

Is First-Order Vector Autoregressive Model Optimal for fMRI Data?

Chee-Ming Ting

cmting@utm.my

Center for Biomedical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Abd-Krim Seghouane

Abd-krim.seghouane@unimelb.edu.au

Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia

Muhammad Usman Khalid

Usman.Khalid@nicta.com.au

College of Engineering and Computer Science, Australian National University, Canberra ACT 0200, Australia

Sh-Hussain Salleh

hussain@fke.utm.my

Center for Biomedical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

We consider the problem of selecting the optimal orders of vector autoregressive (VAR) models for fMRI data. Many previous studies used model order of one and ignored that it may vary considerably across data sets depending on different data dimensions, subjects, tasks, and experimental designs. In addition, the classical information criteria (IC) used (e.g., the Akaike IC (AIC)) are biased and inappropriate for the high-dimensional fMRI data typically with a small sample size. We examine the mixed results on the optimal VAR orders for fMRI, especially the validity of the order-one hypothesis, by a comprehensive evaluation using different model selection criteria over three typical data types—a resting state, an event-related design, and a block design data set—with varying time series dimensions obtained from distinct functional brain networks. We use a more balanced criterion, Kullback’s IC (KIC) based on Kullback’s symmetric divergence combining two directed divergences. We also consider the bias-corrected versions (AICc and KICc) to improve VAR model selection in small samples. Simulation results show better small-sample selection performance of the proposed criteria over the classical ones. Both bias-corrected ICs provide more accurate and consistent model order

choices than their biased counterparts, which suffer from overfitting, with KICc performing the best. Results on real data show that orders greater than one were selected by all criteria across all data sets for the small to moderate dimensions, particularly from small, specific networks such as the resting-state default mode network and the task-related motor networks, whereas low orders close to one but not necessarily one were chosen for the large dimensions of full-brain networks.

1 Introduction

Vector autoregressive (VAR) modeling of fMRI time series data has been widely used to infer directed connectivity networks of distinct brain regions (Harrison, Penny, & Friston, 2003; Valdés-Sosa, 2004; Deshpande, LaConte, James, Peltier, & Hu, 2009; Valdés-Sosa et al., 2005; Gorrostieta, Ombao, Bedard, & Sanes, 2012; Gorrostieta, Fiecas, Ombao, Burke, & Cramer, 2013; Seghouane & Amari, 2012; Ting, Seghouane, Salleh, & Noor, 2014; Ting, Seghouane, Salleh, & Noor, 2015). Determining the optimal model order p for fMRI data modeling is an important practical as well as a theoretical issue when using VAR models. Our focus in this note is the identification of the optimal VAR model orders when modeling fMRI time series and approaches to estimating these orders. This is potentially important in the analysis of directed functional connectivity using constructs like Granger causality (Seghouane & Amari, 2012). Granger causality (usually) rests on autoregressive models, and inferences about Granger causal dependencies depend on optimizing the model order. Although we do not expand on this application, the implications of our work reach into the domain of directed connectivity through estimates of autoregression coefficients, which underlie Granger causality, directed coherence, and estimates of transfer entropy (Seghouane & Amari, 2012).

The estimation of VAR orders is often carried out as model selection based on information criteria (IC) (e.g., Akaike IC (AIC) and Bayesian IC (BIC)) to find an optimal model balancing between the goodness of fit and model complexity. Using these classical criteria, an order of $p = 1$ was suggested by some studies to be optimal, for example, for high-dimensional voxel-wise data of a large-scale brain network during the resting state (Valdés-Sosa, 2004). However, it was also claimed as valid for low-dimensional data sets obtained from specific small networks of only a few functionally linked regions of interest (ROIs), e.g., event-related-design data of a 6-ROI motor network (Deshpande et al., 2009). The suggested optimality of VAR of order one has been used, automatically, without any further tests, in later studies as a true hypothesis for analyzing fMRI data regardless of different data sets, subjects, dimensions of the networks under study, and experimental conditions. These studies include, for example, a whole-brain network of 116 ROIs for a block design visual task (Valdés-Sosa

et al., 2005), and a 9-ROI motor-visual network with an event-related task (Sato et al., 2010), among many others (Lund, Nørgaard, Rostrup, Rowe, & Paulson, 2005).

Our primary concern is whether the optimality of the VAR(1) model for fMRI, as claimed by the studies noted, is too strong. We raise questions on its validity that might be limited to the particular data sets being studied. A contradictory result in Gorrostieta et al. (2012) showed that the optimal order selected by BIC for a 7-ROI event-related motor-visual data set was $p = 2$. Another study of a 5-ROI motor-decision network of stroke patients (Gorrostieta et al., 2013) argued that VAR(1) was not sufficient to capture the temporal dynamics in the fMRI data, based on diagnostic tests on the fitted residuals, which suggested a VAR(2) instead. Order $p = 4$ was chosen by Bayesian evidence for a block-design data from a 6-ROI visual-attentional network (Harrison et al., 2003). Therefore, there has been a lack of consensus in the literature on the optimal model orders for fMRI data. The selected VAR orders may vary considerably with different ICs, even across data sets of varied dimensions, subjects, tasks, and experimental designs.

