

# Wavelet Based Time-varying Vector Autoregressive Modelling

João R. Sato<sup>a,c</sup> Pedro A. Morettin<sup>a,\*</sup> Paula R. Arantes<sup>b,c</sup>  
Edson Amaro Jr.<sup>b,c</sup>

<sup>a</sup>*Institute of Mathematics and Statistics, University of São Paulo, Brazil*

<sup>b</sup>*Department of Radiology, Medical School, University of São Paulo, Brazil*

<sup>c</sup>*NIF: Functional Neuroimage Research, University of São Paulo, Brazil*

---

## Abstract

Vector autoregressive (VAR) modelling is one of the most popular approaches in multivariate time series analysis. The parameters interpretation is simple, and provide an intuitive identification of relationships and Granger causality among time series. However, the VAR modelling requires stationarity conditions which could not be valid in many practical applications. Locally stationary or time dependent modelling seem attractive generalizations, and several univariate approaches have already been proposed. In this paper we propose an estimation procedure for time-varying vector autoregressive processes, based on wavelet expansions of autoregressive coefficients. The asymptotic properties of the estimator are derived and illustrated by computer intensive simulations. We also present an application to brain connectivity identification using functional magnetic resonance imaging (fMRI) datasets.

*Key words:* Wavelets, time-varying, autoregressive, multivariate

---

## 1 Introduction

Since its description (Sims (1980)), vector autoregressive modelling has been successfully applied to the analysis of multivariate time series, focusing on the identification of complex relationships among several time series. The

---

\* Correspondent author: Pedro A. Morettin, Institute of Mathematics and Statistics, Rua do Matão, 1010. 05508-090 - São Paulo - Brazil.

*Email address:* pam@ime.usp.br (Pedro A. Morettin).

parameters of VAR models can be easily interpreted and provide a simple identification of Granger causality (Granger (1969)). Hence, due to its high flexibility and generality, several applications can be found in econometrics (Sims (1982, 1998); Stock and Watson (2001)), natural sciences (Lu et al. (2006)) and neuroscience (Goebel et al. (2003); Baccala and Sameshima (2001); Sato et al. (2006)).

Several studies in econometrics are founded in VAR modelling, which provide a quantitative analysis of relationships between financial assets. Some examples of monetary policies studies based on VAR models can be found in Lawrence et al. (1999); Sack (2000), and many other examples and applications of VAR in econometrics are presented in Enders (2005); Stock and Watson (2001). Considering applications in medical research, Harrison et al. (2003) and Goebel et al. (2003) introduced the VAR models in functional magnetic resonance imaging (fMRI) focusing the brain connectivity modelling. The concept of Granger causality in the context of brain signals may be interpreted as the direction of information flow. Baccala and Sameshima (2001) proposed a frequency domain version of Granger causality: the partial directed coherence measure (PDC), which is based on VAR models. They also showed the usefulness of partial directed coherence in the analysis of brain connectivity modelling using electroencephalography (EEG) data. More details about VAR properties, estimation and applications can be found in Lutkepohl (1993).

Nevertheless, the classical VAR modelling is adequate only in the analysis of stationary time series, and in many cases, stationarity assumptions are too restrictive or not reasonable. In these cases, time-varying parameters or adaptive modelling seem to be attractive alternatives. Time varying autoregressive (TVAR) models have been developed in several directions since the early 1980's, specially in the Bayesian framework. See Prado et al. (2001) for an excellent review. These models led for example to new methods of time series decomposition and analysis. West (1997); Huerta and West (1997) and Krystal et al. (1999) use these models in connection with EEG data. For further applications of TVAR models see Paulik et al. (1994) and Primiceri (2005).

Dahlhaus and co-workers, in a series of articles, developed a theory for locally stationary processes. Basically this is the same idea of Priestley's evolutionary processes, but where an asymptotic theory is feasible. Dahlhaus (Dahlhaus (1997); Dahlhaus et al. (1999); Dahlhaus (2000)) developed many estimators and asymptotic results, considering locally stationary processes and functions projection in adequate spaces.

In this paper, we propose a novel time-varying multivariate estimation, founded on locally stationary processes, vector autoregression and wavelet decompositions. The asymptotic properties of estimators and simulations are also presented. We also illustrate an application of the proposed approach to dynamic

brain connectivity modelling, considering functional magnetic resonance imaging datasets.

The plan of the article is as follows. In section 2 we introduce the basic background on wavelets and locally stationary processes. The proposed approach to estimate time-varying vector autoregressive processes is the subject of section 3. Some simulations are performed in section 4 and an application to brain connectivity modelling is given in section 5. We end the paper with some comments in section 6.

## 2 Background

### 2.1 Wavelets

Wavelet analysis is a very prominent and flexible tool nowadays. In addition, it is the focus of several mathematics, statistics (Vidakovic (1999)) and signal processing research (Percival and Walden (2000)) and many applications can be found in biology (Bullmore et al. (2003); Gackenheim (2006)), geophysics (Jevrejeva et al. (2003); Grinsted et al. (2004)) and image processing (Chambolle et al. (1998)). Similarly to trigonometric functions, wavelets generate basis of function spaces (Meyer (1993)), but wavelet decompositions are well localized in scale and time, allowing the analysis of non-stationary time series. Furthermore, in contrast to Fourier analysis, wavelets also provide high adaptativeness in non-homogeneous smoothness spaces.

The main characteristic of wavelet basis is that the space of functions is generated considering dilations and translations of a unique function  $\psi(t)$  (mother wavelet). Formally, the space is generated by functions of the class

$$\Psi = \left\{ \psi_{a,b}(x) = |a|^{-1/2} \psi \left( \frac{x-b}{a} \right), a, b \in \mathbb{R}, a \neq 0 \right\}, \quad (1)$$

and  $\psi$  must satisfy the following conditions:

- i)  $\int_{-\infty}^{\infty} \psi(t) dt = 0$ ;
- ii)  $\int_{-\infty}^{\infty} |\psi(t)| dt < \infty$ ;
- iii)  $\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\lambda)|}{|\lambda|} d\lambda < \infty$ , where  $\hat{\psi}(\lambda)$  is the Fourier transform of  $\psi(t)$ ;
- iv)  $\int_{-\infty}^{\infty} t^j \psi(t) dt = 0, j = 0, 1, 2, \dots, r-1$ , for at least a  $r \geq 1$  and  $\int_{-\infty}^{\infty} |t^r \psi(t)| dt < \infty$ .

An way to generate wavelets is considering an orthogonal function  $\phi$  (father wavelet), which is the solution of the equation

$$\phi(t) = \sqrt{2} \sum_k l_k \phi(2t - k), \quad (2)$$

where

$$l_k = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \phi(2t - k) dt. \quad (3)$$

Considering binary dilations and dyadic translations of  $\phi(t)$ , we obtain a space generated by the class of function

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k), \quad (4)$$

where  $j, k \in \mathbb{Z}$ . Then, we may obtain  $\psi(t)$  by filtering  $\phi(t)$  adequately, using the relationships

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k), \quad (5)$$

and

$$h_k = \sqrt{2} \int_{-\infty}^{\infty} \psi(t) \phi(2t - k) dt. \quad (6)$$

The quadrature mirror filter relation is given by  $h_k = (-1)^k l_{1-k}$ , and actually, the filters  $l_k$  and  $h_k$  are low-pass and high-pass filters, respectively. A useful result is that any function  $g(t)$  satisfying  $\int_{-\infty}^{\infty} g(t)^2 dt < \infty$  (Hardle et al. (1997)), can be expanded as

$$g(t) = \sum_k c_{j_0,k} \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_k d_{j,k} \psi_{j,k}(t), \quad (7)$$

for some coarse scale  $j_0$ , usually taken as zero. Here we are using the short notation “scale  $j$ ” for  $2^{j-1}$ .

Basically, the coefficient  $d_{j,k}$  are associated with changes of averages on scale  $j$  spaced  $2^j$  units apart, and  $c_{j_0,k}$  are smoothed averages over scale  $j_0$ . Increasing one unit in  $j$  will lead to a increase of  $2^{j+1} - 2^j$  on the number of resolution points. The coefficient  $k$  is the time location coefficient, indicating translations of the wavelet functions.

Wavelets commonly used belong to the wavelet class of Daubechies (Daubechies (1992)), which are obtained by choosing adequate parameters of the low-pass ( $l_k$ ) and high-pass filters ( $h_k$ ).

## 2.2 Locally Stationary Processes

Considering practical applications, the locally stationary concept is very useful, as it allows a formal characterization of time-varying parameters of stochastic processes. Here we follow Dahlhaus (1997).

**Definition 1** *A sequence of multivariate Gaussian stochastic processes  $x_{t,T} = (x_{t,T}^{(1)}, \dots, x_{t,T}^{(s)})'$  with  $t = 1, \dots, T$  is a locally stationary process with transfer covariance matrix  $\mathbf{A}_{t,T}^0(\lambda)$  and mean vector  $\mu(t/T)$  if there is a representation*

$$x_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) \mathbf{A}_{t,T}^0(\lambda) d\xi(\lambda), \quad (8)$$

which must satisfy the following properties:

i)  $\xi(\lambda)$  is a complex orthonormal process vector in  $[-\pi, \pi]$  with  $\overline{\xi_a(\lambda)} = \xi_a(-\lambda)$ ,  $E[\xi_a(\lambda)] = 0$  and

$$E[d\xi_a(\lambda)d\xi_b(\lambda)] = \delta_{ab}\eta(\lambda + \mu)d\lambda d\mu, \quad (9)$$

where  $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$  is the extension of Dirac delta function with period  $2\pi$ ;

ii) There is a constant  $K$  and a function matrix  $2\pi$  periodic  $\mathbf{A} : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}^{s \times s}$  with  $\overline{\mathbf{A}(u, \lambda)} = \mathbf{A}(u, -\lambda)$  and

$$\sup_{t, \lambda} |\mathbf{A}_{t,T}^0(\lambda)_{ab} - \mathbf{A}(t/T, \lambda)_{ab}| \leq KT^{-1} \quad (10)$$

$\forall a, b = 1, \dots, s$  and  $T \in \mathbb{N}$ .  $\mathbf{A}(u, \lambda)$  and  $\mu(u)$  are continuous functions in  $u$ .

