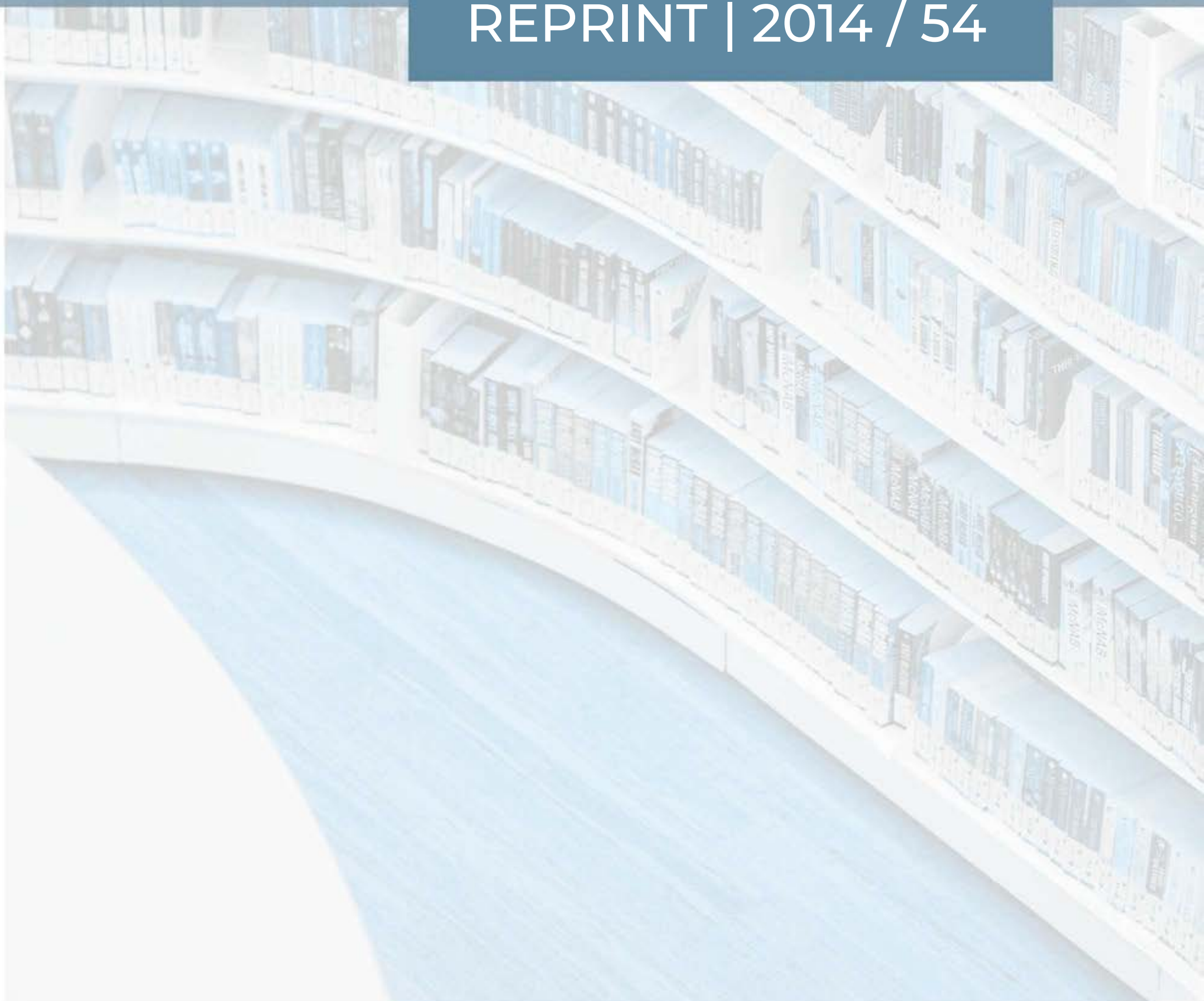
NODEWISE GRAPHICAL MODELING USING THE FOCUSED INFORMATION CRITERION FOR ‘P LARGER THAN N’ SETTINGS

Eugen Pircalabelu, Gerda Claeskens, Sara Jahfari, Lourens Waldorp

REPRINT | 2014 / 54



UCLouvain

LIDAM
ISBA | Louvain Institute of Data Analysis and Modeling in economics and statistics

Nodewise graphical modeling using the Focused Information Criterion for 'p larger than n' settings.