The studies noted used classic criteria that are unreliable for VAR order selection for fMRI data, typically of large dimension N with small sample size T , especially in the full-brain network analysis. The classic AIC is an asymptotically unbiased estimator of Kullback's directed divergence, a dissimilarity measure between the true and the fitted model. However, it becomes strongly biased for small samples or when the number of fitted parameters k is a moderate to large fraction of T . This bias might lead to the selection of overfitted models, as already shown for the univariate AR model selection with small samples (Hurvich & Tsai, 1989). It becomes more critical for the VAR case (Hurvich & Tsai, 1993) when N is large, which involves fitting many more parameters (i.e., $p \times N^2$).

This note aims to examine the validity of the optimality of first-order autoregression for fMRI data in particular and address the mixed results in the literature in general. We present a comprehensive study on selecting the optimal VAR order for modeling connectivity in fMRI data, with particular focus on two issues not fully explored in the literature: (1) the sensitivity of the selected orders to data sets of different tasks, experimental designs, and data dimensions and (2) the performance of different ICs for small-sample selection, particularly in the high-dimensional settings. We performed a careful cross-data set evaluation using a resting-state data set and two motor task-related data sets with event-related and block designs. We compared the selected orders for the large-dimensional data of the whole brain with the small data from specific functional networks, that is, the default mode and the motor networks.

We use a bias-corrected version of classical AIC, AIC_c (Hurvich & Tsai, 1989, 1993), to improve the small-sample VAR order selection. Simulation results showed a significant bias reduction in AIC_c compared to the AIC, and hence better performance in order selection in small samples. We also

consider a new class of criteria, the Kullback information criterion (KIC), which is based on Kullback's symmetric divergence summing up two single directed divergences with alternated roles of the true and the fitted candidate models in the measure (Cavanaugh, 1999). The AIC with unidirected divergence is biased to selecting overfitted models, while the KIC with combined information from both directions is arguably more balanced and able to detect both overfitted and underfitted models. The bias-corrected version KIC_c has been proposed for linear regression model selection in finite samples (Cavanaugh, 2004) and extended for univariate and multivariate AR models in our earlier studies (Seghouane & Bekara, 2004; Seghouane, 2006). Simulation results showed that KICs provide more accurate VAR order choices than their AIC analogues, and KIC_c performs the best in small-sample situations (Seghouane, 2006). In this study, we compare the performance of these different criteria via simulations and across the three real fMRI data sets.

2 VAR Order Selection

2.1 VAR Model. Let $\mathbf{y}_t = (y_{t1}, \dots, y_{tN})^T$ be an $N \times 1$ vector of fMRI time series observations measured from N voxels or ROIs (defining the nodes of a brain network) at time t , for $t = 1, \dots, T$. The dimension of the time series N is usually large, or even larger than sample size T . Suppose the observations $\mathbf{y}_1, \dots, \mathbf{y}_T$ are generated by an N -dimensional VAR(p) process,

$$\mathbf{y}_t = \sum_{\tau=1}^p \mathbf{A}_\tau \mathbf{y}_{t-\tau} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (2.1)$$

where p is the model order, \mathbf{A}_τ is an $N \times N$ matrix of AR coefficients at time lag τ , and $\{\boldsymbol{\varepsilon}_t\}$ are $N \times 1$ independent and identically distributed (i.i.d.) gaussian noise with mean zero and $N \times N$ variance-covariance matrix $\boldsymbol{\Sigma}_\varepsilon$. The AR coefficient matrices $\mathbf{A}_\tau = \{a_{ij}^\tau\}_{1 \leq i, j \leq N}$ can characterize the network of temporal interdependencies between different brain regions for different time lags $\tau = 1, \dots, p$. There exists a causal influence that node j exerts on node i after a time interval of τ if $a_{ij}^\tau > 0$. Model 2.1 can be rewritten in the form of a multivariate linear model,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}, \quad (2.2)$$

where,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_T \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{y}'_0 & \cdots & \mathbf{y}'_{-(p-1)} \\ \vdots & \ddots & \vdots \\ \mathbf{y}'_{(T-1)} & \cdots & \mathbf{y}'_{(T-p)} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mathbf{A}'_1 \\ \vdots \\ \mathbf{A}'_p \end{bmatrix} \text{ and } \mathbf{E} = \begin{bmatrix} \mathbf{e}'_1 \\ \vdots \\ \mathbf{e}'_T \end{bmatrix}.$$

The total number of unknown parameters comprises the AR coefficients in β and noise variance-covariances in Σ_ϵ , $k = pN^2 + N(N + 1)/2$. Note that a small increment in p will cause a rapid increase in the number of parameters by order of N^2 , especially when N is large. The maximum likelihood estimation of VAR model 2.1 is equivalent to the ordinary least squares (OLS) regression of Y on X . The OLS estimators of β and Σ_ϵ are defined by $\hat{\beta} = (X'X)^{-1}X'Y$ and $\hat{\Sigma}_\epsilon = (1/T)(Y - X\hat{\beta})'(Y - X\hat{\beta})$, respectively.

