On Characterizations of Spectral Density Matrices by Cross Validated Log Likelihood Criterion

Yasumasa Matsuda
Faculty of Economics,
Niigata University,
2-8050 Ikarahi, Niigata 950-2181, Japan
matsuda@econ.niigata-u.ac.jp

Yoshihiro Yajima
Faculty of Economics,
University of Tokyo,
7-3-1 Hongo , Bunkyo-ku, Tokyo 113-0033, Japan
yajima@e.u-tokyo.ac.jp

Howell Tong
Department of Statistics,
The London School of Economics and Political Science,
Houghton Street, London, WC2A 2AE, U.K.
and
Department of Statistics and Actual Science,
Hong Kong University
H.Tong@lse.ac.uk

Abstract

We propose the Cross Validated Log Likelihood (CVLL) criterion for selection of models characterizing spectral density matrices in a nonparametric or semiparametric way, which we call characterizations. Characterizations considered in this paper include many interesting special cases such as graphical models for multivariate time series. We show asymptotic properties of the CVLL, which derive the consistency of graphical modelling by the CVLL criterion. We demonstrate the empirical properties of the CVLL selection with both simulated and real data.

Key Words. multivariate time series; nonparametric; semiparametric; model selection; periodogram; conditional independence; partial correlation graph; graphical modelling; consistency.

1 Introduction

Let $\{X_{t,a}, a = 1, \ldots, r\}$ be an r-dimensional stationary time series and $f(\lambda), \ -\pi \le \lambda \le \pi$, be its spectral density matrix. Linear time series models such as multivariate autoregressive moving average (ARMA) models are included in a family, $\wp_1$, of parametric models for the spectral density matrix $f(\lambda)$:

$$\wp_1 \quad = \quad \{\{g(\theta, \lambda), \theta \in \Theta\} : g(\theta, \lambda) \in H[-\pi, \pi]\}, \tag{1}$$

where $H[-\pi, \pi]$ consists of $r \times r$ positive semi-definite Hermite matrix-valued integrable functions on $[-\pi, \pi]$.

Estimation of the parameter θ has been examined by many authors. See, for example, Dunsmuir (1979), Brockwell and Davis (1991, Section 10.8) and Hosoya (1997). Moreover, model selection criteria such as Akaike's Information Criterion (AIC, Akaike 1974) have been proposed for selecting the optimal model from among several candidate parametric models. Especially order determination for ARMA processes by AIC has been investigated extensively in the literature. For brief discussions of linear time series model selection, see Brockwell and Davis (1991, Section 9.3). The AIC is not a consistent order selection procedure for ARMA models, though it is asymptotically efficient for AR models (Shibata (1980), Hurvich and Tsai (1989)).

In this paper, we consider a family, $\wp_2$, of *characterizations* for the spectral density matrix $f(\lambda)$:

$$\wp_2 \quad = \quad \{\{G\left(\theta, \mathrm{vec}(f(\lambda))\right), \theta \in \Theta\} : G\left(\theta, \mathrm{vec}(f(\lambda))\right) \in H[-\pi, \pi]\}, \tag{2}$$

where $G = (G_{ab}),\ a, b = 1, \ldots, r$ is a matrix-valued function, and vec transforms an $r \times r$ matrix into an r^2-dimensional vector by stacking the column vectors. A parameter vector θ and all the components of $f(\lambda)$ are not necessarily included in the arguments of G, which means that the corresponding characterization is not a parametric model but a nonparametric or semiparametric constraint on the spectral density matrix. Hence we use the terminology *characterization* instead of *model* for (2) in order to distinguish (2) from parametric models (1). Several useful constraints on spectral density matrices such as graphical models for multivariate time series can be described in the form of (2), which will be introduced in Section 4. The object of this paper is to propose a method to find a suitable characterization for the spectral density matrix $f(\lambda)$.