Eugen Pircalabelu [1], Gerda Claeskens [1], Sara Jahfari [2], Lourens Waldorp [3]

[1] KU Leuven, ORSTAT and Leuven Statistics Research Center, Belgium
[2] University of Amsterdam, Cognitive Science Center, The Netherlands
[3] University of Amsterdam, Department of Psychological Methods, The Netherlands

E-mail for correspondence: `eugen.pircalabelu@kuleuven.be`

Abstract: We present a new method of estimating graphical models with the clear goal of providing models that minimize the mean squared error of certain parameters of interest to the researcher. The method is applicable to undirected as well as to mixed graphs containing both directed and undirected edges. Quadratic approximations to several well studied penalties deal with problems where the number of nodes is greater than the number of samples. Extensions of the current application include a dynamical image of graphs based on consecutive focus points and estimating graphs where information is borrowed among subjects.

Keywords: Undirected Markov networks; Temporal chain graphs; Focused information criterion; Model selection; fMRI

1 Introduction and motivation

Motivated by a resting-state fMRI study and by the definitory characteristics of such data (a complete dataset of n repeated measurements for each of p brain regions per subject) we propose a new method for model selection for undirected as well as mixed graphical models (with both directed and undirected edges), in situations where the number of nodes in the graph is larger than the number of cases. The method is constructed to have good mean squared error properties and it is based on the focused information criterion (FIC) developed in Claeskens and Hjort (2003).

This paper was published as a part of the proceedings of the 29th International Workshop on Statistical Modelling, Georg-August-Universität Göttingen, 14–18 July 2014.

The reason for modeling the data in this way is two-fold: first, the graphical representation offers insight into the complex relations that might exist between the nodes (brain regions in this case) revealing thus patterns of functional connectivity within the brain (Bullmore and Sporns, 2009). Second, we wish to accommodate distinct research objectives and select models tailored to those interests. Unlike traditional model selection criteria (such as AIC or BIC), the FIC allows to selecting individual models, tailored to a specific research purpose (the *focus*), as opposed to attempting an identification of a single model that should be used for all purposes.

2 Method description

In the context of graphical modeling (see Lauritzen, 1996), given multivariate data, the goal is to estimate plausible positions of edges in the graph, or equivalently conditional independencies between variables. A temporal chain graph $G(E, V)$, includes both directed and undirected edges for which the multivariate random vector Y at time t follows

$$Y_t | Y_{t-1} \sim N(\Gamma Y_{t-1}, \Sigma).$$

A directed edge between nodes i and j belongs to E if $\Gamma_{ij} \neq 0$ and an undirected one is placed if $\Sigma^{-1}_{ij} \neq 0$.
Working under a nodewise local misspecification framework assumes that the true model is in a 'neighborhood' of the least complex model one is willing to assume, that is for a particular node Y, given a sample of n cases, it assumes that

$$Y_k \text{ has density } f(y_k | w_k, z_k, \theta, \gamma_0 + \delta/\sqrt{n}),$$

where f is two times continuously differentiable in a neighborhood of the vector (θ_0, γ_0). We define a *focus parameter*, i.e. $\mu(\theta_0, \gamma_0 + \delta/\sqrt{n})$ as a function of the parameters of the model density, and potentially of a user-specified vector of covariate values, for any particular node in the graph $G(E, V)$. The vector θ corresponds to all parameters that are included in every model (if a node X is known a priori to influence Y, then its effect is included in θ, and we call it a protected node) while γ collects all parameters corresponding to the potential influential set of nodes. W and Z denote the sets of protected and unprotected nodes.
An estimator for (θ, γ) is obtained by optimizing

$$Q(\theta, \gamma) = \frac{1}{n} \sum_{k=1}^{n} \log f(y_k | w_k, z_k, \theta, \gamma) - \frac{\lambda}{n} \sum_{j=1}^{d_\gamma} \psi(|\gamma_j - \gamma_{j0}|), \tag{1}$$

with respect to θ and γ for a given penalty function ψ (that is twice differentiable in 0) and that depends on an external value λ.