2.2 Model-Order Selection Criteria. Let $p(y|\theta_k)$ denote the VAR model of order p that generates the observations $y \triangleq y_1, \dots, y_T$, where θ_k is a vector of the corresponding k model parameters. We use $\theta_p \triangleq \theta_{k_p}$ for simplicity of notation. Given y , our objective is to determine the true model order p_0 . This can be cast as a model selection problem of searching among a set of P fitted candidate VAR models from orders 1 to $P\{p(y|\hat{\theta}_p)\}$, $1 \leq p \leq P$, an optimal model of order \hat{p} that best approximates the true model $p(y|\theta_0)$ of order p_0 , according to some discrepancy measures between $p(y|\hat{\theta}_p)$ and $p(y|\theta_0)$. Various model selection criteria were developed as estimators of these dissimilarity measures. The candidate model corresponding to the minimum value of a criterion is considered optimal.

One such measure is the Kullback's directed divergence (Shibata, 1989),

$$2I(\theta_0, \hat{\theta}_p) = d(\theta_0, \hat{\theta}_p) - d(\theta_0, \theta_0), \tag{2.3}$$

where $d(\theta_i, \theta_j) = E_{\theta_i}\{-2 \ln p(y|\theta_j)\}$ and $E_{\theta_i}\{\cdot\}$ denotes the expectation with respect to $p(y|\theta_i)$. The constant $d(\theta_0, \theta_0)$ can be dropped without affecting the comparison between candidate models. AIC is an asymptotically unbiased estimator of $d(\theta_0, \hat{\theta}_p)$ and for the VAR model is given as

$$\begin{aligned} AIC &= -2 \ln p(y|\hat{\theta}_p) + 2k \\ &= T \ln |\hat{\Sigma}_\epsilon| + 2 \left\{ pN^2 + \frac{N(N + 1)}{2} \right\}. \end{aligned} \tag{2.4}$$

The classic criteria such as AIC and BIC can also be treated as approximations to the Bayesian model evidence, as shown by Penny, Stephan, Mechelli, & Friston (2004) and Penny (2012). The (Bayes) optimal model order is the order that maximizes model evidence. The model evidence is the probability of the data given a particular model order. In other words, it is the integrated or marginal likelihood of the data, having integrated out uncertainty about the (autoregressive) model parameters. It is trivial to show that model evidence comprises accuracy minus complexity (Penny, 2012). Invariably the criteria used to optimize model order approximate this

mixture of accuracy and complexity under various simplifying assumptions. In this note, we consider only the ad hoc criteria, the AIC and BIC as baseline, noting that these criteria can always be expressed in terms of accuracy (first term) and a (computationally cheap) approximation to complexity (second term).

An AIC bias-corrected version has been developed for small-sample-size settings (Hurvich & Tsai, 1993),

$$AIC_c = T \ln |\hat{\Sigma}_\epsilon| + 2b \left\{ pN^2 + \frac{N(N+1)}{2} \right\}, \quad (2.5)$$

where $b = T / \{T - (pN + N + 1)\}$. AIC_c differs from AIC only by a scale factor b , which is nonnegligible when T is small relative to N . Simulation results showed that AIC_c improves VAR order selection in small samples compared to AIC.

An alternative measure is Kullback's symmetric divergence combining two directed divergences with alternated role of true and the approximating models in equation 2.3,

$$\begin{aligned} 2J(\theta_0, \hat{\theta}_p) &= 2I(\theta_0, \hat{\theta}_p) + 2I(\hat{\theta}_p, \theta_0) \\ &= d(\theta_0, \hat{\theta}_p) - d(\theta_0, \theta_0) + d(\hat{\theta}_p, \theta_0) - d(\hat{\theta}_p, \hat{\theta}_p). \end{aligned} \quad (2.6)$$

Cavanaugh (1999) proposed the KIC as an asymptotically unbiased estimator of the quantity $K(\theta_0, \hat{\theta}_p) = d(\theta_0, \hat{\theta}_p) + d(\hat{\theta}_p, \theta_0) - d(\hat{\theta}_p, \hat{\theta}_p)$, defined for the VAR model as (Seghouane, 2006)

$$\begin{aligned} KIC &= -2 \ln p(\mathbf{y}|\hat{\theta}_p) + 3k \\ &= T \ln |\hat{\Sigma}_\epsilon| + 3 \left\{ pN^2 + \frac{N(N+1)}{2} \right\}, \end{aligned} \quad (2.7)$$

which imposes a stronger penalty term on the number of parameters than the AIC to prevent overfitting. The bias-corrected version has been derived for VAR model (Seghouane, 2006),

$$\begin{aligned} KIC_c &= T \ln |\hat{\Sigma}_\epsilon| + \frac{TN(2pN + N + 1)}{T - pN - N - 1} + \frac{TN}{T - Np - (N - 1)/2} \\ &\quad + \frac{2N^2p + N^2 - N}{2}, \end{aligned} \quad (2.8)$$

where the penalty term is a function of T to account for small-sample conditions. KIC_c was shown to outperform KIC and AICs in a simulation study of small-sample order selection studies (Seghouane & Bekara, 2004;

Seghouane, 2006). Note that the constant terms $TN \ln(2\pi) + TN$ in the original formulas, which play no practical role in model selection, have dropped been already in equations 2.4 and 2.5 and equations 2.7 and 2.8.

