Vector Autoregressive Models With Measurement Errors for Testing Granger Causality

Alexandre G. Patriota[1], **João R. Sato**[2,3] and **Betsabé G. Blas Achic**[1]

[1]Departament of Statistics, University of São Paulo - SP - Brazil
Postal code 66281 - CEP 05314-970, São Paulo - SP - Brazil

[2] Center of Mathematics, Computation and Cognition Center,
Federal University of ABC, Santo André - Brazil

[3] Institute of Radiology - Hospital das Clínicas, São Paulo - Brazil
CEP 05403-001, São Paulo - SP - Brazil

Abstract

This paper develops a method for estimating parameters of a vector autoregression (VAR) observed in white noise. The estimation method assumes the noise variance matrix is known and does not require any iterative process. This study provides consistent estimators and the asymptotic distribution of the parameters required for conducting tests of Granger causality. Methods in the existing statistical literature cannot be used for testing Granger causality, since under the null hypothesis the model becomes unidentifiable. Measurement error effects on parameter estimates were evaluated by using computational simulations. The results suggest that the proposed approach produces empirical false positive rates close to the adopted nominal level (even for small samples) and has a satisfactory performance around the null hypothesis. The applicability and usefulness of the proposed approach are illustrated using a functional magnetic resonance imaging dataset.

Key Words: Asymptotic property, errors-in-variables model, Granger causality, multivariate time series.

1 INTRODUCTION

Multivariate time series modeling is an important component for the quantitative assessment of relationships between variables in many applied areas. This issue is essential in financial applications, for example, enabling optimal portfolio allocation, setting trading strategies over sectors of the market, or exchanging rates (Sims, 1980;

Ni and Sun, 2003). In addition, the vector autoregressive model (VAR) is widely used in many fields such as economics (Granger, 1969), geophysics (Liu and Rodríguez , 2005), bioinformatics (Fujita et al., 2007a) and neuroscience (Goebel et al., 2003).

The main reasons for the attractiveness of the VAR model in applied areas are its simplicity and relation with the concept of Granger causality (Granger, 1969). Granger causality has become a prominent concept in connectivity networks modeling, because it provides inferences about the direction of information flow between different time series. Several studies in biological systems emphasize the importance of identification and description of gene regulator networks (Gottesman, 1984; Katoh , 2007), mainly in the study of tumors or structural diseases. Mukhopadhyay and Chatterjee (2007); Fujita et al. (2007a,b) introduced the utilization of VAR-based models to study these issues by applying these models to gene expression datasets. In Neuroscience, the *functional integration* theories highlight that brain functions heavily depend on neural connectivity networks (Cohen and Tong, 2001). Several neuroimaging studies (Goebel et al., 2003; Sato et al., 2006; Abler et al., 2006) suggested that VAR models and Granger causality are suitable to identify the information flow between neural structures. Nevertheless, it is well known that most biological measurements are subject to error, since the precision of acquisition equipments is never absolute. Actually, this limitation is present in most studies involving experimental data, such as chemistry, physics, biometrics, etc.

Although technically incorrect, the most common procedure is simply to ignore the measurement errors, i.e.: to assume that the variables of interest are the observed ones. It is important to highlight that this assumption has serious implications. The conventional VAR model, in this case, would not identify correctly the relationships between the variables of interest (latent variables). It happens because the model white noise will not be independent which leads to misestimations of the model parameters. The usual assumption is acceptable when the errors are negligible. However, it is known that due to acquisition processes limitations, the measurement errors in biology (e.g.: gene expressions or neuro signals) are not negligible in most cases. Thus, the utilization of conventional VAR models may result in biased parameter estimation and as a consequence, unreliable Granger causality detection.

In the following, we define the usual VAR model (for a more detailed description, see for instance, Lütkepohl, 2005). Let $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{pt})^\top$ denotes a $(p \times 1)$ vector of time series variables. The usual VAR(r) model has the form

$$\boldsymbol{z}_t = \boldsymbol{a} + \boldsymbol{B}_1 \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{B}_r \boldsymbol{z}_{t-r} + \boldsymbol{q}_t, \;\; t = 1, \cdots, n \tag{1}$$

where n is the sample size, $\boldsymbol{B}_j$ for $j = 1, \ldots, r$ are $(p \times p)$ coefficient matrices and $\boldsymbol{q}_t$ is a $(p \times 1)$ unobservable zero mean white noise vector process with covariance matrix $\boldsymbol{\Sigma}$. For convenience, we consider that $\boldsymbol{z}_l = \boldsymbol{0}$ for all $l \leq 0$. We are assuming throughout this paper that model (1) satisfies the stability condition defined in Lütkepohl (2005) on page 12. Therefore, under stationarity conditions, the mean and the autocovariance function are given, respectively, by

$$E(\boldsymbol{z}_t) = \boldsymbol{\mu}_{\boldsymbol{z}} = \left(\boldsymbol{I}_p - \sum_{j=1}^{r} \boldsymbol{B}_j\right)^{-1} \boldsymbol{a},$$

$$\boldsymbol{\gamma}(h) = E[(\boldsymbol{z}_t - \boldsymbol{\mu}_{\boldsymbol{z}})(\boldsymbol{z}_{t-h} - \boldsymbol{\mu}_{\boldsymbol{z}})^\top] = \sum_{j=1}^{r} \boldsymbol{B}_j \boldsymbol{\gamma}(h-j), \quad \text{for} \;\; h = 1, 2, 3, \ldots$$

and

$$\boldsymbol{\gamma}(0) = \sum_{j=1}^{r} \boldsymbol{B}_j \boldsymbol{\gamma}(-j) + \boldsymbol{\Sigma}$$

where $\boldsymbol{I}_p$ denotes the $p \times p$ identity matrix and $\boldsymbol{\gamma}(-j) = \boldsymbol{\gamma}(j)^\top$.

Model (1) can be written in short as

$$\boldsymbol{z}_t = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{z}^*_{t-1} + \boldsymbol{q}_t, \;\; t = 1, \cdots, n \tag{2}$$

where $\boldsymbol{B} = (\boldsymbol{B}_1 \; \boldsymbol{B}_2 \; \ldots \; \boldsymbol{B}_r)$ is a $p \times pr$ matrix and $\boldsymbol{z}^*_{t-1} = (\boldsymbol{z}_{t-1}^\top, \boldsymbol{z}_{t-2}^\top, \ldots, \boldsymbol{z}_{t-r}^\top)^\top$.

Therefore, if the white noise has Normal distribution, the conditional Maximum Likelihood (ML) estimators of $\boldsymbol{a}$, $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are equal to the ordinary least squares estimators. They are given, respectively, by

$$\widehat{\boldsymbol{a}}_{ML} = \bar{\boldsymbol{z}}_t - \widehat{\boldsymbol{B}}_{ML}\bar{\boldsymbol{z}^*}_{t-1}, \quad \widehat{\boldsymbol{B}}_{ML} = (\boldsymbol{S}^{-1}_{\boldsymbol{z}^*_{t-1}} \boldsymbol{S}_{\boldsymbol{z}^*_{t-1}\boldsymbol{z}_t})^\top \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{ML} = n^{-1}\sum_{i=1}^{n} \widehat{\boldsymbol{q}}_i \widehat{\boldsymbol{q}}_i^\top \tag{3}$$

where $\bar{\boldsymbol{z}^*}_{t-1} = n^{-1}\sum_{i=1}^{n} \boldsymbol{z}^*_{i-1}$, $\bar{\boldsymbol{z}}_t = n^{-1}\sum_{i=1}^{n} \boldsymbol{z}_i$, $\widehat{\boldsymbol{q}}_i = \boldsymbol{z}_i - \widehat{\boldsymbol{a}}_{ML} - \widehat{\boldsymbol{B}}_{ML}\boldsymbol{z}^*_{i-1}$, $\boldsymbol{S}_{\boldsymbol{z}^*_{t-1}} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{z}^*_{i-1} - \bar{\boldsymbol{z}^*}_{t-1}){\boldsymbol{z}^*_{i-1}}^\top$ and $\boldsymbol{S}_{\boldsymbol{z}^*_{t-1}\boldsymbol{z}_t} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{z}^*_{i-1} - \bar{\boldsymbol{z}^*}_{t-1}){\boldsymbol{z}_i}^\top$.

The consistency of those conditional ML estimators is assured under the stationary conditions (see Lütkepohl, 2005, for further details). The consistency is shown using the fact that

$$\bar{\boldsymbol{z}}_t \xrightarrow{\mathcal{P}} \boldsymbol{\mu}_{\boldsymbol{z}}, \quad \bar{\boldsymbol{z}^*}_{t-1} \xrightarrow{\mathcal{P}} \boldsymbol{\mu}_{\boldsymbol{z}^*} = \boldsymbol{1}_r \otimes \boldsymbol{\mu}_{\boldsymbol{z}}, \quad \boldsymbol{S}_{\boldsymbol{z}^*_{t-1}} \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}_r(0) \quad \text{and} \quad \boldsymbol{S}_{\boldsymbol{z}^*_{t-1}\boldsymbol{z}_t} \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}_r(0)\boldsymbol{B}^\top$$

where "$\xrightarrow{\mathcal{P}}$" denotes convergence in probability when the sample size increases, $\otimes$ denotes the Kronecker product, $\boldsymbol{1}_r$ is a $r-$dimensional column vector of ones, and the covariance function of $\boldsymbol{z}_{t-1}^{*}$ is given by

$$\begin{aligned}\boldsymbol{\Gamma}_r(h) &= E[(\boldsymbol{z}_{t-1}^{*}-\boldsymbol{\mu}_{\boldsymbol{z}^*})(\boldsymbol{z}_{t-h-1}^{*}-\boldsymbol{\mu}_{\boldsymbol{z}^*})^{\top}] \\ &= \begin{bmatrix} \boldsymbol{\gamma}(h) & \boldsymbol{\gamma}(h+1) & \dots & \boldsymbol{\gamma}(h+r-1) \\ \boldsymbol{\gamma}(h-1) & \boldsymbol{\gamma}(h) & \dots & \boldsymbol{\gamma}(h+r-2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\gamma}(h-r+1) & \boldsymbol{\gamma}(h-r+2) & \dots & \boldsymbol{\gamma}(h) \end{bmatrix}.\end{aligned}$$

As described previously, VAR modeling is commonly applied for detecting Granger causality relationships. The basic idea of Granger causality is the evaluation of temporal information founded on the assumption that the cause always precedes its effect (Granger, 1969). Let x_t and y_t be two time series. From the statistical perspective, x_t is said to Granger-cause y_t if the prediction error of y_t, conditioning on the past values of both series, is less than considering solely the past values of y_t. In other words, the past values of x_t contains relevant information to improve the predictions of y_t. Note that Granger causality concept is not equivalent to classical Aristotelian causality, since the former is based solely on prediction errors. However, due to its simplicity, it is more tractable in scientific experiments and may suggest possible causal relationships.

One possible approach of using VAR models for Granger causality detection is by performing statistical tests on $\boldsymbol{B}_j$'s coefficients. Considering y_t equation, if there is at least one coefficient multiplying the past values of x_t which is not equal to zero, then x_t is said to Granger-cause y_t. Thus, this procedure involves the estimation of $\boldsymbol{B}_j$, their respective covariance matrices, and the application of hypothesis testing.

In general, many physical, biological and chemical variables have the measurement process subject to noise effects and it is very common to analyze them by using models assuming that these measurement errors are negligible. It may bring up undesirable features as biased estimates as well as their standard errors and, as a consequence, dangerously false confidence intervals and unreliable hypotheses testing. Thus, it is necessary to consider the measurement error in the modeling of these type of data.