The evolutionary spectral density matrix is defined as  $f(u, \lambda) := \mathbf{A}(u, \lambda) \overline{\mathbf{A}(u, \lambda)'}.$

A simple example of an univariate locally stationary process is a time-varying coefficients moving average process of order  $q$ , i.e.

$$x_{t,T} = u(t/T) + \sum_{j=1}^q a_j(t/T) \epsilon_{t-j,T} + \epsilon_{t,T}, \quad (11)$$

where  $\{\epsilon_{t,T}\}$  is a sequence of independent Gaussian i.i.d. random variables with mean zero and  $u(t)$  and  $a(t)$  are functions on  $[0, 1]$ . Furthermore, if a time-varying coefficients autoregressive model can be represented by

$$x_{t,T} = u(t/T) + \epsilon_{t,T} + \sum_{i=1}^{\infty} c_i(t/T)\epsilon_{t-i,T}, \quad (12)$$

assuming

$$\sup_t \sum_{i=1}^{\infty} |c_i(t/T)| < \infty. \quad (13)$$

then it is also a locally stationary process (Dahlhaus et al. (1999)). In this paper, time-varying multivariate autoregressive models which admit an infinite moving average representation are considered.

### 2.3 Time-varying Vector Autoregressive Processes

For some positive integer  $T$  (number of observations), Let  $x_t, T$  be a multivariate time series with dimension  $s$ ,  $x_{t,T} = (x_{1t,T}, x_{2t,T}, \dots, x_{st,T})'$ . The time-varying vector autoregressive model is defined by,

$$x_{t,T} = u(t/T) + \sum_{l=1}^p \mathbf{A}_l(t/T)x_{t-l,T} + \epsilon_{t,T}, \quad (14)$$

$$\text{Var}(\epsilon_{t,T}) = \boldsymbol{\Sigma}(t/T) = \begin{pmatrix} \sigma_1^2(t/T) & \sigma_{12}(t/T) & \dots & \sigma_{1s}(t/T) \\ \sigma_{21}(t/T) & \sigma_2^2(t/T) & \dots & \sigma_{2s}(t/T) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{s1}(t/T) & \sigma_{s2}(t/T) & \dots & \sigma_s^2(t/T) \end{pmatrix},$$

where  $\epsilon_{t,T}$  is the innovations vector (independent and with mean zero),  $u(t/T)$  is the vector of intercepts, and  $\mathbf{A}_l(t/T)$  ( $l = 1, 2, \dots, p$ ) are the autoregressive coefficients matrices, given by

$$u(t/T) = \begin{pmatrix} u_1(t/T) \\ u_2(t/T) \\ \vdots \\ u_s(t/T) \end{pmatrix}, \quad \mathbf{A}_l(t/T) = \begin{pmatrix} a_{11}^{(l)}(t/T) & a_{12}^{(l)}(t/T) & \dots & a_{1s}^{(l)}(t/T) \\ a_{21}^{(l)}(t/T) & a_{22}^{(l)}(t/T) & \dots & a_{2s}^{(l)}(t/T) \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1}^{(l)}(t/T) & a_{s2}^{(l)}(t/T) & \dots & a_{ss}^{(l)}(t/T) \end{pmatrix},$$

$l = 1, \dots, p$ , respectively.

In this model, the intercept vector, the autoregressive coefficient matrices and the innovations covariance matrices are time-varying parameters. Further, despite the fact of being a discrete process, in order to establish asymptotic properties, all these functions are defined in the support  $[0, 1]$ . Hence, this model is a particular extension of classical vector autoregression to the locally stationary case. The time-varying properties are very useful, because they relax stationarity assumptions, and they also provide a simple interpretation of functional coefficients.

The applicability and reliability of time-varying models rely on estimation procedures. Most of the proposed approaches are based on adaptive filters or windowed estimation, that provide smooth and reliable estimates in cases of long time series. However, they are not adequate in cases of short or cyclical time series, where window width should not be long. An alternative approach is founded on function expansions in adequate spaces. Considering the univariate case, Dahlhaus et al. (1999) and Chiann and Morettin (1999, 2005) suggested the wavelet expansions for time-varying linear systems. In this paper, we present the generalization of these approaches to the multivariate case.

### 3 The Wavelet Approach

As there is a large number of parameters in the multivariate case, in order to simplify the notation, assume that  $x_t \equiv x_{t,T}$  and  $f(t) \equiv f(t/T)$ , where  $x_{t,T}$  is a stochastic process,  $f(t/T)$  is a function with support in  $[0, 1]$ ,  $\psi_{-1,0}(x) = \phi_{0,0}(x)$  and  $\alpha_{0,0} = \beta_{-1,0}$ . Thus, for any function  $f(x) \in L^2$ , its wavelet expansion is given by

$$f(x) = \sum_{j=-1}^{\infty} \sum_{k=0}^{\infty} \beta_{j,k} \psi_{j,k}(x), \quad (15)$$

where

$$\beta_{j,k} = \langle f(x); \psi_{j,k} \rangle. \quad (16)$$

Hence, considering that the multivariate time-varying autoregression parameters are functions of time and assuming they belong to  $L^2$ , the model can be represented as

$$\begin{aligned} x_t &= \sum_{j=-1}^{\infty} \sum_{k=0}^{\infty} u_{j,k} \psi_{j,k}(t) + \sum_{l=1}^p \sum_{j=-1}^{\infty} \sum_{k=0}^{\infty} \mathbf{A}_{j,k}^{(l)} \psi_{j,k}(t) x_{t-l} + \epsilon_t, \\ &= \sum_{j=-1}^J \sum_{k=0}^{2^j-1} u_{j,k} \psi_{j,k}(t) + \sum_{l=1}^p \sum_{j=-1}^J \sum_{k=0}^{2^j-1} \mathbf{A}_{j,k}^{(l)} \psi_{j,k}(t) x_{t-l} + \\ &\quad \sum_{j>J}^{\infty} \sum_{k=0}^{\infty} u_{j,k} \psi_{j,k}(t) + \sum_{l=1}^p \sum_{j>J}^{\infty} \sum_{k=0}^{\infty} \mathbf{A}_{j,k}^{(l)} \psi_{j,k}(t) x_{t-l} + \epsilon_t, \\ &= \sum_{j=-1}^J \sum_{k=0}^{2^j-1} u_{j,k} \psi_{j,k}(t) + \sum_{l=1}^p \sum_{j=-1}^J \sum_{k=0}^{2^j-1} \mathbf{A}_{j,k}^{(l)} \psi_{j,k}(t) x_{t-l} + \nu_t, \end{aligned} \quad (17)$$

where

$$\nu_t = \sum_{j>J}^{\infty} \sum_{k=0}^{\infty} u_{j,k} \psi_{j,k}(t) + \sum_{l=1}^p \sum_{j>J}^{\infty} \sum_{k=0}^{\infty} \mathbf{A}_{j,k}^{(l)} \psi_{j,k}(t) x_{t-l} + \epsilon_t = \eta_t + \epsilon_t, \quad (18)$$

$u_{j,k}$  are vectors and  $\mathbf{A}_{j,k}^{(l)}$  ( $l = 1, \dots, p$ ,  $j = 1, 2, \dots$  and  $k = 1, 2, \dots$ ) are matrices containing the wavelet expansion coefficients. The vector  $\eta_t = o(T)$  is the truncation error, considering a finite expansion from  $j = 0, 1, \dots, J = J(T)$ .

The main idea of time-varying coefficients wavelet expansion is to provide a simple and consistent estimator. An approximated maximum likelihood estimation could be applied in this case, but depending on the dimension, truncation scale and order of the model, there will be a really large number of parameters to be estimated and optimization algorithms may even not converge. Note that considering projections in adequate spaces, we obtain a linear model in the parameters  $u_{j,k}$  and  $\mathbf{A}_{j,k}^{(l)}$ . Hence, we propose an interactive generalized least square procedure, as an extension of Cochrane and Orcutt algorithm to the multivariate case.

### 3.1 Estimation

Firstly, assume that the covariance matrix  $\Sigma(t)$  is known. The model is given by

$$x_t = \sum_{j=-1}^J \sum_{k=0}^{2^j-1} u_{j,k} \psi_{j,k}(t) + \sum_{l=1}^p \sum_{j=-1}^J \sum_{k=0}^{2^j-1} \mathbf{A}_{j,k}^{(l)} \psi_{j,k}(t) x_{t-l} + \nu_t, \quad (19)$$

where  $x_t$  is a  $s$ -dimensional locally stationary process and  $u_{j,k}$  and  $\mathbf{A}_{j,k}^{(l)}$  contain all the coefficients to be estimated. Assuming the truncation in the scale  $J = J(T)$ , the model can be represented as follows.

**Definition 2** *A wavelet time-varying vector autoregressive (WTVAR) is defined as*

$$x_t = \mathbf{\Lambda}[Z \otimes \Psi(t)]' + \nu_t, \quad (20)$$

where  $\otimes$  indicates the Kronecker product,  $Z$  is a  $1 \times (sp+1)$  vector containing lagged values of  $x_t$ , given by

$$Z = [1x'_{t-1}, x'_{t-2}, \dots, x'_{t-p}], \quad (21)$$

$\Psi(t)$  is a  $1 \times 2^J$  vector containing wavelets functions,

$$\Psi(t) = [\psi_{-1,0}(t), \psi_{0,0}(t), \dots, \psi_{J,2^J-1}(t)], \quad (22)$$

and  $\mathbf{\Lambda}$  is a  $s \times (sp+1)2^J$  matrix containing all the wavelet expansion coefficients.