3 Experimental Results

In this section, we first compare the small-sample performance of various selection criteria in estimating the orders of VAR models, based on simulated data. We aim to demonstrate the effectiveness of our proposed bias-corrected and symmetric Kullback’s divergence criteria in providing more accurate estimates of the model orders under the small-sample-size settings. Having established this, we then generalize the simulation findings by applications to real data to verify the appropriateness of order one in VAR modeling of fMRI signals.

3.1 Simulations. We consider two VAR model structures for simulated data generation: low-dimensional models with high orders and higher-dimensional models with low orders. The first is bivariate VAR models with a range of known orders $p_0 = 2, 4, 6, 8$. We fixed the coefficient matrices for the first two lags for $p_0 = 2$,

$$\mathbf{A}_1 = \begin{bmatrix} 0.50 & -0.30 \\ 0.20 & 0.65 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} -0.50 & 0.30 \\ 0.0 & -0.40 \end{bmatrix},$$

which Hurvich and Tsai (1993) used to evaluate the bias-corrected AIC; based on these, we set recursively $\mathbf{A}_\tau = 0.8\mathbf{A}_{\tau-2}$ for $\tau = 3, \dots, 8$ to emulate the decaying temporal dependencies at the longer time lags for the higher-order models $p_0 > 2$. The second is 20-dimensional VAR models with $p_0 = 1$ and $p_0 = 2$, with block-diagonal coefficient matrices each formed by two 10×10 subblocks along the main diagonal. The nonzero entries of the subblocks were randomly generated from a scaled uniform distribution—for i and j in the same block— $\lambda^{p_0} a_{ij} \sim U[-0.5, 0.5]$ with a scaling constant λ , set here as 1.35. All other off-diagonal entries are zeros. A similar structure with a lower dimension has been used in a simulation study of order selection for fMRI using the Bayesian evidence (Harrison et al., 2003). The above AR parameters were chosen to ensure stability of the simulated processes. Noise covariance matrix $\Sigma_\epsilon = \mathbf{I}$ was used for all models.

A thousand simulated realizations of sample size T were generated from each model with the known order p_0 . The values of T were chosen to represent small-sample scenarios, compared to the large number of parameters k . We also investigated the impact of increased ratios k/T on the selection performance, using the values of ratio 0.16 for all the bivariate models (see Table 1), 2 and 4 for the 20-dimensional VAR(1) and VAR(2) models (see Table 2), respectively. For each realization, candidate VAR models of orders

Table 1: Frequency of Order Selected by Various Criteria over 1000 Realizations for Bivariate VAR Models with Known Orders $p_0 = 2, 4, 6, 8$ and Different Sample Sizes T .

Model	T	Criterion	Selected Model Order											
			1	2	3	4	5	6	7	8	9	10	11	12
VAR(2)	50	BIC	30	934	24	1	1	1	0	0	0	2	3	4
		AIC	0	211	40	14	11	20	9	11	15	35	78	556
		AIC _c	11	858	89	23	8	6	2	1	0	2	0	0
		KIC	18	776	56	12	9	6	4	5	5	10	18	81
VAR(4)	100	KIC _c	23	945	30	1	1	0	0	0	0	0	0	0
		BIC	34	7	31	919	9	0	0	0	0	0	0	0
		AIC	0	0	3	639	128	60	40	23	25	16	24	42
		AIC _c	0	0	3	870	89	24	9	4	1	0	0	0
VAR(6)	150	KIC	0	0	6	917	56	16	5	0	0	0	0	0
		KIC _c	0	2	9	949	34	4	2	0	0	0	0	0
		BIC	92	67	121	9	23	686	2	0	0	0	0	0
		AIC	0	0	0	1	4	708	143	53	31	25	15	20
VAR(8)	200	AIC _c	0	0	2	1	7	861	98	17	8	5	1	0
		KIC	0	2	7	3	11	907	60	5	3	2	0	0
		KIC _c	0	3	13	7	14	932	29	1	1	0	0	0
		BIC	78	59	518	68	60	9	5	203	0	0	0	0
		AIC	0	0	0	1	4	0	5	722	141	57	36	34
		AIC _c	0	0	0	3	8	2	7	843	99	21	7	10
		KIC	0	0	14	14	26	8	14	860	49	9	3	3
		KIC _c	0	1	29	24	49	10	17	843	25	2	0	0

Note: Numbers in bold are the highest counts.