Yajima and Matsuda (2004) proposed a method to test for one characterization G,

$$\mathrm{H}_0: \ G\left(\theta_0, \mathrm{vec}(f(\lambda))\right) \ = f(\lambda) \quad \text{vs.} \quad \mathrm{H}_1: \ G\left(\theta_0, \mathrm{vec}(f(\lambda))\right) \ \neq f(\lambda),$$

and derived the asymptotic null distribution. However, we often encounter the case when several alternatives are given. Graphical modelling stated in Section 5 is a typical case for which the test is not easy to apply because of several alternatives. Hence a criterion is required to evaluate the goodness of a characterization which would enable us to select the best characterization from among several candidates.

The existing criteria proposed for (1), however, cannot be applied to characterization selection in (2), since "the number of parameters" which penalizes model complexity is not easily defined for characterizations. The Cross Validation (CV), which was originally proposed by Stone (1974), has been used as an alternative criterion for parametric or nonparametric model selection. See, e.g., Shao (1993) for linear model selection, Kavalieris (1989) for order selection of AR models, Härdle *et al.*(1988) for a smoothing parameter determination of nonparametric regression models. CV penalizes the complexity of a model by the *leave-one-out*

approach instead of an explicit expression on model complexity, which makes it applicable to nonparametric model selection as well as parametric model selection.

In the context of nonlinear time series modelling, Cheng and Tong (1992) used the cross validated version of residual sum of squares as a criterion to estimate the order of nonparametric time series autoregressions, and showed the consistency of the order selection procedure. In this paper, we adopt the cross validated log (quasi) likelihood (CVLL) as a criterion for characterization selection. The distinctive feature is that our approach provides a consistency to some characterizations selection, e.g. graphical modelling for multivariate time series. In contrast to the inconsistency of CV selection for standard parametric models (Kavalieris, 1989, Shao, 1993), it should be emphasized that CV can provide consistent selection procedures in the nonparametric context: the order selection of nonparametric autoregressions based on the cross validated residual sum of squares in Cheng and Tong (1992) and graphical modelling based on the CVLL criterion in Section 5.

We introduce the CVLL criterion in Section 2, and show its asymptotic properties in Section 3. In Section 4, we give several examples of characterizations within the framework of (2). In Section 5, we apply the CVLL criterion to graphical modelling and prove that the CVLL selection of graphical models is consistent. In Section 6, we use simulation and real data to show empirical properties of the criterion. In the Appendix, we finally present proofs of Theorems and Lemmas.

2 The CVLL criterion

Throughout this paper, A_{ab} and A^{ab} are generic symbols for the (a, b)th element of the matrices A and A^{-1} respectively, and A' is the matrix transpose of A. Let $\{\mathbf{X}_t = (X_{t,a}, a = 1, \ldots, r)', -\infty < t < \infty\}$ be an r-dimensional stationary time series with the spectral density matrix $f(\lambda), -\pi \le \lambda \le \pi$. We define $f(\lambda)$ outside the region $[-\pi, \pi]$ to have a period of 2π.

Suppose we have observed $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, and a set, S, of candidate characterizations in (2) for the spectral density matrix $f(\lambda)$ is given. Define the discrete Fourier transform $W(\lambda)$ and the periodogram matrix $I(\lambda)$ of $\mathbf{X}_t$ as

$$\begin{aligned} W(\lambda) &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \mathbf{X}_t \exp(-i\lambda t), \\ I(\lambda) &= W(\lambda)\bar{W}(\lambda)'. \end{aligned}$$

$I(\lambda)$ is defined periodically with the period 2π for $\lambda \notin [-\pi, \pi]$. Let $\lambda_j = \frac{2\pi j}{n}, j = 1, \ldots, [n/2]$ be the jth Fourier frequency.