In this paper, we study a VAR model with main concern on including measurement errors. Let $\boldsymbol{z}_t$ be the true (latent) variable that is not directly observed, instead a substitute variable $\boldsymbol{Z}_t$ is observed. The relation between the latent and observed variables is given by the following additive structure

$$\boldsymbol{Z}_t = \boldsymbol{z}_t + \boldsymbol{e}_t, \ \ t = 1, \cdots, n \tag{4}$$

where $\boldsymbol{Z}_t = (Z_{1t}, Z_{2t}, \cdots, Z_{pt})^\top$ is the observed vector and $\boldsymbol{e}_t = (e_{1t}, e_{2t}, \cdots, e_{pt})^\top$ is the measurement error vector with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}_e$. In most cases, if the usual conditional ML estimator is adopted for the observations subject to errors, i.e., replacing $\boldsymbol{z}_t$ with $\boldsymbol{Z}_t$ in equation (1), the estimator of $\boldsymbol{B}$ will not be consistent (as can be seen in (6)). Therefore, measurement error equation (4) should be included in the estimation procedure. Nevertheless, model (1) with equation (4) is not identifiable, since the covariance matrices of $\boldsymbol{q}_t$ and $\boldsymbol{e}_t$ are confounded when $\boldsymbol{B} = \boldsymbol{0}$. It is easy to see that in the univariate AR(1), note that when $r = p = 1$ and $b = 0$ we have: $Z_i = a + q_i + e_i$ with $\mathrm{E}(Z_i) = a$, $\gamma(0) = \sigma^2 + \sigma_e^2$ and $\gamma(h) = 0$ for all $h \neq 0$. It is impossible to estimate σ^2 and σ_e^2 separately by observing only $Z_1, \ldots, Z_n$. This problem can be avoided by using previous knowledge about the variance of $\boldsymbol{e}_t$.

This paper is organized as follows. Section 2 proposes consistent estimators for the VAR model with measurement errors and also presents the asymptotic distribution of the estimator of the elements of $\boldsymbol{B}$. In Section 3, simulation studies are undertaken to investigate some aspects of the proposed estimators (rejection rates for a test of hypothesis, biases and mean square errors) also it is verified the impact by erroneously considering the usual model. We applied the models in a functional magnetic resonance imaging dataset in Section 4 and we finish the paper with conclusions and remarks in Section 5.

2 VAR WITH MEASUREMENT ERRORS

In the presence of measurement errors, the conventional ML estimation of VAR models produces biased estimators and they may lead to wrong statistical inferences (see Fuller, 1987, in which it is found a discussion over errors-in-variables in regression models). Andersson (2005) warns for the problems in testing Granger causality by using a VAR model when the variables are subject to measurement errors. However, the author does not propose any approach to overcome such problems. There are some studies to treat time series observed in white noise in the literature, those studies use Kalman filtering methodology and an Expectation and Maximization algorithm that requires intensive iterative procedures, (e.g., Geweke, 1977; Aigner et al., 1984). Maravall and Aigner (1977) have provided a careful exposition of the identifiability of some time series models with errors in variables. Beck (1990) describes approaches based on state space modeling and Kalman filtering and demonstrates the usefulness of these tools in dynamic models. Kellstedt et al. (1996) show the efficiency gains adopting errors-in-variables models, and the precision of Kalman filter estimates in the face of autocorrelation. These measurement techniques have been applied to a variety of substantive problems, including dynamic representation, social problems (such as racial inequality), monetary policy and public entrepreneurship (Williams and McGinnis, 1992).

These state space models can be attractive alternatives to conventional VAR modeling. However, in practice, the implementation of the estimators are not described in analytical form, but by interactive algorithms or numerical optimization solutions. In addition, the derivation of the asymptotic distribution of those estimators may be complex. In Shumway and Stoffer (2000), the section on state space methods shows an alternative procedure for how to estimate $\boldsymbol{B}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_e$ under model (1) with error equations (4), using the EM algorithm. Hannan et al. (2003) proposed another iterative procedure to estimate these parameters. Nevertheless, as the main goal of this paper is to test Granger causality and the effect of the autoregressive coefficientes, e.g., the coefficient that relates $z_{t,j} \to z_{t,j+r}$, these approaches cannot be used, since the model becomes unidentifiable under the hypothesis $\boldsymbol{B} = \mathbf{0}$.

In this study, we provide simple and closed forms for the estimators when $\boldsymbol{\Sigma}_e$ is known, which allows the direct derivation of their respective asymptotic properties. Since the main concern of several practical applications is Granger causality testing, this information is essential to data analysis. In this section, the main concern is the parameter estimation and its asymptotic properties. Theorem 1 states consistent estimators for the model parameters and Theorem 2 establishes the asymptotic distribution for the estimator of $\mathrm{vec}(\boldsymbol{B}^\top)$ given in Theorem 1, where $\mathrm{vec}(\boldsymbol{C})$ is an operator that heaps the columns of the matrix $\boldsymbol{C}$.

The methodology presented in this section is based on correcting the asymptotic bias of conventional ML estimator caused by the measurement error effect. The outcome is a consistent estimator with good asymptotic properties such as normality. The estimators and the asymptotic covariance matrix for the proposed estimator of $\mathrm{vec}(\boldsymbol{B}^\top)$ are computed easily and no iterative procedure is required. We must remark that those estimators are not the conditional ML estimators nor the ML estimators taking into account the measurement errors which are very complicated to reach by maximizing the likelihood, even under normality of the errors.

Theorem 1. *If $\boldsymbol{e}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e)$ with $\boldsymbol{\Sigma}_e$ known. Then, the parameters of model (1)* *under measurement errors as in (4) have consistent estimators given by*

$$\widehat{\boldsymbol{a}} = \bar{\boldsymbol{Z}}_t - \widehat{\boldsymbol{B}}\bar{\boldsymbol{Z}}^*{}_{t-1}, \quad \widehat{\boldsymbol{B}} = \left[(\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} - \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)^{-1}\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t}\right]^\top \tag{5}$$

and

$$\widehat{\boldsymbol{\Sigma}} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}_i - \widehat{\boldsymbol{a}} - \widehat{\boldsymbol{B}}\boldsymbol{Z}^*_{i-1})(\boldsymbol{Z}_i - \widehat{\boldsymbol{a}} - \widehat{\boldsymbol{B}}\boldsymbol{Z}^*_{i-1})^\top - \boldsymbol{\Sigma}_e - \widehat{\boldsymbol{B}}(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\widehat{\boldsymbol{B}}^\top$$

where $\bar{\boldsymbol{Z}}^*{}_{t-1} = n^{-1}\sum_i \boldsymbol{Z}^*_{i-1}$, $\bar{\boldsymbol{Z}}_t = n^{-1}\sum_i \boldsymbol{Z}_i$, $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} = n^{-1}\sum_i(\boldsymbol{Z}^*_{i-1} - \bar{\boldsymbol{Z}}^*{}_{t-1})\boldsymbol{Z}^{*\top}_{i-1}$ *and* $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t} = n^{-1}\sum_i(\boldsymbol{Z}^*_{i-1} - \bar{\boldsymbol{Z}}^*{}_{t-1})\boldsymbol{Z}_i^\top$.

The proof of Theorem 1 can be found in Appendix A.1. Notice that, if $\boldsymbol{\Sigma}_e = \boldsymbol{0}_{p\times p}$, that is, when there is no measurement error, then the estimators of Theorem 1 become the conditional ML estimators presented in (3). Also, it can be seen that the conditional ML estimator of $\boldsymbol{B}$ from model (1), without considering errors (4), is given by

$$\widehat{\boldsymbol{B}}_{ML} = \left[\boldsymbol{S}^{-1}_{\boldsymbol{Z}^*_{t-1}}\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t}\right]^\top,$$

which is not consistent, since

$$\widehat{\boldsymbol{B}}_{ML} \xrightarrow{\mathcal{P}} \boldsymbol{B}[\boldsymbol{I}_{pr} + (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\boldsymbol{\Gamma}_r(0)^{-1}]^{-1}. \tag{6}$$

The main steps to demonstrate (6) is given in Appendix A.1, in which is sufficient to compute the limit of $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}}$ and $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t}$. The quantity $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}}$ has two sources of variations, one that refers to the unobservable variable $\boldsymbol{z}^*_{t-1}$ and another one that refers to the measurement error.

If the measurement error is huge and the sample size is not large enough, the quantity $(\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} - \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)$ may not be positive definite and the estimator $\widehat{\boldsymbol{B}}$, presented in (5), will be inadmissible. If the quantity $(\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} - \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)$ has at least one eigenvalue close to zero the estimator $\widehat{\boldsymbol{B}}$, presented in (5), will be unstable (because the computation of a matrix inverse requires all eigenvalues to be different from zero). If the matrix $\boldsymbol{\Sigma}_e$ is well specified, one way to avoid such inadmissibility and instability is increasing the sample size.

In many practical applications, there is some interest on testing some elements of the matrix $\boldsymbol{B}$ (e.g., the so called Granger causality test). However, the exact distribution of $\mathrm{vec}(\widehat{\boldsymbol{B}}^\top)$ is difficult to compute. Thus, one can use its asymptotic distribution to build confidence regions and hypothesis testing as an approximation when the sample size is finite. The Theorem below gives us the asymptotic distribution of $\mathrm{vec}(\widehat{\boldsymbol{B}}^\top)$.

Theorem 2. *If* $\boldsymbol{e}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e)$ *with* $\boldsymbol{\Sigma}_e$ *known and* $E(q_{ij_1}q_{ij_2}q_{ij_3}q_{ij_4}) < \infty$ *for all* $j_1, j_2, j_3, j_4 \in \{1, \ldots, p\}$, *where* q_{ij} *is the* j^{th} *element of* $\boldsymbol{q}_i$. *Then, the asymptotic distribution of* $vec(\widehat{\boldsymbol{B}}^\top)$ *obtained in Theorem 1 is given by*

$$\sqrt{n}(vec(\widehat{\boldsymbol{B}}^\top) - vec(\boldsymbol{B}^\top)) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}), \tag{7}$$

where the $p^2r \times p^2r$ *matrix* $\boldsymbol{\Phi}$ *is given by*

$$\boldsymbol{\Phi} = \boldsymbol{\Sigma}_\vartheta \otimes \boldsymbol{\Gamma}_r(0)^{-1} + (\boldsymbol{I}_p \otimes \boldsymbol{\Gamma}_r(0)^{-1})\boldsymbol{A}_r(\boldsymbol{I}_p \otimes \boldsymbol{\Gamma}_r(0)^{-1})$$

where

$$\begin{aligned}\boldsymbol{A}_r &= \boldsymbol{\Sigma}_\vartheta \otimes (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e) + \boldsymbol{B}^\top \otimes [\boldsymbol{\Sigma}_e \boldsymbol{B}(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)] + \\ &\quad - \sum_{h=1}^{r} \left\{ (\boldsymbol{B}_h \boldsymbol{\Sigma}_e) \otimes \boldsymbol{\Gamma}_r(h) + (\boldsymbol{\Sigma}_e \boldsymbol{B}_h^\top) \otimes \boldsymbol{\Gamma}_r(-h) \right\} + \\ &\quad + \sum_{h=1-r}^{r-1} [\boldsymbol{B}(\boldsymbol{J}_{-h} \otimes \boldsymbol{\Sigma}_e)\boldsymbol{B}^\top] \otimes \boldsymbol{\Gamma}_r(h).\end{aligned}$$

and $\boldsymbol{\Sigma}_\vartheta = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_e + \boldsymbol{B}(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\boldsymbol{B}^\top$, *where* $\boldsymbol{J}_l$ *is a* $(r \times r)$ *matrix of zeros with one's in the* $|l|^{th}$ *diagonal above (below) the main diagonal if* $l > 0$ $(l < 0)$ *and* $\boldsymbol{J}_0$ *is a* $(r \times r)$ *matrix of zeros.*

The proof of Theorem 2 can be seen in Appendix A.2. For all r and $\boldsymbol{\Sigma}_e = 0$ we have $\boldsymbol{\Phi} = \boldsymbol{\Sigma} \otimes \boldsymbol{\Gamma}_r(0)^{-1}$, as given in Lütkepohl (2005). The Normal distribution assumption for the measurement error is required to compute the expectation of polynomial functions (until forth degrees) of the elements of $\boldsymbol{e}_t$.