$p = 1, \dots, P$ were fitted by the OLS method, and various criteria were used to select among the candidates an estimator \hat{p} to recover the true model order p_0 . The results are presented as a confusion matrix of frequencies of the selected orders for the 1000 realizations, producing an approximate posterior model distribution $p(p|y)$ over p . Tables 1 and 2 show the estimation results by various criteria with the maximum model order considered, $P = 12$ and $P = 8$, for VAR model structures 2.1 and 2.2, respectively.

From both tables, Kullback’s symmetric divergence-based criteria, KICs generally show the best performance in selecting the correct model orders, followed by AIC_c and BIC, for the true models considered under the small-sample settings. The KIC_c consistently achieves the highest accuracy among all criteria with the percentage of correct selections above 90%, except for the bivariate VAR(8), which reveals a mild overfitting tendency of the criterion. These are consistent with the results in Seghouane (2006) in terms of the prediction error (ME) of the fitted models. Table 1 shows that BIC initially performs better than the AIC_c and KIC for small orders ($p_0 = 2$ and $p_0 = 4$);

Table 2: Frequency of Order Selected by Various Criteria over 1000 Realizations for 20-Dimensional VAR Models with Known Orders $p_0 = 1$ and $p_0 = 2$.

Model	T	Criterion	Selected Model Order							
			1	2	3	4	5	6	7	8
VAR(1)	200	BIC	1000	0	0	0	0	0	0	0
		AIC	0	0	0	0	0	0	0	1000
		AIC _c	1000	0	0	0	0	0	0	0
		KIC	713	0	0	0	0	0	0	287
		KIC _c	1000	0	0	0	0	0	0	0
VAR(2)	200	BIC	807	193	0	0	0	0	0	0
		AIC	0	0	0	0	0	0	0	1000
		AIC _c	0	1000	0	0	0	0	0	0
		KIC	0	20	0	0	0	0	0	980
		KIC _c	1	999	0	0	0	0	0	0

Note: Numbers in bold are the highest counts.

however, it is surpassed by the AIC_c and KIC for high orders. The BIC is more prone to underfitting compared to other criteria.

The bias-corrected KIC_c and AIC_c clearly outperform their original versions KIC and AIC, which tend to choose overfitted models, with gross overparameterization by the AIC especially evident for the bivariate VAR(2) model. When the parameter counts k increase rapidly compared to T with the larger dimensions N , the performance of the biased criteria further deteriorates with a more substantial amount of overfitting, as shown in Table 2. This is due to the increased bias of the criteria as estimators of the Kullback divergences. The results reconfirm that the bias reduction in AIC_c and KIC_c, which implies stronger penalty functions to prevent overfitting under finite samples, can provide more accurate model orders than their original versions, as demonstrated in Hurvich and Tsai (1989, 1993; Seghouane & Bekara, 2004; Seghouane, 2006). Moreover, we can see that the bias-corrected criteria, particularly the KIC_c, perform better in terms of consistency (i.e., convergence of the selected orders in probability to true values) compared to the larger variance of the distribution of the biased AIC and KIC estimates.

3.2 Evaluation on fMRI Data. Here we check the validity of previous findings that order one is sufficient for VAR modeling of fMRI data. We performed a comprehensive evaluation on real fMRI data using different selection criteria over (1) three typical data types—a resting state, an event-related design, and a block-design motor task data set—and (2) varying dimensions of time series obtained from distinct functional brain networks of different sizes.

3.2.1 Resting-State Data. We used resting-state fMRI data of the first subject from the first session obtained from Shehzad et al. (2009). The data were acquired with a Siemens Allegra 3-Tesla scanner. During scans, subjects were asked to relax with eyes open, with the word *relax* projected in white on a black background. Functional MR images were obtained using a blood oxygenation level-dependent BOLD-weighted echo planar imaging (EPI) sequence (TR/TE = 2000/25 ms, flip angle (FA) = 90°, voxel size = 3 × 3 × 3 mm³, matrix size = 64 × 64, 39 contiguous slices). $T = 197$ EPI volumes were collected for each scan.

3.2.2 Event-Related Finger-Tapping Data. We used an event-related finger-tapping task data set of a single subject provided by Lee, Tak, and Ye (2011). The subject was instructed to perform right finger flexion. The session began with a 30 s preparation and a 30 s resting period, followed by 40 repeated trials of a task period and a resting period with a duration of 28 s each trial, always starting with the task period and ending with an additional 30 s resting period. Within the repeated trials, the resting period consisted of an interstimulus interval between 4 s and 20 s with an average of 12 s. Functional images were acquired on a 3.0-T ISOL system, Korea (TR/TE = 2000/35 ms, FA = 80°, voxel size 3.44 × 3.44 × 4 mm³, matrix = 64 × 64, 24 slices). A total of $T = 326$ EPI volumes were collected.

