invariant over the time t . Let $\sigma_{y_j}^2 = \text{Var}(y_{tj})$ and $\sigma_{x_i}^2 = \text{Var}(x_{ti})$ denote the variances of the corresponding components, and assume that $\sigma_{y_j} > 0$ and $\sigma_{x_i} > 0$ for all $j = 1, \dots, d$, and $i = 1, \dots, dp$. We define the correlation part of ϕ_{ij} by $\psi_{ij} = \phi_{ij} \sigma_{x_i} / \sigma_{y_j}$. That is, ϕ_{ij} is separated as

$$\phi_{ij} = \psi_{ij} \frac{\sigma_{y_j}}{\sigma_{x_i}}. \quad (3)$$

Once we obtain shrinkage estimates $\hat{s}_{y_j}^{*2}$, $\hat{s}_{x_i}^{*2}$ and $\hat{\psi}_{ij}^*$ for $\sigma_{y_j}^2$, $\sigma_{x_i}^2$ and ψ_{ij} , respectively, we can obtain the estimate of ϕ_{ij} by

$$\hat{\phi}_{ij}^* = \hat{\psi}_{ij}^* \frac{\hat{s}_{y_j}^*}{\hat{s}_{x_i}^*}.$$

2.3.1. Estimation of the variance parts.

For notational convenience, let $\mathbf{z}_t = [\mathbf{x}_t', \mathbf{y}_t']'$, $t = p+1, \dots, T$, denote the $((dp+d) \times 1)$ vectors and $\sigma_i^2 = \text{Var}(z_{ti})$, $i = 1, \dots, dp+d$, denote the variance parts. Let $\bar{z}_i = \frac{1}{T-p} \sum_{t=p+1}^T z_{ti}$ and $\hat{s}_i^2 = \frac{1}{T-p-1} \sum_{t=p+1}^T (z_{ti} - \bar{z}_i)^2$ represent the sample mean and the sample variance. The shrinkage estimator, \hat{s}_i^{*2} , of σ_i^2 , is obtained by shrinking the sample variances toward their median as in

$$\hat{s}_i^{*2} = \lambda_v \hat{s}_{\text{med}}^2 + (1 - \lambda_v) \hat{s}_i^2, \quad i = 1, \dots, dp+d \quad (4)$$

where $0 \leq \lambda_v \leq 1$ is a shrinkage parameter and \hat{s}_{med}^2 is the median of the sample variances, i.e., $\hat{s}_{\text{med}}^2 = \text{median}(\hat{s}_1^2, \dots, \hat{s}_{dp+d}^2)$. The optimal shrinkage level, $\hat{\lambda}_v^*$, is determined by

minimizing the mean of the sum of squared error losses, $E \left[\sum_{i=1}^{dp+d} (\hat{s}_i^{*2} - \sigma_i^2)^2 \right]$. Assuming that the first two moments of distributions of \hat{s}_i^2 and \hat{s}_{med}^2 exist, Lee, Choi and Kim (2011) provide $\hat{\lambda}_v^*$ in the closed form by

$$\hat{\lambda}_{\text{EB},v}^* = \frac{\sum_{i=1}^{dp+d} \widehat{\text{Var}}_{\text{EB}}(\hat{s}_i^2)}{\sum_{i=1}^{dp+d} (\hat{s}_{\text{med}}^2 - \hat{s}_i^2)^2}$$

where $\widehat{\text{Var}}_{\text{EB}}(\hat{s}_i^2)$ is calculated from data as follows: let $w_{ti} = (z_{ti} - \bar{z}_i)^2$, $\bar{w}_i = \frac{1}{T-p} \sum_{t=p+1}^T w_{ti}$, then

$$\widehat{\text{Var}}_{\text{EB}}(\hat{s}_i^2) = \frac{1}{(T-p-1)^2} \sum_{t=p+1}^T \sum_{\tau=p+1}^T \widehat{\text{Cov}}_{\text{EB}}(w_{ti}, w_{\tau i})$$

where

$$\widehat{\text{Cov}}_{\text{EB}}(w_{ti}, w_{t+k,i}) = \frac{1}{T-p} \sum_{\tau=p+1}^{T-|k|} (w_{\tau i} - \bar{w}_i)(w_{\tau+k,i} - \bar{w}_i), \quad |k| \leq T-p-1.$$

To keep the value of $\hat{\lambda}_{\text{EB},v}^*$ within $[0, 1]$, we use $\hat{\lambda}_{\text{EB},v}^{**} = \max(0, \min(1, \hat{\lambda}_{\text{EB},v}^*))$.

2.3.2. Estimation of the correlation parts.

From Eqs. (2) and (3), we get

$$y_{tj}^s = c_j^s + \sum_{i=1}^{dp} \psi_{ij} x_{ti}^s + \varepsilon_{tj}^s, \quad j = 1, \dots, d, \quad (5)$$

where $y_{tj}^s = y_{tj}/\sigma_{y_j}$, $x_{ti}^s = x_{ti}/\sigma_{x_i}$, $c_j^s = c_j/\sigma_{y_j}$, and $\varepsilon_{tj}^s = \varepsilon_{tj}/\sigma_{y_j}$. We assume that the variances, $\sigma_{y_j}^2$ and $\sigma_{x_i}^2$, are estimated by the sample variances, $\hat{s}_{y_j}^2$ and $\hat{s}_{x_i}^2$, so that y_{tj}^s and x_{ti}^s are obtainable from data. Let \hat{R}_1 and \hat{R}_2 be the sample correlation matrices defined by

$$\hat{R}_1 = \frac{1}{T-p-1} \sum_{t=p+1}^T (\mathbf{x}_t^s - \bar{\mathbf{x}}^s) (\mathbf{x}_t^s - \bar{\mathbf{x}}^s)'$$

$$\hat{R}_2 = \frac{1}{T-p-1} \sum_{t=p+1}^T (\mathbf{x}_t^s - \bar{\mathbf{x}}^s) (\mathbf{y}_t^s - \bar{\mathbf{y}}^s)'$$

Let $\Psi = (\psi_{ij})$ be the matrix of the correlation parts. From (5), the ordinary least squares estimate of Ψ is given by $\hat{\Psi} = \hat{R}_1^{-1} \hat{R}_2$, which cannot be calculated if \hat{R}_1 is not invertible. On the other hand, the shrinkage estimator, $\hat{\Psi}^*$, of Ψ is obtained by shrinking the sample correlation matrices as in

$$\hat{\Psi}^* = (\hat{R}_1^*)^{-1} \hat{R}_2^* \quad (6)$$

where

$$\hat{R}_1^* = \lambda I + (1-\lambda) \hat{R}_1$$

$$\hat{R}_2^* = (1-\lambda) \hat{R}_2$$

with a common shrinkage parameter $0 \leq \lambda \leq 1$. That is, the off-diagonal entries of \widehat{R}_1 and the entries of \widehat{R}_2 are shrunk toward zero. We note that $\widehat{\Psi}^* = \theta$ if $\lambda = 1$ and $\widehat{\Psi}^* = \widehat{\Psi}$ if $\lambda = 0$.

Lee, Choi and Kim (2011) have shown that the optimal value, $\widehat{\lambda}^*$, of λ for the correlation parts has the following parametric form:

$$\widehat{\lambda}_{\text{EB}}^* = \frac{d^2 p}{\nu(T - p - 1) + d^2 p} \quad (7)$$

where the unknown parameter ν is dependent only on the model parameters (Φ, V) in the model (1) ($\Phi = [A_1, \dots, A_p]'$ is the coefficient matrix and $V = \text{E}[\varepsilon_t \varepsilon_t']$ is the covariance matrix for noise ε_t).

The parameter ν can be estimated by applying a modification to the standard k -fold cross-validation: the standard k -fold cross-validation procedure randomly divides the given set of data pairs, $(\mathbf{x}_t^s, \mathbf{y}_t^s), t = p + 1, \dots, T$, into k sets of almost equal sizes. For the i th set among the k sets, we use the rest $k - 1$ sets to calculate the coefficient matrix, $\widehat{\Psi}^*(i)$, from (6) for some λ . An optimal value $\widehat{\lambda}^*(i)$ is obtained by minimizing the sum of the squares of prediction errors on the i th set with respect to λ . That is, if $(\mathbf{x}_t^{\text{s,val}}, \mathbf{y}_t^{\text{s,val}}), t = 1, \dots, M_i$, denote the data pairs in the i th set, then

$$SSPE(i, \lambda) = \sum_{t=1}^{M_i} \left\| \mathbf{y}_t^{\text{s,val}} - \bar{\mathbf{y}}^s(i) - (\widehat{\Psi}^*(i))'(\mathbf{x}_t^{\text{s,val}} - \bar{\mathbf{x}}^s(i)) \right\|^2 \quad (8)$$

where $\bar{\mathbf{x}}^s(i)$ and $\bar{\mathbf{y}}^s(i)$ are the sample mean vectors calculated using the data pairs in the rest $k - 1$ sets.

Note that each of the obtained estimates, $\widehat{\lambda}^*(i), i = 1, \dots, k$, is the optimal shrinkage parameter for the shrinkage estimates, $\widehat{\Psi}^*(i)$, which were calculated using only the rest $i - 1$ sets. Let $N_i = T - p - M_i$ denote the number of data pairs in the rest $i - 1$ sets. Then, N_i corresponds to the term $T - p$ in (7), and the i th estimate of ν is obtained from (7) by using $\widehat{\lambda}^*(i)$ as

$$\widehat{\lambda}^*(i) = \frac{d^2 p}{\widehat{\nu}^*(i)(N_i - 1) + d^2 p} \Leftrightarrow \widehat{\nu}^*(i) = \left(\frac{\widehat{\lambda}^*(i)}{1 - \widehat{\lambda}^*(i)} \frac{N_i - 1}{d^2 p} \right)^{-1}.$$

Finally, the estimate, $\widehat{\nu}_{\text{EB}}^*$, of ν is determined by

$$-\log \widehat{\nu}_{\text{EB}}^* = \frac{1}{k} \sum_{i=1}^k (-\log \widehat{\nu}^*(i)) = \frac{1}{k} \sum_{i=1}^k \left(\log \left(\frac{\widehat{\lambda}^*(i)}{1 - \widehat{\lambda}^*(i)} \right) + \log \left(\frac{N_i - 1}{d^2 p} \right) \right).$$

2.4. Modified Local False Discovery Rate Procedure

For checking the Granger-causality, it is preferred to run a statistical test using partial correlations instead of using model coefficients in case of data sets with small T and large d (Schäfer and Strimmer, 2005a). Let $\text{corr}(y_j, x_i | x_{\text{rest}})$ denote the partial correlation between two variables y_j and x_i . Whittaker (1990) shows that in the multivariate normal linear regression model, $\phi_{ij} = 0$ if and only if $\text{corr}(y_j, x_i | x_{\text{rest}}) = 0$. For the calculation of the partial correlations from shrinkage estimates, see Ongen-Rhein and Strimmer (2007).

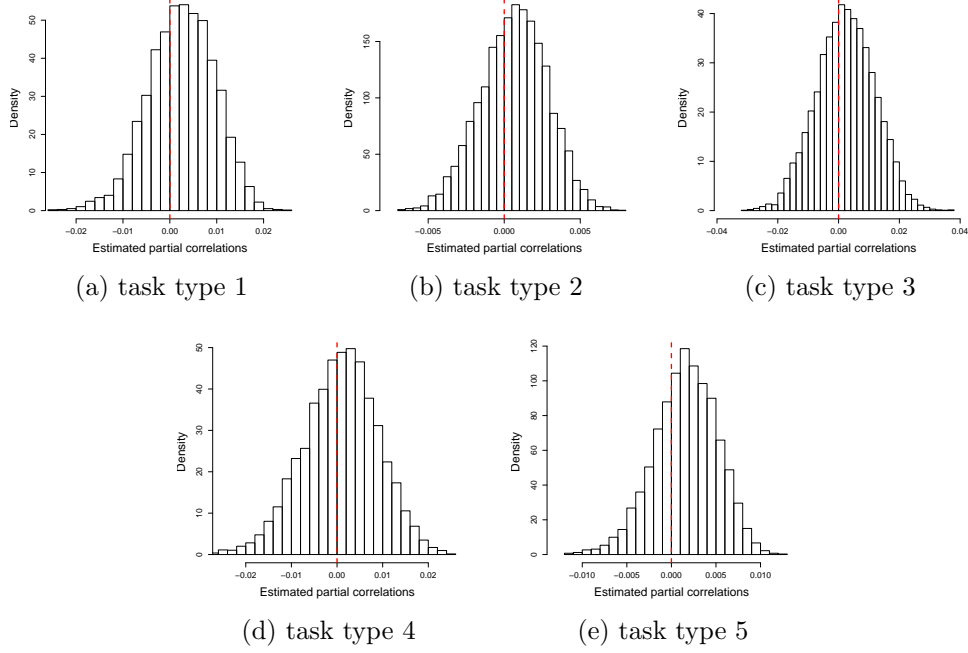


Fig. 1. Empirical distributions of the estimated partial correlations for the fMRI data of the 16th subject

The obtained estimates of partial correlations are used in a local FDR procedure to detect nonzero values of ϕ_{ij} . In the standard local FDR procedure suggested by Strimmer (2008) we assume that distributions of the estimated partial correlations are symmetric. However, many real world fMRI data result in skewed distributions of estimated partial correlations. Fig. 1 shows an example of the empirical distributions of the estimated partial correlations for the given fMRI data of the 16th subject. In the figure we can see that all of the five distributions obtained for the five problem types are not symmetric. In Section 3 we will show that skewed distributions of sample or estimated partial correlations are well observed in time series data which are generated from VAR models whose nonzero coefficients were drawn from skewed distributions.