The assumption of known measurement error variance is usually considered in many fields; such as, astrophysics (Akritas and Bershady, 1996; Kelly, 2007; Kelly et al., 2008), epidemiology (Kulathinal et al., 2002; Patriota et al., 2009), analytical chemistry (Cheng and Riu, 2006), among others. However, in real datasets this measurement error variance is, in general, estimated. If $\widehat{\boldsymbol{\Sigma}}_e$ is a consistent estimator for $\boldsymbol{\Sigma}_e$, then we have usually that $\widehat{\boldsymbol{\Sigma}}_e = \boldsymbol{\Sigma}_e + O_p(m^{-1/2})$, where m is the sample size used in the previous experiment, and $O_p(m^{-1/2})$ means limited in probability even multiplying by $m^{1/2}$. Then, provided that $\lim_{n\to\infty} n/m = 0$, all asymptotic results derived in this section remain valid. However, note that if $\lim_{n\to\infty} n/m = \infty$, then it is not possible to compute the asymptotic distribution for $\text{vec}(\widehat{\boldsymbol{B}}^\top)$, since its covariance matrix will diverge. We remark that, although if $\lim_{n\to\infty} n/m \in (0, \infty)$ the asymptotic distribution derived in this paper will not be valid, our results can also be used here with some caution. Our simulation studies (see Section 3) show that the rejection rates under the null hypothesis are controlled even when $\boldsymbol{\Sigma}_e$ is replaced by an estimator built by using a previous sample (m) proportional to the sample size (n).

In some cases, the partitioner can just specify the covariance matrix $\boldsymbol{\Sigma}_e$ rather than estimating it through previous experiments. In such cases, a misspecification in this covariance matrix may occur. For the sake of simplicity, suppose that $\boldsymbol{\Sigma}_e$ is the true covariance matrix and the misspecified one is $\boldsymbol{\Sigma}_e^{(mis)} = \delta\boldsymbol{\Sigma}_e$. Let $\widehat{\boldsymbol{B}}^{(mis)}$ be the estimator of $\boldsymbol{B}$ built by using the misspecified covariance matrix $\boldsymbol{\Sigma}_e^{(mis)}$ instead of $\boldsymbol{\Sigma}_e$. Then, using the results of the Appendix, we have that

$$\widehat{\boldsymbol{B}}^{(mis)} \xrightarrow{\mathcal{P}} \boldsymbol{B}\big[\boldsymbol{I}_{pr} + (1-\delta)(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\boldsymbol{\Gamma}_r(0)^{-1}\big]^{-1},$$

i.e., the estimator $\widehat{\boldsymbol{B}}^{(mis)}$ is not consistent for $\boldsymbol{B}$. However, if $0 < \delta < 2$ the matrix $\big[\boldsymbol{I}_{pr} + (1-\delta)(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\boldsymbol{\Gamma}_r(0)^{-1}\big]^{-1}$ will have eigenvalues closer to 1, in absolute value, than the ones of $\big[\boldsymbol{I}_{pr} + (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\boldsymbol{\Gamma}_r(0)^{-1}\big]^{-1}$ (notice that, if λ_C is the eigenvalue of $\boldsymbol{C}$, then $1 + \gamma\lambda_C$ is the eigenvalue of $\boldsymbol{I} + \gamma\boldsymbol{C}$). In this sense, the estimator $\widehat{\boldsymbol{B}}^{(mis)}$ will have lesser asymptotic bias than $\widehat{\boldsymbol{B}}_{ML}$, for $0 < \delta < 2$. In other words, even if we underestimate the true matrix $\boldsymbol{\Sigma}_e$ or if we overestimate by up to two times, the multiplicative term of the asymptotic bias will be closer to the identity matrix than the one produced by the naive estimator (i.e., considering that $\boldsymbol{\Sigma}_e = \boldsymbol{0}$).

Notice that, if $r = 1$ we have the VAR(1) model and the asymptotic covariance simplifies to

$$\boldsymbol{\Phi} = \boldsymbol{\Sigma}_\vartheta \otimes \boldsymbol{\gamma}(0)^{-1} + (\boldsymbol{I}_p \otimes \boldsymbol{\gamma}(0)^{-1})\boldsymbol{A}_1(\boldsymbol{I}_p \otimes \boldsymbol{\gamma}(0)^{-1})$$

where

$$\boldsymbol{A}_1 = \boldsymbol{\Sigma}_\vartheta \otimes \boldsymbol{\Sigma}_e + \boldsymbol{B}^\top \otimes (\boldsymbol{\Sigma}_e\boldsymbol{B}\boldsymbol{\Sigma}_e) - [(\boldsymbol{B}\boldsymbol{\Sigma}_e) \otimes (\boldsymbol{\gamma}(0)\boldsymbol{B}^\top) + (\boldsymbol{\Sigma}_e\boldsymbol{B}^\top) \otimes (\boldsymbol{B}\boldsymbol{\gamma}(0))].$$

The i^{th} element of $\mathrm{vec}(\widehat{\boldsymbol{B}}^\top)$, is asymptotically normally distributed with standard error given by the square root of i^{th} diagonal element of $\boldsymbol{\Phi}$. Thus, we can obtain hypothesis tests on the individual coefficients, or more general form of contrasts

$$H_0 : \boldsymbol{C}\mathrm{vec}(\boldsymbol{B}^\top) = \boldsymbol{d} \qquad \mathit{Versus} \qquad H_1 : \boldsymbol{C}\mathrm{vec}(\boldsymbol{B}^\top) \neq \boldsymbol{d},$$

which involve coefficients across different equations of the VAR model. Thus, Granger causality testing can be carried out by adequately specifying this contrasts matrix. An illustrative example is the case of series x_t and y_t, in which we are interested in evaluating the Granger causality from x_t to y_t in an r-order VAR model. The matrix $\boldsymbol{C}$ has r rows, one for each coefficient related to the past values of x_t in the y_t equation. Considering that each column of $\boldsymbol{C}$ refers to each VAR coefficient, the contrast matrix is specified by simply setting 1 to the cell at the respective column and row for the x_t coefficients in y_t equation. This may be tested using the Wald-type statistic conveniently expressed as

$$n(\boldsymbol{C}\text{vec}(\widehat{\boldsymbol{B}}^\top) - \boldsymbol{d})^\top \left[\boldsymbol{C}\boldsymbol{\Phi}\boldsymbol{C}^\top\right]^{-1}(\boldsymbol{C}\text{vec}(\widehat{\boldsymbol{B}}^\top) - \boldsymbol{d}) \tag{8}$$

Under the null hypothesis, (8) has a $\chi^2(c)$ distribution in the limit, where $c = rank(\boldsymbol{C})$ gives the number of linear restrictions.

The previous procedure can also be developed to include the intercept by applying the delta method (Lehmann and Casella, 1998) in the asymptotic distribution of $(\bar{\boldsymbol{Z}}_t^\top,\ \bar{\boldsymbol{Z}}_{t-1}^{*\top},\ \text{vec}(\widehat{\boldsymbol{B}}^\top)^\top)$, since $\widehat{\boldsymbol{a}} = \bar{\boldsymbol{Z}}_t - (\boldsymbol{I} \otimes \bar{\boldsymbol{Z}}_{t-1}^{*\top})\text{vec}(\widehat{\boldsymbol{B}}^\top)$. Although, this asymptotic distribution is important to test hypotheses regarding the model intercept, it is outside the main scope of this article and does not have any impact on the Granger causality.

3 SIMULATION RESULTS

In this section, some simulation studies were conducted in order to evaluate the adequacy of the asymptotic distribution of $\text{vec}(\widehat{\boldsymbol{B}}^\top)$ for small and moderate samples sizes. Computations were performed using the software R (www.r-project.org).

For each setup of parameters and sample sizes, it was considered 15,000 Monte Carlo samples generated from a VAR(1) model with measurement errors, given by

$$\begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} + \begin{pmatrix} q_{1t} \\ q_{2t} \end{pmatrix}, \tag{9}$$

$$\begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \tag{10}$$

In all samples, the following setup of parameters was considered: $a_1 = a_2 = 1$, $b_{11} = b_{22} = 0.5$,

$$\boldsymbol{\Sigma} = \begin{bmatrix} 10 & 5 \\ 5 & 5 \end{bmatrix},$$

where the vector parameters values of (b_{12}, b_{21}) were the values of the set $\{(b_{12}, b_{21}); b_{12} \in S \text{ and } b_{21} \in S\}$, where $S = \{-0.4, -0.2, 0.0, 0.2, 0.4\}$, the variance of the measurement error $\boldsymbol{e}_t$ was $\boldsymbol{\Sigma}_e = 2\boldsymbol{I}_2$, and the sample sizes $n = 50, 100, 250, 500$. As in actual datasets $\mathbf{\Sigma}_e$ is usually estimated, we simulated $m = 0.6n$ identically and independent random variables from a Normal distribution with mean zero and variance two. Then, we estimate $\widehat{\mathbf{\Sigma}}_e = \widehat{\sigma}_e^2 \boldsymbol{I}_2$, where $\widehat{\sigma}_e^2$ is the sample variance computed from this random variables.

The rejection rates of the hypothesis $H_0 : b_{12} = b_{21} = 0$ (i.e., $z_{2,t-1}$ does not help to explain $z_{1,t}$ and $z_{1,t-1}$ does not help to explain $z_{2,t}$) are shown in Table 1, in which the test sizes are the rejection rates under the null hypothesis (that appears in bold). Wald-type statistic (8) is used at 5% nominal level. From this table we conclude that the test sizes from the proposed model are closer to the nominal level (5%), as compared to the usual approach for all sample sizes. Furthermore, when n increases the test sizes for the usual model also increase and, consequently, they do not converge to the adopted nominal level. This is an expected behavior because the usual approach produces biased estimates and standard errors. Table 1 depicts the power of the test in each methodology, which shows a good performance of the proposed approach. Nevertheless, it is not possible to compare the power between the two methods because they have different empirical test sizes.

[[Table 1]]

In addition, the results shown in Table 1 are similar for other values of the parameters $\boldsymbol{a}$ and $\boldsymbol{B}$, if the same proportionality of $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_e$ is set as defined above. But, other simulations suggest that the larger the measurement error, the larger the sample size required to have a good asymptotic approximation for Wald-type statistic (8).

Further, simulation studies were also conducted for testing the hypothesis $H_0 : b_{12} = 0$ at 5% nominal level. In this study, we consider $b_{21} = 0.2$. Other simulations were built considering other values for b_{21}, however, the results are close to each other and, for this reason, they were omitted. As can be seen, Tables 1 and 2 present similar behaviors, i.e., the proposed model has always empirical size test closer to the nominal level than the usual one.

[[Table 2]]

In Tables 1 and 2, the usual approach seems to be most powerful than the proposed approach when $b_{21} = 0.2$ and $b_{21} = 0.4$. However, as aforementioned, they cannot be compared directly, just because the real nominal level used to compute that powers are not the same. Thus, a descriptive measure was defined in order to analyze both methodologies around the null hypothesis. Let $a_n(\alpha)$ be the probability of the error type I using the true distribution of (8) when the sample size is n and

α is the adopting nominal level based on its asymptotic distribution. For instance, in Table 2 $\widehat{a_{100}}(0.05) = 0.0541$ for the proposed approach and $\widehat{a_{100}}(0.05) = 0.0851$ for the usual one (i.e., $\widehat{a_n}(\alpha)$ is the test size for a given n and α). An expected behavior for good statistics is $a_n(\alpha) \overset{n\to\infty}{\longrightarrow} \alpha$ which means that the quantiles of the true distribution of (8) will be close to the quantiles of the asymptotic distribution, $\chi^2(c)$, when the sample size is sufficiently large. Thus, the relation $a_n(\alpha)/\alpha$ measures how far is the α-quantile of the asymptotic distribution from the true distribution of (8) for each n. Therefore, a corrected power may be defined by

$$P_n^{(c)}(\alpha) = \frac{P_n(a_n(\alpha))}{(a_n(\alpha)/\alpha)}$$

where $P_n(a(\alpha))$ is the power using the true probability of the error type I, namely $a_n(\alpha)$. The main idea is penalizing the power by the ratio between $a_n(\alpha)$ and α. Note that, the power under the null hypothesis has to be the nominal level and the comparison of powers from different statistics must be done adopting the same nominal level. Under the null hypothesis, we have that

$$P_{1n}^{(c)}(\alpha) = P_{2n}^{(c)}(\alpha) = \alpha,$$

since under the null hypothesis $P_n(a_n(\alpha)) = a_n(\alpha)$. Hence, the corrected powers $P_{1n}^{(c)}$ and $P_{2n}^{(c)}$ are comparable. Moreover, under an alternative hypothesis and when n increases, an expected behavior of $P_n^{(c)}(\alpha)$ is to converge towards one. Although, this corrected power is not a monotonic function of the sample size nor of the nominal level, we believe that it can be used as a descriptive measure to evidence how unsuitable is the usual model when compared with the proposed one outside the null hypothesis. Furthermore, the proposed corrected power varies between 0 and infinity. Figure 1 shows the corrected power for both approaches, the null hypothesis was $H_0 : b_{12} = 0$. The full line refers to the proposed approach and the dashed line refers to the usual one. The panels (a.1), (b.1), (c.1) and (d.1) refer to the corrected power when the alternative hypothesis are $b_{12} = -0.4$, $b_{12} = -0.2$, $b_{12} = 0.2$ and $b_{12} = 0.4$, respectively at $\alpha = 0.01$. The panels (a.2), (b.2), (c.2) and (d.2) refer to the corrected power when the alternative hypothesis are $b_{12} = -0.4$, $b_{12} = -0.2$, $b_{12} = 0.2$ and $b_{12} = 0.4$, respectively at $\alpha = 0.05$. The panels (a.3), (b.3), (c.3) and (d.3) refer to the corrected power when the alternative hypothesis are $b_{12} = -0.4$, $b_{12} = -0.2$, $b_{12} = 0.2$ and $b_{12} = 0.4$, respectively at $\alpha = 0.10$. We observe in all graphs that, the usual approach has the worst performance (going to zero when the sample size increases) while the proposed one have an expected behavior for a good statistic (going to one when the sample size increases). In general, the corrected power under the usual methodology goes to zero because the distance between $a_n(\alpha)$ and α increases much faster than the uncorrected power, $P_n(a_n(\alpha))$, when n increases. This behavior is still true for other setups of parameters.