3.2.3 Block-Design Hand-Movement Data. A single-subject block-design hand-pressing data set acquired from the <http://www.psychology.gatech.edu> website was used. The subject was asked to press the left hand at some time points and the right hand at other times. The same hand was pressed continuously for a time block of 12 s with a resting block of 36 s. The experimental paradigm started with the left hand press, with a 12 s resting period between the left and the right hand presses. Only the right-hand data were considered here. Functional images were acquired on a Siemens TrioTim system (TR/TE = 2000/32 ms, FA = 90°, voxel size 3 × 3 × 3 mm³, matrix size = 64 × 64, 36 slices). A total of $T = 188$ EPI volumes were collected.

3.2.4 Data Preprocessing and Analysis. The data sets were preprocessed using SPM8 and Matlab with the following steps (Khalid & Seghouane, 2014): realignment to the first image for motion correction, normalization to a standard Talarach template, and spatial smoothing with an 8 mm full-width half maximum (FWHM) gaussian kernel. The low-frequency drifts in the data were removed using a discrete cosine transform (DCT) filter (with a cut-off frequency of 1/128 Hz). The resting-state data were temporally bandpass-filtered (0.01–0.08 Hz) to retain their low-frequency fluctuations, which reflect spontaneous neuronal activity. For the task-related data, the high-frequency components were temporally smoothed by a 1.5 s FWHM gaussian filter.

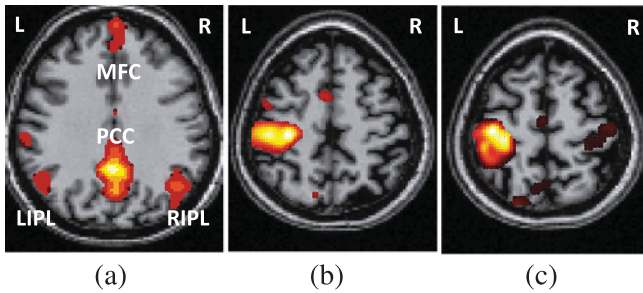


Figure 1: Spatial maps of functional networks being modeled. (a) PCC-seeded correlation map of the DMN during rest (converted to z-statistics and threshold at $p < 0.05$). (b, c) F -statistics activation maps of the motor regions, threshold respectively at $p < 0.0001$ and $p < 0.00001$ for the event-related right-finger-tapping task (b) and the block design right-hand pressing task (c).

We investigated the sensitivity of the selected orders to the increased time series dimensions, from small-scale-specific functional networks to the whole-brain networks. For the resting-state data, 79 time series belonging to the default mode network (DMN) were selected based on the strongest correlations with a seed ROI based on the posterior cingulate cortex (PCC). For both event-related and block-design data sets respectively, 29 time series from the activated motor regions were selected. Figure 1 shows the spatial maps for the areas of functional networks under study. Figure 1a shows the major regions of the resting-state DMN, which are strongly correlated with the PCC as a hub, comprising the medial prefrontal cortex (MFC), left and right inferior parietal lobe (IPL). Both the right finger-tapping and the right hand-pressing tasks induced the activation of the left primary motor cortex areas, as shown in Figures 1b and 1c. For each data set, we also evaluated two whole-brain networks of 116 and 200 ROIs parcellated based on automated anatomical labeling (AAL) (Tzourio-Mazoyer et al., 2002) and Craddock's spectral clustering (Craddock, James, Holtzheimer, Hu, & Mayberg, 2012) atlases, respectively. The AAL and Craddock-based ROIs were prepared using the DPARSF toolbox (Yan & Zang, 2010). One hundred random selections of half of the ROI time series were carried out for each data type and each network. The sampling interval and a small sample size are fixed at $TR = 2$ s and $T = 180$ across all data sets for consistent comparison. The VAR model order selection analysis was performed on the preprocessed time series; for the block design data, it was performed on their residuals, having regressed out the task-related mean levels of activation, modeled by a general linear model (GLM).

3.2.5 Results. Table 3 shows the frequencies of VAR orders selected by each criterion over the 100 replications for the different data types and the

Table 3: Frequency of VAR Order Selected by Various Criteria for Resting-State, Event-Related, and Block-Design Hand Movement fMRI Data from Functional Brain Networks of Distinct Sizes (over 100 Independent Runs of Random Selection of N Time Series for Each Network).