We shall introduce the CVLL of a candidate G in S. The matrix defined by

$$g(\lambda) \;:=\; G(\theta_0, \mathrm{vec}(f(\lambda))), -\pi \le \lambda \le \pi, \tag{3}$$

is called *the characterized density matrix* in the following. Then the CVLL for the G is defined by

$$\mathrm{CVLL}(G) = \sum_{j=1}^{[n/2]} \log\det\left(\hat{g}_{-j}(\lambda_j)\right) + \mathrm{tr}\left(I(\lambda_j)\hat{g}_{-j}^{-1}(\lambda_j)\right), \tag{4}$$

where

$$\begin{aligned} \hat{g}_{-j}(\lambda_j) &= G\left(\hat{\theta}, \mathrm{vec}(\hat{f}_{-j}(\lambda_j))\right), \\ \hat{f}_{-j}(\lambda_j) &= \frac{1}{m} \sum_{k=-m/2, k\neq 0}^{m/2} I(\lambda_{j+k}), \end{aligned} \tag{5}$$

and $\hat{\theta}$ is an estimator for θ_0. The $\hat{g}_{-j}(\lambda_j)$ and $\hat{f}_{-j}(\lambda_j)$ are the leave-one-out (or cross-validated) versions of $\hat{g}(\lambda_j)$ and $\hat{f}(\lambda_j)$, which include the term $k = 0$ and in which the divisor is replaced by $m+1$. Hurvich (1985) and Beltrao and Bloomfield (1987) used the CVLL criterion to determine the optimal bandwidth for univariate kernel spectrum estimates. The Hurvich's definition may be viewed as a special case of the CVLL but for a very different purpose of ours.

For the given set of candidate characterizations, we evaluate the CVLL of each candidate and minimize CVLL(G) over the set of candidates. We propose to choose the minimizer $\hat{G}$ as an estimator of the best characterization for $f(\lambda)$ among the candidates.

3 Asymptotic properties of the CVLL criterion

We shall show which candidate the CVLL criterion selects from the given set of candidates S asymptotically. Following the notation of Shao (1993), we classify S into two categories:

- Category I: $\{G \in S | G(\theta_0, \mathrm{vec}(f(\lambda))) \neq f(\lambda) \text{ for some } \lambda \in [-\pi, \pi]\}$.
- Category II: $\{G \in S | G(\theta_0, \mathrm{vec}(f(\lambda))) = f(\lambda) \text{ for all } \lambda \in [-\pi, \pi]\}$.

We shall prove two Theorems: Theorem 1 shows that the CVLL criterion selects a candidate in Category II from among S asymptotically, and Theorem 2 specifies the characterization in Category II which the CVLL criterion selects asymptotically.

These assumptions are required to prove the Theorems.

A1 $\{\mathbf{X}_t\}$ is an r-dimensional Gaussian stationary process.

A2 $f(\lambda)$ is positive definite for $-\pi \leq \lambda < \pi$.

A3 $f_{ab}(\lambda), a, b = 1, \ldots, r, -\pi \leq \lambda < \pi$ is twice continuously differentiable.

A4 $m = O(n^{\beta})$, $1/2 < \beta < 2/3$.

A5 $\hat{\theta} - \theta_0 = O_p\left(n^{-1/2}\right)$.

Theorem 1 *Under A1-A5, if $G_{ab}(\theta, y)$, which may be in either Category I or II, is continuously differentiable with respect to θ and y,*

$$\mathrm{CVLL}(G) = \sum_{j=1}^{[n/2]} (\log\det f(\lambda_j) + r) + \frac{n}{2\pi}\mathrm{KL}(f, g) + o_p(n), \tag{6}$$

where $\mathrm{KL}(f, g)$ is the Kullback-Leibler distance between the two matrix functions f and g given by

$$\mathrm{KL}(f, g) = \int_0^{\pi} \left[\mathrm{tr}\left(f(\lambda)g(\lambda)^{-1}\right) - \log\det\left(f(\lambda)g(\lambda)^{-1}\right) - r\right] d\lambda.$$

See Appendix for the proof of Theorem 1. Corollary I follows immediately from Theorem 1.