[[Figure 1]]

[[Table 3]]

Table 3 shows that the biases of the estimators of b_{ij} $(i, j = 1, 2)$ from the proposed model are almost always smaller (in absolute value) than the value supplied by the usual model (except only for the parameter b_{21} when $n = 50$). Moreover, the larger the sample size, the smaller the bias and MSE under the proposed model (this does not happen for the usual approach). For this specific table, the true parameters are $b_{21} = 0.2$ and $b_{12} = -0.4$, all other parameters were chosen as previously described.

Table 4 presents the rejection rates for testing univariate hypotheses. In this table, the model was generated by considering $p = 4$, $a_1 = a_2 = a_3 = a_4 = 1$, $b_{11} = 0.9$, $b_{22} = 0.6$, $b_{33} = 0.4$, $b_{44} = 0.5$, $b_{41} = 0.5$, $b_{14} = -0.3$, $b_{12} = b_{13} = b_{21} = b_{23} = b_{24} = b_{31} = b_{32} = b_{34} = b_{42} = b_{43} = 0$. The measurement error variance was 0.60 and the variance of $\boldsymbol{q}_t$ was

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.80 & 0.20 & 0.20 & 0.05 \\ 0.20 & 0.80 & 0.05 & -0.05 \\ 0.20 & 0.05 & 1.00 & 0.10 \\ 0.05 & -0.05 & 0.10 & 0.90 \end{pmatrix}.$$

Notice that, these parameters and hypothesis tests were chosen to mimic our application (see next section for further details). We test the univariate hypotheses in each Monte Carlo simulation, say $H_0 : b_{12} = 0$, $H_0 : b_{13} = 0$, $H_0 : b_{21} = 0$, $H_0 : b_{23} = 0$, $H_0 : b_{24} = 0$, $H_0 : b_{31} = 0$, $H_0 : b_{32} = 0$, $H_0 : b_{42} = 0$ and $H_0 : b_{43} = 0$ by using the usual and proposed approaches. The measurement error variance was estimated through replications ($m = 0.6n$). Here, $n = 100, 200, 400$ in which for the real data $n = m = 200$.

[[Table 4]]

Notice that, for the proposed model the test sizes are, in average, closer to 5% than the usual one. Next section presents a comparison between of the results of Table 4 and the application.

4 APPLICATION

As previously described, the models including measurement errors have great relevance in applied sciences, since equipment imprecisions are inherent to data acquisition. Actually, the usual models are commonly applied ignoring these errors. Nowadays, the scientific community started to pay enough attention to the fact that these procedures may lead to spurious results. In this section, we illustrate the concepts introduced in the present study with an application embedded in Neuroscience research, with the utilization of VAR modeling for the characterization of brain networks.

The dataset explored in this application is proceeding from a functional magnetic resonance imaging (fMRI) experiment. Basically, fMRI acquisition is based on monitoring the BOLD signal (blood oxygenation level dependent) at several brain regions through time. One of the main advantages of fMRI over other imaging techniques is its non-invasive protocol and relative high spatial resolution. The BOLD signal is related to oxygen consumption and blood flow, being considered as an indirect measure of local neural activity (Logothetis et al. (2001)). Regarding this property, BOLD signal is used to quantify and locate the brain activity in humans.

In this study, the BOLD signals at four brain regions from a subject in a resting state (eyes closed) experiment were considered. The data was collected in a Siemens 3Tesla MR system (TR=1800ms, TA=900ms, TE=30ms). The selected brain regions were: left primary motor cortex (left M1), right primary motor cortex (right M1), supplementary motor area (SMA) and right cerebellum. For this volunteer, these regions were previously mapped by using a fingertap motor experiment. The anatomical location of these areas are shown in Figure 2. These regions are frequently involved in active and planned right hand fingertapping, and their role is already established in motor execution. However, we aim to evaluate the default connectivity network between these areas, which can be depicted by the information flow during a resting state run, which may be identified using VAR models and Granger causality concept.

A well described limitation inherent to all fMRI acquisition is the high level of scanner noise. Thus, the signals observed mirror not only the physiological variations but also includes measurement errors. For this specific dataset, it was estimated that the error composed approximately 57.10% of the observed time series standard deviation. This estimate was obtained by considering the squared root of the median variance of BOLD time series from extracranial voxels (i.e., we used 2,354 auxiliar time series of length 200), with baseline signal (mean) greater than 75. Voxels with baseline below this threshold are too far from tissue (image corners) and have minimal variance, which may lead to an underestimate of noise level. For simplicity, each observed series were normalized to have mean zero and variance one. The measurement error was considered to be serially uncorrelated, independent of the latent variables and with a standard deviation estimated at 0.571.

The model considered for the latent variable is given by

$$\boldsymbol{z}_t = \boldsymbol{a} + \boldsymbol{B}_1 \boldsymbol{z}_{t-1} + \boldsymbol{q}_t, \;\; t = 1, \cdots, n \tag{11}$$

where $n = 200$ is the time series length, $\boldsymbol{z}_t = (z_{1t}, z_{2t}, z_{3t}, z_{4t})^\top$ with z_{1t} : *Left M1* signal, z_{2t} : *SMA* signal, z_{3t} : *Right M1* signal and z_{4t} : *Right cerebellum* signal; $\boldsymbol{B}_1$ is the (4×4) autoregressive coefficients matrix

$$\boldsymbol{B}_1 = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{pmatrix}, \tag{12}$$

and $\boldsymbol{q}_t$ is an (4×1) unobservable zero mean white noise vector. The observed variables are given by

$$\boldsymbol{Z}_t = \boldsymbol{z}_t + \boldsymbol{e}_t, \;\; t = 1, \cdots, n \tag{13}$$

where $\boldsymbol{Z}_t = (Z_{1t}, Z_{2t}, Z_{3t}, Z_{4t})^\top$ and $\boldsymbol{e}_t = (e_{1t}, e_{2t}, e_{3t}, , e_{4t})^\top$ is the measurement error vector with $\mathrm{Var}(\boldsymbol{e}_t) = 0.571^2 \boldsymbol{I}_4$.

The time series plots corresponding to the respective observed BOLD signal at each brain region are represented in Figure 3. Since we are interested in identifying the links of connectivity networks using Granger causality, the statistical inferences are related to the parameters b_{ij} $(i, j = 1, 2, 3, 4)$. If $b_{ij} \neq 0$, then there is a information flow from brain area j to area i (Baccala and Sameshima (2001)). The coefficient estimates, standard errors and p-values ($H_0 : b_{ij} = 0$ vs $H_1 : b_{ij} \neq 0$) for both usual and proposed approaches are shown in Tables 5 and 6, respectively.

[[Figure 2]]

[[Figure 3]]

[[Figure 4]]

[[Figure 5]]

[[Table 5]]

[[Table 6]]

The estimate of $\boldsymbol{\Sigma}$ is

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.81 & 0.16 & 0.18 & 0.04 \\ 0.16 & 0.76 & 0.05 & -0.05 \\ 0.18 & 0.05 & 0.95 & 0.09 \\ 0.04 & -0.05 & 0.09 & 0.87 \end{pmatrix}.$$

The results described in Tables 5 and 6 suggest the existence of bidirectional information flow between Left M1 and Cerebellum. However, the application of usual approach indicates also that Left M1 sends information to SMA and Right M1, and that the latter sends to SMA. For both usual and proposed approaches, the diagrams of the networks at the significance level of 5% are shown in Figure 4. As highlighted by the simulations results, the utilization of usual VAR estimation, ignoring the measurement errors, may result in wrong test nominal sizes. In this context, it is important to mention that the main differences between the usual and proposal results were on standard deviation estimates. Further, the proposal estimates are almost twice the values resulting from usual approach. The theory and simulations suggest the existence of biases in the latter. Consequently, the p-values from the usual method tend to be underestimated, resulting in high rejection rates. Note that these connections may possibly exist, but since the nominal level of the test is "incorrect", the type I error is not under control. In addition, note that some coefficients were considerably underestimated, for example b_{11}, b_{22} and b_{33}. See, the qq-plots represented in Figure 5, which suggest that the probability density of residuals $\boldsymbol{Z}_t - \hat{\boldsymbol{Z}}_t$ are reasonably approximated by the Normal distribution.

In what follows we compare the results of Tables 5 and 6 with Table 4. Note that, for the real data, at a 5% nominal level, the proposed approach does not detect difference from zero for the following coefficients b_{12}, b_{13}, b_{21}, b_{23}, b_{24}, b_{31}, b_{32}, b_{34}, b_{42} and b_{43}. In contrast, the usual approach does not detect such differences only for the coefficients b_{12}, b_{13}, b_{24}, b_{32}, b_{34}, b_{42} and b_{43}. That is, the results agree for these coefficientes, however for b_{21} (Left M1 $\to$ SMA, p-value for the usual and proposed methods are 0.008 and 0.332, respectively), b_{23} (Right M1 $\to$ SMA, p-value for the usual and proposed methods are 0.002 and 0.053, respectively), b_{31} (Left M1 $\to$ Right M1, p-value for the usual and proposed methods are 0.030 and 0.073, respectively) they do not coincide. Futhermore, the hypothesis $b_{21} = 0$ presents the greatest difference between the p-values, which keeps different conclusions even if we set a 10% nominal level. While, for the hypotheses $b_{23} = 0$ and $b_{31} = 0$ the conclusions become the same at a 10% nominal level. Thus, looking at the results of Table 4 we can find a possible explanation for this fact. Notice that, for the usual approach and $n = 200$, the empirical false positive rates under the hypothesis $b_{21} = 0$ is 7.55% (the proposed approach is 4.75%); under the hypothesis $b_{23} = 0$ is 5.72% (the proposed approach is 4.75%) and under the hypothesis $b_{31} = 0$ is 5.73% (the proposed approach is 5.25%). As can be seen, the usual method is rejecting more than the proposed one for the hypothesis $b_{21} = 0$, whereas for the hypotheses $b_{23} = 0$ and $b_{32} = 0$ the usual method is still rejecting more than the proposed one, but a little less pronouced. The same behavior can be seen in the application.

Some studies (Biswal et al (1995)) suggest the existence of functional networks between motor areas even in resting state condition. These studies are based on correlation analysis between the BOLD signal at different brain sites. First, it is important to note that Granger causality is conceptually different from correlation, which is symmetric (it does not provide the direction of information flow), evaluated in a pairwise fashion (and not in the full multivariate sense) and it does not take into account temporal information. In fact, correlation analysis is more closely related to instantaneous Granger Causality concept, which can be useful to quantify simultaneity between time series but it is unsuitable in the context of information flow detection. Second, the usual correlation analysis does not consider the presence of measurement errors, which may also affect the statistical significance of the results. The nature of functional networks in resting state is still unclear and is the subject of several studies (Long et al. (2008)). Nevertheless, we have demonstrated in this study that the inclusion of measurement errors can considerably influence the final results. Thus, the development of novel approaches dealing with this artifact is necessary.