Assuming that we have observed  $x_t$  for  $t = 1, 2, \dots, T$ , define the following matrices

$$\mathbf{X}_t = \begin{pmatrix} x_{1,(p+1)} & x_{2,(p+1)} & \dots & x_{s,(p+1)} \\ x_{1,(p+2)} & x_{2,(p+2)} & \dots & x_{s,(p+2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,T} & x_{2,T} & \dots & x_{s,T} \end{pmatrix},$$

$$\mathbf{X}_{t-l} = \begin{pmatrix} x_{1,(p-l+1)} & x_{2,(p-l+1)} & \dots & x_{s,(p-l+1)} \\ x_{1,(p-l+2)} & x_{2,(p-l+2)} & \dots & x_{s,(p-l+2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,(T-l)} & x_{2,(T-l)} & \dots & x_{s,(T-l)} \end{pmatrix},$$

$l = 1, \dots, p$  and also

$$\Psi = \begin{pmatrix} \psi_{-1,0}(p+1) & \psi_{0,0}(p+1) & \dots & \psi_{J,2^J-1}(p+1) \\ \psi_{-1,0}(p+2) & \psi_{0,0}(p+2) & \dots & \psi_{J,2^J-1}(p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{-1,0}(T) & \psi_{0,0}(T) & \dots & \psi_{J,2^J-1}(T) \end{pmatrix}.$$

**Definition 3** If  $a_i$  and  $b_i$  are row vectors,  $i = 1, \dots, N$ , the row-Kronecker product  $\otimes^R$  is defined as

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} \otimes^R \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} = \begin{pmatrix} a_1 \otimes b_1 \\ a_2 \otimes b_2 \\ \vdots \\ a_N \otimes b_N \end{pmatrix}.$$

Further, define the following matrices

$$\mathbf{W} = [1_{N-p} \otimes^R \Psi, \mathbf{X}_{t-1} \otimes^R \Psi, \dots, \mathbf{X}_{t-l} \otimes^R \Psi], \quad (23)$$

$$\mathbf{M} = \mathbf{I}_s \otimes \mathbf{W}, \quad (24)$$

and the vector

$$Y = \text{vec}(\mathbf{X}_t). \quad (25)$$

The innovation covariance matrix  $\nu_t$  is defined as

$$\Sigma = \begin{pmatrix} \text{diag}[\sigma_1^2(t)] & \text{diag}[\sigma_{12}(t)] & \dots & \text{diag}[\sigma_{1s}(t)] \\ \text{diag}[\sigma_{21}(t)] & \text{diag}[\sigma_2^2(t)] & \dots & \text{diag}[\sigma_{2s}(t)] \\ \vdots & \vdots & \ddots & \vdots \\ \text{diag}[\sigma_{s1}(t)] & \text{diag}[\sigma_{s2}(t)] & \dots & \text{diag}[\sigma_s^2(t)] \end{pmatrix},$$

where  $\text{diag}[h(t)]$  stands for the diagonal matrix of  $h(t)$ , for  $t = (p+1), (p+2), \dots, T$ .

Hence, the WTVAR model of order  $p$  can be represented in the linear model form

$$Y = \mathbf{M}\beta + \nu. \quad (26)$$

The generalized least square estimator (GLS) of  $\beta = \text{vec}(\Lambda)$  is given by

$$\begin{aligned}\hat{\beta} &= (\mathbf{M}'\Sigma^{-1}\mathbf{M})^{-1}\mathbf{M}'\Sigma^{-1}Y \\ &= (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'V\end{aligned}\tag{27}$$

with

$$\mathbf{D} = \Sigma^{-\frac{1}{2}}\mathbf{M},\tag{28}$$

$$V = \Sigma^{-\frac{1}{2}}Y.\tag{29}$$

Assume the following conditions:

**(C1)** (Dahlhaus et al. (1999)) Assume that the functions to be estimated in the WTVAR model  $u(t)$ ,  $\mathbf{A}_i(t)$  and  $\Sigma(t)$  are real, bounded and belong to

$$\mathcal{F}_i = \left\{ f_i(x) : f_i(x) = \sum_{j=-1}^{\infty} \sum_{k=0}^{\infty} \beta_{j,k}^{(i)} \psi_{j,k}(x) \mid \|\beta_{\cdot,\cdot}^{(i)}\|_{m_i,p_i,q_i} < \infty \right\},\tag{30}$$

where

$$\|\beta_{\cdot,\cdot}^{(i)}\|_{m_i,p_i,q_i} = \left( \sum_{j \geq l} \left[ 2^{j s_i p_i} \sum_{k \in I_j} |\beta_{j,k}^{(i)}|^{p_i} \right]^{q_i/p_i} \right)^{1/q_i},\tag{31}$$

where  $s_i = m_i + 1/2 + 1/p_i$  and  $I_j = \{k : k = 0, \dots, 2^j - 1\}$ . Here  $m_i$  denote the degree of smoothness of  $\mathcal{F}_i$  and  $1 \leq p_i, q_i \leq \infty$  specify the norm of  $\mathcal{F}_i$ . Examples of such spaces are Besov, Sobolev and Hölder spaces. See Triebel (1992) for details.

**(C2)**  $\phi$  and  $\psi$  are compact supported and they belong to  $\mathcal{C}^r[0, 1]$ , the set of all continuous functions, with continuous derivatives up to order  $r$ , with  $r > m = \max\{m_i\}$ . Further, assume that  $\int \phi(t)dt = 1$  and  $\int \psi(t)t^k dt = 0$  for  $0 \leq k \leq r$ .

**(C3)** There is  $\gamma \geq 0$  where  $|\text{cum}_n(\nu_t)| \leq C^n(n!)^{1+\gamma} \forall n, t$ . Here  $C$  is a positive constant.

**(C4)**  $x_t$  is a locally stationary process, with an infinite moving average representation,

$$x_t = u(t) + \sum_{i=0}^{\infty} \mathbf{C}_i(t)\epsilon_{t-i},$$

and

$$\sup_t \left( \sum_{i=0}^{\infty} |\mathbf{C}_i(t)| \right) \leq \infty. \quad (32)$$

(C5)  $\|\Sigma^{-1}\|_2 < \infty$ , where  $\|\cdot\|_2$  indicates the spectral norm.

*Remarks:* (i) Under C1, the wavelet coefficients satisfy

$$\sup_{f_i \in \mathcal{F}_i} \left\{ \sum_{j \geq J} \sum_k |\beta_{jk}^{(i)}|^2 \right\} = O\left(2^{-2J\tilde{s}_i}\right),$$

where  $\tilde{s}_i = m_i + 1/2 - 1/\tilde{p}_i$ , with  $\tilde{p}_i = \min\{p_i, 2\}$ . It can be shown (Donoho et al. (1995)) that the loss in the reconstruction of  $x_t$  in (17) is of order  $T^{-2m_i/(2m_i+1)}$ , if we choose  $J$  such that  $2^{J-1} \leq T^{1/2} \leq 2^J$ . This will be used in the proof of Proposition 1.

(ii) The Gaussian distribution satisfies C3. For  $\gamma = 0$ , this condition is valid for the Gamma and Inverse Gaussian distributions. It also holds for  $\gamma > 0$  and some fatter tail distributions.

(iii) In practice, these assumptions require a certain degree of smoothness of the function to be estimated and also constraint the random errors distribution. For most applications, the estimation of smooth functions and Gaussian distributed errors are reasonable assumptions.

**Proposition 1** *Assuming that the conditions (C1) – (C5) are satisfied, the order  $p$  is fixed and we assume Gaussian innovations with known covariance matrix  $\Sigma$ , then*

$$\sqrt{N}\mathbf{H}(\hat{\beta} - \beta) \stackrel{a}{\sim} Z \quad (33)$$

where  $Z$  is a  $k$ -dimensional multivariate normal with zero mean and covariance matrix  $\mathbf{\Gamma}$ ,

$$\mathbf{\Gamma} = \lim_{N \rightarrow \infty} N\mathbf{H}E[\mathbf{D}'\mathbf{D}]^{-1}\mathbf{H}', \quad (34)$$

$\mathbf{H}$  is a  $k$  row matrix with the same number of columns of  $\mathbf{D}$ , and  $N = (T - p)$ .

*Remark:* The matrix  $\mathbf{H}$  is used to obtain only a finite subpartition of  $\beta$ .

Proof: See Appendix.

Let  $\hat{\xi}$  be a subpartition of  $\hat{\beta}$  containing all the estimators of coefficients of the wavelet expansion of a function  $f(t)$ . Thus, a natural estimator of  $f(t)$  at each  $t$  is

$$\hat{f}(t) = \hat{\xi}'\psi(t) = \sum_{j=0}^J \sum_k \hat{\beta}_{j,k} \psi_{j,k}(t), \quad (35)$$

where  $\psi(t)$  is a vector containing the respective wavelet functions of  $\hat{\xi}$  evaluated on the time  $t$ . Let  $\Sigma_{\hat{\xi}}$  be the covariance matrix of  $\hat{\xi}$ , a subpartition of the covariance matrix of  $\hat{\beta}$ , then

**Proposition 2** *Under the conditions of Proposition 1, for each  $t$  we have*

$$(\hat{f}(t) - f) \stackrel{a}{\sim} N(0, \psi \Sigma_{\hat{\xi}} \psi'). \quad (36)$$

Proof: The result is an immediate consequence of the previous proposition.

### 3.2 Hypothesis Testing

Define the following estimator of the innovations variance in the transformed model

$$\hat{\sigma}^2 = \left( \frac{1}{s[(T-p) - p2^J]} \right) [\mathbf{D}'\mathbf{D} - \hat{\beta}'\mathbf{D}'Y]. \quad (37)$$

$$(38)$$

The estimated covariance matrix of  $\hat{\beta}$  is given by

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{M}'\Sigma^{-1}\mathbf{M})^{-1}. \quad (39)$$

Thus, the Wald statistic for contrasts is

$$W = \frac{(\mathbf{C}\hat{\beta} - m)'[\mathbf{C}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta} - m)}{\hat{\sigma}}, \quad (40)$$

where  $m$  is a vector and  $\mathbf{C}$  is the contrast matrix of the following test:

$$H_0: \mathbf{C}\beta = m,$$

$H_A: \mathbf{C}\beta \neq m$ .

Under the null hypothesis, it follows from *Proposition 1* and the Cochran theorem that  $W$  has an asymptotic chi-square distribution with  $\text{rank}(C)$  degrees of freedom.