Corollary 1. *If G_1 is in Category I and G_2 is in Category II , then under the conditions in Theorem 1,*

$$\lim_{n\to\infty} P(\mathrm{CVLL}(G_1) > \mathrm{CVLL}(G_2)) = 1.$$

Corollary I assures us that the CVLL criterion always chooses a characterization in Category II asymptotically among S. Theorem 1, however, cannot specify which candidate in Category II the CVLL criterion selects asymptotically, since $\mathrm{KL}(f,g) = 0$ for all the characterizations in Category II. Before introducing Theorem 2, a criterion to distinguish among candidates in Category II is defined.

Definition 1(Asymptotic Mean Integrated Squared Error). *For a candidate characterization G in Category II, the asymptotic mean integrated squared estimation error of G ($\mathrm{AMISE}(G)$) is defined by*

$$\mathrm{AMISE}(G) \quad := \quad E_\infty \left[\frac{m}{n} \sum_{j=1}^{[n/2]} \mathrm{tr} \left\{ \left(\hat{g}_{-j}(\lambda_j) - f(\lambda_j) \right) f^{-1}(\lambda_j) \right\}^2 \right],$$

where E_∞ is the expectation with respect to the asymptotic distribution.

Theorem 2. *Suppose that a characterization G is in Category II and that $G_{ab}(\theta, y)$ is three times continuously differentiable with respect to θ and y. Then under A1-A5,*

$$\mathrm{CVLL}(G) \quad = \quad \sum_{j=1}^{[n/2]} \left[\log\det f(\lambda_j) + \mathrm{tr}\left(I(\lambda_j) f(\lambda_j)^{-1} \right) \right] + \frac{n}{m}\frac{1}{2}\mathrm{AMISE}(G) + o_p\left(\frac{n}{m} \right).$$

Moreover, $\mathrm{AMISE}(G)$ *is evaluated as*

$$\mathrm{AMISE}(G) \quad = \quad \frac{1}{2\pi} \int_0^\pi \sum_{\alpha,\beta,\gamma,\nu=1}^{r} \mu_{\alpha\beta\gamma\nu}(\lambda) f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda) d\lambda, \tag{7}$$

where

$$\begin{aligned} \mu_{\alpha\beta\gamma\nu}(\lambda) \quad &= \quad \sum_{a,b=1}^{r} \sum_{e_1,e_2=1}^{r} h_{ae_1,\alpha\beta}(\lambda) h_{be_2,\gamma\nu}(\lambda) f^{e_1 b}(\lambda) f^{e_2 a}(\lambda), \\ h_{ae_1,\alpha\beta}(\lambda) \quad &= \quad \left. \frac{\partial G_{ae_1}(\theta_0, y)}{\partial y_{\alpha\beta}} \right|_{y=\mathrm{vec}(f(\lambda))}. \end{aligned}$$

The proof of Theorem 2 is shown in Appendix. Corollary 2 is a direct consequence of Theorem 2.

Corollary 2. *If both G_1 and G_2 are in Category II and* $\mathrm{AMISE}(G_1) > \mathrm{AMISE}(G_2)$*, then under the conditions in Theorem 2,*

$$\lim_{n\to\infty} P(\mathrm{CVLL}(G_1) > \mathrm{CVLL}(G_2)) = 1.$$

Corollaries 1 and 2 establish that it is the characterization which has the smallest AMISE in Category II that the CVLL criterion selects among S asymptotically.

Remark. In choosing the bandwidth m for the nonparametric spectral estimator (5), any bandwidth satisfying Assumption A4 can be used theoretically. In practice, it seems natural to adopt the $m(G)$ minimizing $\mathrm{CVLL}(G)$ for a characterization G.

4 Examples of characterizations

In this section, we show some typical examples of characterizations G for an r-dimensional time series $\{X_{t,a}, a = 1, \ldots, r\}$. Moreover the values of $\mathrm{AMISE}(G)$ are derived according to

(7), which can be used to show an asymptotic comparison among characterizations in Category II by Corollary 2. In the following, for a condition C, let I_C take 1 if C is satisfied and 0 otherwise.