Subsequent steps involve specifying a collection of models S (based on which some γ elements are set at 0, while others are freely estimated) and (1) is optimized for parameters corresponding to the prespecified models. For $n \to \infty$ the quantity $\sqrt{n}(\hat{\mu}_S - \mu_{true}) \xrightarrow{d} \Lambda_S$ for which Λ_S is a normal random variable with a certain mean and variance.
Adding squared bias and variance, we immediately obtain the MSE expression of the estimator $\hat{\mu}_S$.
The FIC estimates the MSE for each of a collection of models S at each node, and selects the model with the smallest MSE value. A nodewise decomposable FIC score pertaining to the entire graph is then constructed as:

$$\mathrm{FIC}(G(E_S, V)) = \sum_{l=1}^{p} \widehat{\mathrm{MSE}}(\hat{\mu}_{l;S_l}).$$

Since our above argumentation was nodewise related, we construct an estimated graph as follows: at each node both the contemporaneous and dynamic effects of other nodes are used in order to determine a low-MSE model, and none of them is on a priori grounds protected. Once all nodewise models are selected (some might include only contemporaneous or dynamic effects, while others might contain both) we apply the following 'OR' rule adapted from Meinshausen and Bühlmann (2006):

$$\begin{aligned}
\hat{E}^{\lambda,\mathrm{OR}}_{i-j} &= \left\{(i,j) \cup (j,i) : i_t \in \hat{ne}^{\lambda}_{j_t} \text{ OR } j_t \in \hat{ne}^{\lambda}_{i_t}\right\} \\
\hat{E}^{\lambda,\mathrm{OR}}_{i\to j} &= \left\{(i,j) : i_{t-1} \in \hat{ne}^{\lambda}_{j_t}\right\} \\
\hat{E}^{\lambda,\mathrm{OR}} &= \left\{\hat{E}^{\lambda,\mathrm{OR}}_{i-j} \cup \hat{E}^{\lambda,\mathrm{OR}}_{i\to j}\right\},
\end{aligned}$$

where $\hat{ne}^{\lambda}$ denotes the neighborhood of the considered node for a certain value of λ. The focus of the research (i.e. the purpose of the model), directs the selection and different focuses may lead to different selected models. In this way, one can obtain better selected models in terms of MSE, than obtained from a global model search. For this application of FIC for graphs, the focus is the expected value of a variable, reflecting interest in discovering a topology of the graph that produces a low MSE for this focus. To deal with situations where $n < p$ a quadratic approximation to penalties such ℓ_1, 'SCAD' or 'Bridge' is applied on the γ vector. We propose further the selection of the regularization level as the quantity optimizing the MSE expression as

$$\lambda_S = \frac{\omega^t G_S \delta 1_q^t (J^{11,S,0})^t \omega - \omega^t \delta 1_q^t (J^{11,S,0})^t \omega}{\omega^t J^{11,S,0} 1_q 1_q^t (J^{11,S,0})^t \omega} \frac{\sqrt{n}}{\psi''(0)},$$

where G_S and $J^{11,S,0}$ can easily be obtained from the Fisher information matrix, while ω involves both the derivatives of the focus parameter and parts of the Fisher information matrix and $\psi^{''}(0)$ is the second derivative of the penalty function evaluated at 0.

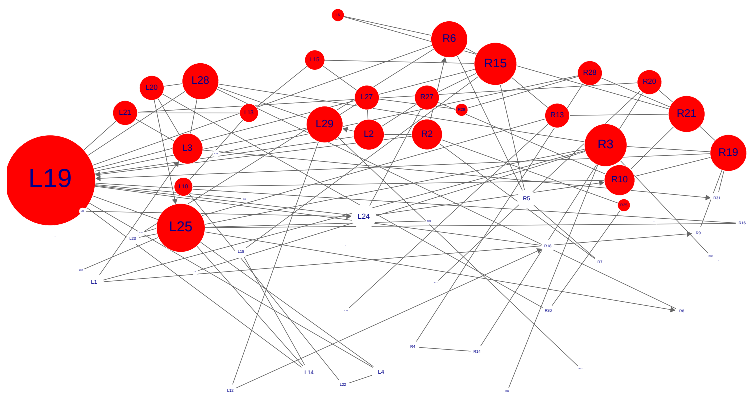


FIGURE 1. fMRI data. FIC graph based on ℓ_2 penalty. Undirected edges denote contemporaneous effects between brain regions while directed edges denote autoregressive dynamic effects. L/R denote the left/right hemisphere and larger labels correspond to high-degree regions.