### 3.3 Innovations Covariance Matrix Estimation

In most applications to real data, the innovations covariance matrix  $\Sigma$  used in the generalized least square estimator is unknown. Assume that the innovations  $\epsilon_{lt}$  have zero mean and time-varying variance  $\sigma_l^2(t)$ , then

$$\begin{aligned} E(\epsilon_{lt}^2) &= \text{Var}(\epsilon_{lt}^2) + E(\epsilon_{lt})^2 \\ &= \text{Var}(\epsilon_{lt}^2) = \sigma_l^2(t). \end{aligned} \quad (41)$$

Thus, a reasonable estimator for  $\sigma_l^2(t)$  is the squared residuals, since

$$\begin{aligned} E(r_{lt}^2) &= E[(S_{lt} + \epsilon_{lt})^2] + o(1) \\ &= E(\epsilon_{lt}^2) + 2E(S_{lt}\epsilon_{lt}) + E(S_{lt}^2) + o(1), \end{aligned} \quad (42)$$

where  $S_{lt} = o(1)$  is a truncation error. Hence

$$\lim_{T \rightarrow \infty} E(r_{lt}^2) = \sigma_l^2(t). \quad (43)$$

Analogously a reasonable estimator to the time-varying covariance  $\sigma_{lm}^2(t)$  of the innovations of two time series  $x_{lt}$  and  $x_{mt}$  is given by  $r_{lt}r_{mt}$ . Then, considering the wavelet expansion, we have

$$\hat{\sigma}_l^2(t) = \sum_j \sum_k v_{j,k} \psi_{j,k}(x), \quad (44)$$

$$\hat{\sigma}_{lm}(t) = \sum_j \sum_k c_{j,k} \psi_{j,k}(x), \quad (45)$$

where the coefficients  $v_{j,k}$  and  $c_{j,k}$  can be obtained by a classical wavelet smoothing of squared and cross residual, resulting in a consistent estimator of  $\Sigma$ .

Hence, we propose an interactive generalized least square algorithm given by:

Step 1. Assume  $\Sigma = \mathbf{I}$  and apply GLS;

Step 2. Obtain estimates of the innovations variances and covariances;

Step 3. Apply the GLS using the estimated innovations covariance matrix;

Step 4. Return to step 2 until convergence.

A practical problem to the application of this algorithm is that in some cases, the estimated innovation's covariance matrix is almost singular or not positive definite. Hence, one may have problems in matrix inversion or Cholesky decompositions steps. By simulations, we noticed that it could happen cases when the innovations variances is too small. For these cases, we suggest a normalization of the time series variance multiplying the time series by a constant, or including an inferior bound constraint in the diagonal of the estimated covariance matrix.

**Proposition 3** *Let  $\hat{\Sigma}$  be a consistent estimator of  $\Sigma$  (i.e.,  $\|\hat{\Sigma} - \Sigma\|_2 \xrightarrow{P} 0$ ), and define*

$$\hat{\beta}_{\hat{\Sigma}} = (\mathbf{M}'\hat{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{M}'\hat{\Sigma}^{-1}Y. \quad (46)$$

*Then the asymptotic distribution of  $\hat{\beta}_{\hat{\Sigma}}$  is the same as that of  $\hat{\beta}$ .*

Proof:  $\hat{\Sigma} - \Sigma \xrightarrow{P} 0$ . Thus,

$$\begin{aligned} T(\mathbf{M}'\hat{\Sigma}^{-1}\mathbf{M})^{-1} &= T(\mathbf{M}'[\hat{\Sigma} - \Sigma + \Sigma]^{-1}\mathbf{M})^{-1} \\ &= T(\mathbf{M}'\Sigma^{-1}\mathbf{M})^{-1} + o_p(1). \end{aligned} \quad (47)$$

The result follows from the Slutsky theorem.  $\square$

## 4 Simulations

Computational simulations were performed focusing the evaluation of the proposed estimators and the small sample approximation by asymptotic results. All simulations consist on 500 experiments of bivariate time series with length  $T = 128$ , Gaussian errors, wavelet D16 and function expansions considering 4 wavelets (24 coefficients). The innovations covariance matrix was also estimated, as the iterated GLS algorithm was applied in all simulations. The theoretical asymptotic properties describe coefficients with normal distri-

bution and small biases. Based on all simulations, the average and standard deviation of the estimated curves (at each time-point) were calculated.

#### 4.1 Group 1

The aim of this first group of simulations is the evaluation of estimator performance in cases of smooth continuous curves,

$$x_t = u(t) + \mathbf{A}(t)x_{t-1} + \epsilon_t, \quad (48)$$

where

$$\mathbf{u}(t) = \begin{pmatrix} \sin(\frac{2\pi t}{128} + \pi) \\ -0.6 \end{pmatrix}, \mathbf{A}(t) = \begin{pmatrix} 0.5 - \frac{\sin(\frac{2\pi t}{128})}{4} & -0.5 + \frac{\cos(\frac{2\pi t}{128})}{4} \\ 0.5 - \frac{\sin(\frac{2\pi t}{128})}{4} & -0.3 \end{pmatrix}$$

and

$$\mathbf{\Sigma}(t) = \begin{pmatrix} 1.4 & 1.4 \left(1 + \frac{\cos(\frac{2\pi t}{128})}{4}\right) \\ 1.4 \left(1 + \frac{\cos(\frac{2\pi t}{128})}{4}\right) & 2 \left(1 + \frac{\cos(\frac{2\pi t}{128})}{4}\right)^2 \end{pmatrix},$$

where  $u(t)$  is the intercept vector,  $\mathbf{A}(t)$  is the autoregressive coefficient matrix and  $\mathbf{\Sigma}(t)$  is the innovations covariance matrix. The results of these simulations are presented in Figures 1, 2 and 3.

#### 4.2 Group 2

In this second group of simulation, we would like to evaluate the estimators performance and small sample approximations in the case where the true curves cannot be described using few wavelet functions. In this case, the number of coefficients estimated is not enough to describe the true curves in details. This suggests that a good selection of wavelet basis is necessary in order to avoid bias. However, this issue is not simple, as priori knowledge about the target curves is required. Focusing on the aims of these simulations, we used few smooth functions (D16) to estimate non-continuous curves, although Haar wavelets might lead to better results. Two threshold functions  $f_1(t)$  and  $f_2(t)$

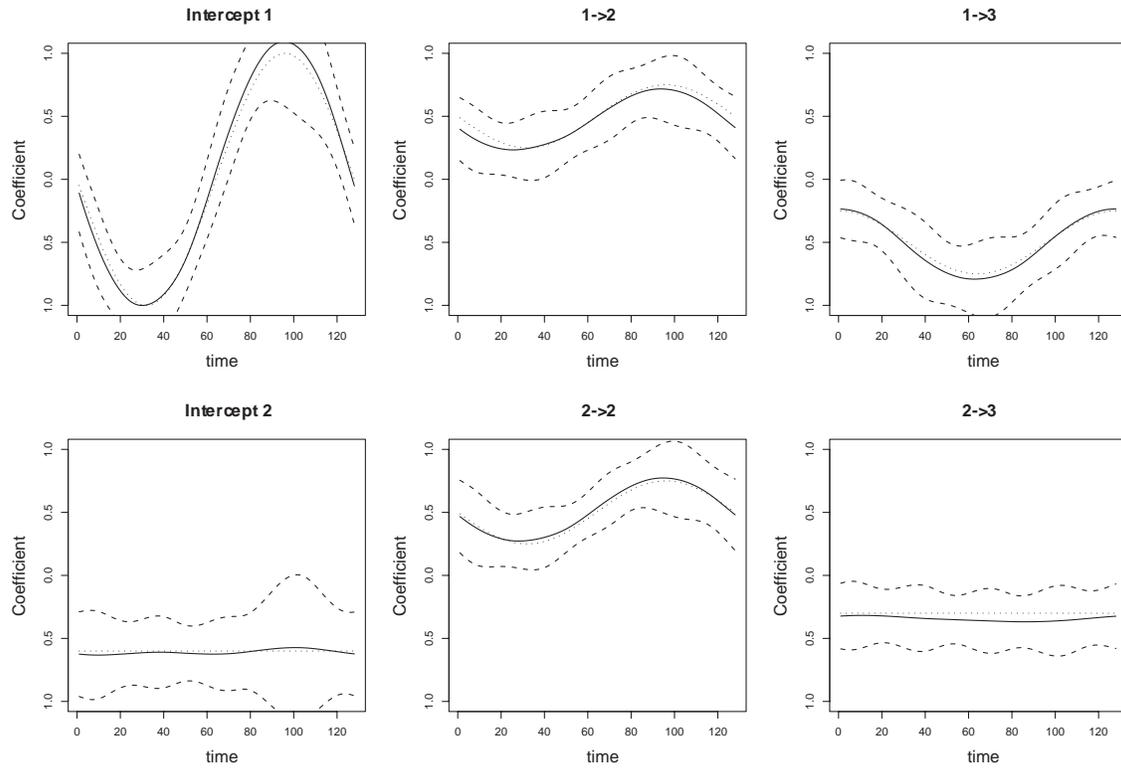


Fig. 1. Simulation 1: The theoretical autoregressive functions and average of estimates are presented in dotted and solid lines, respectively. The label  $i \rightarrow j$  describes the influence curve from series  $i$  to  $j$  ( $i$ -th column at  $j$ -th row of the matrix  $\mathbf{A}(t)$ ). The dashed lines show the interval of one standard deviation. At each time point, the average and standard deviation were calculated over 500 experiments.

were considered, and they are defined by,

$$f_1(t) = \begin{cases} 1, & \text{if } \frac{T}{2} < t \leq T, \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_2(t) = \begin{cases} 1, & \text{if } \frac{T}{4} < t < \frac{3T}{4}, \\ 0, & \text{otherwise} \end{cases}$$