Therefore we propose a modified local FDR procedure which is effective for skewed distributions. In local FDR procedures, the empirical distribution of estimated partial correlations is approximated by a mixture distribution

$$f(x) = \eta_0 f_0(x; \kappa_0) + (1 - \eta_0) f_A(x), \quad 0 \leq \eta_0 \leq 1,$$

where f_0 is called a null distribution and f_A an alternative distribution. The parameter η_0 represents the portion of zero true partial correlations. The parametric form of f_0 is provided by Hotelling (1953) under the normality assumption. That is, if a true partial correlation coefficient is zero, then the distribution of the corresponding sample partial correlation coefficient is given by

$$f_0(x; \kappa) = (1 - x^2)^{(\kappa-3)/2} \frac{\Gamma(\kappa/2)}{\pi^{1/2} \Gamma((\kappa-1)/2)}, \quad -1 \leq x \leq 1,$$

where κ is the degree of freedom. Note that $\text{Var}(X_0) = 1/\kappa$ for a random variable X_0 having the distribution $f_0(x; \kappa)$.

Local FDR procedures should estimate η_0 and κ_0 from the given estimated partial correlations. In this paper, we assume that the null distribution f_0 is not biased. Considering that the given empirical distribution of the partial correlations has a shifted mode, $x_0 \neq 0$, as in Fig. 1, we can infer that $f(0) \leq f(x_0)$ and $f_0(0; \kappa_0) > f_0(x_0; \kappa_0)$, which yield

$$(1 - \eta_0)f_A(x_0) = f(x_0) - \eta_0 f_0(x_0; \kappa_0) > f(0) - \eta_0 f_0(0; \kappa_0) = (1 - \eta_0)f_A(0) \geq 0,$$

so that $f_A(x_0) > 0$. In the modified local FDR procedure, we use either the positive part or the negative part of the estimated partial correlations, in an effort to avoid including estimated partial correlations whose true partial correlations are not zero. For instance, in Fig. 1, the modes of the empirical distributions are positive and not far from zero. Since the alternative distributions are less mixed on the other side of the horizontal axis, the modified procedure decides to use the other side to estimate the null distribution.

Next, the modified procedure uses a standard kernel density estimation algorithm such as the Parzen windows estimation with Gaussian kernels and a rule-of-thumb bandwidth selection method to obtain a smoothed distribution, $f^s(x)$, for the selected side of the estimated partial correlations. That is, the selected side of the empirical distribution is smoothed and scaled to get

$$f^s(x) = f^s(-x), \quad \forall x \in [-1, 1] \quad \text{and} \quad \int_{-1}^1 f^s(x) dx = 1.$$

Then, with an appropriately selected cutoff value, $0 < \rho \leq 1$, the modified procedure picks up a quantile value $r = r(\rho)$ such that $100\rho\%$ of the partial correlations belong to the interval between 0 and r . For instance, if the negative side is to be used, then the ratio of the numbers of the estimated partial correlations used for estimating κ_0 satisfies approximately $\rho = \#\{r \leq x \leq 0\} / \#\{-1 \leq x \leq 0\}$. Optimal values for η_0 and κ_0 are then obtained by minimizing the sum of squares of differences between $f^s(x)$ and $\eta f_0(x; \kappa)$. That is, if the negative side is used, then

$$\left(\hat{\eta}, \hat{\kappa} \right) = \underset{0 < \eta \leq 1, \kappa > 1}{\text{argmin}} \int_{r \leq x \leq 0} (f^s(x) - \eta f_0(x; \kappa))^2 dx. \quad (9)$$

Note that $\hat{\eta} f_0(x; \hat{\kappa})$ is an approximation to $f^s(x)$, and $f^s(x)$ has been obtained by scaling $f(x)$ to have $\int_{-1}^1 f^s = 1$, i.e., $f(x) \approx c_1 f^s(x)$ for some $c_1 > 0$. Since $\int_{-1}^0 f^s = 0.5$ and $\int_{-1}^0 f = \frac{\#\{-1 \leq x \leq 0\}}{\#\{-1 \leq x \leq 1\}} = \frac{\#\{-1 \leq x \leq 0\}}{d^2 p} = c_1 \int_{-1}^0 f^s$, $\hat{\eta}$ should further be re-scaled as

$$\hat{\eta} = \hat{\eta} \times c_1 = \hat{\eta} \times \frac{\int_{-1}^0 f}{\int_{-1}^0 f^s} = \hat{\eta} \times \frac{2\#\{-1 \leq x \leq 0\}}{d^2 p}.$$

The factor, c_1 , multiplied to $\hat{\eta}$ is the ratio at which f^s should be re-scaled to f .

Finally, the local area-based FDR score (Strimmer, 2008) is obtained as follows:

$$\text{fdr}(x) = \Pr(\text{zero partial correlation} | x) = \frac{\hat{\eta} f_0(x; \hat{\kappa})}{f(x)}$$

where $f(x)$ represents the empirical distribution of the partial correlations which is also smoothed by a kernel density estimation method for robustness. A standard way of selecting

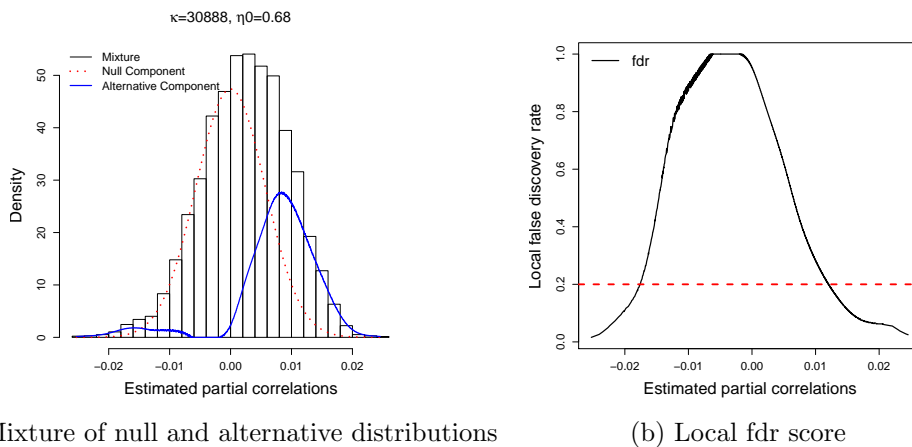


Fig. 2. Estimated mixture of null and alternative distributions (a), and estimated local fdr score (b) for the fMRI data in Fig. 1 (a)

nonzero partial correlations is to select x with $\text{fdr}(x) \leq 0.2$. Fig. 2 (a) and (b) illustrate the results for the estimation of the null and alternative distributions and the local FDR scores obtained based on the data in Fig. 1 (a).

2.5. Selection of the Cutoff Value

The cutoff value, $0 < \rho \leq 1$, determines the amount of the data, that is, the estimated partial correlations, which are used in the estimation of $\hat{\eta}$ and $\hat{\kappa}$. If ρ is too small, then only a small quantity of data is used for the estimation and the estimates $\hat{\eta}$ and $\hat{\kappa}$ become vulnerable to noise fluctuation in data. On the other hand, if ρ is too large, then a large part of data corresponding to nonzero true partial correlations is included in the estimation. Therefore we propose the following steps to choose a proper cutoff value. For simplicity, we suppose that the negative side of the given data is used.

Cutoff selection procedure:

- (a) Step 1. Estimate $\hat{\kappa}$ for each value of ρ in a grid, e.g., $\rho \in \{0.1, 0.15, 0.2, \dots, 1\}$. We denote the estimates by $\hat{\kappa}_1, \dots, \hat{\kappa}_n$ and the grid points by ρ_1, \dots, ρ_n .
- (b) Step 2. Calculate the differences in the values of $1/\hat{\kappa}$, i.e., let $\delta_i = 1/\hat{\kappa}_{i+1} - 1/\hat{\kappa}_i$. δ_i represents variability in estimation.
- (c) Step 3. Select large enough values of ρ so that the estimates are not largely variable due to noise. Specifically, pick up i^* such that $i^* = \max(i^{*+}, i^{*-})$ where $i^{*+} = \text{argmax}_i \{\delta_i : \delta_i > 0\}$ and $i^{*-} = \text{argmin}_i \{\delta_i : \delta_i < 0\}$. This process excludes an unreliable part, and the interval $\rho_{i^*+1} \leq \rho \leq \rho_n$ is considered as valid candidate values in the next step.
- (d) Step 4. Find the minimum point of $1/\hat{\kappa}_i$ in the valid interval, i.e., let $i^{**} = \text{argmin}_i \{1/\hat{\kappa}_i : i^* + 1 \leq i \leq n\}$. Finally, $\rho_{i^{**}}$ is the selected cutoff value.

The reason why we select the minimum point of $1/\hat{\kappa}_i$ at Step 4 is partly due to the intuition that the more the interval $[r(\rho), 0]$ contains the area under $(1 - \eta_0)f_A(x)$ relatively to

$\eta_0 f_0(x; \kappa_0)$, the larger the variance, $1/\hat{\kappa}$, of $f_0(x; \hat{\kappa})$ becomes. This intuition is justified by the following theorem:

THEOREM 1. *Suppose that $f^s(x) = c_1(\eta_0 f_0(x; \kappa_0) + (1 - \eta_0)f_A(x))$ for $-1 \leq x \leq 0$ for some $c_1 > 0$, and $f^s(x) = f^s(-x)$ for $0 < x \leq 1$. Suppose that there exists a value $-1 \leq r_1 < 0$ such that $f_A(x) = 0$ for $r_1 \leq x \leq 0$. Let $\hat{\eta} = \hat{\eta}(r)$ and $\hat{\kappa} = \hat{\kappa}(r)$ be the estimates determined by (9). Then each of the following holds:*

- (a) *If $r_1 \leq r < 0$, then $\hat{\eta}(r) = c_1 \eta_0$ and $\hat{\kappa}(r) = \kappa_0$.*
- (b) *If $-1 \leq r < r_1$ and $\hat{\eta} f_0(x; \hat{\kappa}) \leq f^s(x)$ for some $x \in [r_1, 0]$, then $\hat{\kappa}(r) \leq \kappa_0$.*
- (c) *If $-1 \leq r < r_1$ and $\hat{\eta} f_0(x; \hat{\kappa}) > f^s(x)$ for all $x \in [r_1, 0]$, then $\hat{\kappa}(r) \leq \kappa_0$ provided $2c_1(1 - \eta_0)f_A(x_1) \leq \frac{r}{r_1}(\hat{\eta} f_0(x_1; \hat{\kappa}) - c_1 \eta_0 f_0(x_1; \kappa_0))$ for all $x_1 \in (r, r_1)$.*

PROOF. See Appendix A.

The condition in statement (b) of Theorem 1 is not a strong one in the sense that we usually expect the estimated null density to be smaller than the empirical density, i.e., $\hat{\eta} f_0(x; \hat{\kappa}) \leq f^s(x)$. Theorem 1 can be extended to cases including mild noise conditions, in other words, a noise process with mean zero may be added to $f^s(x)$ and we will get the same results. In Section 3 we show that the above cutoff selection procedure performs well for a wide range of simulated data sets generated from various VAR models.

3. EXPERIMENTS

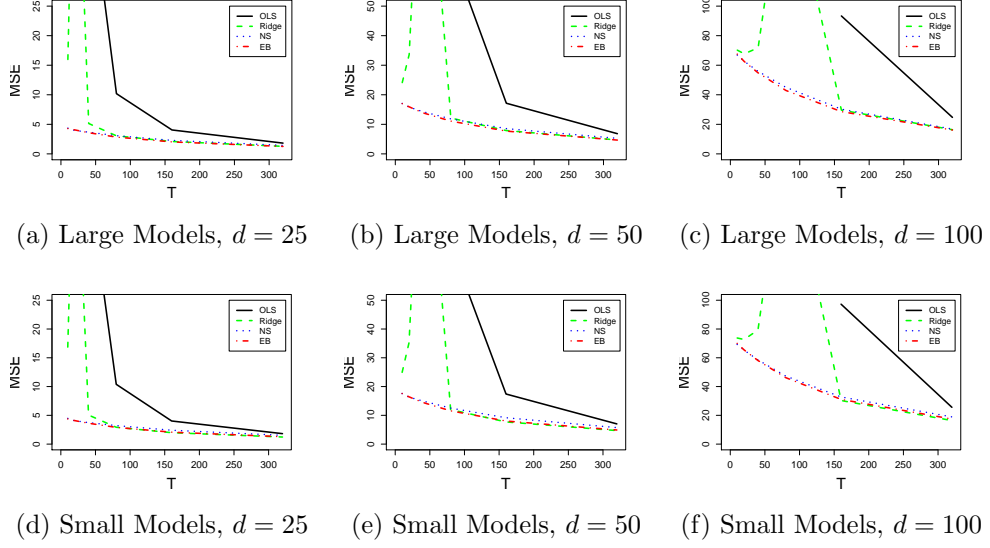
3.1. Simulated Data Sets

To compare performances of the proposed procedures with other competing procedures, we generated VAR models with $p = 1$ and $d \in \{25, 50, 100\}$. Based on the way the nonzero coefficients are created, the VAR models are classified into two groups as follows:

- (a) **Large Models:** Among the total of $d^2 p$ coefficients in the VAR models, 30% of the coefficients were randomly selected and set to 0.1, 10% of the coefficients were randomly selected and set to -0.2 , and the rest were set to zero.
- (b) **Small Models:** Among the total of $d^2 p$ coefficients in the VAR models, 4% of the coefficients were randomly selected and set to 0.3, 1% of the coefficients were randomly selected and set to -0.6 , and the rest were set to zero.