3 fMRI data application

Figure 1 illustrates that for a new subject, for which certain brain regions (colored in red) have a large signal compared to the average, the FIC performs well in discovering the activated regions. Moreover, studying the evolution of the network over time by considering each time the measured signal as a focus, one can conclude that, for example, based on Figure 2 the small-worldness feature of the network which quite often is an interesting aspect of the network to look at, seems to be implausible for the two subjects in the analysis. Most of the estimated values seem to be below a threshold line at 1 and only in a relatively small number of cases the estimated values are larger than the cut-off value. Interestingly, it seems that for subject 1 the estimated networks at later time points produce larger values, while for subject 2 the networks estimated in the first and third part of the measurements produce larger index values.
The basic method is then expanded towards three types of situations:

- where one wishes to bring in all information from all the subjects in the analysis and construct one estimated network;
- where one wishes to construct related networks for subjects by allowing for dependence between the measurements for one region of interest across multiple subjects;
- or where one is pooling all data and wishes to estimate networks taking into account subject specific effects.

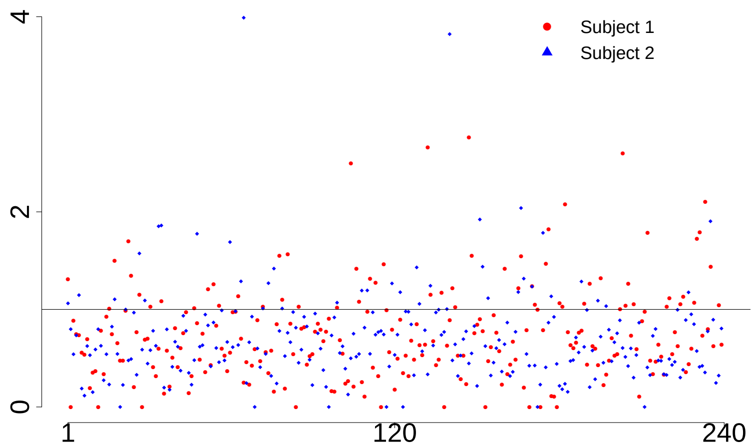


FIGURE 2. fMRI data. Estimated Small-World indexes for two subjects. At each time point, for each of the subjects, a graph based on the ℓ_2 penalty is estimated and based on the graph structure, the Small-World index is computed.

4 Simulation study

To show the method's performance in a controlled study, we have generated data from 42 different settings using various sample sizes, number of nodes or true graph model and inspected the empirical MSE, sparsity index and a 'structure closeness' score F_1 that measures how close the estimated graph is to the true graph for 3 different focuses. Figure 3 presents averages of the empirical MSE value when the data obtained from 42 different simulation settings each with 300 simulation runs is pooled together. The FIC procedures with three types of penalties provide low MSE with the ℓ_2 penalty being the best performing one.
Depending on the focus and the simulation setting sometimes the FIC provides sparser graphs than competitors (sometimes denser) and sometimes the graphs estimated by the FIC are closer to the true data generating process, sometimes further away, but most importantly the estimated graphs perform well with respect to MSE as intended from construction.

5 Discussion

Using the FIC one has at disposal a powerful method to study evolutions over time of networks constructed to study the functional connectivity between brain regions. The method clearly identifies important brain regions which seem to be highly connected with others, acting as 'hubs' or 'informational gateways' and performs well on simulated data.

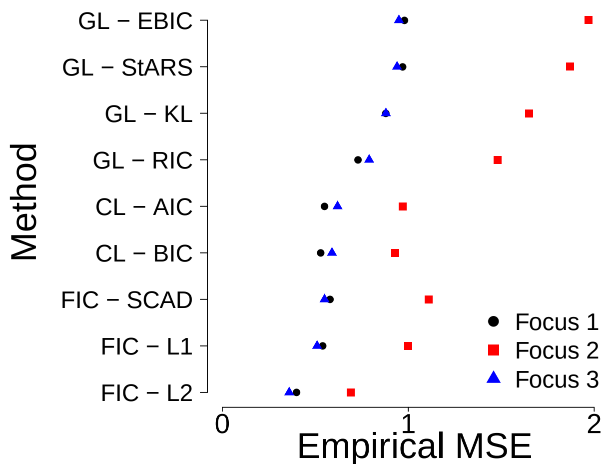


FIGURE 3. Simulated data. Averages of empirical MSE estimated across 300 simulation runs and 42 different settings for three focus points.

Acknowledgments: This research was partially supported by KU Leuven grant GOA FlexStatRob.

References

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, **10(3)**, 186 – 198.

Claeskens, G. and Hjort, N. (2003). The focused information criterion (with discussion and a rejoinder by the authors). *JASA*, **98**, 900 – 916.

Lauritzen, S. (1996). *Graphical Models*. Oxford University Press.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, **34(3)**, 1436 – 1462.