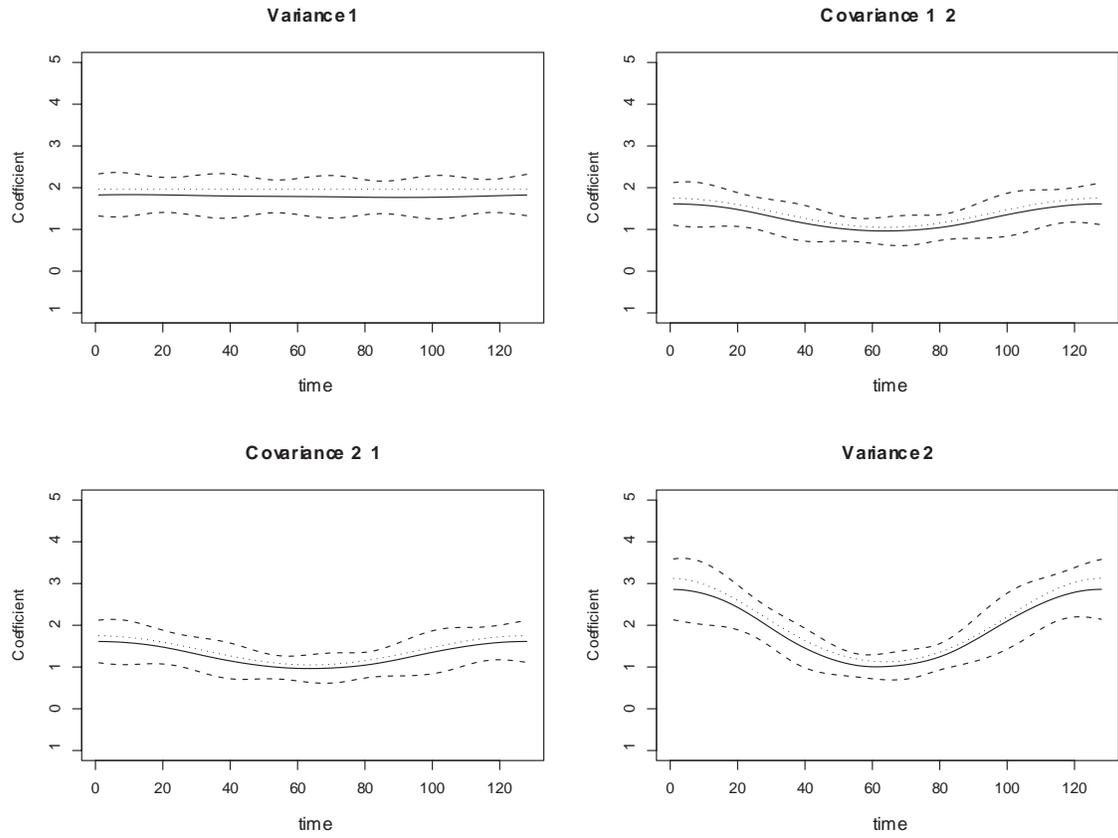


Fig. 2. Simulation 1: The theoretical covariance functions and average of estimates are presented in dotted and solid lines, respectively. The dashed lines show the interval of one standard deviation. At each time point, the average and standard deviation were calculated over 500 experiments.

The time-varying multivariate autoregressive model considered in the simulations is also of order 1, with parameters given by

$$\mathbf{u}(t) = \begin{pmatrix} 0.25f_1(t) \\ 0.3 \end{pmatrix}, \mathbf{A}(t) = \begin{pmatrix} -0.3g(t) & -0.4f_2(t) \\ 0.2f(t) & -0.3 \end{pmatrix}$$

and

$$\Sigma(t) = \begin{pmatrix} 0.5 + 0.5f_1(t) & (0.5 + 0.5f_1(t))(0.5 + 0.5f_2(t)) \\ (0.5 + 0.5f_1(t))(0.5 + 0.5f_2(t)) & 2(0.5 + 0.5f_2(t)) \end{pmatrix}.$$

The results are presented in Figures 4, 5 and 6.

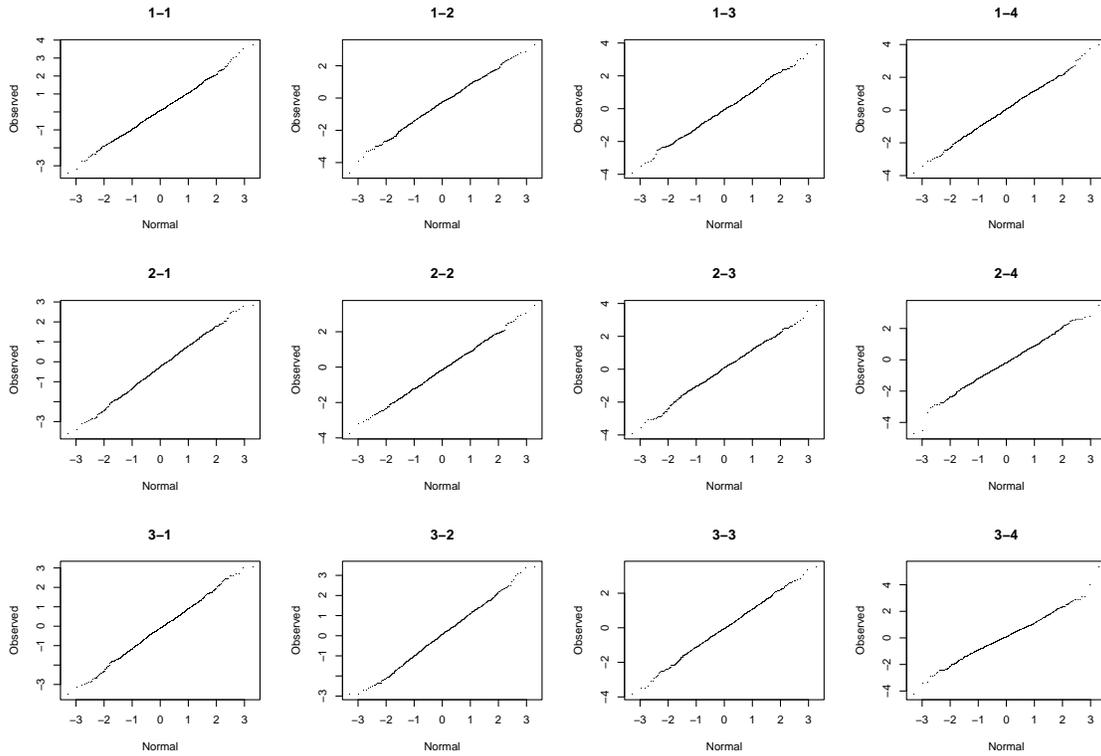


Fig. 3. Normal QQ-plots of estimated coefficients (1st to 12th). The label  $i$ - $j$  describes the QQ-plot of the  $j$ -th coefficient of the  $i$ -th curve. The results for coefficients 13th to 24th are analogous.

### 4.3 Conclusions

The first group of simulations shows a satisfactory performance of the proposed approach. The average of estimated curves for both intercept, autoregressive and covariance functions are similar to the true curves (Figures 1 and 2). Finally even in a small sample, we verify a satisfactory approximation of the estimators distribution to the expected Gaussian distribution (Figure 3).

Additionally, the second group of simulations also points toward the validity of the asymptotic distribution even in cases where the number of functions is not large. Further, considering the scale of detail chosen ( $J = 2, 4$  wavelet functions), the average of estimated and true curves are similar (Figures 4 and 5). This means that despite the fact of being biased, the curves estimators still capture the long-run variations. In addition, the QQ-plots of the estimated wavelet coefficients in Figure 6 yields evidence of a satisfactory approximation to normal distribution.

In conclusion, the simulations results point toward a high similarity between the estimator distribution in finite samples and the expected Gaussian one.

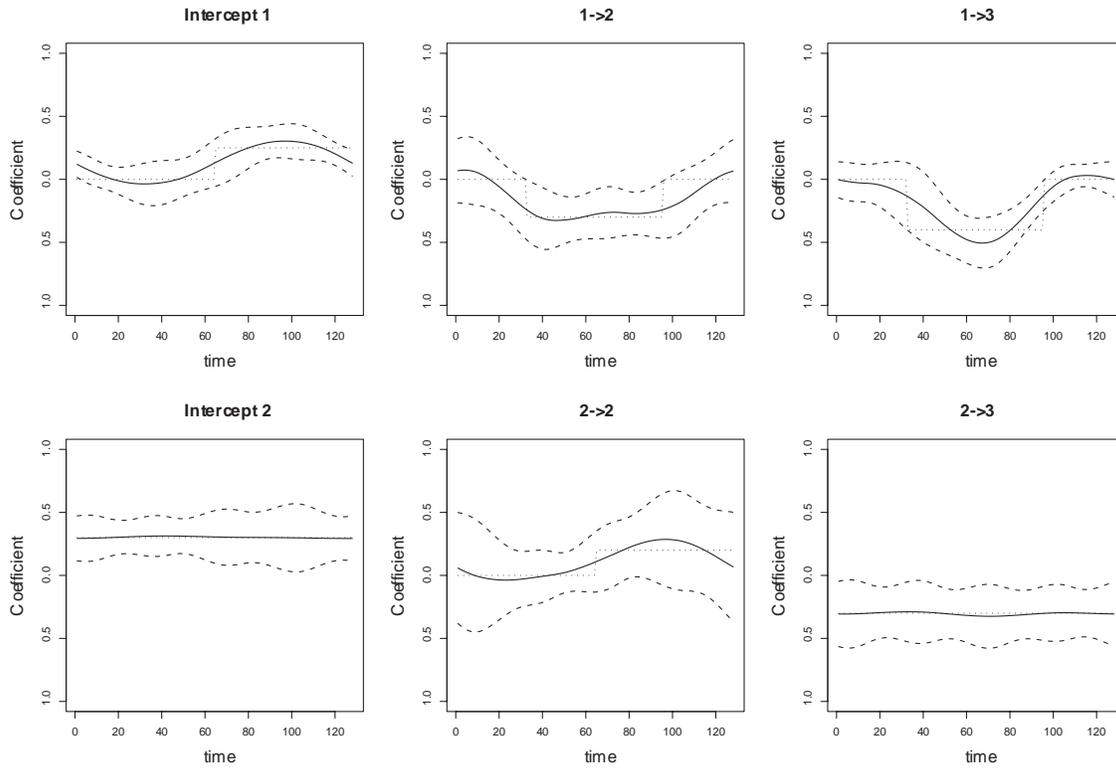


Fig. 4. Simulation 2: The theoretical autoregressive functions and average of estimates are presented in dotted and solid lines, respectively. The label  $i \rightarrow j$  describes the influence curve from series  $i$  to  $j$  ( $i$ -th column at  $j$ -th row of the matrix  $\mathbf{A}(t)$ ). The dashed lines show the interval of one standard deviation. At each time point, the average and standard deviation were calculated over 500 experiments.

## 5 Application to Real Data

In this section, an application of the time-varying autoregressive models to brain connectivity modelling, using functional magnetic resonance time series (fMRI) is presented. The fMRI data consist on temporal acquisition of BOLD (blood oxygenation level dependent) signal of several brain areas, which can be interpreted as an indirect measure of neuronal activation (Ogawa et al. (1990)). Thus, the dataset is a multivariate time series representing the neuronal activity in the brain. The aim of this study is the identification of cortical/subcortical connectivity differences between healthy subjects and patients with Parkinson disease.