Using these models, we generated multivariate time series data with $T \in \{10, 20, 40, 80, 160, 320\}$ and $V = 0.01I$. Note that the nonzero coefficients of the models are drawn from skewed distributions. That is, the locations and the numbers of the positive and the negative nonzero coefficients are different. In Fig. 6 we can see that the data sets generated in this way form distributions of the estimated partial correlations of fMRI data sets as anticipated.

We repeatedly selected VAR models and their data sets 50 times for performance evaluation. For evaluation of the empirical Bayesian shrinkage estimation (EB) procedure, we also ran other procedures such as ordinary least squares regression (OLS)(Seber and Lee, 2003), ridge regression (RID)(Golub, Heath and Wahba, 1979), and nonparametric shrinkage estimation (NS)(Opgein-Rhein and Strimmer, 2007). We calculated the mean squared

**Fig. 3.** Parameter estimation performance by MSE

error (MSE) between true model coefficients and estimated coefficients, that is,

$$\widehat{MSE}(l) = \left\| \Phi - \tilde{\Phi}(l) \right\|^2 = \sum_{i=1}^{dp} \sum_{j=1}^d (\phi_{ij} - \tilde{\phi}(l)_{ij})^2$$

where $\tilde{\Phi}(l) = (\tilde{\phi}(l)_{ij})$ is the estimated coefficient matrix for the l th data set. Fig. 3 shows the MSE values averaged over the 50 data sets. In the figure the EB and the NS procedures give low MSE values while the EB procedure produces the smallest MSEs for small T values. The RID procedure gives small MSE for large T but it is very unstable for small T . We get similar results for both small and large models.

Next, we compared the modified local FDR procedure with the standard local FDR procedure by Strimmer (2008). We denote the modified one by "ebfdr" and the standard one by "fdrtool". We compared the performances of selecting nonzero true partial correlations from the estimated partial correlations obtained by either the EB or the NS procedures. The performances are measured through the numbers of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). The scores of precision, recall and F1 score are defined accordingly by

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$\frac{1}{F1score} = \frac{Precision^{-1} + Recall^{-1}}{2}$$

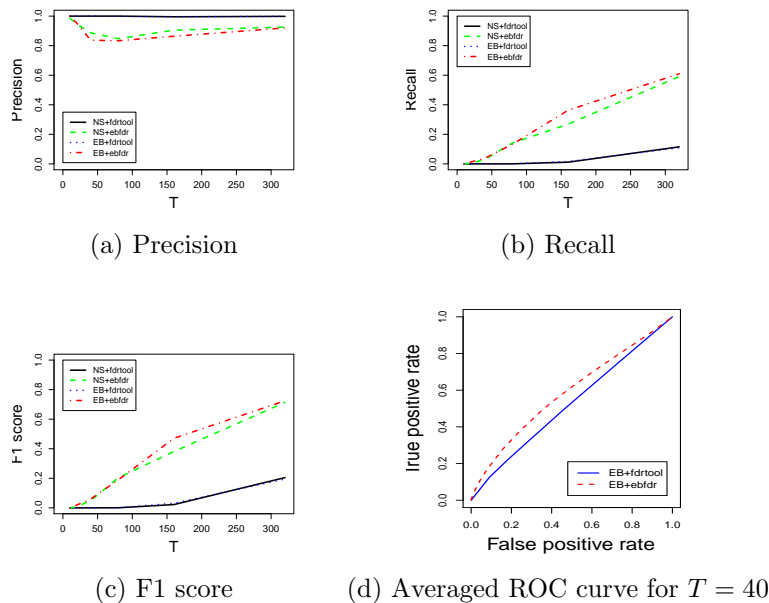


Fig. 4. Nonzero model coefficients detection performances by (a) precision, (b) recall, (c) F1 score and (d) ROC curve for the large models with $p = 1$ and $d = 50$

The receiver operating characteristic (ROC) curves were also drawn based on local FDR scores provided by either the ebfdr or the fdrtool. The 50 ROC curves were averaged vertically, i.e., averaged with respect to true positive rates.

Figs. 4 and 5 illustrate the performance scores of the ebfdr and the fdrtool procedures when they are provided data obtained by either the EB or the NS procedures. We can find that the fdrtool, colored in black and blue, selects very small number of positives as candidate nonzero coefficients so that it has very small recall scores together with high precisions. On the other hand, the ebfdr, colored in green and red, selects a larger number of positives so that its precision is slightly lower than that of the fdrtool but its recall is much larger than that of the fdrtool. We note that the precision of the ebfdr is kept around or above 0.8 which is as anticipated by the threshold $\text{fdr}(x) \leq 0.2$. And the results of averaged ROC curves show that the ebfdr is superior to the fdrtool, which is due to the increments in the number of true positives while keeping the number of false positives low. The NS and EB procedures do not make notable differences but the precision of the NS is slightly larger than that of the EB, and the recall of the EB is slightly larger than that of the NS.

Fig. 6 illustrates the cutoff selection procedure. We used the large models with $p = 1$ and $d = 50$, and the estimated partial correlations were obtained by the EB procedure. We note that the empirical distributions in Fig. 6 (a) and (b) look similar to those in Fig. 1 for the fMRI data sets. Especially, we note that the shifted modes are caused by the mixture of null and alternative distributions.

Moreover, comparing the histograms for $T = 40$ and $T = 160$, we note that the alternative distribution for $T = 160$ is more clearly identified than $T = 40$. This observation is

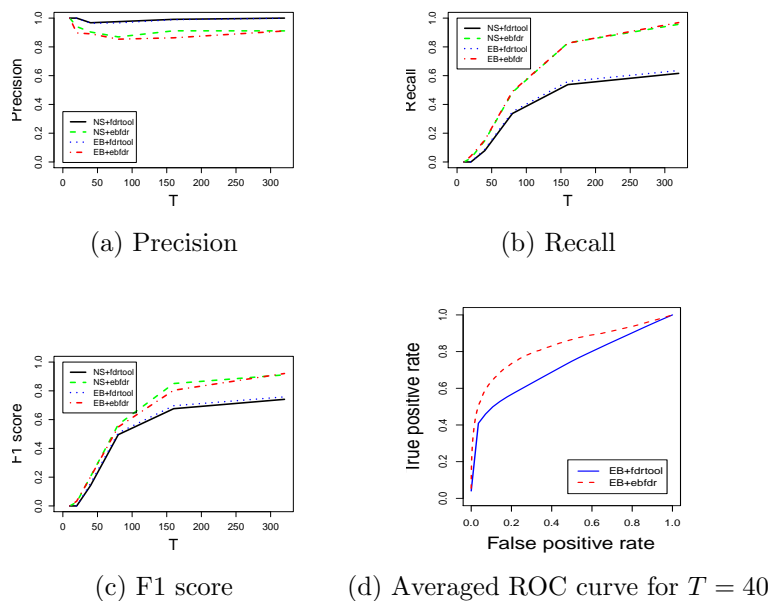


Fig. 5. Nonzero model coefficients detection performances by (a) precision, (b) recall, (c) F1 score and (d) ROC curve for the small models with $p = 1$ and $d = 50$

reflected by the selected cutoff values: when T is small, an alternative distribution is mixed with null distribution on a region close to zero, and a smaller cutoff value is selected. But when T is large, an alternative distribution is less mixed with null distribution on a region close to zero, and a larger cutoff value is selected. In Fig. 6, $\rho = 0.5$ was selected for $T = 40$ whereas $\rho = 0.7$ was selected for $T = 160$. In Fig 6 (c) and (d) the red dotted lines represent the differences in the values of $1/\hat{\kappa}$, i.e., $\delta_i = 1/\hat{\kappa}_{i+1} - 1/\hat{\kappa}_i$, as was defined in Step 2 in the cutoff selection procedure. We note that $1/\hat{\kappa}$ is unstable for small cutoff values and the proposed procedure safely avoids the unstable regions.

The performance of the cutoff selection procedure was compared with methods of using fixed cutoff values by calculating the scores of precision and recall. Fig. 7 shows that the proposed cutoff selection procedure keeps precisions greater than 0.8 while it gives high recall scores, that is, it gives the second largest recall scores for large models and the largest recall scores for small models. On the other hand, the fixed cutoff value of 0.2, depicted by black lines, gives the least values of precision, while it gives the highest recall scores for large models and comparable recall scores for small models. The fixed cutoff values of 0.6 and 1.0, depicted by green and blue dotted lines, give high precision scores and low recall scores, which implies that they produce a smaller number of true positives compared with other cutoff values. Therefore, we conclude that the fixed cutoff values are unreliable and they are subject to the data size, while the proposed cutoff selection procedure shows stable and reliable performance results.

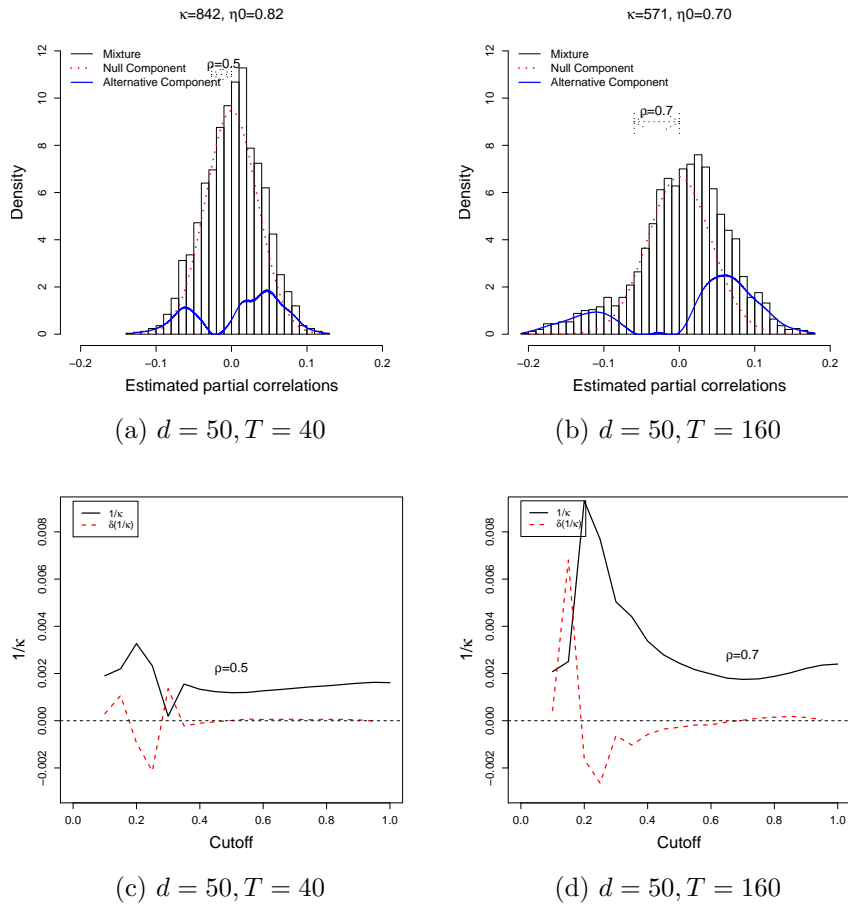
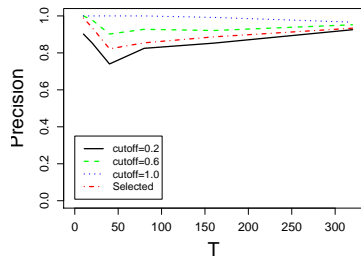
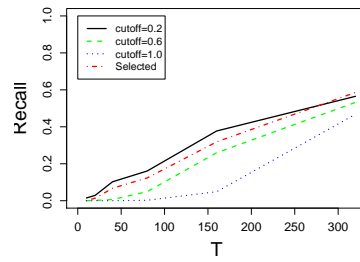


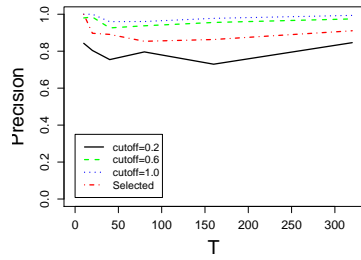
Fig. 6. Illustration of the procedure for cutoff value selection based on the changes in $1/\hat{\kappa}$ values, when $T = 40$ (a and c) and $T = 160$ (b and d)