Example 1.(no constraint) This is the case when no constraints on a spectral density matrix are imposed.

$$G_{ab}(\mathrm{vec}(f(\lambda))) := f_{ab}(\lambda),\ a,b=1,\ldots,r. \tag{8}$$

This is the nonparametric case when no parameter vectors are included in the variables of G. We consider *no constraint* as the kind of characterization which is always in Category II. Let us derive the value of AMISE(G). Since $h_{ae_1,\alpha\beta}(\lambda) = I_{(a,e_1)=(\alpha,\beta)}$, we have $\mu_{\alpha\beta\gamma\nu}(\lambda) = f^{\beta\gamma}(\lambda)f^{\nu\alpha}(\lambda)$. Hence

$$\begin{aligned}
\mathrm{AMISE}(G) &= \frac{1}{2\pi}\int_0^\pi \sum_{\alpha,\beta,\gamma,\nu=1}^{r} f^{\beta\gamma}(\lambda)f^{\nu\alpha}(\lambda)f_{\alpha\nu}(\lambda)f_{\gamma\beta}(\lambda)d\lambda \\
&= \frac{1}{2}\sum_{\alpha=1}^{r} 1 \sum_{\beta=1}^{r} 1 = r^2/2.
\end{aligned}$$

Example 2.(mutual independence among sub-series) Suppose an r-dimensional time series is grouped into p independent sub-series, namely $\mathbf{X}_t = (X_{t,M_1},\ldots,X_{t,M_p})'$ for p disjoint sets $M_1 \oplus \ldots \oplus M_p = \{1,2,\ldots,r\}$ such that $\{X_{t,M_i}\}$ and $\{X_{t,M_j}\}$ are mutually independent for $i \neq j$. The characterization for the case is described as

$$G_{ab}(\mathrm{vec}(f(\lambda))) := \begin{cases} f_{ab}(\lambda), & \text{if } a \in M_i \text{ and } b \in M_i \text{ for some } i, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Put $E := \{(a,b), a \in M_i \text{ and } b \in M_i \text{ for some } i\}$. Since

$$h_{ae_1,\alpha\beta}(\lambda)f^{e_1b}(\lambda) = I_{(\alpha,\beta)\in E}I_{a=\alpha}I_{e_1=\beta}f^{e_1b}(\lambda),$$

we have $\mu_{\alpha\beta\gamma\nu}(\lambda) = I_{(\alpha,\beta)\in E}I_{(\gamma,\nu)\in E}f^{\beta\gamma}(\lambda)f^{\nu\alpha}(\lambda)$. By noting that the spectral density matrix is block diagonal,

$$\begin{aligned}
\mathrm{AMISE}(G) &= \frac{1}{2\pi}\int_0^\pi \sum_{(\alpha,\beta)\in E}\sum_{(\gamma,\nu)\in E} f^{\beta\gamma}(\lambda)f^{\nu\alpha}(\lambda)f_{\alpha\nu}(\lambda)f_{\gamma\beta}(\lambda)d\lambda \\
&= \sharp\{E\}/2 \\
&= \frac{\sharp\{M_1\}^2 + \ldots + \sharp\{M_p\}^2}{2},
\end{aligned}$$

where $\sharp$ means the cardinality of a set. In the special case when $M_i = \{i\},\ i=1,\ldots,r$, which means that the r components are mutually independent, AMISE(G) $= r/2$, since $\sharp\{M_i\} = 1$ for $i = 1,\ldots,r$.

These results establish the consistency of independent sub-series selection. Suppose that we have a set of candidate characterizations in the form of (9), which is classified into Categories I and II as in Section 3. Then it is easily seen by Corollaries 1 and 2 that the CVLL selects the characterization whose number of the independent components is the largest in Category II asymptotically.