In summary, since the proposal and usual results differ, we conclude that the presence of measurement error cannot be ignored. An adequate treatment for this artifact is essential for the adequate description and modeling of brain networks. It is surprising that this important limitation received proper attention only recently. We believe that a preliminary analysis of this problem points toward the demand for the development of new estimation procedures regarding scanner noise characterization, physiological noise and computational implementation.

5 CONCLUSION

This paper has introduced a new approach to model multivariate times series when measurement errors are present. The simulation studies indicate that the proposed approach provides coherent results (test size close to the nominal level even for small samples, power increasing with the sample size under alternative hypotheses, biases and mean square errors decreasing when the sample size increases) under small and moderate measurement error. Such features seem no to be shared by the conventional maximum likelihood estimators which present a much inferior performance. Furthermore, the proposal is easily attained and iterative procedures are not required. The theory, simulations and application showed that the presence of measurement error cannot be neglected and a proper model has to be considered for an adequate description and modeling of brain networks. We expect to report extensions of the proposed model (for elliptical errors, heteroscedasticity situations, also trying to incorporate the variability of the measurement error variance estimation in the asymptotics), a residual study and more simulation studies for large measurement errors on incoming papers.

Acknowledgments

We gratefully acknowledge grants from FAPESP (Brazil). The authors are also grateful to the Editor-in-Chief Professor G.J. Babu, an associate editor and two anonymous referees for helpful comments and suggestions.

A PROOF OF THEOREMS

A.1 Proof of Theorem 1

In order to prove the consistence of the estimators stated in Theorem 1, namely

$$\widehat{\boldsymbol{a}} = \bar{\boldsymbol{Z}}_t - \widehat{\boldsymbol{B}}\bar{\boldsymbol{Z}}^*{}_{t-1}, \qquad \widehat{\boldsymbol{B}} = \left[(\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} - \boldsymbol{I}_r \otimes \Sigma_e)^{-1}\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t}\right]^\top$$

and

$$\widehat{\boldsymbol{\Sigma}} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}_i - \widehat{\boldsymbol{a}} - \widehat{\boldsymbol{B}}\boldsymbol{Z}^*_{i-1})(\boldsymbol{Z}_i - \widehat{\boldsymbol{a}} - \widehat{\boldsymbol{B}}\boldsymbol{Z}^*_{i-1})^\top - \boldsymbol{\Sigma}_e - \widehat{\boldsymbol{B}}(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\widehat{\boldsymbol{B}}^\top,$$

we must study the limits of the quantities $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}}$, $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t}$, $\bar{\boldsymbol{Z}}^*{}_{t-1}$ and $\bar{\boldsymbol{Z}}^*{}_t$ when the sample size goes to infinity. Note that $\boldsymbol{Z}^*_{t-1} = \boldsymbol{z}^*_{t-1} + \boldsymbol{e}^*_{t-1}$, where $\boldsymbol{e}^*_{t-1} =$

$(\boldsymbol{e}_{t-1}^\top, \ldots, \boldsymbol{e}_{t-r}^\top)^\top$, and under the stationary conditions of a VAR(r) model we have that

$$\begin{aligned}
\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} &= n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}^*_{i-1} - \bar{\boldsymbol{Z}}^*{}_{t-1})\boldsymbol{Z}^{*\top}_{i-1} \\
&= n^{-1}\sum_{i=1}^{n}(\boldsymbol{z}^*_{i-1} + \boldsymbol{e}^*_{i-1} - \bar{\boldsymbol{z}}^*{}_{t-1} - \bar{\boldsymbol{e}}^*{}_{t-1})(\boldsymbol{z}^*_{i-1} + \boldsymbol{e}^*_{i-1})^\top \\
&= \boldsymbol{S}_{\boldsymbol{z}^*_{t-1}} + \boldsymbol{S}_{\boldsymbol{e}^*_{t-1}} + O_p(n^{-1/2}) \\
&= \boldsymbol{\Gamma}_r(0) + \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e + O_p(n^{-1/2}),
\end{aligned}$$

where $\boldsymbol{S}_{\boldsymbol{e}^*_{t-1}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{e}^*_{i-1}\boldsymbol{e}^{*\top}_{i-1}$, and $O_p(n^{-1/2})$ means limited in probability even multiplying by $n^{1/2}$ (it happens with the crossing product in the above expression). That is, $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}_r(0) + \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e$. Following the same scheme, we have that

$$\begin{aligned}
\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t} &= n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}^*_{i-1} - \bar{\boldsymbol{Z}}^*{}_{t-1})\boldsymbol{Z}_i^\top \\
&= n^{-1}\sum_{i=1}^{n}(\boldsymbol{z}^*_{i-1} + \boldsymbol{e}^*_{i-1} - \bar{\boldsymbol{z}}^*{}_{t-1} - \bar{\boldsymbol{e}}^*{}_{t-1})(\boldsymbol{z}_i + \boldsymbol{e}_i)^\top \\
&= \boldsymbol{S}_{\boldsymbol{z}^*_{t-1}\boldsymbol{z}_t} + O_p(n^{-1/2}) \\
&= \boldsymbol{\Gamma}_r(0)\boldsymbol{B}^\top + O_p(n^{-1/2}),
\end{aligned}$$

and finally, both the quantities $\bar{\boldsymbol{Z}}^*{}_{t-1}$ and $\bar{\boldsymbol{Z}}^*{}_t$ converge in probability to $\boldsymbol{\mu}^*$. Hence,

$$(\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} - \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)^{-1} \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}_r(0)^{-1} \qquad \text{and} \qquad \boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\boldsymbol{Z}_t} \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}_r(0)\boldsymbol{B}^\top,$$

thus, the probability convergence of $\widehat{\boldsymbol{a}}$, $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\Sigma}}$ to $\boldsymbol{a}$, $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ follow, respectively.

A.2 Proof of Theorem 2

The proof idea has three steps. The first step consists in showing that $\mathrm{vec}(\widehat{\boldsymbol{B}}^\top) - \mathrm{vec}(\boldsymbol{B}^\top)$ can be written as linear combinations of a vectorial mean. The second one, we must demonstrate that this vectorial mean has an asymptotic Normal distribution. The last step must conclude that $\mathrm{vec}(\widehat{\boldsymbol{B}}^\top) - \mathrm{vec}(\boldsymbol{B}^\top)$ also has an asymptotic Normal distribution. In order to prove Theorem 2, we need some auxiliary results, which are exposed in two propositions below.

Proposition 1. *Under the model (1) and (4), the proposed estimator $\widehat{\boldsymbol{B}}$ has the following relationship*

$$vec(\widehat{\boldsymbol{B}}^\top) - vec(\boldsymbol{B}^\top) = (\boldsymbol{I}_p \otimes \boldsymbol{\Gamma}_r(0)^{-1})\bar{\boldsymbol{W}} + O_p(n^{-1}),$$

where

$$\bar{\boldsymbol{W}} = n^{-1}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{W}_{1i}\\ \vdots \\ \boldsymbol{W}_{qi}\end{pmatrix} = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i$$

with $\boldsymbol{W}_i = (\boldsymbol{q}_i + \boldsymbol{e}_i - \boldsymbol{B}\boldsymbol{e}^*_{i-1}) \otimes (\boldsymbol{z}^*_{i-1} - \boldsymbol{\mu}^* + \boldsymbol{e}^*_{i-1}) - \boldsymbol{\Psi}$ *and* $\boldsymbol{\Psi} = [\boldsymbol{I}_p \otimes (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)]vec(\boldsymbol{B}^\top)$.

Proof: Define $\boldsymbol{B}_{.k}$ as a vector $(rp \times 1)$ of coefficients associated with the k^{th} element of the vector $\boldsymbol{z}_t$, that is

$$z_{kt} = a_k + \boldsymbol{B}^\top_{.k}\boldsymbol{z}^*_{t-1} + q_{kt}.$$

Thus, we have that $\text{vec}(\boldsymbol{B}^\top) = (\boldsymbol{B}^\top_{.1}, \boldsymbol{B}^\top_{.2}, \cdots, \boldsymbol{B}^\top_{.p})^\top$ and the estimator of Theorem 1 for it can be written as $\text{vec}(\widehat{\boldsymbol{B}}) = (\widehat{\boldsymbol{B}}^\top_{.1}, \widehat{\boldsymbol{B}}^\top_{.2}, \cdots, \widehat{\boldsymbol{B}}^\top_{.p})^\top$, where $\widehat{\boldsymbol{B}}_{.k} = (\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} - \boldsymbol{I} \otimes \Sigma_e)^{-1}\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}Z_{kt}}$ and $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}Z_{kt}} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}^*_{i-1} - \bar{\boldsymbol{Z}}^*{}_{t-1})Z_{kt}$ for $k = 1, \ldots, p$. Moreover, the model (2) may be rewritten in terms of the observed variables as

$$\begin{aligned}\boldsymbol{Z}_t &= \boldsymbol{a} + \boldsymbol{B}\boldsymbol{Z}^*_{t-1} + \boldsymbol{\vartheta}_t,\\ \boldsymbol{\vartheta}_t &= \boldsymbol{q}_t + \boldsymbol{e}_t - \boldsymbol{B}\boldsymbol{e}^*_{t-1},\end{aligned} \tag{14}$$

and for the k^{th} element of $\boldsymbol{Z}_t$ we have

$$\begin{aligned}Z_{kt} &= a_k + \boldsymbol{B}^\top_{.k}\boldsymbol{Z}^*_{t-1} + \vartheta_{kt},\\ \vartheta_{kt} &= q_{kt} + e_{kt} - \boldsymbol{B}^\top_{.k}\boldsymbol{e}^*_{t-1}.\end{aligned} \tag{15}$$

Then, it follows that

$$\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}Z_k} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}^*_{i-1} - \bar{\boldsymbol{Z}}^*{}_{t-1})(a_k + \boldsymbol{B}^\top_{.k}\boldsymbol{Z}^*_{i-1} + \vartheta_{ki}) = \boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}}\boldsymbol{B}_{.k} + \boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\vartheta_k},$$

where $\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}\vartheta_k} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}^*_{i-1} - \bar{\boldsymbol{Z}}^*{}_{t-1})\vartheta_{ki} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{z}^*_{i-1} - \boldsymbol{\mu}^* + \boldsymbol{e}^*_{i-1})\vartheta_{ki} + O_p(n^{-1})$. Hence, denoting $\boldsymbol{S}_{\boldsymbol{z}^*_{t-1}\vartheta_k} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{z}^*_{i-1} - \boldsymbol{\mu}^* + \boldsymbol{e}^*_{i-1})\vartheta_{ki}$ we have that

$$\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}Z_k} = (\boldsymbol{S}_{\boldsymbol{Z}^*_{t-1}} - \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\boldsymbol{B}_{.k} + \boldsymbol{S}_{\boldsymbol{z}^*_{t-1}\vartheta_k} - \boldsymbol{\Psi}_k + O_p(n^{-1}),$$

with $\boldsymbol{\Psi}_k = -(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_e)\boldsymbol{B}_{.k}$. As a result, we have

$$\widehat{\boldsymbol{B}}_{.k} = \boldsymbol{B}_{.k} + \boldsymbol{\Gamma}^{-1}_r(0)\bar{\boldsymbol{W}}_k + O_p(n^{-1})$$

where $\bar{\boldsymbol{W}}_k = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_{ki}$ and $\boldsymbol{W}_{ki} = (\boldsymbol{z}^*_{i-1} - \boldsymbol{\mu}^* + \boldsymbol{e}^*_{i-1})\vartheta_{ki} - \boldsymbol{\Psi}_k$. Hence, it follows that

$$\text{vec}(\widehat{\boldsymbol{B}}^\top) - \text{vec}(\boldsymbol{B}^\top) = (\boldsymbol{I}_p \otimes \boldsymbol{\Gamma}_r(0)^{-1})\bar{\boldsymbol{W}} + O_p(n^{-1}),$$

where

$$\bar{\boldsymbol{W}} = n^{-1}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{W}_{1i}\\ \vdots \\ \boldsymbol{W}_{qi}\end{pmatrix} = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i$$

with $\boldsymbol{W}_i = (\boldsymbol{q}_i+\boldsymbol{e}_i-\boldsymbol{B}\boldsymbol{e}^*_{i-1})\otimes(\boldsymbol{z}^*_{i-1}-\boldsymbol{\mu}^*+\boldsymbol{e}^*_{i-1})-\boldsymbol{\Psi}$ and $\boldsymbol{\Psi} = [\boldsymbol{I}_p\otimes(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_e)]\text{vec}(\boldsymbol{B}^\top)$.