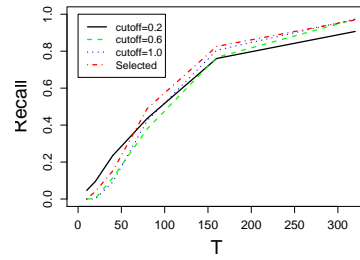
(a) Precision, large model



(b) Recall, large model



(c) Precision, small model



(d) Recall, small model

Fig. 7. Cutoff selection procedure performance by precision and recall for large and small models with $p = 1$ and $d = 50$

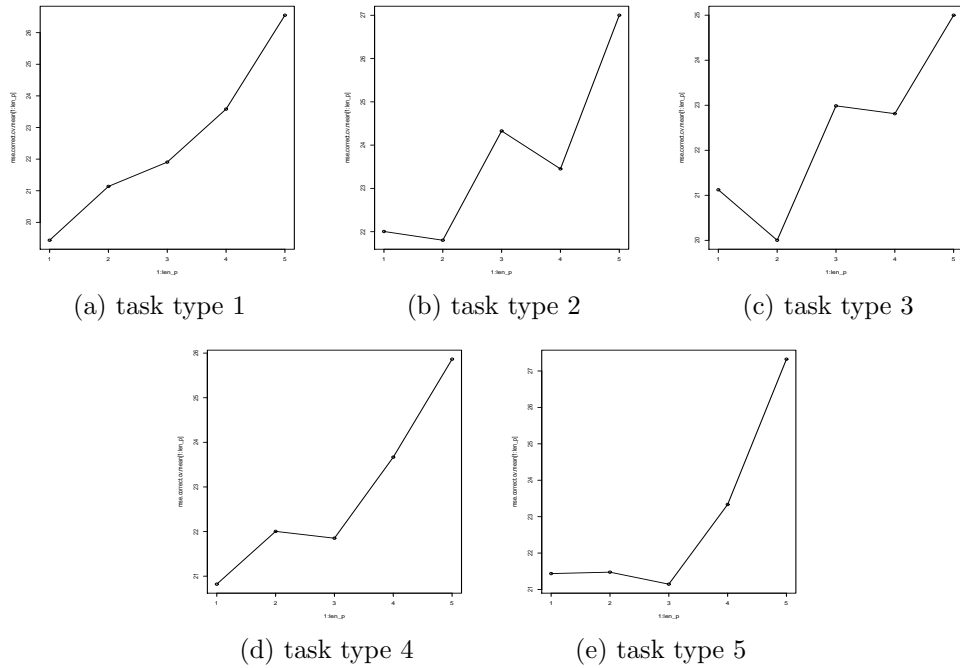


Fig. 8. Model order p versus cross validation error plot for each tasks to help determine the most proper value of p for the data set of the 16th subject

3.2. Multitask fMRI Data Set

We applied the proposed procedures to the multitask fMRI data sets introduced in Section 2. To determine the order p of VAR models, we analyzed the average of the k prediction errors given in (8) obtained by the k -fold cross validation procedure. Fig. 8 illustrates the averaged cross validation error values for data set of the 16th subject over different values of $p \in \{1, 2, 3, 4, 5\}$ where we set k to 5. We observe that for all tasks the optimal value of p is determined at $p = 1, 2$, or 3, and $p = 1$ results in small cross validation errors on average over all tasks. Data sets of other subjects show almost similar results. Therefore in this real fMRI data experiments we use $p = 1$. We note that the actual time difference between two consecutive scans is 3 seconds, and we conjecture that the time difference is long enough for two brain regions to transfer information.

Table 2 summarizes information on the VAR networks obtained for each of the 5 tasks for the 16th subject. All the five networks have the 82 nodes connected. Total degrees denote the sum of the degrees of all the nodes in the network, which is exactly twice the number of edges. The hub nodes were identified based on the degree of each node, which will be explained later. We note that among the hub nodes, the nodes representing right hemisphere regions of the brain dominate for task types 1 and 2 which are relatively easy tasks, while the nodes representing left hemisphere regions of the brain dominate for task types 3, 4 and 5 which are relatively difficult tasks.

Such a difference of connectivity between right and left hemispheres in relations with the task types are more investigated in Fig. 9 for the data sets of all the 27 subjects. The figure shows the ratio of the number of hubs in the right hemisphere to the number of all

Table 2. Summary on the obtained networks for the 16th subject. Note that among the hub nodes, the nodes representing right hemisphere regions dominate for task types 1 and 2 which are relatively easy tasks, while the nodes representing left hemisphere regions dominate for task types 3, 4 and 5 which are relatively difficult tasks.

task type	nodes	edges	total degrees	hub nodes
1	82	589	1178	R26, L21, R23, R22, R42, L39, R17, L23, R21, R43, R44, R9
2	82	793	1586	L36, R30, R21, L28, R22, R20, R35, L21, L35, R19, R7, R1, R28, L17, L19
3	82	412	824	R41, L28, L34, L18, R39, R20, L37
4	82	585	1170	L26, L5, R29, R5, L34, L43, L29, L27, L20
5	82	1244	2488	L29, R3, L18, R4, L41, L9, L27, R5, L3, L44, R6

the hubs for subjects whose number of hubs is at least 5. We can see that for task types 1 and 2, more than half of the hubs are located on the right hemisphere for most of the subjects. On the other hand, for tasks 3, 4 and 5, we can see that the ratio varies more largely between subjects and the median of the ratio is not larger than 0.5. We can conclude that the effective connectivity for brain networks adaptively changes according to the task types.

Figs. 11 to 15 in Appendix B depict subgraphs with 50 edges of the obtained VAR networks for the 5 task types for the data set of the 16th subject. The solid and dotted lines indicate positive and negative partial correlation coefficients, respectively. Each node is labeled by "L" or "R" which represents the left or the right hemisphere of the brain, and the number in each node represents the Brodmann area number. Some of the nodes are colored in yellow to indicate that they are hub nodes.

To identify important brain regions which are related with the other regions the most, we first calculated the degree of each node in the networks. The degree of a node is the number of its adjacent edges. It is the most fundamental measure of graph centrality and the other measures such as closeness and betweenness are closely related with it (Bullmore and Sporns, 2009). Figs. 16 to 20 in Appendix B illustrate the degrees of all the nodes in the networks of the 16th subject in a descending order. We classified the nodes with degrees greater than one standard deviation above the network mean as hub nodes (Sporns, Honey and Kötter, 2007). The threshold value is indicated by red dotted lines crossing the plots vertically.

Figs. 16 to 20 in Appendix B also present the histograms for the degrees of nodes in the networks. If the networks were random networks each pair of nodes would be connected with equal probability and the networks would result in symmetrically centered degree distributions. However, the obtained 5 networks clearly show skewed heavy-tailed degree distributions, which is one of the basic properties of complex networks (Bullmore and Sporns, 2009).

To show that the networks obtained by the proposed procedures for the given fMRI

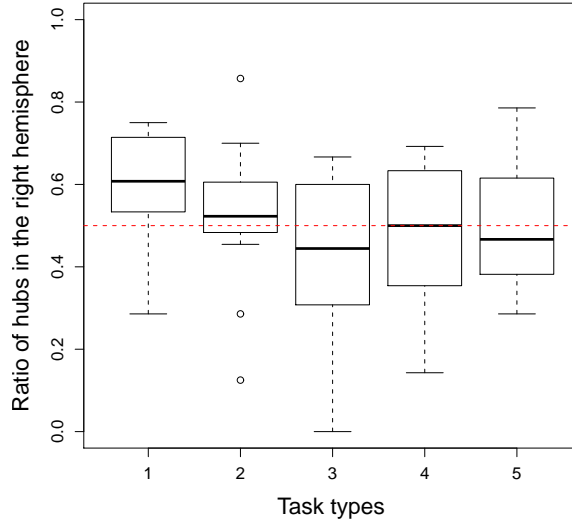


Fig. 9. The ratio of the number of hubs in the right hemisphere to the number of whole hubs for subjects whose number of hubs is at least 5. For task types 1 and 2, more than half of the hubs are located on the right hemisphere for most of the subjects.

data sets yield significantly skewed degree distributions, we conducted statistical tests by comparing the skewness of degree distribution of each of the fMRI networks to those of random networks. The skewness of a degree distribution is calculated by

$$Skewness = \frac{\frac{1}{d} \sum_{i=1}^d (C^D(i) - \bar{C}^D)^3}{\left(\frac{1}{d} \sum_{i=1}^d (C^D(i) - \bar{C}^D)^2\right)^{3/2}}$$

where $C^D(i)$ represents the degree of i th node and \bar{C}^D is the mean of the degrees. A positive skewness indicates a long right tail of the degree distribution, and vice versa. Moreover, we created each of the random networks as a directed network with a given number of edges in such a way that loop edges are allowed but multiple edges are not allowed as networks for VAR models.

Fig. 10 illustrates the skewnesses of the degree distributions of the random networks and the networks obtained from fMRI data sets of all subjects for each of the 5 task types. As for the random networks, we generated 30 random networks for each number of edges and calculated the mean of the skewnesses and the plus/minus one standard deviation around the mean, which is represented by the black error bar in the figure. On the other hand, each point in the figure represents the skewness for the network from the fMRI data set from a subject for each task type. We can see that most of the skewnesses for the fMRI networks are significantly larger than those of random networks. Specifically, we can further estimate p -values for each of the fMRI networks for the statistical test whose alternative hypothesis is that the skewness for the fMRI network is larger than those for random networks. After generating 100 random networks with the same number of edges to that of the fMRI network,

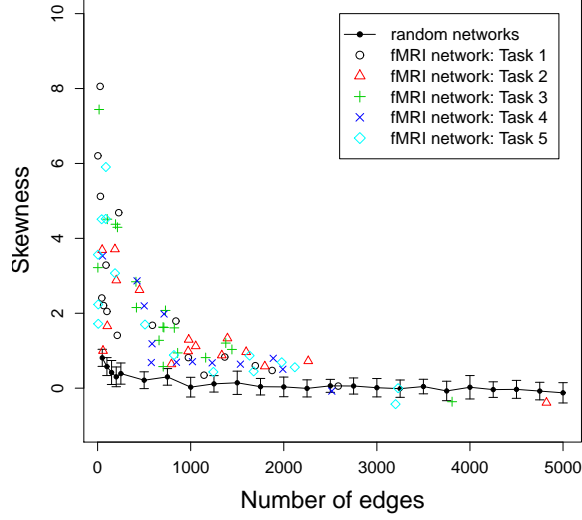


Fig. 10. Skewnesses of degree distributions of random networks and the networks obtained by the proposed procedures for each of the subjects and task types. Most of the skewnesses for the obtained networks of brain regions are significantly larger than those for random networks.

the p -value is estimated by calculating the number of the random networks whose degree distributions have skewnesses larger than that of the fMRI network, i.e.,

$$p\text{-value} = \frac{\#\{\text{random networks with skewnesses} > \text{skewness of fMRI network}\}}{\#\{\text{random networks}\}}.$$

As for the 16th subject, we obtained the skewnesses 1.68, 0.64, 2.84, 1.19, and 0.43 for task types 1, 2, 3, 4 and 5, respectively. The number of edges of the networks are 589, 793, 412, 585 and 1244, respectively. The estimated p -values are 0.00, 0.04, 0.00, 0.00 and 0.10. Therefore, all the 5 networks have sufficiently small p -values to reject the null hypothesis with a significance level 0.1. Since a large skewness is usually caused by long right tail of the degree distribution, the high degree nodes are key factors to the nonrandomness of the networks.

To further investigate characteristics of each brain region, we calculated betweenness centrality of each node in the networks (Bullmore and Sporns, 2009). The betweenness centrality is defined by the number of shortest paths going through a node. Specifically, it is defined by

$$C^B(i) = \sum_{s \neq v, s \neq i \neq v} \frac{g_{siv}}{g_{sv}}$$

where g_{sv} is the number of shortest paths between nodes s and v , and g_{siv} is the number of shortest paths between nodes s and v passing through the node i . Since the determination of shortest paths is largely dependent on whether the network edges are directed or not, providing directed networks of brain regions is crucial to discovering the betweenness of brain

Table 3. Brodmann areas with high betweenness centralities according to the causal networks obtained for the 16th subject. Only top six Brodmann areas are presented in the order of their ranks for each of the five task types. The betweenness centrality measures are also presented in parenthesis. Brodmann areas located at the medial part of the brain closely to hippocampus are emphasized with bold type.

Rank	Type 1	Type 2	Type 3	Type 4	Type 5
1	R26 (740.18)	L36 (671.94)	R41(1298.34)	L26 (740.30)	L29 (678.50)
2	L21(446.74)	R30 (463.56)	L21(1180.58)	R29 (639.48)	R4(430.10)
3	L38(423.57)	L19(441.28)	L36 (903.80)	L27 (454.00)	R3(320.41)
4	R35 (359.02)	L35 (349.28)	L28 (673.52)	L5((399.07)	L9(289.28)
5	R22(340.84)	R35 (303.73)	L30 (615.57)	L45(318.73)	L27 (285.54)
6	R23 (313.39)	L28 (266.75)	L1(533.00)	R2(317.25)	R29 (274.82)

regions. Moreover, a node with high betweenness centrality is considered to be essential to efficient communication between nodes.