Data Set	Brain Network	N	Criterion	Selected Model Order									
				1	2	3	4	5	6	7	8	9	10
Resting state	Default mode network	30	BIC	3	0	10	4	27	3	28	16	5	4
			AIC	2	0	10	4	26	3	28	16	7	4
			AIC _c	0	0	3	0	42	3	27	15	5	5
			KIC	2	0	10	4	26	3	29	17	5	4
	Whole brain	50	KIC _c	0	0	3	0	42	3	28	14	5	5
			BIC	0	17	28	31	16	1	4	2	1	0
			AIC	0	17	25	33	17	1	4	2	1	0
			AIC _c	0	4	44	27	14	3	4	3	0	1
		100	KIC	0	17	26	32	17	1	4	2	1	0
			KIC _c	0	5	44	27	14	3	4	2	1	0
			BIC	76	16	4	4	0	0	0	0	0	0
			AIC	69	17	7	6	1	0	0	0	0	0
Event-related-design motor task	Motor network	15	AIC _c	79	11	4	5	1	0	0	0	0	
			KIC	71	18	6	4	1	0	0	0	0	
			KIC _c	79	12	4	4	1	0	0	0	0	
			BIC	24	1	3	0	29	11	3	5	13	11
	Whole brain	50	AIC	12	1	3	0	29	11	3	5	13	23
			AIC _c	13	16	3	0	36	11	3	3	10	5
			KIC	12	1	3	0	29	11	3	5	13	23
			KIC _c	14	15	3	0	36	11	3	3	10	5
		100	BIC	15	5	22	46	7	3	0	1	1	0
			AIC	15	3	20	48	8	3	1	1	1	0
			AIC _c	12	2	24	49	8	3	0	1	1	0
			KIC	15	3	20	49	7	3	1	1	1	0
Block-design motor task	Motor network	15	KIC _c	12	2	24	49	8	3	0	1	1	0
			BIC	20	68	9	3	0	0	0	0	0	
			AIC	19	69	9	3	0	0	0	0	0	
			AIC _c	21	67	9	3	0	0	0	0	0	
	Whole brain	50	KIC	20	68	9	3	0	0	0	0	0	
			KIC _c	21	67	9	3	0	0	0	0	0	
			BIC	19	3	0	1	27	12	8	2	21	7
			AIC	16	3	0	1	27	12	6	2	21	12
		100	AIC _c	16	7	0	1	29	12	8	2	20	5
			KIC	16	3	0	1	27	12	7	2	21	11
			KIC _c	18	5	0	1	29	12	8	2	20	5
			BIC	2	14	30	24	20	4	3	0	1	2
100	AIC	2	14	28	24	20	4	4	1	1	2		
	AIC _c	0	1	43	25	17	5	5	1	1	2		
	KIC	2	14	29	24	19	4	4	1	1	2		
	KIC _c	0	1	43	26	17	4	5	1	1	2		
	BIC	37	47	14	2	0	0	0	0	0	0		
	AIC	36	46	16	2	0	0	0	0	0	0		
	AIC _c	37	44	16	3	0	0	0	0	0	0		
	KIC	36	47	15	2	0	0	0	0	0	0		
KIC _c	37	46	15	2	0	0	0	0	0	0			

Notes: The sampling rate and sample size are fixed as $TR = 2s$, $T = 180$. Numbers in bold are the highest counts.

increasing time series dimensions from distinct functional networks. The selection is performed among candidate models of orders $p = 1, \dots, 10$. It is clear that as the dimension N increases, all criteria tend to select lower model orders for all three data sets due to the penalty imposed on the model complexity to prevent overfitting. For the relatively low dimensions $N = 15$ and $N = 30$ of the small functional networks, orders of much higher than one are suggested consistently by all criteria for all data sets, with the highest $p = 7$ for the resting-state data from the DMN, followed by $p = 5$ for both the event-related and block design data from the motor networks. This explains why higher orders were found to be more optimal by the studies (Harrison et al., 2003; Gorrostieta et al., 2012, 2013), which fitted only a few ROI time series. For the whole-brain networks of moderate dimensions $N = 50$, orders of $p = 3$ to $p = 4$ are selected for all data sets. For the high-dimensional case of $N = 100$, $p = 1$ is suggested as optimal for the resting-state data, consistent with the findings of Valdés-Sosa et al. (2005), who studied similar dimensions of 116 time series. However, an order close to one but not one (i.e., $p = 2$) is chosen for both the event-related and block design data. Among the ICs, AIC and KIC tend to select much higher orders than their bias-corrected counterparts (e.g., for the DMN). As illuminated by the simulation results, these are possibly overfitted models that fail to be detected due to the bias of the classic criteria in finite-sample cases. Both bias-corrected ICs perform comparably in detecting the overfitted models, with slightly better performance by the KIC_c .

4 Discussion

We have presented a careful interrogation of the optimal temporal orders of VAR for modeling fMRI data using various information criteria across different tasks, experimental designs, and dimensions of the brain networks being modeled. Based on our results, we suggest, albeit with due caution, a lower bound of the optimal VAR orders for fMRI data (with small sample size $T \sim 200$) as $p \geq 5$ for low dimensions $N \leq 15$, across all type of data sets with higher orders for the non-task-related resting-state data. Orders of $p = 3$ to $p = 4$ are suggested for the moderate dimensions $30 \leq N \leq 50$, while orders close to one but not necessarily one for the high dimensions comparable to T , $N > 100$ depending on the data types. Bias-corrected criteria AIC_c and KIC_c should be used in favor of AIC and KIC to improve the small-sample order selection. Our results imply that Valdés-Sosa's assertion, "For this data, the GCV criterion indicated that the most appropriate model order for the spatio-temporal multivariate autoregressive model (STMAR) was $p = 1$ " (Valdés-Sosa, 2004), which was adopted indiscriminately by subsequent studies (Valdés-Sosa et al., 2005; Satoa et al., 2010; Lund et al., 2005) among others as a conclusively true hypothesis for fMRI data, is no longer valid across different conditions. Therefore, we conclude that when modeling fMRI time series with VAR, the optimal orders, which have been

shown to vary considerably with different data sets, experiments, and time series dimensions, should always be carefully analyzed using appropriate model selection criteria (Hurvich & Tsai, 1993; Seghouane, 2006), and not taken as $p = 1$ automatically.