Example 3. (separable correlations) A separable correlation is characterized by

$$\mathrm{cov}(X_{t,a}, X_{s,b}) = \sigma_{ab}\rho(t-s), \tag{10}$$

where $(\sigma_{ab}), a, b = 1, \ldots, r$ is an $r \times r$ positive definite matrix and ρ is a positive definite function on integers (Haslett and Raftery (1989), Martin (1990) and Matsuda and Yajima (2004)). Applying an inverse Fourier transform to the separable correlation, the spectral density matrix is

$$f_{ab}(\lambda) = \sigma_{ab}\tilde{f}(\lambda), \tag{11}$$

where $\tilde{f}(\lambda)$ is a scalar-valued nonnegative integrable function on $[-\pi, \pi]$. Hence the characterization is described as

$$G_{ab}(\sigma, \mathrm{vec}(f(\lambda))) := \frac{\sigma_{ab}}{r}\sum_{i=1}^{r}\frac{f_{ii}(\lambda)}{\sigma_{ii}},\ a, b = 1, \ldots, r. \tag{12}$$

This is the semiparametric case when the parameter vector $\sigma = (\sigma_{ab}, a, b = 1, \ldots, r)$ is included in the variables of G. Noting that

$$h_{ae_1,\alpha\beta}(\lambda)f^{e_1b}(\lambda) = I_{\alpha=\beta}\frac{\sigma_{ae_1}f^{e_1b}(\lambda)}{r\sigma_{\alpha\alpha}},$$

from (11) we have

$$\mu_{\alpha\beta\gamma\nu}(\lambda) = I_{\alpha=\beta}I_{\gamma=\nu}\frac{1}{r\sigma_{\alpha\alpha}\sigma_{\gamma\gamma}\tilde{f}(\lambda)^2}.$$

Hence

$$\mathrm{AMISE}(G) = \frac{1}{2r}\sum_{\alpha,\gamma=1}^{r}\frac{\sigma_{\alpha\gamma}^2}{\sigma_{\alpha\alpha}\sigma_{\gamma\gamma}}. \tag{13}$$

The interesting special case when $\sigma_{\alpha\gamma} = 0$ for $\alpha \neq \gamma$, which means that the components series are mutually independent with an identical autocorrelation, has AMISE(G) reducing to $1/2$.

Example 4. (graphical models) Originally, graphical modelling was developed for independent multivariate series. See Lauritzen (1996) for basic notations and definitions of graphical models. Recently, Dahlhaus (2000) extended the concept of undirected conditional independence graphs to multivariate time series. Since a graphical model is described by a kind of spectral constraint, the CVLL criterion can be applied to graphical modelling. In the next section, we discuss this application specifically.

Remark 1. The CVLL criterion can be applied to the following testing problem for a given characterization G:

$$\mathrm{H}_0:\ G(\theta_0, \mathrm{vec}(f(\lambda))) = f(\lambda) \quad \text{vs.} \quad \mathrm{H}_1:\ G(\theta_0, \mathrm{vec}(f(\lambda))) \neq f(\lambda).$$

Adopt H_0 if the CVLL of the G is smaller than that of *no constraint* (8), and adopt H_1 otherwise. Then the testing procedure is asymptotically consistent: when H_0 is true and AMISE(G) is smaller than $r^2/2$, the CVLL of the G is asymptotically smaller than that of *no constraint* by Theorem 2, and when H_1 is true, the CVLL of the G is larger than that of *no constraint* asymptotically by Theorem 1.

Remark 2. The expression for a characterization is sometimes not unique. When a characterization G is not expressed uniquely, the expression having the smallest AMISE(G) should be adopted. If an inefficient expression for the G was adopted, it could happen that the CVLL was not able to distinguish between the G and another characterization more inefficient than the

G. Let us give an example. A characterization for a separable correlation is expressed either as (12) or as

$$G_{ab}(\sigma, \mathrm{vec}(f(\lambda))) \quad := \quad \sigma_{ab}\frac{f_{11}(\lambda)}{\sigma_{11}}, \ a, b = 1, \dots, r, \tag{14}$$

and AMISE(G) of (14) is $r/2$. The expression (12) is better than (14) for discriminating a case where a separable correlation is embedded within other characterizations in Category II. Suppose that the characterizations (12), (14) and (9) with $M_i = \{i\}$, $i = 1, \dots, r$ are all in Category II, where the AMISE(G)s are $1/2$, $r/2$ and $r/2$, respectively. Then the CVLLs of (14) and (9) are asymptotically equivalent, while the CVLL of (12) is smaller than that of (9) asymptotically if $r > 1$. Though (12) and (14) are both expressions for a separable correlation, the latter is powerless to distinguish mutual independence with identical autocorrelations from mutual independence.