Parkinson's disease (PD) is a movement disorder resulting from the lack of dopamine, a neurotransmitter which has an important role in motor control. The clinical result is bradykinesia/akinesia, rigidity, and also some involuntary movements as tremors. The motor circuitry connection in PD works abnormally; the exact extension of damage and the strategies to get more from it to improve movement are still unknown. There are electrical evidences of

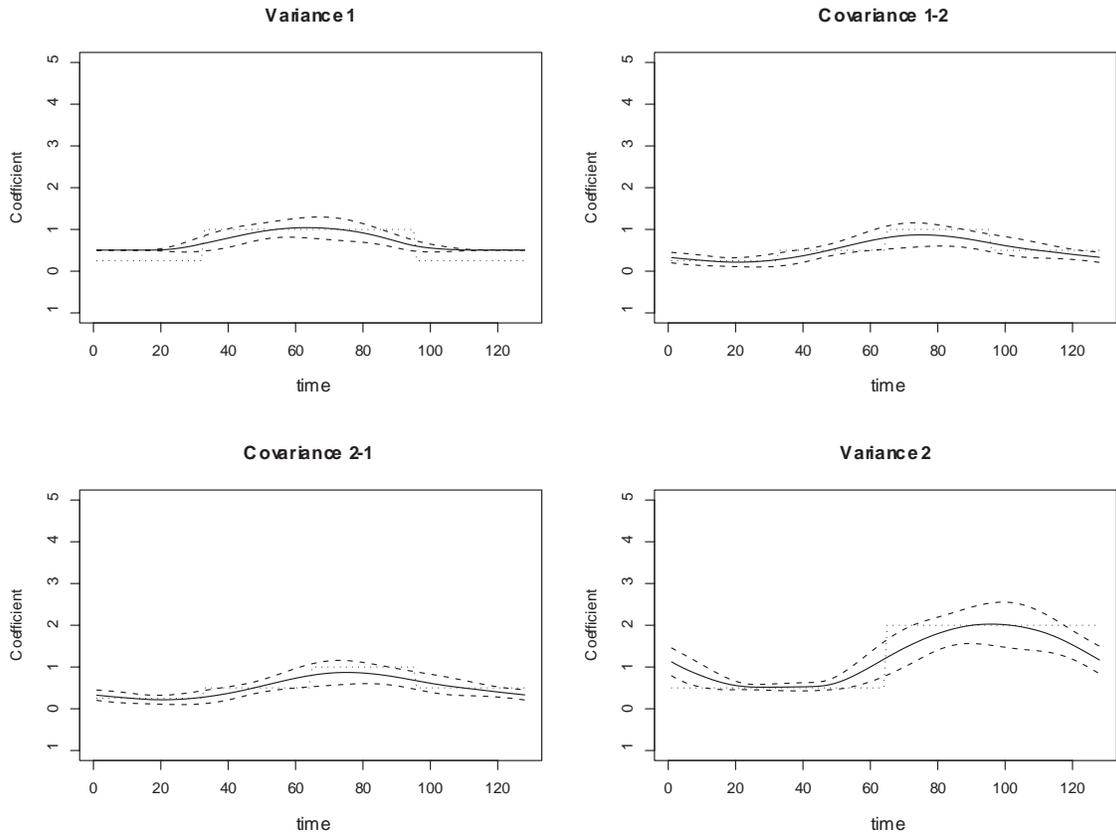


Fig. 5. Simulation 2: The theoretical covariance functions and average of estimates are presented in dotted and solid lines, respectively. The dashed lines show the interval of one standard deviation. At each time point, the average and standard deviation were calculated over 500 experiments.

abnormal activity in the secondary motor areas, due to a delay in movement programming.

Seven right handed healthy volunteers and seven patients with PD participated in this fMRI study. Both groups performed a simple motor task: self-paced finger tapping movement of dominant hand, in an cyclical event-related paradigm. The whole dataset have been acquired in the Radiology Institute of the Clinics Hospital (University of São Paulo, Brazil), in a 1.5T Signa LX scanner (GE, Milwaukee, USA), equipped with 23 mT/m gradients and echo-planar capability, with head coil. The acquisitions were based on a 2D gradient eco EPI,  $TR = 2sec$ ,  $TE = 40ms$ ,  $FA = 90^\circ$ , bandwidth=64kHz,  $FOV = 20cm$ ,  $64 \times 64$  voxels, 7 mm thick slices with 0.7mm gap, 15 slices oriented according to the bicomissural plane. The subjects were instructed to perform the finger tapping sequence only once, in response to a verbal command, with a constant inter-stimulus interval of 20 seconds.

The fMRI images were preprocessed applying motion realignment, slice time correction, and spatial normalization (SPM2, <http://www.fil.ion.ucl.ac.uk/spm/>) to the stereotatic space of Talairach and Tournoux (Talairach and Tournoux

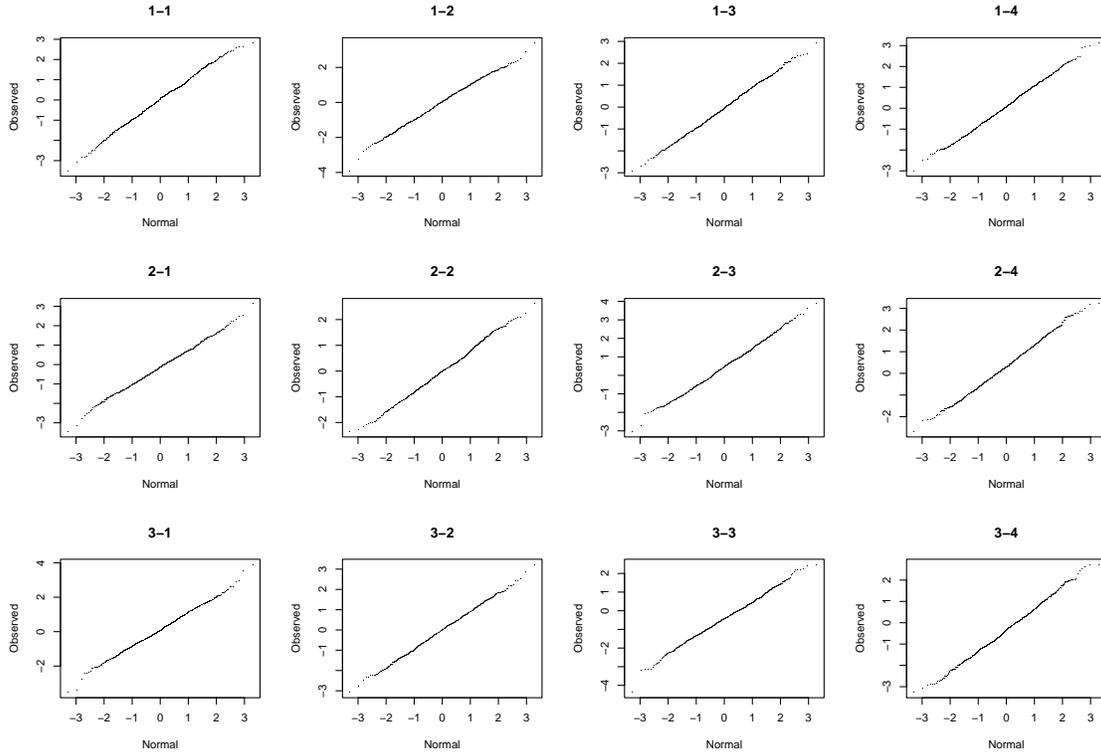


Fig. 6. Normal QQ-plots of estimated coefficients (1st to 12th). The label  $i$ - $j$  describes the qq-plot of the  $j$ -th coefficient of the  $i$ -th curve. The results for coefficients 13th to 24th are analogous.

(1998)). Focusing on the connectivity analysis, we selected three regions of interest (ROI's) in the local maxima of activation (SPM2) in cerebellum (CER), supplementar motor area(SMA) and primary motor cortex (M1) (Figure 7), obtaining trivariate time series for each subject.

In the stationary case, the vector autoregressive model were successfully applied in brain information flow identification (Goebel et al. (2003)). Here, the time-varying vector autoregressive modelling was applied focusing the identification of a dynamic connectivity structure, considering the autoregressive functions as information flow intensity. Further, as the paradigm design is periodic, we assume that the connectivity structure in each cycle is the same. Thus, the connectivity structure is assumed periodic (replications), allowing a more precise estimation. The autoregressive functions (connectivity) were estimated for each subject separately, considering the D16 wavelets ( $p = 1$ ), as smooth variations are expected. The group connectivity function was defined as the autoregressive functions average between all subjects. In order to detect connectivity differences between healthy and Parkinson groups, the Mann-Whitney test was applied for each time point. The results are presented in Figure 8.

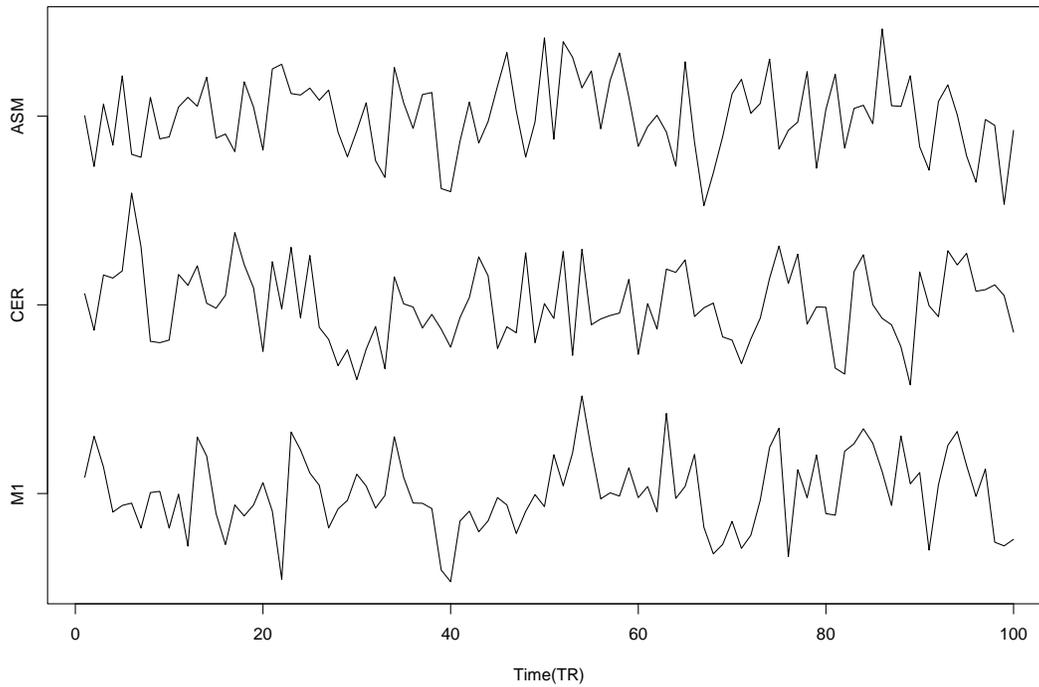


Fig. 7. ROI's BOLD signal for subject 1.