Acknowledgments

This work was supported by the Universiti Teknologi Malaysia and the Ministry of Education Malaysia under Fundamental Research Grant Scheme R.J130000.7809.4F668.

References

- Cavanaugh, J. E. (1999). A large-sample model selection criterion based on Kullback's symmetric divergence. *Stat. Prob. Lett.*, *42*, 333–343.
- Cavanaugh, J. E. (2004). Criteria for linear model selection based on Kullback's symmetric divergence. *Austral. New Zealand J. Stat.*, *46*(2), 257–274.
- Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, *33*, 1914–1928.
- Deshpande, G., LaConte, S., James, G. A., Peltier, S., & Hu, X. (2009). Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping*, *30*, 1361–1373.
- Gorrostieta, C., Fiecas, M., Ombao, H., Burke, E., & Cramer, S. (2013). Hierarchical vector auto-regressive models and their applications to multi-subject effective connectivity. *Frontiers in Computational Neuroscience*, *7*(159), 1–11.
- Gorrostieta, C., Ombao, H., Bedard, P., & Sanes, J. N. (2012). Investigating brain connectivity using mixed effects vector autoregressive models. *Neuroimage*, *59*, 3347–3355.
- Harrison, L., Penny, W. D., & Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage*, *19*, 1477–1491.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.
- Hurvich, C. M., & Tsai, C. L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *J. Time Series Anal.*, *14*, 271–279.
- Khalid, M. U., & Seghouane, A. K. (2014). A single SVD sparse dictionary learning algorithm for fMRI data analysis. In *Proceedings of the IEEE Statistical Signal Process Workshop* (pp. 65–68). Piscataway, NJ: IEEE.
- Lee, K., Tak, S., & Ye, J. C. (2011). A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion. *IEEE Trans. Medical Imaging*, *30*(5), 1076–1089.
- Lund, T. E., Nørgaard, M. D., Rostrup, E., Rowe, J. B., & Paulson, O. B. (2005). Motion or activity: Their role in intra- and inter-subject variation in fMRI. *Neuroimage*, *26*(3), 960–964.
- Penny, W. D. (2012). Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage*, *59*, 319–330.

- Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Comparing dynamic causal models. *Neuroimage*, *22*, 1157–1172.
- Satoa, J. R., Fujitab, A., Cardoso, E. F., Thomazd, C. E., Brammere, M. J., & Amaro, Jr., E. (2010). Analyzing the connectivity between regions of interest: An approach based on cluster Granger causality for fMRI data analysis. *Neuroimage*, *52*(4), 1444–1455.
- Seghouane, A. K. (2006). Vector autoregressive model-order selection from finite samples using Kullback's symmetric divergence. *IEEE Trans. Circuits and Systems-I Regular Papers*, *53*(10), 2327–2335.
- Seghouane, A. K., & Amari, S. (2012). Identification of directed influence: Granger causality, Kullback-Leibler divergence and complexity. *Neural Computation*, *24*, 1722–1739.
- Seghouane, A. K., & Bekara, M. (2004). A small-sample model selection criterion based on the Kullback symmetric divergence. *IEEE Trans. Signal Process.*, *52*(12), 3314–3323.
- Shehzad, Z., Kelly, A. M., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., . . . Milham, M. P. (2009). The resting brain: Unconstrained yet reliable. *Cerebral Cortex*, *19*, 2209–2229.
- Shibata, R. (1989). Statistical aspect of model selection. In C. Willems (Ed.), *From data to model* (pp. 215–240). New York: Springer-Verlag.
- Ting, C.-M., Seghouane, A.-K., Salleh, S.-H., & Noor, A. B. M. (2014). Estimation of high-dimensional brain connectivity from fMRI data using factor modeling. In *Proceedings of the IEEE Statistical Signal Process. Workshop* (pp. 73–76). Piscataway, NJ: IEEE.
- Ting, C.-M., Seghouane, A.-K., Salleh, S.-H., & Noor, A. B. M. (2015). Estimating effective connectivity from fMRI data using factor-based subspace autoregressive models. *IEEE Signal Process. Letters*, *22*(6), 757–761.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M., (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289.
- Valdés-Sosa, P. A. (2004). Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics*, *2*, 239–250.
- Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., & Canales-Rodríguez, E. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philos. Trans. R. Soc. B*, *360*, 969–981.
- Yan, C., & Zang, Y. (2010). DPARSF: A MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. *Frontiers in Systems Neuroscience*, *4*(13), 1–7.

Received February 26, 2015; accepted May 13, 2015.