We also calculated other graph centrality measures of each node such as out-closeness and in-closeness, which are also distinctive for directed networks. The closeness centrality measures the number of steps required from a node to another in the network. Specifically, it is defined by the inverse of the average length of the shortest paths to or from the other nodes in the network:

$$C^C(i) = \frac{d-1}{\sum_{j \neq i} d_{ij}}$$

where d is the number of nodes in the network and d_{ij} is the length of the shortest path between nodes i and j . The out-closeness and the in-closeness centralities are further defined by considering the direction of the shortest paths. A node with high out-closeness centrality is considered to influence many other nodes through short or direct paths, while a node with high in-closeness centrality is considered to be influenced by many other nodes.

Tables 5 to 9 in Appendix B list the nodes classified as hub nodes and their graph centrality measures for each task type for the 16th subject. Each number in parenthesis represents the rank according to the measure. We note that high-degree nodes have either high out-closeness or high in-closeness measures together with high betweenness measures in general. This is because those measures are related with the degree centrality. Moreover, we note that many of the hub nodes have high out-closeness centralities compared to the in-closeness centralities, especially for task types 1, 3 and 4. Based on this observation we can guess that many hub regions of the brain are taking important roles as information senders to other regions.

There are some more findings. First, we note that many of the Brodmann areas with high betweenness centralities have memory functions and, anatomically, occupy the medial part of the brain which is close to hippocampus. Table 3 presents a list of Brodmann areas with the highest betweenness centralities in each of the five networks obtained for the 16th subject. According to Table 4, Brodmann area 23 is labeled as Posterior cingulate cortex, areas 26, 29 and 30 as Retrosplenial cingulate cortex, area 27 as Piriform cortex, areas 28 and 34 as Entorhinal cortex, and areas 35 and 36 as Perirhinal cortex. Those Brodmann areas are anatomically located in the medial part of the brain, close to the hippocampus, and supposed to be related to the memory functions. In Table 3 those Brodmann areas are emphasized with bold type, which cover more than half of the Brodmann areas listed in the table.

Second, the other Brodmann areas in Table 3 which are not written in bold are supposed

to characterize the adaptive changes of brain networks to different cognitive loads for efficient information processing. Task type 1 requires Brodmann areas L21, R22, and L38, which are in the temporal lobe. Task type 2 requires L19 in the occipital lobe. Task type 3 requires L21 and R41 in the temporal lobe and L1 in the parietal lobe. Task type 4 requires L45 in the frontal lobe and R2 and L5 in the parietal lobe. Task type 5 requires R4 and L9 in the frontal lobe and R3 in the parietal lobe.

These changes in network structures are supposed to be related with the changes in the subject's specific strategies depending on the cognitive loads of the five task types. Based on the categorization of strategies by Glück and Fitting (2003), the holistic strategies involve visualization and other ways of representing spatial relations, whereas the analytic strategies, sometimes labeled as verbal strategies, are considered to reduce spatial information to an essentially nonspatial, listlike format. We note that people can solve easy tasks by holistic strategies, whereas people use analytic strategies more for tasks of higher difficulty.

In the fMRI experiment, the task types 1 and 2 are considered easy and the task types 3, 4 and 5 are considered difficult. We note that Brodmann areas in frontal and parietal lobes are more included among the high betweenness Brodmann areas and high degree Brodmann areas as well for task types 4 and 5. On the other hand, the visual areas of the brain in the occipital lobe are more included for task type 2. This observation leads us to the conclusion that analyzing high betweenness Brodmann areas and high degree Brodmann areas helps to find the strategies used by the subjects.

4. CONCLUSION

In this paper we proposed an improved process of inferring large-scale causal networks for modeling high dimensional fMRI data with VAR models. We proposed empirical Bayesian shrinkage estimation procedure to improve the existing nonparametric shrinkage estimation algorithm by incorporating the dependence assumption in time series data appropriately, so that we can handle data of both small and large numbers of observations, i.e., scans, while a large number of brain regions is included in data. We also proposed a modified local FDR procedure for pruning unlikely connections especially for the skewed empirical distribution of the estimated partial correlations. We also proposed a reliable cutoff selection algorithm with mathematical justification. In a nutshell, we can say that the proposed procedures for the inference of large-scale causal networks are applicable to a wide variety of real fMRI data.

We applied the proposed procedures to the fMRI data. In the experiment, subjects took psychometric tests to assess different aspects of spatial abilities. The discovered directed networks of Brodmann areas revealed important brain regions involved in the given cognitive tasks in terms of the graph theoretic measures such as degree and betweenness centralities. We have shown that the discovered networks are far from randomly connected networks, implicating dynamical behaviors of brain regions in visual information processing.

We could further observe that the brain networks change adaptively according to the given cognitive tasks by analyzing the level of involvement of brain regions in each hemisphere in terms of the number of identified hubs over all the subjects. Due to the directed nature of the obtained networks, we could further analyze such graph theoretic measures as out-closeness, in-closeness, and betweenness centralities. For analyzing the betweenness centralities, we selected the 16th subject who scored almost equally high on all of the 5 tasks. By analyzing the betweenness centralities of the 16th subject, we could discover that

the memory regions of the brain which are located closely to the hippocampus are taking important roles in efficient information processing over all types of spatial tasks. Moreover, we could discover that the adaptive changes in the brain regions with high betweenness or high degree centralities are closely related to the changes in the strategy uses of the subjects, holistic versus analytic strategies.

The application of graph theoretic measures such as the betweenness centrality in this paper suggested that graph theoretical analysis methods can serve as promising tools for understanding networks of brain regions. Other measures such as modularity and motif are also known to be related to complex behaviors of networks. Moreover, group based analysis can be further conducted based on the graph theoretic measures to investigate differences in strategy uses of subjects more specifically in various cognitive tasks.

A. PROOF OF THEOREM 1

First, we show that (a) holds. If $r_1 \leq r < 0$, then we have

$$\int_r^0 (f^s(x) - \eta f_0(x; \kappa))^2 dx = \int_r^0 (c_1 \eta_0 f_0(x; \kappa_0) - \eta f_0(x; \kappa))^2 dx . \quad (10)$$

If $\eta = c_1 \eta_0$ and $\kappa = \kappa_0$, then (10) is zero. On the other hand, if (10) is zero, then, since $f_0(x; \kappa)$ is a continuous function for any κ , we have $c_1 \eta_0 f_0(x; \kappa_0) = \eta f_0(x; \kappa)$ for all $x \in [r, 0]$. So, it follows that $c_1 \eta_0 f_0(r; \kappa_0) = \eta f_0(r; \kappa)$ and $c_1 \eta_0 f_0(0; \kappa_0) = \eta f_0(0; \kappa)$, which implies $\eta = c_1 \eta_0$ and $\kappa = \kappa_0$. Therefore, the minimization of (10) leads to the unique solution ($\hat{\eta} = c_1 \eta_0, \hat{\kappa} = \kappa_0$).

Next, we show that (b) holds. If $-1 \leq r < r_1$, then we have

$$\int_{r_1}^0 (f^s(x) - \hat{\eta} f_0(x; \hat{\kappa}))^2 dx \geq \int_{r_1}^0 (f^s(x) - c_1 \eta_0 f_0(x; \kappa_0))^2 dx \quad (11)$$

and

$$\int_r^0 (f^s(x) - \hat{\eta} f_0(x; \hat{\kappa}))^2 dx \leq \int_r^0 (f^s(x) - c_1 \eta_0 f_0(x; \kappa_0))^2 dx . \quad (12)$$

By subtracting Eq. (11) from Eq. (12), we get

$$\int_r^{r_1} (f^s(x) - \hat{\eta} f_0(x; \hat{\kappa}))^2 dx \leq \int_r^{r_1} (f^s(x) - c_1 \eta_0 f_0(x; \kappa_0))^2 dx ,$$

which is further expanded to

$$\int_r^{r_1} (c_1 \eta_0 f_0(x; \kappa_0) - \hat{\eta} f_0(x; \hat{\kappa}))^2 dx \leq 2c_1(1 - \eta_0) \int_r^{r_1} f_A(x) (\hat{\eta} f_0(x; \hat{\kappa}) - c_1 \eta_0 f_0(x; \kappa_0)) dx . \quad (13)$$

By the mean value theorem for integration, there exists $x_1 \in (r, r_1)$ such that

$$\int_r^{r_1} f_A(x) (\hat{\eta} f_0(x; \hat{\kappa}) - c_1 \eta_0 f_0(x; \kappa_0)) dx = (r_1 - r) f_A(x_1) (\hat{\eta} f_0(x_1; \hat{\kappa}) - c_1 \eta_0 f_0(x_1; \kappa_0)) . \quad (14)$$

From Eqs. (13) and (14) we can conclude that

$$\hat{\eta} f_0(x_1; \hat{\kappa}) \geq c_1 \eta_0 f_0(x_1; \kappa_0) \quad (15)$$

for some $x_1 \in (r, r_1)$ unless $f_A(x) = 0$ for all $x \in [r, r_1]$. On the other hand, from the assumption of (b), there exists $x_2 \in [r_1, 0]$ such that

$$\hat{\eta}f_0(x_2; \hat{\kappa}) \leq f^s(x_2) = c_1\eta_0f_0(x_2; \kappa_0). \quad (16)$$

By dividing Eq. (15) by (16), we get

$$\frac{f_0(x_1; \hat{\kappa})}{f_0(x_2; \hat{\kappa})} \geq \frac{f_0(x_1; \kappa_0)}{f_0(x_2; \kappa_0)}$$

where $|x_1| > |x_2|$. Since $f_0(x_1; \kappa)/f_0(x_2; \kappa) = ((1 - x_1^2)/(1 - x_2^2))^{(\kappa-3)/2}$, we can conclude that $\hat{\kappa} \leq \kappa_0$.

Finally, we show that (c) holds. By expanding Eq. (12), we get

$$\int_r^0 (c_1\eta_0f_0(x; \kappa_0) - \hat{\eta}f_0(x; \hat{\kappa}))^2 dx \leq 2c_1(1 - \eta_0) \int_r^{r_1} f_A(x)(\hat{\eta}f_0(x; \hat{\kappa}) - c_1\eta_0f_0(x; \kappa_0)) dx.$$

The above inequality is further simplified from $\int_r^0 = \int_r^{r_1} + \int_{r_1}^0$ by

$$\begin{aligned} & \int_{r_1}^0 (c_1\eta_0f_0(x; \kappa_0) - \hat{\eta}f_0(x; \hat{\kappa}))^2 dx \\ & \leq \int_r^{r_1} \left(2c_1(1 - \eta_0)f_A(x) - (\hat{\eta}f_0(x; \hat{\kappa}) - c_1\eta_0f_0(x; \kappa_0)) \right) (\hat{\eta}f_0(x; \hat{\kappa}) - c_1\eta_0f_0(x; \kappa_0)) dx. \end{aligned}$$

By the mean value theorem for integration, there exists $x_1 \in (r, r_1)$ and $x_2 \in (r_1, 0)$ such that

$$\begin{aligned} & -r_1(c_1\eta_0f_0(x_2; \kappa_0) - \hat{\eta}f_0(x_2; \hat{\kappa}))^2 \\ & \leq (r_1 - r) \left(2c_1(1 - \eta_0)f_A(x_1) - (\hat{\eta}f_0(x_1; \hat{\kappa}) - c_1\eta_0f_0(x_1; \kappa_0)) \right) \\ & \quad \times (\hat{\eta}f_0(x_1; \hat{\kappa}) - c_1\eta_0f_0(x_1; \kappa_0)). \end{aligned} \quad (17)$$

Define

$$A(x) = \frac{\hat{\eta}f_0(x; \hat{\kappa}) - c_1\eta_0f_0(x; \kappa_0)}{c_1\eta_0f_0(x; \kappa_0)}$$

and note that $f_0(x_1; \kappa_0) < f_0(x_2; \kappa_0)$ for $|x_1| > |x_2|$. By dividing each side of Eq. (17) by $c_1\eta_0f_0(x_2; \kappa_0)$ and $c_1\eta_0f_0(x_1; \kappa_0)$, we get

$$-r_1A(x_2)^2 < (r_1 - r) \left(\frac{2(1 - \eta_0)f_A(x_1)}{\eta_0f_0(x_1; \kappa_0)} - A(x_1) \right) A(x_1). \quad (18)$$

Note that $A(x_2) > 0$ from the assumption in (c). Moreover, we can derive that $A(x_1) > 0$ from Eq. (18): If $A(x_1) = 0$, then it contradicts to Eq. (18). If $A(x_1) < 0$, then, from Eq. (18), we get $(2(1 - \eta_0)f_A(x_1)/\eta_0f_0(x_1; \kappa_0) - A(x_1)) < 0$, which implies $0 \leq 2(1 - \eta_0)f_A(x_1)/\eta_0f_0(x_1; \kappa_0) < A(x_1)$. This is also a contradiction. Therefore $A(x_1) > 0$.