Applying the proposed approach for connectivity modelling, we observed that the information flow intensity from M1 to SMA is different during a short time interval after the finger tap execution. Furthermore, there are no evidences that the connectivity intensities from CER to SMA and from M1 to CER are different, at some short time after the task. Basically, one may infer that parkinsonians have a basal connectivity (CER to SMA and M1 to CER) different from healthy subjects. However, when patients perform a task involving cortical activations, this connectivity structure comes close to the normal one. Furthermore, motor programming network involves interactions between SMA and M1 (Ito et al. (1999)). Parkinsonians have alterations in motor planning, as a result of disorders in thalamic/cortex network (DeLong (1990)). Hence, it is expected alterations in this network, when patients perform motor tasks, as evidenced by the panel (M1 to SMA) in Figure 8. These results have a reasonable neurophysiological interpretation and they are important to the characterization of Parkinson's disease.

Finally, we conclude that the results obtained via wavelet time-varying autoregressive modelling are satisfactory, and they also emphasize the connectivity alterations present in parkinsonians, which is in agreement with theories found in literature.

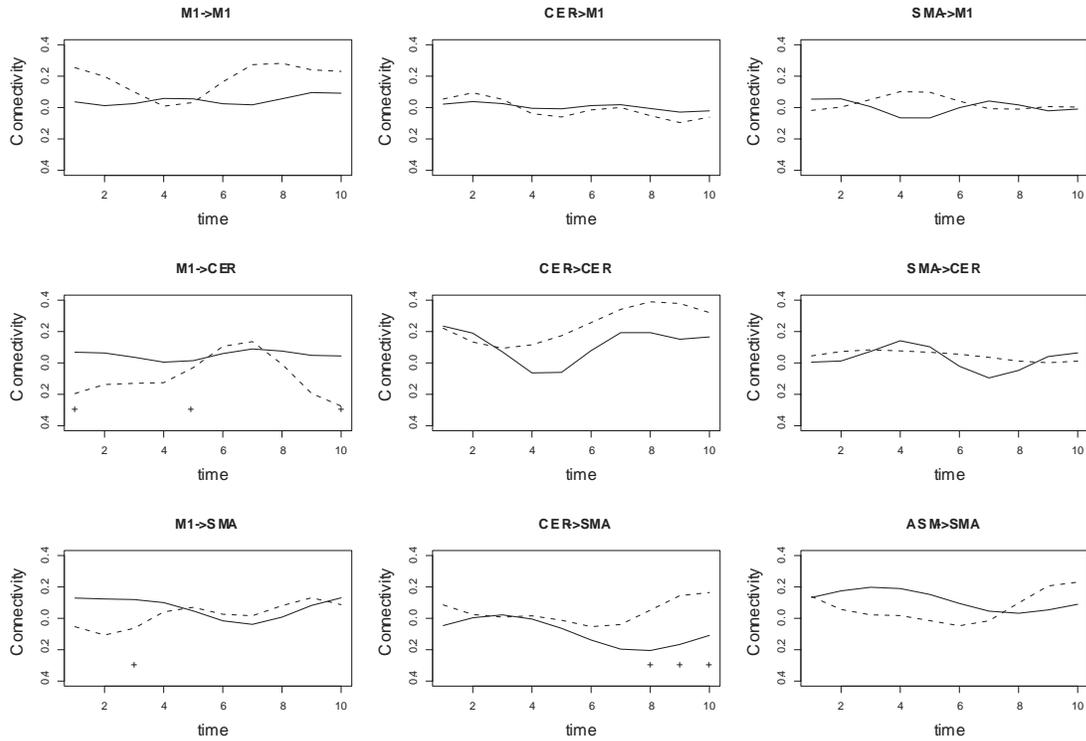


Fig. 8. The black and dashed curves indicate the connectivity strength of normal and Parkinson patients, respectively. The points in (+) indicate the timepoints where the curves are significantly different ( $p - value < 0.05$ ). The finger tapping task is performed in  $t=1$ .

## 6 Conclusions

Most studies which focus on the relationships or causality structure identification among time series are founded in vector autoregressive modelling. However, in many practical applications, stationarity conditions are not reasonable. In this paper, we proposed the wavelet time-varying autoregressive modelling, which is based on multivariate locally stationary processes. The usefulness of the proposed approach is illustrated via simulations and applications to functional magnetic resonance datasets.

## 7 Acknowledgements

We are grateful to CNPq(142616/2005-2) and FAPESP(03/10105-2) - Brazil for the support.

## A Appendix

**Lemma 1** *Assuming that the conditions (C1)-(C5) are satisfied,*

$$NH(\mathbf{D}'\mathbf{D})^{-1}\mathbf{H}' \xrightarrow{P} \Gamma, \quad (\text{A.1})$$

where

$$\Gamma = \lim_{N \rightarrow \infty} NHE[\mathbf{D}'\mathbf{D}]^{-1}\mathbf{H}'. \quad (\text{A.2})$$

**Remark:** *Note that each cell in  $\lim_{N \rightarrow \infty} E[\mathbf{D}'\mathbf{D}]/N$  may be written as*

$$\lim_{N \rightarrow \infty} \sum_{t=1}^N f(t/N)/N = \int_0^1 f(t)dt.$$

Proof: Firstly, note that  $\mathbf{D}$  is a block diagonal matrix. Let  $d_{ij}$  be the  $i$ -th row, and  $j$ -th column cell of a non-null block of  $\mathbf{D}$ . Then,

$$d_{tj} = f_{tj}X_{tj}, \quad (\text{A.3})$$

where  $f_{tj}$  is a function obtained by the product of a wavelet function and the generalized least square transformation and  $X_{tj}$  is a local stationary process. Let  $(dd)_{lm}$  the  $l$ -th row,  $m$ -th column cell of any subpartition of  $\mathbf{D}'\mathbf{D}$ . Then  $(dd)_{lm}$  can be written as

$$\begin{aligned} (dd)_{lm} &= \sum_{t=1}^N d_{tl}d_{tm} = \sum_{t=1}^N f_{tl}X_{tl}f_{tm}X_{tm} \\ &= \sum_{t=1}^N f_{tl}f_{tm} \sum_{j=0}^{\infty} \sum_{v=0}^{\infty} \theta_{lj}(t)\theta_{mv}(t)\epsilon_{l,(t-j)}\epsilon_{m,(t-v)} \\ &\quad + \sum_{t=1}^N f_{tl}f_{tm}\mu_l(t) \sum_{v=0}^{\infty} \theta_{mv}(t)\epsilon_{m,(t-v)} + \sum_{t=1}^N f_{tl}f_{tm}\mu_m(t) \sum_{j=0}^{\infty} \theta_{lj}(t)\epsilon_{l,(t-j)} \\ &\quad + \sum_{t=1}^N f_{tl}f_{tm}\mu_m(t)\mu_l(t). \end{aligned}$$

If we denote

$$z_{lmt}^{(1)} = \sum_{t=1}^N f_{tl}f_{tm} \sum_{j=0}^{\infty} \sum_{v=0}^{\infty} \theta_{lj}(t)\theta_{mv}(t) \left[ \epsilon_{l,(t-j)}\epsilon_{m,(t-v)} - E[\epsilon_{l,(t-j)}\epsilon_{m,(t-v)}] \right],$$

$$z_{lmt}^{(2)} = \sum_{t=1}^N f_{tl} f_{tm} \mu_l(t) \sum_{v=0}^{\infty} \theta_{mv}(t) \epsilon_{m,(t-v)},$$

$$z_{lmt}^{(3)} = \sum_{t=1}^N f_{tl} f_{tm} \mu_m(t) \sum_{j=0}^{\infty} \theta_{lj}(t) \epsilon_{l,(t-j)},$$

then

$$E \left[ z_{lmt}^{(1)} | \mathcal{F}_{t-h} \right] = \sum_{t=1}^N f_{tl} f_{tm} \sum_{j=h}^{\infty} \sum_{v=h}^{\infty} \left( \theta_{lj}(t) \theta_{mv}(t) \epsilon_{l,(t-j)} \epsilon_{m,(t-v)} - E[\epsilon_{l,(t-j)} \epsilon_{m,(t-v)}] \right),$$

$$E \| E \left[ z_{lmt}^{(1)} | \mathcal{F}_{t-h} \right] \| \leq \sum_{t=1}^N |f_{tl} f_{tm}| \sum_{j=h}^{\infty} \sum_{v=h}^{\infty} |\theta_{lj}(t)| |\theta_{mv}(t)| E \| [\epsilon_{l,(t-j)} \epsilon_{m,(t-v)} - E[\epsilon_{l,(t-j)} \epsilon_{m,(t-v)}]] \|$$

$$\leq \sum_{j=h}^{\infty} \sum_{v=h}^{\infty} |\theta_{lj}(t)| |\theta_{mv}(t)| K = \xi_{lmh}.$$

Analogous results hold for  $z_{lmt}^{(2)}$  and  $z_{lmt}^{(4)}$  and from the MA representation, the sequence  $\theta_j(t)_{j=0}^{\infty}$  converges, then  $\lim_{h \uparrow \infty} \xi_{lmh} = 0$ , and taking  $c_t = 1$ , we obtain that  $z_{lmt}^{(1)}$ ,  $z_{lmt}^{(2)}$  and  $z_{lmt}^{(3)}$  are  $L^1$ -Mixingales. The result follows from the Weak Law for  $L^1$ -Mixingales (Andrews (1988)).  $\square$

**Lemma 2** *Assuming that the conditions (C1)-(C5) are satisfied,*

$$(1/\sqrt{N})\mathbf{G}(\mathbf{D}'e) \xrightarrow{D} Z, \quad (\text{A.4})$$

where  $Z$  is a random variable with multivariate normal distribution, with null mean and covariance matrix  $\mathbf{\Gamma}^{*-1} = \lim_{N \rightarrow \infty} \frac{\mathbf{G}E[\mathbf{D}'\mathbf{D}]\mathbf{G}'}{N}$ .