5 Application to graphical models

We will show in this section that the CVLL criterion is an effective tool for identifying the undirected graphical models for multivariate time series. First we define the conditional independence $X_{t,a}$ and $X_{t,b}$ given the rest of components. Let $Y_{t,ab} = \{X_{t,j}, j \neq a, b\}$.

$$X_a \perp X_b | Y_{ab} \Longleftrightarrow \mathrm{cov}(\varepsilon_{a|\{a,b\}^c}(s), \varepsilon_{b|\{a,b\}^c}(t)) = 0 \text{ for all } s, t \in Z,$$

where

$$\varepsilon_{a|\{a,b\}^c}(t) \quad = \quad X_{t,a} - \mu_a - \sum_u d_a(t-u)Y_{ab}(u),$$

which minimizes

$$E\left(X_{t,a} - \mu_a - \sum_{u=-\infty}^{\infty} d_a(t-u)Y_{ab}(u)\right)^2.$$

Now let us recall the definition of an undirected graph (V, E) due to Dahlhaus (2000).

Definition 2(partial correlation graph, Dahlhaus, 2000). *Let $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,r})'$ be a multivariate stationary time series and $V = \{1, \dots, r\}$ be the corresponding set of vertices. Let $(a, b) \notin E$ if and only if $X_a \perp X_b | Y_{ab}$. Then $G = (V, E)$ is called a partial correlation graph for time series.*

Dahlhaus(2000, Theorem 2.4) proved the following simple criterion for building graphical models. Let $f(\lambda)$, $-\pi \leq \lambda \leq \pi$ be the spectral density matrix of $\mathbf{X}_t$.

$$\begin{aligned} X_a \perp X_b | Y_{ab} \quad &\Longleftrightarrow \quad \mathrm{cov}(\varepsilon_{a|\{a,b\}^c}(s), \varepsilon_{b|\{a,b\}^c}(t)) = 0 \text{ for all } s, t \in Z \\ &\Longleftrightarrow \quad \frac{f^{ab}(\lambda)}{\sqrt{f^{aa}(\lambda)f^{bb}(\lambda)}} \equiv 0 \text{ for } -\pi \leq \lambda \leq \pi. \end{aligned} \tag{15}$$

By this result, the missing edges in the partial correlation graph can uniquely be identified from the zeros in the rescaled inverse of the spectral matrix. The spectral matrix is, however, unknown in practice and should be replaced with some suitable estimator $\hat{f}(\lambda)$. Dahlhaus (2000) proposed the test statistic

$$\max_{1 \leq j \leq [n/2]} \left| \frac{\hat{f}^{ab}(\lambda_j)}{\sqrt{\hat{f}^{aa}(\lambda_j)\hat{f}^{bb}(\lambda_j)}} \right|^2$$

to detect the zeros. As he mentioned, it is very difficult to determine the null distribution under the conditional independence, which is required to test whether (15) is true or not.

Instead we apply the CVLL criterion to this problem. Let (V, E) be a partial correlation graph of $\{X_{t,a}, a = 1, \ldots, r\}$. Since the graph is equivalent to the spectral constraint (15) for $(a, b) \notin E$, the characterized density $g(\lambda)$ by the graph (V, E) should satisfy

$$\begin{aligned} g^{ab}(\lambda) &= 0, && \text{for } (a,b) \notin E, \\ g_{ab}(\lambda) &= f_{ab}(\lambda), && \text{for } (a,b) \in E. \end{aligned} \tag{16}$$