On the other hand, the condition in (c), i.e., $2c_1(1 - \eta_0)f_A(x_1) \leq \frac{r}{r_1}(\hat{\eta}f_0(x_1; \hat{\kappa}) - c_1\eta_0f_0(x_1; \kappa_0))$, is equivalent to the following one:

$$\frac{r_1 - r}{-r_1} \left(\frac{2(1 - \eta_0)f_A(x_1)}{\eta_0f_0(x_1; \kappa_0)} - A(x_1) \right) \leq A(x_1). \quad (19)$$

Then, from Eq. (18), it follows that $A(x_2) < A(x_1)$. $A(x_2) < A(x_1)$ implies $f_0(x_2; \hat{\kappa})/f_0(x_1; \hat{\kappa}) < f_0(x_2; \kappa_0)/f_0(x_1; \kappa_0)$, and finally we obtain $\hat{\kappa} < \kappa_0$.

B. SUPPLEMENTAL MATERIALS

Table 4. Summary of the anatomical labels and functions of the 41 Brodmann areas

Area No.	Anatomical Labels and Functions
1, 2, 3	[Primary somatosensory cortex] somatosensory information processing
4	[Primary motor cortex] control of voluntary movements
5,7	[Somatosensory association cortex] somatosensory processing and association, integration of visual and motor information
6	[Supplementary motor cortex (medial) and premotor cortex (lateral)] Planning of complex, coordinated movements
8	[Frontal eye field, part of frontal cortex] planning complex movements, the management of uncertainty
9, 46	[Dorsolateral prefrontal cortex] motor planning, organization and regulation
10	[Anterior prefrontal cortex] involved in strategic processes of memory retrieval and executive function
11	[Orbitofrontal cortex] cognitive processing of decision-making
17	[Primary visual cortex (V1)] visual information processing
18	[Secondary visual cortex (V2)] shape recognition and visual attention, storage of object recognition memory
19	[Associative visual cortex (V3)] shape recognition and visual attention, processing of global motion
20	[Inferior temporal gyrus] high-level visual processing and recognition
37	[Fusiform gyrus, occipitotemporal cortex] face and word recognition, within-category identification
23	[Posterior cingulate cortex] involved in emotion system
24, 32	[Anterior cingulate cortex] error detection, anticipation of tasks, attention, motivation, and modulation of emotional responses
25	[Subgenual cingulate cortex] influences a vast network involved in changes in appetite and sleep; the mood and anxiety; memory formulation; self-esteem
26, 29, 30	[Retrosplenial cingulate cortex] recall of episodic information
28, 34	[Entorhinal cortex] memory and navigation
35, 36	[Perirhinal cortex] visual perception and memory
38	[Temporopolar area, part of the temporal cortex] important area in self representation, semantic (left) and autobiographic (right)
27	[Piriform cortex] part of olfactory system, odor identification
43	[Primary gustatory cortex] perception of taste
48	[Retrosubicular area, a small part of the medial surface of the temporal lobe]
41	[Primary auditory cortex] auditory information processing
42	[Auditory association cortex] auditory information processing
21	[Middle temporal gyrus] language and auditory processing
22	[Superior temporal gyrus, of which the posterior part contains Wernicke's area] language and auditory processing
39	[Angular gyrus, part of Wernicke's area] related to language, mathematics and cognition
40	[Supramarginal gyrus, part of Wernicke's area] involved in reading both in regards to meaning and phonology
44, 45	[Broca's area] syntactic and semantic tasks
47	[Inferior prefrontal gyrus] processing of syntax in spoken and signed languages, and more recently in musical syntax.

Table 5. Summary of graph centrality measures for identification and classification of hub nodes of the obtained network for the task type 1 of the 16th subject. Each number in parenthesis represents the rank according to the measure. Note that high-degree nodes have either high out-closeness or high in-closeness measures together with high betweenness measures in general.

hub node	degree	out-closeness	in-closeness	betweenness
R26	53(1)	.382(3)	.0315(37)	740.18(1)
L21	50(2)	.386(2)	.0312(56)	446.74(2)
R23	46(3)	.329(10)	.0316(32)	313.39(6)
R22	40(4)	.360(4)	.0313(54)	340.84(5)
R42	33(5)	.352(5)	.0306(79)	20.67(44)
L39	31(6)	.333(8)	.0308(75)	56.33(31)
R17	31(7)	.305(17)	.0313(49)	145.22(16)
L23	28(8)	.279(29)	.0315(33)	93.23(25)
R21	27(9)	.288(24)	.0313(48)	71.93(28)
R43	27(10)	.340(7)	.0308(73)	198.78(10)
R44	27(11)	.316(14)	.0313(51)	102.46(23)
R9	26(12)	.278(31)	.0315(39)	107.51(22)

REFERENCES

- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, **57**(1), 289–300.
- Brodmann, K. (1909) Vergleichende lokalisationslehre der großhirnrinde in ihren prinzipien dargestellt auf grund des zellenbaues. Leipzig: Barth.
- Bullmore, E. and Sporns, O. (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, **10**(3), 186–198.
- Doan, T., Litterman, R. and Sims, C. A. (1984) Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**(1), 1–100.
- Efron, B. (2003) Robbins, empirical Bayes and microarrays. *Ann. Statist.*, **31**(2), 366–378.
- Efron, B. (2004) Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Statist. Ass.*, **99**(465), 96–104.
- Eickhoff, S. B., Heim, S., Zilles, K. and Amunts, K. (2006) Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *NeuroImage*, **32**(2), 570–582.
- Ekstrom, R. B., French, J. W., Harman, H. H. and Dermen, D. (1976) *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, New Jersey: Educational Testing Service.
- French, J. W., Ekstrom, R. B. and Price, L. A. (1963) *Kit of Reference Tests for Cognitive Factors*. Princeton, New Jersey: Educational Testing Service.
- Friston, K. J., Harrison, L. and Penny, W. (2003) Dynamic causal modelling. *NeuroImage*, **19**, 1273–1302.

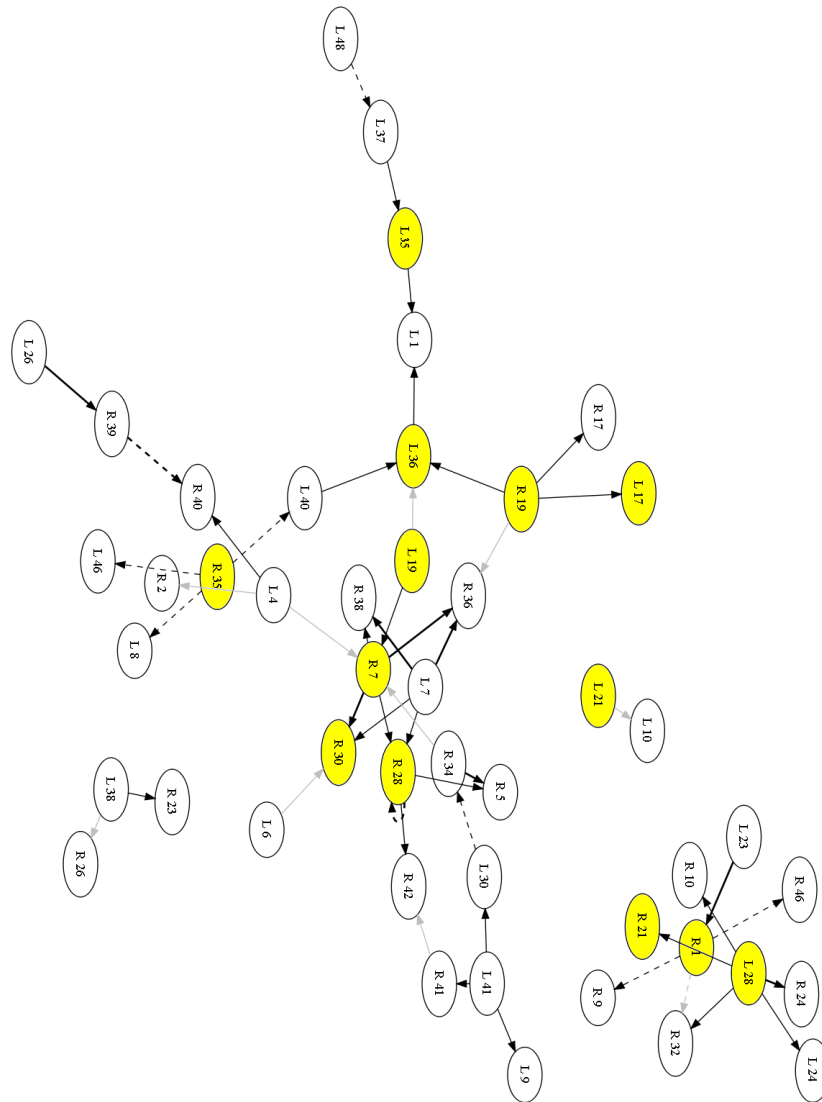


Fig. 12. Subgraph with 50 edges obtained by the suggested procedures for the task type 2. The solid and dotted lines indicate positive and negative partial correlation coefficients, respectively, and the line intensity denotes their strength. Hub nodes are colored in yellow.

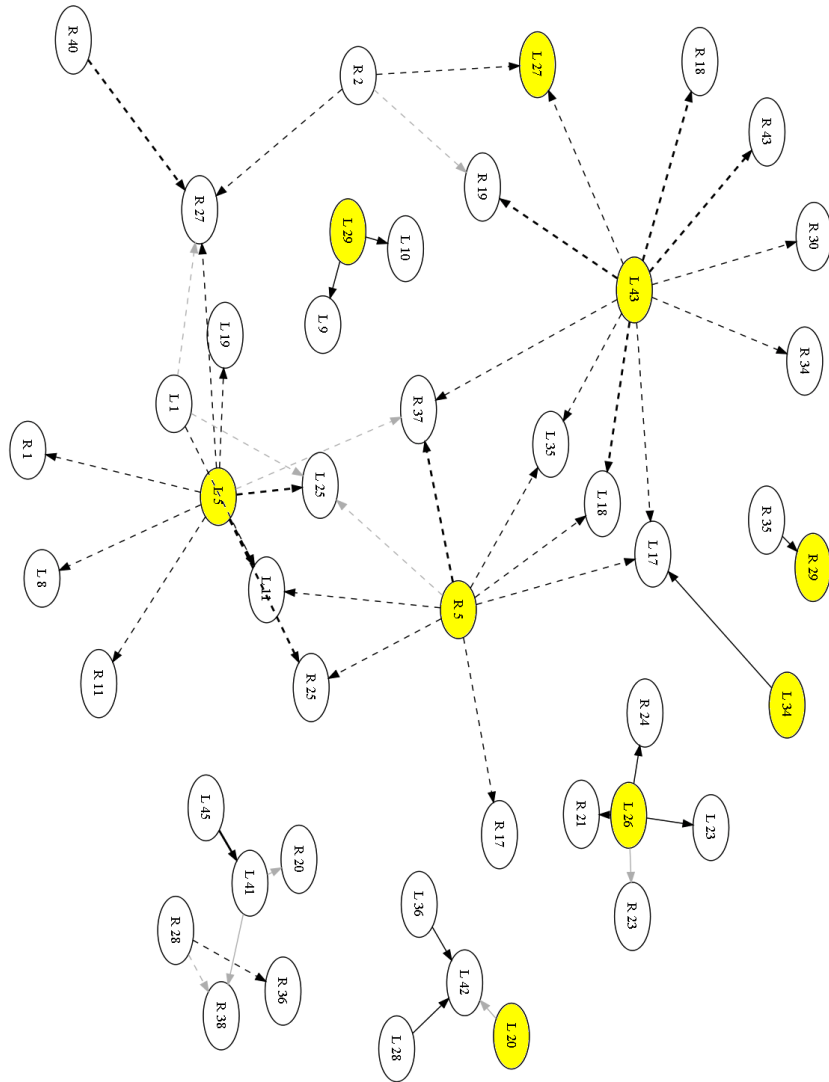


Fig. 14. Subgraph with 50 edges obtained by the suggested procedures for the task type 4. The solid and dotted lines indicate positive and negative partial correlation coefficients, respectively, and the line intensity denotes their strength. Hub nodes are colored in yellow.