Proof: Let  $(de)_j$  the  $j$ -th element of the vector  $(\mathbf{D}'e)$ . It can be written as

$$(de)_j = \sum_{t=1}^N (de)_{jt} = \sum_{t=1}^N h_{tj} X_{tj} e_t, \quad (\text{A.5})$$

where  $(de)_{jt} = h_{tj} X_{tj} e_t$  and  $\{e_t\}_{t=1}^N$  are independent processes,  $X_{tj}$  is a locally stationary process and  $h_{tj}$  is a function obtained by the product of a wavelet function and the generalized least square transformation function.

The vector  $\mathbf{D}'e$  may be written as  $\mathbf{D}'e = \sum_{t=1}^N b_t$ , where  $b_t$  is a vector whose elements are  $(de)_{jt}$ .

Further define  $\mathcal{F}_{t-1}$  as denoting the information up to time  $(t-1)$ , we have

$$E[b_t | \mathcal{F}_{t-1}] = 0, \quad (\text{A.6})$$

and thus,  $\{b_t\}_{t=1}^{\infty}$  is a vector martingale differences sequence. Further,

$$E \left[ \frac{\sum_{t=1}^N b_t b_t'}{N} \right] = E \left[ \frac{\mathbf{D}' e e' \mathbf{D}}{N} \right] = E \left[ E \left[ \frac{\mathbf{D}' e e' \mathbf{D}}{N} | \mathbf{D} \right] \right] = E \left[ \frac{\mathbf{D}' \mathbf{D}}{N} \right] \quad (\text{A.7})$$

$$\mathbf{G} E \left[ \frac{\sum_{t=1}^N b_t b_t'}{N} \right] \mathbf{G}' = \mathbf{G} E \left[ \frac{\mathbf{D}' \mathbf{D}}{N} \right] \mathbf{G}' \rightarrow \mathbf{\Gamma}^{*-1}, \quad (\text{A.8})$$

where

$$\mathbf{\Gamma}^{*-1} = \lim_{N \rightarrow \infty} \frac{\mathbf{G} E[\mathbf{D}' \mathbf{D}] \mathbf{G}'}{N}. \quad (\text{A.9})$$

The result follows from the central limit theorem for vector martingale differences.  $\square$

**Proof of Proposition 1:** Note that

$$\begin{aligned} \hat{\beta} &= (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}' V \\ &= (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}' (\mathbf{D} \beta + S + e) \\ &= \beta + (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}' e + (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}' S \\ &= \beta + T_1 + T_2. \end{aligned} \quad (\text{A.10})$$

Considering  $T_2$ , analogously to Donoho et al.(1995) and Dahlhaus et al. (1999),

$$\begin{aligned} \sup \left( \sum_{j \geq J} \sum_k \|\beta_{j,k}^{(i)}\|^2 \right) &= O(T^{-\frac{2m_i}{2m_i+1}}), \quad (\text{A.11}) \\ T_2 &= (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}' S, \\ \|T_2\|_2 &\leq \|(\mathbf{D}' \mathbf{D})^{-1}\|_2 \|\mathbf{D}' S\|_2 \\ &= O_p \left( \left( 2^{-J \min\{\bar{s}\}_i} + T^{-1/2} 2^{-J \min\{m_i-1/2-1/(2p_i)\}} \right) \sqrt{\log(T)} \right) \\ &= O_p(T^{-1/2-\tau(J)}), \quad (\text{A.12}) \end{aligned}$$

where  $\tau(J) > 0$ , and then  $\sqrt{T} \|T_2\|_2 = O_p(T^{-\tau(J)}) = o_p(1)$ . Further, from Lemma 1,

$$\sqrt{T} T_1 = N \mathbf{H} (\mathbf{D}' \mathbf{D})^{-1} \mathbf{H}' \xrightarrow{P} \mathbf{\Gamma}, \quad (\text{A.13})$$

where

$$\mathbf{\Gamma} = \lim_{N \rightarrow \infty} N \mathbf{H} \mathbf{E}[(\mathbf{D}' \mathbf{D}^{-1})] \mathbf{H}' \quad (\text{A.14})$$

Applying Lemma 2 we have

$$\begin{aligned} \sqrt{T} \mathbf{H}(\hat{\beta} - \beta) &= \sqrt{T} T_1 + o_p(1), \\ \sqrt{T} \mathbf{H}(\hat{\beta} - \beta) &= Z + o_p(1), \end{aligned}$$

where  $Z$  is a random vector, which is multivariate normally distributed, with zero mean and covariance matrix  $\mathbf{\Gamma}$ . Hence, the result follows from the Slutsky theorem.  $\square$

## References

- Andrews, D.W. (1988). Laws of large number for dependent non-identically distributed random variables. *Econometric Theory*, **4**:458-67.
- Baccala L.A., Sameshima K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*. **8**:463-474.
- Bullmore E., Fadili J., Breakspear M., Salvador R., Suckling J., Brammer M. (2003). Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Stat Methods Med Res.*, **12**:375-99.
- Chambolle A., DeVore R.A., Lee N., Lucier B.J.(1998). Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, **7**:319–335.
- Chiann, C. and Morettin, P.A. (1999). Estimation of time-varying linear systems. *Statistical Inference for Stochastic Processes*, **2**:253-285.
- Chiann, C. and Morettin, P.A. (2005). Time Domain Nonlinear Estimation of Time Varying Linear Systems. *Journal of Nonparametric Statistics*, **17**:365-383.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.*, **25**:1-37.
- Dahlhaus, R., Neumann, M.H. and von Sachs, R. (1999). Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli*, **5**:873-906.

- Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics*, **28**:1762-94.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia. SIAM.
- DeLong M.R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends Neuroscience*, **13**:281-5.
- Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? (with discussion). *Journal of the Royal Statistical Society, Series B*, **57**: 301-337.
- Enders, W. (2005). *Applied Econometric Times Series*. Wiley Series in Probability and Statistics. John Wiley and Sons.
- Gackenheim C., Cayon L., Reifenberger R. (2006). Analysis of scanning probe microscope images using wavelets. *Ultramicroscopy*, **106**:389-97.
- Goebel R., Roebroek A., Kim D.S., Formisano E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Resonance Imag.*, **21**:1251-61.
- Granger C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**:424-38.
- Grinsted A., Jevrejeva S., Moore J. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Proc. Geophys.*, **11**:561-66.
- Hardle, W., Kerkyacharian, G., Picard, D., Tsybakov, A. (1997). *Wavelets approximations and statistical applications*. New York. Springer.
- Harrison L., Penny W.D., Friston K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage*, **19**:1477-91.
- Huerta G and West M (1997). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Association, Series B*, **61**:1-19.
- Ito K., Morrish P.K., Rakshi J.S., Uema T., Ashburner J., Bailey D.L., Friston K.J., Brooks D.J. (1999). Statistical Parametric Mapping with 18F-dopa PET shows bilaterally reduced striatal and nigral dopaminergic function in early Parkinson's disease. *Journal Neurol Neurosurg Psychiatry*, **66**:754-8.

- Jevrejeva S., Moore J.C., Grinsted A. (2003). Influence of the Arctic Oscillation and El Nin-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach. *J. Geophys. Res.*, **108**(D21), 4677.
- Krystal AD, Prado R, and West M (1999). New methods of time series analysis for nonstationary EEG data: Eigenstructure decompositions of time varying autoregressions. *Clinical Neurophysiology*, **110**: 1-10.
- Lawrence, C.J., Eichenbaum, M., Evans, C.L. (1999). *Monetary police shocks: what have we learned and to what end?*. Handbook of Macroeconomics, vol.1A:65:148. Amsterdam, Elsevier Science Ltd.
- Lu, P.J., Yogo, M., Marshall, C.R. (2006). Phanerozoic marine biodiversity dynamics in light of the incompleteness of the fossil record. *PNAS*, **13**:2736-39.
- Lutkepohl, H.(1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag. New York.
- Meyer, Y. (1993). *Wavelets: Algorithms and Applications*. Philadelphia: SIAM.
- Ogawa S, Lee TM, Nayak AS, Glynn P (1990). Oxigenation- sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Journal Magnetic Resonance in Medicine*, **14**:68-78.
- Paulik, M.J., Mohankrishnan, N., and Nikiforuk, M., (1994) A time varying vector autoregressive model for signature verification, *Proceedings of the 37th Midwest Symposium on Circuits and Systems*, **2**:3-5.
- Percival, DB and Walden, AT (2000). *Wavelet Methods for Time Series Analysis*. Cambridge, England: Cambridge University Press.
- Prado, G. Huerta and M. West. (2001) Bayesian time-varying autoregressions: Theory, methods and applications. *Resenhas Journal*, **4**:405-422.
- Primiceri GE (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies*, **3**:821-852.
- Sack, B. (2000). Does the FED act gradually? A VAR analysis. *Journal of Monetary economics*, **46**:229-56.
- Sato JR, Junior EA, Takahashi DY, de Maria Felix M, Brammer MJ, Morettin PA. (2006). A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. *Neuroimage*, **31**:187-96.

- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, **48**:1-48.
- Sims, C.A. (1982). Policy analysis with econometric models. *Brooking papers on economic activities*, **48**:1-48.
- Sims, C.A. (1998). Comment on Glenn Rudebush's 'Do measures of monetary policy in a VAR makes sense?'. *International Economic Review*, **39**:933-48.
- Stock, J.H. and Watson, W.M. (2001). Vector autoregressions. *Journal of Economic Perspectives*, **15**:101-15.
- Talairach J, Tornoux P (1988). *A co-planar stereotatic atlas of the human brain*. Stuttgart: Thieme.
- Triebel, H. (1992). *Theory of Function Spaces II*. Birkhauser.
- Vidakovic, B (1999). *Statistical Modeling by Wavelets*. Wiley Series in Probability and Statistics.
- West M. (1997). Time series decomposition. *Biometrika*, **84**:489-494.