Proposition 2. *If $\boldsymbol{e}_t \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}_e)$ with $\boldsymbol{\Sigma}_e$ known and $E(q_{ij_1}q_{ij_2}q_{ij_3}q_{ij_4}) < \infty$ for all $j_1, j_2, j_3, j_4 \in \{1,\dots,p\}$, where q_{ij} is the j^{th} element of $\boldsymbol{q}_i$. The mean, $\bar{\boldsymbol{W}}$, of Proposition 1 has an asymptotic distribution given by*

$$\sqrt{n}\bar{\boldsymbol{W}} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0},\boldsymbol{T}_r),$$

where

$$\begin{aligned}\boldsymbol{T}_r \;=\;& \boldsymbol{\Sigma}_\vartheta\otimes\boldsymbol{\Gamma}_r(0) + \boldsymbol{\Sigma}_\vartheta\otimes(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_e) + \boldsymbol{B}^\top\otimes[\boldsymbol{\Sigma}_e\boldsymbol{B}(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_e)] + \\ & -\sum_{h=1}^{r}\Big\{(\boldsymbol{B}_h\boldsymbol{\Sigma}_e)\otimes\boldsymbol{\Gamma}_r(h) + (\boldsymbol{\Sigma}_e\boldsymbol{B}_h^\top)\otimes\boldsymbol{\Gamma}_r(-h)\Big\} + \\ & + \sum_{h=1-r}^{r-1}[\boldsymbol{B}(\boldsymbol{J}_{-h}\otimes\boldsymbol{\Sigma}_e)\boldsymbol{B}^\top]\otimes\boldsymbol{\Gamma}_r(h).\end{aligned}$$

where $\boldsymbol{J}_l$ is a $(r\times r)$ matrix of zeros with one's in the $|l|^{th}$ diagonal above (below) the main diagonal if $l > 0$ $(l < 0)$ and $\boldsymbol{J}_0$ is a $(r\times r)$ matrix of zeros.

Proof: Notice that the expectation of $\boldsymbol{W}_i$ is equal to zero for all i. Shumway and Stoffer (2000) state a central limit theorem to a univariate M-dependent sequence of random variables with mean zero. We say that a time series x_t is M-dependent if the set of values $x_s, s \le t$ is independent of the set of values $x_s, s \ge t+M+1$ (Shumway and Stoffer, 2000, on pg. 66). Then, assuming that $\text{E}(q_{ij_1}q_{ij_2}q_{ij_3}q_{ij_4}) < \infty$ for all $j_1, j_2, j_3, j_4 \in \{1,\dots,p\}$ where q_{ij} is the j^{th} element of $\boldsymbol{q}_i$ and defining $\bar{x} = n^{-1}\sum_{i=1}^{n}x_i$, where $x_i = \boldsymbol{\delta}^\top\boldsymbol{W}_i$ we have that $\text{E}(x_i) = 0$, $\text{Cov}(x_i, x_{i-h}) = \boldsymbol{\delta}^\top\text{Cov}(\boldsymbol{W}_i,\boldsymbol{W}_{i-h}^\top)\boldsymbol{\delta} = \boldsymbol{\delta}^\top E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top)\boldsymbol{\delta}$ and

$$\begin{aligned}E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top) \;=\;& E[\boldsymbol{F}_{ih}\otimes(\boldsymbol{z}^*_{i-1}-\boldsymbol{\mu}^*)(\boldsymbol{z}^*_{i-h-1}-\boldsymbol{\mu}^*)^\top] + E[\boldsymbol{F}_{ih}\otimes\boldsymbol{e}^*_{i-1}\boldsymbol{e}^{*\top}_{i-h-1}] + \\ +\;& E[\boldsymbol{F}_{ih}\otimes\boldsymbol{e}^*_{i-1}(\boldsymbol{z}^*_{i-h-1}-\boldsymbol{\mu}^*)^\top] + E[\boldsymbol{F}_{ih}\otimes(\boldsymbol{z}^*_{i-1}-\boldsymbol{\mu}^*)\boldsymbol{e}^{*\top}_{i-h-1}] - \\ -\;& \boldsymbol{\Psi}\boldsymbol{\Psi}^\top\end{aligned}$$

with $\boldsymbol{F}_{ih} = (\boldsymbol{q}_i+\boldsymbol{e}_i-\boldsymbol{B}\boldsymbol{e}^*_{i-1})(\boldsymbol{q}_{i-h}+\boldsymbol{e}_{i-h}-\boldsymbol{B}\boldsymbol{e}^*_{i-h-1})^\top$. Thus, using some matricial results and simple expectation rules we can solve these expectations as follows

$$E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top) = \boldsymbol{0} \qquad \text{for} \quad |h| < r,$$

$$E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top) = -(\boldsymbol{B}_r\boldsymbol{\Sigma}_e)\otimes\boldsymbol{\Gamma}_r(h) \qquad \text{for} \quad h=r,$$

$$E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top) = -(\boldsymbol{\Sigma}_e\boldsymbol{B}_{|r|}^\top)\otimes\boldsymbol{\Gamma}_r(h) \qquad \text{for} \quad h=-r,$$

$$E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top) = [\boldsymbol{B}(\boldsymbol{J}_{-h}\otimes\boldsymbol{\Sigma}_e)\boldsymbol{B}^\top]\otimes\boldsymbol{\Gamma}_r(h)-(\boldsymbol{B}_h\boldsymbol{\Sigma}_e)\otimes\boldsymbol{\Gamma}_r(h) \qquad \text{for} \quad h=1,\dots,r-1,$$

$$E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top) = [\boldsymbol{B}(\boldsymbol{J}_{-h}\otimes\boldsymbol{\Sigma}_e)\boldsymbol{B}^\top]\otimes\boldsymbol{\Gamma}_r(h)-(\boldsymbol{\Sigma}_e\boldsymbol{B}_{|h|}^\top)\otimes\boldsymbol{\Gamma}_r(h) \qquad \text{for} \quad h=-1,\dots,1-r,$$

$$E(\boldsymbol{W}_i\boldsymbol{W}_{i-h}^\top) = \boldsymbol{\Sigma}_\vartheta\otimes\boldsymbol{\Gamma}_r(0)+\boldsymbol{\Sigma}_\vartheta\otimes(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_e)+\boldsymbol{B}^\top\otimes[\boldsymbol{\Sigma}_e\boldsymbol{B}(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_e)] \qquad \text{for} \quad h=0,$$

where $\boldsymbol{J}_l$ is a $(r\times r)$ matrix of zeros with one's in the $|l|^{th}$ diagonal above (below) the main diagonal if $l>0$ ($l<0$) and $\boldsymbol{J}_0$ is a $(r\times r)$ matrix of zeros. That is, $x_1\dots,x_n$ is a strictly M-dependent sequence of random variables with mean zero (where $M=r$) and, therefore, we can use the result stated in Shumway and Stoffer (2000), which says that

$$\sqrt{n}\bar{x}\xrightarrow{\mathcal{D}}\mathcal{N}(0,V_r)$$

where

$$V_r=\sum_{h=-r}^{r}\text{Cov}(\boldsymbol{\delta}^\top\boldsymbol{W}_i,\boldsymbol{\delta}^\top\boldsymbol{W}_{i-h})=\boldsymbol{\delta}^\top\boldsymbol{T}_r\boldsymbol{\delta}$$

with

$$\begin{aligned}\boldsymbol{T}_r \;=\;& \boldsymbol{\Sigma}_\vartheta\otimes\boldsymbol{\Gamma}_r(0)+\boldsymbol{\Sigma}_\vartheta\otimes(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_e)+\boldsymbol{B}^\top\otimes[\boldsymbol{\Sigma}_e\boldsymbol{B}(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_e)]+\\ & -\sum_{h=1}^{r}\left\{(\boldsymbol{B}_h\boldsymbol{\Sigma}_e)\otimes\boldsymbol{\Gamma}_r(h)+(\boldsymbol{\Sigma}_e\boldsymbol{B}_h^\top)\otimes\boldsymbol{\Gamma}_r(-h)\right\}+\\ & +\sum_{h=1-r}^{r-1}[\boldsymbol{B}(\boldsymbol{J}_{-h}\otimes\boldsymbol{\Sigma}_e)\boldsymbol{B}^\top]\otimes\boldsymbol{\Gamma}_r(h).\end{aligned}$$

As $\sqrt{n}\boldsymbol{\delta}^\top\bar{\boldsymbol{W}}$ is asymptotically normally distributed for all $\boldsymbol{\delta}\neq\boldsymbol{0}_r$ then, by the Cramer-Wold device (see Theorem 10.4.5 on page 336 in Athreya and Lahiri, 2006), we have that

$$\sqrt{n}\bar{\boldsymbol{W}}\xrightarrow{\mathcal{D}}\mathcal{N}(\boldsymbol{0},\boldsymbol{T}_r).$$

Then, by the Propositions 1 and 2, the prove of Theorem 2 follows

$$\sqrt{n}(\text{vec}(\widehat{\boldsymbol{B}}^\top)-\text{vec}(\boldsymbol{B}^\top))\xrightarrow{\mathcal{D}}\mathcal{N}(\boldsymbol{0},[\boldsymbol{I}_p\otimes\boldsymbol{\Gamma}_r(0)^{-1}]\boldsymbol{T}_r[\boldsymbol{I}_p\otimes\boldsymbol{\Gamma}_r(0)^{-1}]).$$

References

B. Abler, A. Roebroeck, R. Goebel, A. Hose, C. Schonfeldt-Lecuona, G. Hole, H. Walter, Investigating directed influences between activated brain areas in a motor-response task using fMRI, *Magnetic Resonance Imaging*, **24** (2006) 181–185.

D.J. Aigner, C. Hsiao, A. Kapteyn, T. Wansbeek, Latent Variables in Econometric Time-Series. In Handbook of Econometrics, Z. Griliches and M. Intriligator (eds.), Amsterdam: North-Holland, 1984.

M.G. Akritas, M.A. Bershady, Linear regression for astronomical data with measurement errors and intrinsic scatter, *The Astrophysical Journal*, **470** (1996) 706–714.

J. Andersson, Testing for Granger causality in the presence of measurement errors, *Economics Bulletin*, **3** (2005) 1–13.

K.B. Athreya, S.N. Lahiri, *Measure Theory and Probability Theory*, Springer, 2006.

L.A. Baccala, K. Sameshima, Partial directed coherence: a new concept in neural structure determination, *Biological Cybernetics*, **84** (2001) 463–474.

N. Beck, Estimating Dynamic Models Using Kalman Filtering, *Political Analysis*, **1** (1990) 121–156.

B. Biswal, F.Z. Yetkin, V.M. Haughton, J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI, *Magnetic Resonance in Medicine*, **34** (1995) 537–541.

C.L. Cheng, J. Riu, On estimating linear relationships when both variables are subject to heteroscedastic measurement errors, *Technometrics*, **48** (2006) 511–519.

J.D. Cohen, F. Tong, The face of controversy, *Science*, **293** (2001) 2405–2407.

A. Fujita, J.R. Sato, H.M. Garay-Malpartida, R. Yamaguchi, S. Miyano, M.C. Sogayar, C.E. Ferreira, Modeling gene expression regulatory networks with the sparse vector autoregressive model, *BMC Systems Biology*, **30** (2007a) 1-39.

A. Fujita, J.R. Sato, H.M. Garay-Malpartida, P.A. Morettin, M.C. Sogayar, C.E. Ferreira, Time-varying modeling of gene expression regulatory networks using the wavelet dynamic vector autoregressive method, *Bioinformatics*, **23** (2007b) 1623–1630.