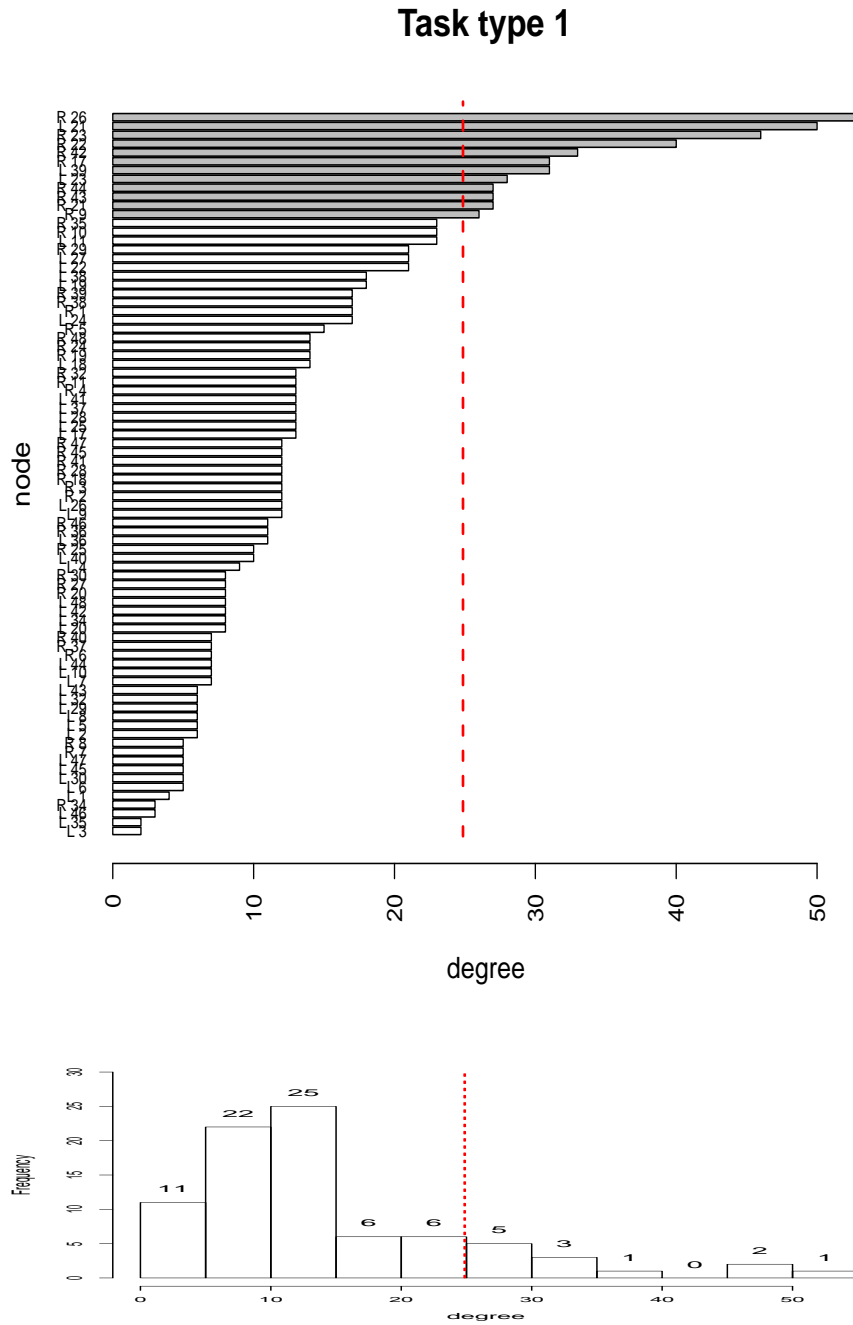
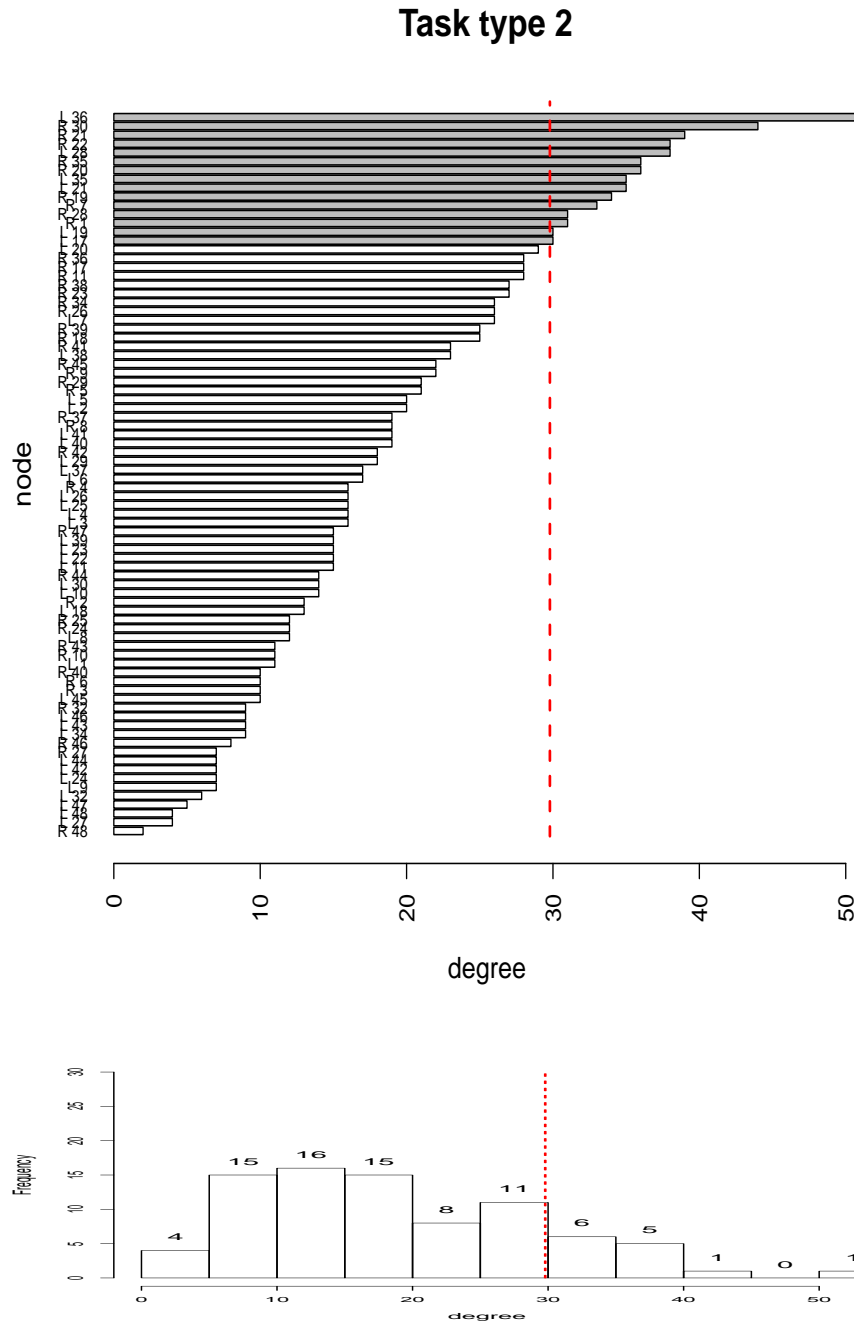


Fig. 16. Degree of each node in the network for task type 1. The histogram in the bottom shows that the degree distribution is far from a Gaussian distribution, which indicates that the obtained causal network is far from random networks. The red dotted line separates hub nodes whose degree is greater than the network mean plus one standard deviation.



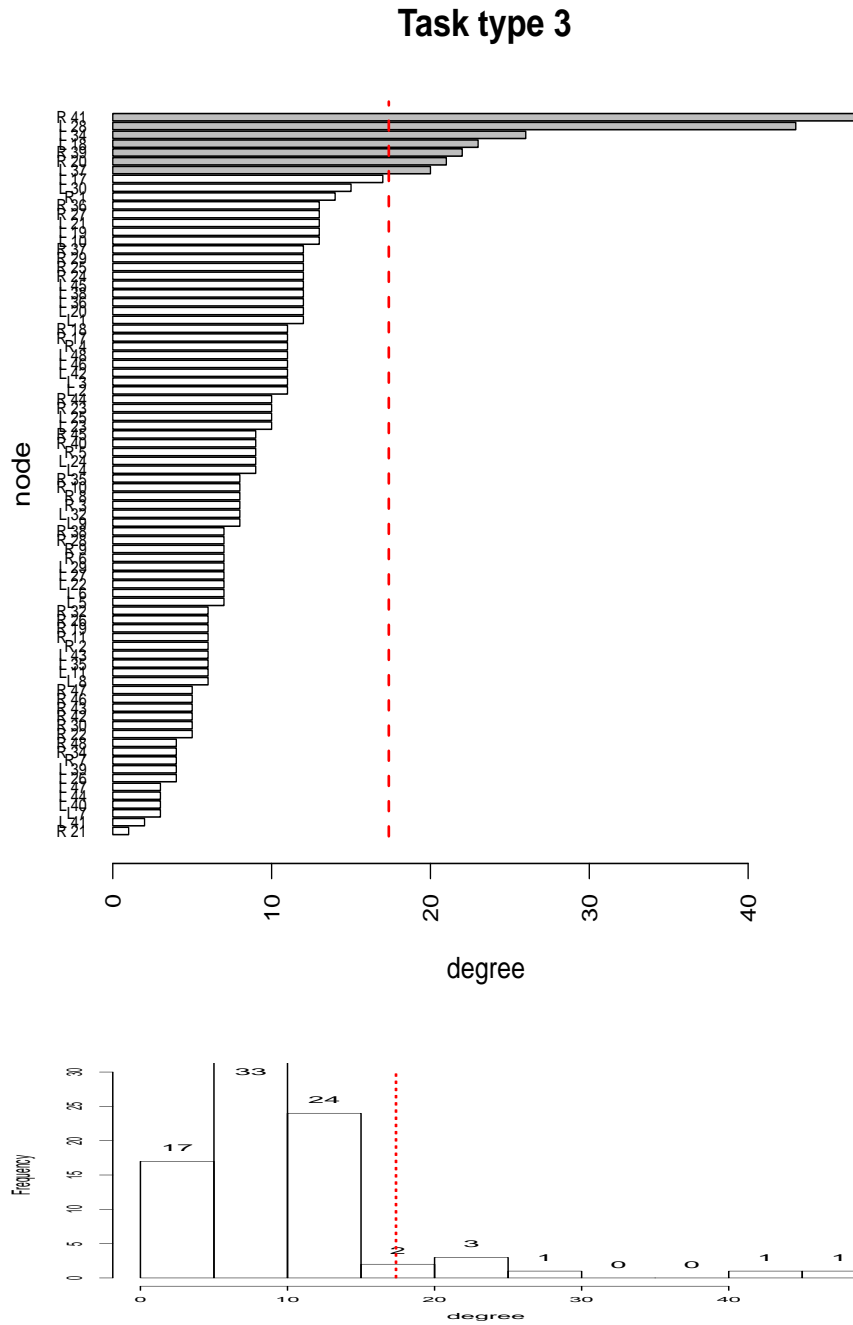


Fig. 18. Degree of each node in the network for task type 3. The histogram in the bottom shows that the degree distribution is far from a Gaussian distribution, which indicates that the obtained causal network is far from random networks. The red dotted line separates hub nodes whose degree is greater than the network mean plus one standard deviation.

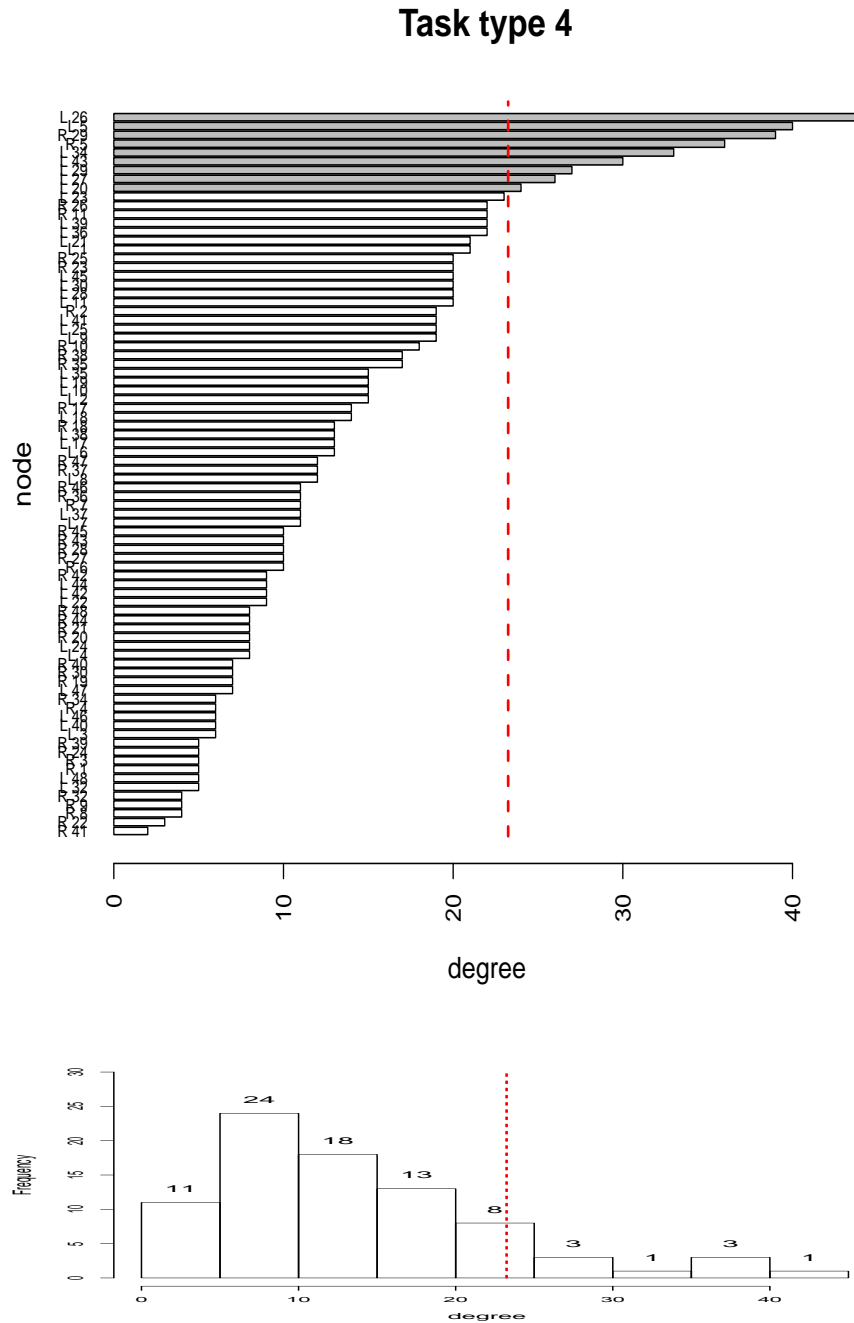


Fig. 19. Degree of each node in the network for task type 4. The histogram in the bottom shows that the degree distribution is far from a Gaussian distribution, which indicates that the obtained causal network is far from random networks. The red dotted line separates hub nodes whose degree is greater than the network mean plus one standard deviation.