W. Fuller, *Measurement Error Models*, Wiley: Chichester, 1987.

J. Geweke, "The Dynamic Factor Analysis of Econometric Time-Series." In Latent Variables in Socio-Economic Modelse, Dennis J. Aigner and Arthur S. Goldberger (eds.), Amsterdam: North-Holland, 1977.

R. Goebel, A. Roebroeck, D.S. Kim, E. Formisano, Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping, *Magnetic Resonance Imaging*, **21** (2003) 1251–1261.

S. Gottesman, Bacterial regulation: global regulatory networks, *Annual Review of Genetics*, **18** (1984) 415–441.

C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37** (1969) 424–438.

K. Hasan, J. Hossain, A. HaqueGranger, Parameter estimation of multichannel autoregressive processes in noise, *Signal Processing*, **83** (2003) 603–610.

M. Katoh, Networking of WNT, FGF, Notch, BMP, and Hedgehog signaling pathways during carcinogenesis, *Stem Cell Review*, **3** (2007) 30–38.

P. Kellstedt, G.E. McAvoy, J.A. Stimson, Dynamic Analysis with Latent Constructs, *Political Analysis*, **5** (1996) 113–150.

B.C. Kelly, Some aspects of measurement error in linear regression of astronomical data, *The Astrophysical Journal*, **665** (2007) 1489–1506.

B.C. Kelly, J. Bechtold, J.R. Trump, M. Vertergaard, A. Siemiginowska, Observational constraints on the dependence of ratio-quiet quasar X-ray emission on black hole mass and accretion rate, *Astrophysical Journal Supplement Series*, **176** (2008) 355–373.

S.B. Kulathinal, K. Kuulasmaa, D. Gasbarra, Estimation of an errors-in-variables regression model when the variances of the measurement error vary between the observations, *Statistics in Medicine*, **21** (2002) 1089–1101.

E.L. Lehmann, G. Casella, *Theory of Point Estimation*, 2nd ed. Springer-Verlag: New York, 1998.

H. Liu, G. Rodríguez, Human activities and global warming: a cointegration analysis, *Environmental Modeling & Software*, **20** (2005) 761–773.

N.K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, Neurophysiological investigation of the basis of the fMRI signal, *Nature*, **412** (2001) 150–157.

X.Y. Long, X.N. Zuo, V. Kiviniemi, Y. Yang, Q.H. Zou, C.Z. Zhu, T.Z. Jiang, H. Yang, Q.Y. Gong, L. Wang, K.C. Li, S. Xie, Y.F. Zang, Default mode network as revealed with multiple methods for resting-state functional MRI analysis, *Journal Neuroscience Methods*, **171** (2008) 349–355.

H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin, 2005.

N.D. Mukhopadhyay, S. Chatterjee, Causality and pathway search in microarray time series experiment, *Bioinformatics*, **23** (2007) 442–449.

A. Maravall, D.J. Aigner, "Identification of the Dynamic Shock-Error Model: The Case of Dynamic Regression." In Latent Variables in Socio-Economic Models, Dennis J. Aigner and Arthur S. Goldberger (eds.), Amsterdam: North-Holland, 1977.

S. Ni, D. Sun, Noninformative priors and frequentist risks of Bayesian estimators of vector-autoregressive models, *Journal of Econometrics*, **115** (2003) 159–197.

A.G. Patriota, H. Bolfarine, M. de Castro, A heteroscedastic structural errors-in-variables model with equation error, *Statistical Methodology*, **6** (2009) 408–423.

R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org, 2008.

J.R. Sato, J.E. Amaro, D.Y. Takahashi, M. de Maria Felix, M.J. Brammer, P.A. Morettin, A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality, *Neuroimage*, **31** (2006) 187–196.

C.A. Sims, Macroeconomics and reality, *Econometrica*, **48** (1980) 1–48.

R.H. Shumway, D.S. Stoffer, *Time Series Analysis and Its Applications*, Springer-Verlag, New York, 2000.

J.T. Williams, M.D. McGinnis, The Dimension of Superpower Rivalry: A Dynamic Factor Analysis, *Journal of Conflict Resolution*, **36** (1992) 68–118.

Table 1: Rejection rates (%) of the hypothesis $H_0 : b_{12} = b_{21} = 0$ (at 5% nominal level) using the Wald statistics (8) for $n = 50$, $n = 100$, $n = 250$ and $n = 500$. The bold numbers at the center are test sizes (they are expected to be 5%) and the numbers around them are empirical powers.

		Corrected approach					Usual approach (OLS)				
		b_{12}					b_{12}				
		-0.4	-0.2	0.0	0.2	0.4	-0.4	-0.2	0.0	0.2	0.4
	$n = 50$										
	-0.4	88.79	43.95	21.55	44.37	83.03	83.17	31.40	19.27	52.19	88.55
	-0.2	82.95	27.38	7.75	22.52	61.89	75.11	17.19	10.37	39.04	77.59
b_{21}	0.0	82.31	21.89	**4.70**	17.44	54.01	75.81	15.33	**13.06**	46.38	81.33
	0.2	87.81	27.67	8.41	24.10	59.58	84.42	25.11	24.30	65.27	92.08
	0.4	93.98	42.31	16.19	34.09	67.33	93.36	43.33	38.86	81.89	97.09
	$n = 100$										
	-0.4	99.81	81.63	44.50	70.98	98.06	99.47	62.79	34.27	79.88	99.35
	-0.2	99.23	57.85	12.33	35.84	87.49	98.05	32.81	12.62	64.40	97.10
b_{21}	0.0	99.15	45.22	**5.03**	28.26	83.06	98.01	26.66	**19.03**	76.63	98.53
	0.2	99.50	49.03	11.12	42.27	88.25	99.03	41.35	42.17	93.19	99.83
	0.4	99.91	69.40	28.78	61.56	92.37	99.89	70.19	67.81	98.92	100.0
	$n = 250$										
	-0.4	100.0	99.75	85.97	97.86	100.0	100.0	96.74	71.29	99.41	100.0
	-0.2	100.0	95.11	26.42	69.13	99.79	100.0	71.89	21.48	95.93	99.99
b_{21}	0.0	100.0	85.09	**5.63**	59.33	99.69	100.0	58.01	**39.23**	99.21	100.0
	0.2	100.0	87.79	19.95	80.56	99.85	100.0	80.67	80.51	99.99	100.0
	0.4	100.0	97.21	62.91	94.57	99.91	100.0	97.65	97.66	100.0	100.0
	$n = 500$										
	-0.4	100.0	100.0	99.29	99.99	100.0	100.0	99.97	95.71	100.0	100.0
	-0.2	100.0	99.93	48.41	93.75	100.0	100.0	95.21	37.71	99.95	100.0
b_{21}	0.0	100.0	99.11	**5.34**	88.17	100.0	100.0	88.13	**67.50**	100.0	100.0
	0.2	100.0	99.45	36.79	98.01	100.0	100.0	98.17	98.19	100.0	100.0
	0.4	100.0	99.99	90.51	99.88	100.0	100.0	99.99	99.99	100.0	100.0

Table 2: Rejection rates (%) of the hypothesis $H_0 : b_{12} = 0$ (at 5% nominal level) using the Wald statistics (8) for $n = 50$, $n = 100$, $n = 250$ and $n = 500$.

	b_{12}				
Model	-0.4	-0.2	0.0	0.2	0.4
$n = 50$					
Proposed Model	42.65	13.51	**4.93**	7.35	14.10
Usual Model	36.54	9.79	**6.53**	17.78	34.56
$n = 100$					
Proposed Model	71.01	21.85	**5.41**	11.09	25.52
Usual Model	61.75	12.97	**8.51**	31.87	61.65
$n = 250$					
Proposed Model	97.57	43.83	**5.28**	21.50	55.67
Usual Model	94.40	22.07	**14.17**	69.65	95.53
$n = 500$					
Proposed Model	99.99	70.28	**5.37**	40.33	85.12
Usual Model	99.94	36.90	**24.39**	94.27	99.93

Table 3: Empirical bias and mean squared error for the proposed and usual model. Note that, the biases

	Proposed model		Usual model	
	Bias	MSE	Bias	MSE
$n = 50$				
b_{11}	-0.0034	1.4357	-0.1446	0.0454
b_{12}	-0.0713	1.7832	0.1098	0.0434
b_{21}	0.0323	0.1504	0.0250	0.0143
b_{22}	-0.0813	0.2005	-0.1589	0.0461
$n = 100$				
b_{11}	-0.0032	0.0218	-0.1313	0.0290
b_{12}	-0.0334	0.0352	0.1209	0.0293
b_{21}	0.0140	0.0120	0.0165	0.0067
b_{22}	-0.0364	0.0195	-0.1265	0.0258
$n = 250$				
b_{11}	-0.0004	0.0079	-0.1252	0.0203
b_{12}	-0.0143	0.0119	0.1299	0.0224
b_{21}	0.0039	0.0043	0.0112	0.0027
b_{22}	-0.0125	0.0066	-0.1086	0.0156
$n = 500$				
b_{11}	-0.0008	0.0040	-0.1235	0.0175
b_{12}	-0.0072	0.0058	0.1326	0.0203
b_{21}	0.0023	0.0021	0.0097	0.0013
b_{22}	-0.0068	0.0031	-0.1024	0.0124

Table 4: Rejection rates under null univariate hypothesis (at 5% nominal level). The model is generated considering $b_{12} = b_{13} = b_{21} = b_{23} = b_{24} = b_{31} = b_{32} = b_{34} = b_{42} = b_{43} = 0$ and the other parameters were taken similar to which estimated for the application. Each cell depicts the nominal level for univariatly testing if $b_{ij} = 0$. The closer to 5% the better is the result.

	Proposed model			Usual model		
	$n = 100$	$n = 200$	$n = 400$	$n = 100$	$n = 200$	$n = 400$
$H_0 : b_{12} = 0$	4.67	4.65	4.60	6.21	7.21	9.78
$H_0 : b_{13} = 0$	5.15	4.78	4.79	5.95	5.43	5.86
$H_0 : b_{21} = 0$	5.35	4.75	4.83	7.30	7.55	8.99
$H_0 : b_{23} = 0$	5.15	5.20	4.74	5.55	5.72	4.89
$H_0 : b_{24} = 0$	5.48	5.18	4.81	6.08	5.17	4.93
$H_0 : b_{31} = 0$	5.61	5.25	4.91	6.42	5.73	5.67
$H_0 : b_{32} = 0$	4.94	5.13	5.09	5.55	5.56	5.30
$H_0 : b_{34} = 0$	5.28	5.21	5.36	5.75	5.30	5.68
$H_0 : b_{42} = 0$	5.06	4.73	4.89	4.91	4.59	5.21
$H_0 : b_{43} = 0$	5.11	5.25	4.96	5.09	5.26	5.33

Table 5: **Application to real data - usual approach:** coefficient estimates, standard deviations and respective p-values (H_0 : coefficient is equal to zero).

Parameter	**Estimate**	**Standard Deviation**	**p−value**
b_{11}	0.537	0.065	<0.001
b_{12}	0.105	0.063	0.097
b_{13}	0.003	0.060	0.967
b_{14}	-0.181	0.059	0.002
b_{21}	0.179	0.068	0.008
b_{22}	0.378	0.066	<0.001
b_{23}	0.145	0.063	0.002
b_{24}	0.047	0.062	0.442
b_{31}	0.165	0.076	0.030
b_{32}	-0.074	0.074	0.319
b_{33}	0.242	0.071	<0.001
b_{34}	-0.061	0.069	0.378
b_{41}	0.294	0.070	<0.001
b_{42}	-0.060	0.068	0.381
b_{43}	0.092	0.065	0.154
b_{44}	0.350	0.064	<0.001

Table 6: **Application to real data - proposed approach:** coefficient estimates, standard deviations and respective p-values (H_0 : coefficient is equal to zero).

Parameter	**Estimate**	**Standard Deviation**	**p−value**
b_{11}	0.935	0.137	<0.001
b_{12}	-0.032	0.127	0.803
b_{13}	-0.095	0.103	0.357
b_{14}	-0.287	0.091	0.002
b_{21}	0.132	0.137	0.332
b_{22}	0.581	0.126	<0.001
b_{23}	0.199	0.103	0.053
b_{24}	0.027	0.092	0.765
b_{31}	0.279	0.156	0.073
b_{32}	-0.184	0.143	0.201
b_{33}	0.346	0.117	0.004
b_{34}	-0.111	0.106	0.294
b_{41}	0.538	0.147	<0.001
b_{42}	-0.252	0.135	0.063
b_{43}	0.044	0.110	0.687
b_{44}	0.528	0.099	<0.001

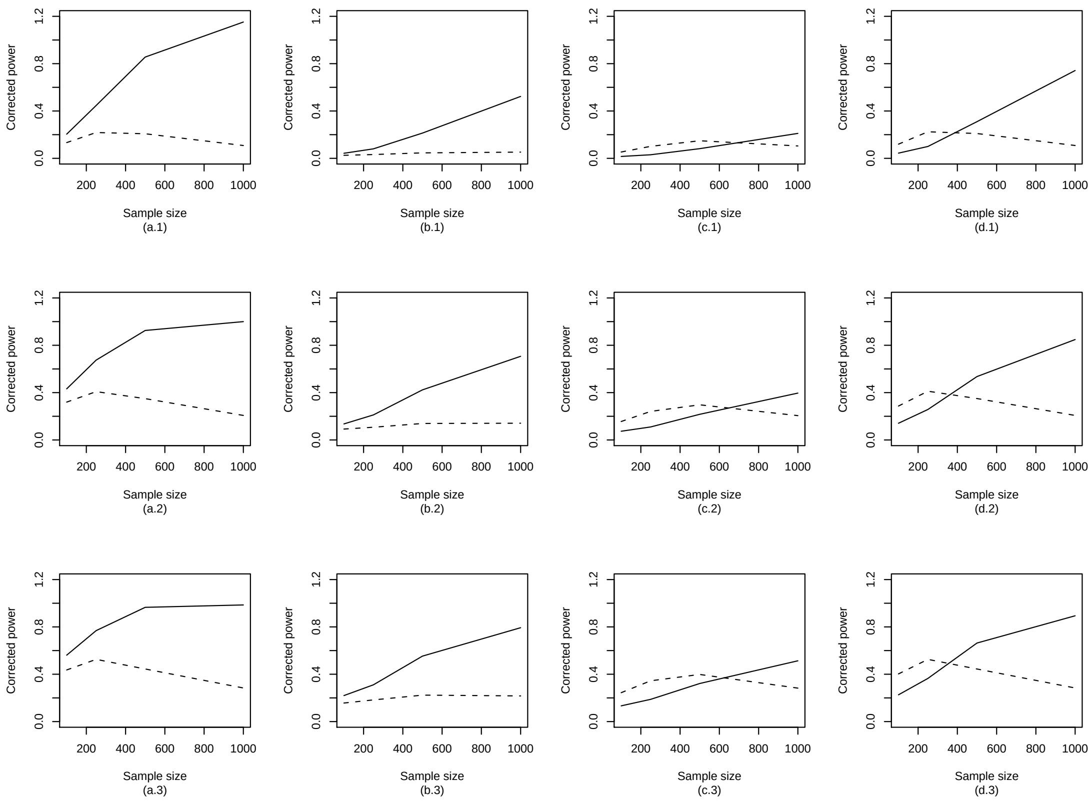

Figure 1: Corrected power *versus* sample size. The full line refers to the proposed approach and the dot line refers to the usual one. It is expected that the corrected power converges to one.

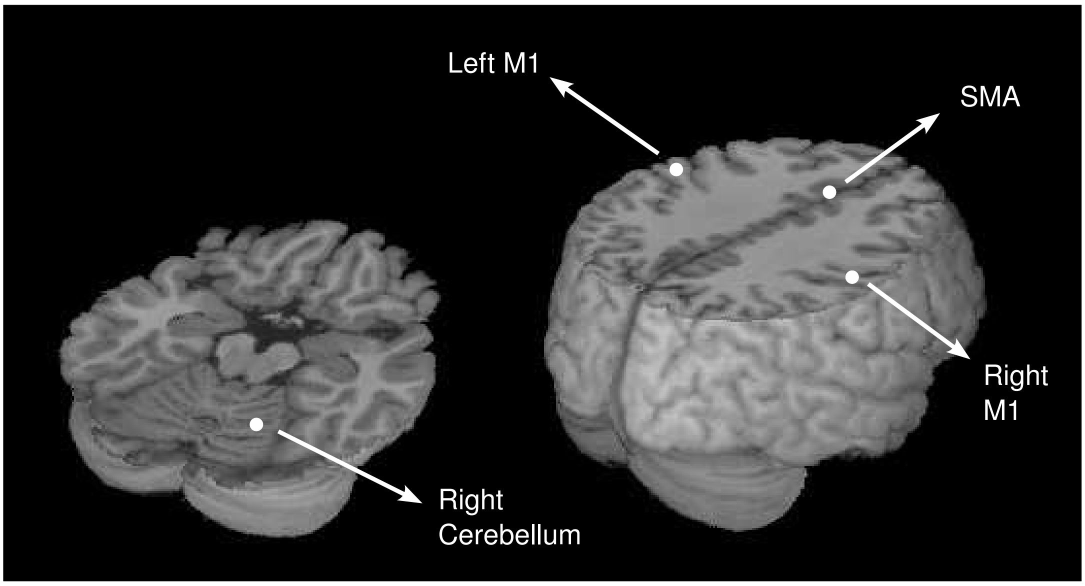

Figure 2: Four areas were selected for connectivity evaluation using the VAR model: **Left M1:** left primary motor cortex, **Right M1:** right primary motor cortex, **SMA:** supplementary motor area and **Right Cerebellum.**

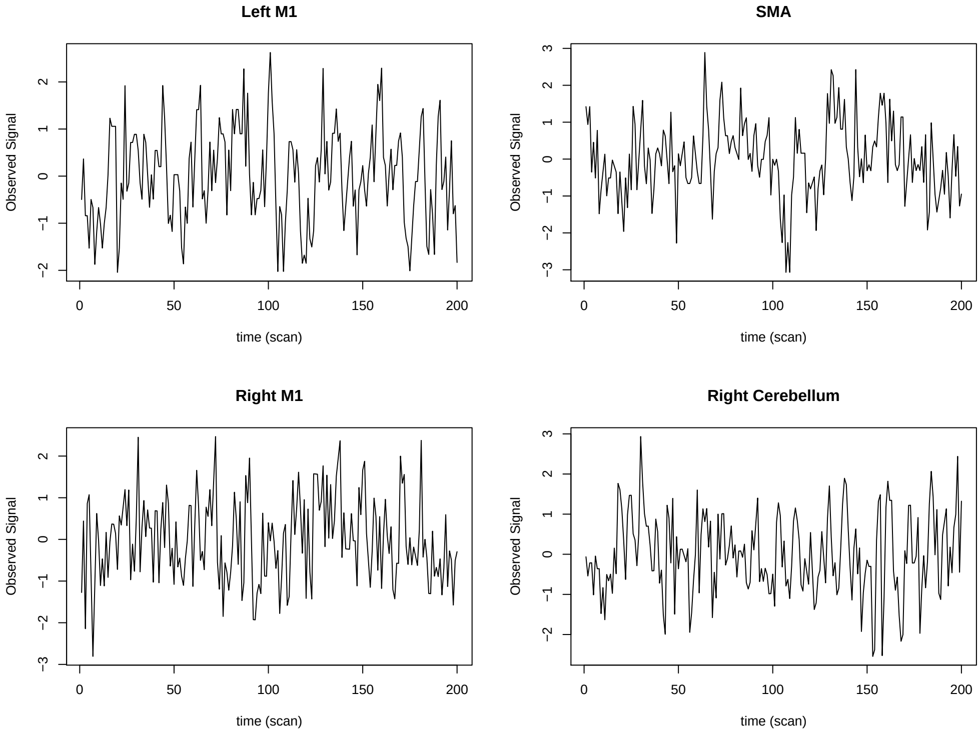

Figure 3: Observed signal at each brain region.

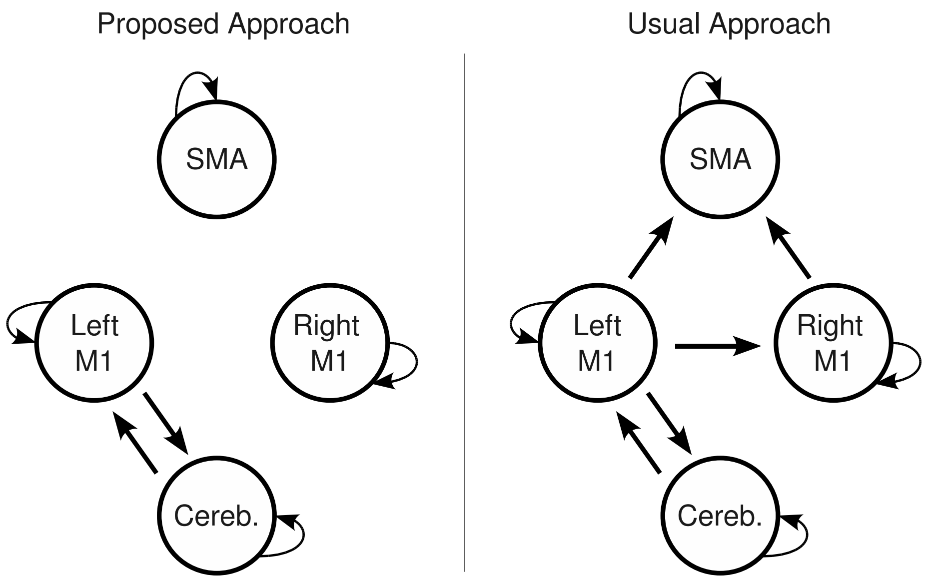

Figure 4: Identified network of information flow by testing the parameters of VAR model ($\alpha = 5\%$)

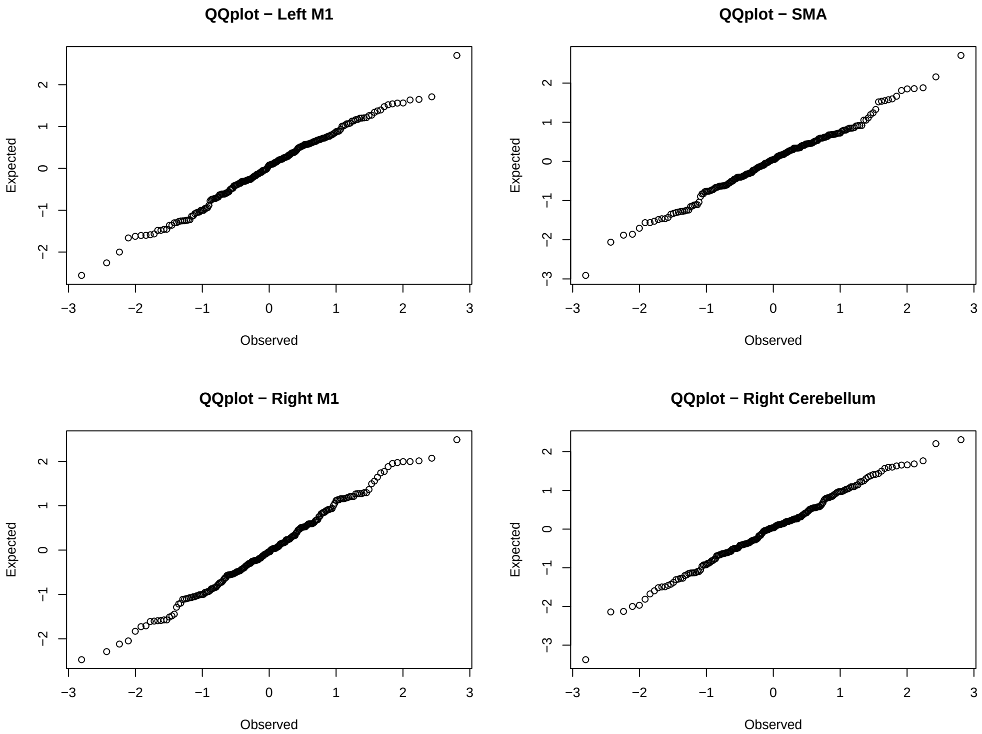

Figure 5: **QQplot for Normal distribution:** Residuals (Observed values - Predicted) at each brain region.