Task type 5

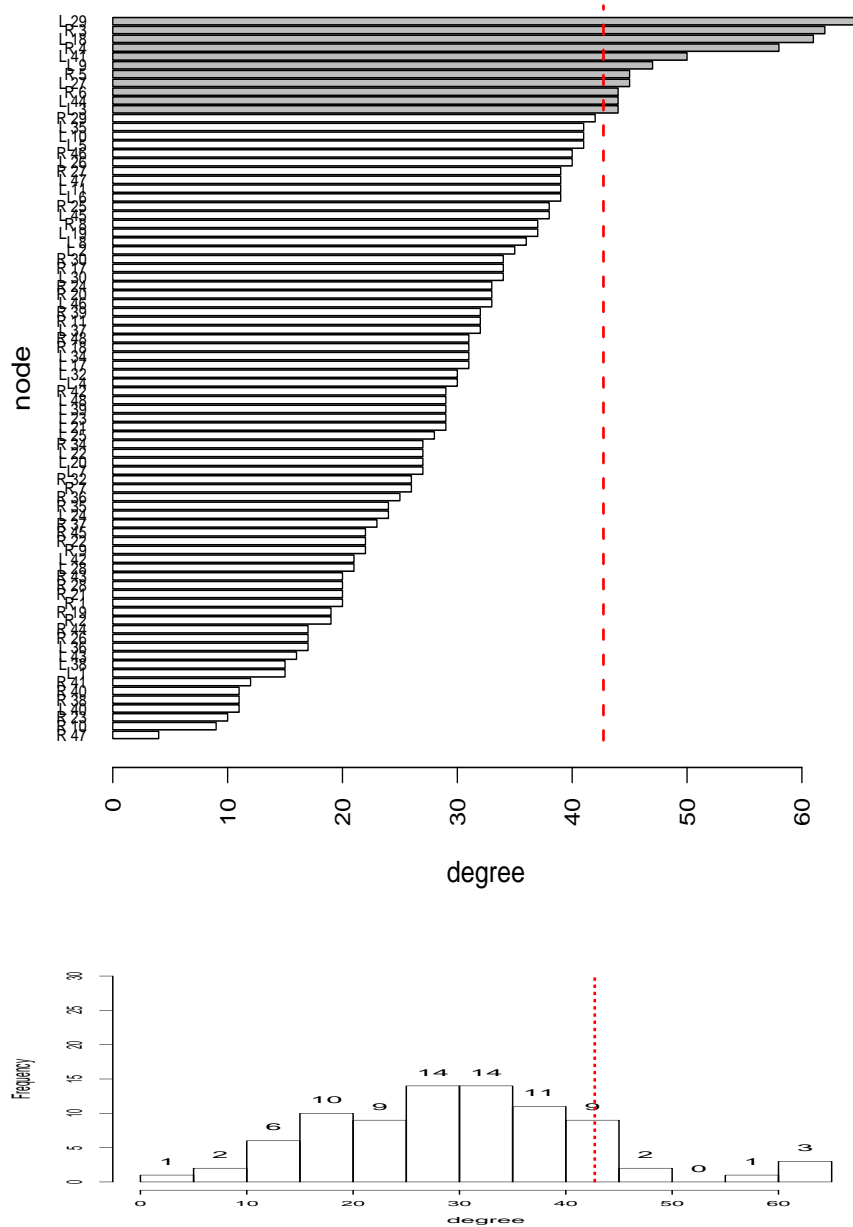


Fig. 20. Degree of each node in the network for task type 5. The histogram in the bottom shows that the degree distribution is far from a Gaussian distribution, which indicates that the obtained causal network is far from random networks. The red dotted line separates hub nodes whose degree is greater than the network mean plus one standard deviation.

Table 6. Summary of graph centrality measures for identification and classification of hub nodes of the obtained network for the task type 2 of the 16th subject. Each number in parenthesis represents the rank according to the measure. Note that high-degree nodes have either high out-closeness or high in-closeness measures together with high betweenness measures in general.

hub node	degree	out-closeness	in-closeness	betweenness
L36	51(1)	.276(2)	.249(22)	671.94(1)
R30	44(2)	.264(7)	.254(14)	463.56(2)
R21	39(3)	.243(26)	.259(8)	123.46(24)
L28	38(4)	.275(3)	.238(39)	266.75(6)
R22	38(5)	.247(19)	.254(13)	126.99(21)
R20	36(6)	.243(25)	.259(7)	145.90(19)
R35	36(7)	.235(38)	.262(4)	303.73(5)
L21	35(8)	.272(4)	.226(53)	197.99(14)
L35	35(9)	.233(41)	.267(3)	349.28(4)
R19	34(10)	.262(8)	.239(38)	193.54(15)
R7	33(11)	.249(17)	.244(29)	262.35(8)
R1	31(12)	.228(52)	.258(9)	263.15(7)
R28	31(13)	.260(9)	.233(44)	175.89(17)
L17	30(14)	.225(56)	.261(5)	59.95(43)
L19	30(15)	.264(6)	.246(23)	441.28(3)

Table 7. Summary of graph centrality measures for identification and classification of hub nodes of the obtained network for the task type 3 of the 16th subject. Each number in parenthesis represents the rank according to the measure. Note that high-degree nodes have either high out-closeness or high in-closeness measures together with high betweenness measures in general.

hub node	degree	out-closeness	in-closeness	betweenness
R41	47(1)	.395(1)	.0527(48)	1298.34(1)
L28	43(2)	.375(2)	.0519(69)	673.52(4)
L34	26(3)	.313(7)	.0525(55)	342.39(13)
L18	23(4)	.306(8)	.0525(54)	279.86(15)
R39	22(5)	.340(4)	.0499(81)	190.01(22)
R20	21(6)	.323(5)	.0522(63)	254.93(18)
L37	20(7)	.290(12)	.0523(62)	195.36(21)

Table 8. Summary of graph centrality measures for identification and classification of hub nodes of the obtained network for the task type 4 of the 16th subject. Each number in parenthesis represents the rank according to the measure. Note that high-degree nodes have either high out-closeness or high in-closeness measures together with high betweenness measures in general.

hub node	degree	out-closeness	in-closeness	betweenness
L26	44(1)	.354(6)	.0766(23)	740.30(1)
L5	40(2)	.367(2)	.0751(48)	399.07(4)
R29	39(3)	.354(7)	.0767(20)	639.48(2)
R5	36(4)	.367(3)	.0743(62)	209.36(15)
L34	33(5)	.352(8)	.0721(78)	155.71(17)
L43	30(6)	.373(1)	.0723(76)	206.66(13)
L29	27(7)	.362(4)	.0722(77)	268.38(12)
L27	26(8)	.314(13)	.0777(11)	454.00(3)
L20	24(9)	.316(12)	.0751(51)	284.25(9)

Table 9. Summary of graph centrality measures for identification and classification of hub nodes of the obtained network for the task type 5 of the 16th subject. Each number in parenthesis represents the rank according to the measure. Note that high-degree nodes have either high out-closeness or high in-closeness measures together with high betweenness measures in general.

hub node	degree	out-closeness	in-closeness	betweenness
L29	65(1)	.600(7)	.281(3)	678.50(1)
R3	62(2)	.686(2)	.254(32)	320.74(3)
L18	61(3)	.743(1)	.236(60)	129.99(16)
R4	58(4)	.628(3)	.257(27)	340.10(2)
L41	50(5)	.587(8)	.266(15)	242.92(7)
L9	47(6)	.519(32)	.277(5)	289.28(4)
L27	45(7)	.563(19)	.266(14)	285.54(5)
R5	45(8)	.466(50)	.269(11)	83.35(29)
L3	44(9)	.579(9)	.247(44)	68.39(34)
L44	44(10)	.604(5)	.253(34)	139.24(15)
R6	44(11)	.628(4)	.240(53)	93.32(24)

- Friston, K. (2009) Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biol.*, **7**(2):e1000033. doi:10.1371/journal.pbio.1000033
- Friston, K. (2011) Dynamic causal modeling and Granger causality Comments on: The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *NeuroImage*, **58**(2), 303–305.
- Glück, J., Machat, R., Jirasko, M. and Rollett, B. (2002) Training-related changes in solution strategy in a spatial test: An application of item response models. *Learning and Individual Differences*, **13**(1), 1–22.
- Glück, J. and Fitting, S. (2003) Spatial strategy selection: Interesting incremental information. *International Journal of Testing*, **3**(3), 293–308.
- Goebel, R., Roebroeck, A., Kim, D. S. and Formisano, E. (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging*, **21**(10), 1251–1261.
- Golub, G. H., Heath, M. and Wahba, G. (1979) Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, **21**(2), 215–223.
- Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.
- Hamilton, J. D. (1994) *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.
- Harrison, L., Penny, W. D. and Friston, K. (2003) Multivariate autoregressive modeling of fMRI time series. *NeuroImage*, **19**(4), 1477–1491.
- Hotelling, H. (1953) New light on the correlation coefficient and its transforms. *J. R. Statist. Soc. B*, **15**, 193–232.
- James, W. and Stein, C. (1961) Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 361–379. Berkeley: University of California Press.
- Lee, N., Choi, H. and Kim, S.-H. (2011) Empirical Bayes approach to shrinkage estimation for vector autoregressive models. *BK21 Research Report 11-01*, Department of Mathematical Sciences, KAIST, Daejeon, South Korea. http://mathsci.kaist.ac.kr/bk21/morgue/research_report_pdf/11-01.pdf.
- Lohman, D. F. (1994) Spatial ability. In *Encyclopedia of Human Intelligence* (ed R. J. Sternberg), **2**, 1000–1007. New York: Macmillan.
- Opgen-Rhein, R. and Strimmer, K. (2007) Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics*, **8**(Suppl 2):S3. doi:10.1186/1471-2105-8-S2-S3.
- Roebroeck, A., Formisano, E. and Goebel, R. (2011) The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *NeuroImage*, **58**(2), 296–302.

- Schäfer, J. and Strimmer, K. (2005a) An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, **21**(6), 754–764. doi:10.1093/bioinformatics/bti062.
- Schäfer, J. and Strimmer, K. (2005b) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, **4**(1):32. doi:10.2202/1544-6115.1175.
- Schwartzman, A., Dougherty, R. F., Lee, J., Ghahremani, D. and Taylor, J. E. (2009) Empirical null and false discovery rate analysis in neuroimaging. *NeuroImage*, **44**, 71–82.
- Seber, G. A. F. and Lee, A. J. (2003) *Linear Regression Analysis*, 2nd edn. Hoboken, New Jersey: John Wiley & Sons.
- Sims, C. A. (1980) Macroeconomics and reality. *Econometrica*, **48**(1), 1–48.
- Sporns, O., Honey, C.J. and Kötter, R. (2007) Identification and classification of hubs in brain networks. *PLoS ONE*, **2**(10):e1049. doi:10.1371/journal.pone.0001049
- Stein, C. (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (ed J. Neyman), **1**, 197–206. Berkeley: University of California Press.
- Strimmer, K. (2008) fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics*, **24**(12), 1461–1462.
- Valdes-Sosa P. A. (2004) Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics*, **2**(2), 239–250.
- Valdes-Sosa, P. A., Sanchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernandez, M., Bosch-Bayard, J., Melie-Garcia, L. and Canales-Rodriguez, E. (2005) Estimating brain functional connectivity with sparse multivariate autoregression. *Phil. Trans. R. Soc. B*, **360**, 969–981.
- Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*. New York: John Wiley & Sons.
- Wikipedia contributors. "Brodmann area". *Wikipedia: The Free Encyclopedia*. http://en.wikipedia.org/wiki/Brodmann_area (accessed November 12, 2011).