The characterization G satisfying (16) cannot be expressed explicitly as a function of $f(\lambda)$. Wermuth and Scheidt (1977) introduced the following recursion. Speed and Kiiveri (1986) showed that the recursion was a special case of the *First Cyclic Algorithm* and proved that $g(\lambda)$ was the limit of the recursion. Suppose that the graph misses M edges $F_i, i = 0, \ldots, M-1$, i.e. $E^c = F_0 \oplus \cdots \oplus F_{M-1}$. Let $g_{n,ab,\lambda}$ and $g_{n,\lambda}^{ab}$ be the abth element of $g_n(\lambda)$ and $g_n^{-1}(\lambda)$ of the nth recursion, respectively. Set $g_{0,ab,\lambda} = f_{ab}(\lambda), a, b = 1, \ldots, r$, and $n' = n(\text{mod } M)$. Then

$$\begin{aligned} g_{n,ab,\lambda} &= g_{n-1,ab,\lambda} + \frac{g^{ab}_{n-1,\lambda}}{g^{aa}_{n-1,\lambda} g^{bb}_{n-1,\lambda} - g^{ab}_{n-1,\lambda} g^{ba}_{n-1,\lambda}}, && \text{for } (a,b) \in F_{n'}, \\ g_{n,ab,\lambda} &= g_{n-1,ab,\lambda}, && \text{for } (a,b) \notin F_{n'}, \end{aligned} \tag{17}$$

for $n = 1, 2, \ldots$.

Suppose that we have a set of candidate graphs for a spectral density matrix $f(\lambda)$ of an r-dimensional time series, which is classified into Categories I and II in the same way as in Section 3. Let us evaluate AMISE(G) for a candidate graph G in Category II.

Theorem 3. *If a characterization G for a partial correlation graph (V, E) of the r-dimensional time series is in Category II,* CVLL(G) *is evaluated as in Theorem 2 under A1-A5 and*

$$\text{AMISE}(G) \quad = \quad \frac{r^2 - 2\sharp\{E^c\}}{2},$$

where $\sharp\{E^c\}$ means the number of missing edges of the graph (V, E).

As a result of Theorems 1 and 3, Corollary 3 follows immediately.

Corollary 3. *Let (V_i, E_i), $i = 1, 2$ be two partial correlation graphs whose characterizations are G_i,, $i = 1, 2$. Under A1-A5,*

1. *if G_1 is in Category I and G_2 is in Category II, then*

$$\lim_{n\to\infty} P(\text{CVLL}(G_1) > \text{CVLL}(G_2)) = 1;$$

2. *if the two characterizations are in category II and $E_1 \supset E_2$, then*

$$\lim_{n\to\infty} P(\text{CVLL}(G_1) > \text{CVLL}(G_2)) = 1.$$

Corollary 3 establishes the consistency of the CVLL selection for graphical modelling: it is the optimal graph having the smallest $\sharp\{E\}$ in Category II that the CVLL criterion selects asymptotically. In practice, either a backward or a forward elimination by the CVLL criterion can detect the optimal graph asymptotically. The backward elimination is, however, introduced for use in the next section, since it is easily seen that the backward one is more effective and accurate in the recursion (17) than the forward one.

Step 0. Put $k = 0$, set G_0 as the *no constraint* given in (8) and calculate CVLL(G_0).

Step 1. Calculate the CVLLs of candidates $G^i_{k+1},\ i = 1, \ldots, \begin{pmatrix} r \\ 2 \end{pmatrix} - k$, each of which misses one edge from the existing edges of the model G_k.

Step 2. Select the best model G_{k+1} minimizing the CVLL from among the candidates.

Step 3. If CVLL$(G_{k+1}) <$ CVLL(G_k), set $k = k + 1$ and return to Step 1. Otherwise, stop the procedure and G_k is our detected model.

Remark 1. In the *First Cyclic Algorithm*(FCA), Speed and Kiiveri (1986) set $F_i, i = 0, \ldots, M-1$ as the off-diagonal elements of cliques of the complimentary graph E^c instead of the missing edges. The algorithm (17) by Wermuth and Scheidt (1977) is a special case of the FCA (Speed and Kiiveri, 1986, 146-147). The FCA is more effective and accurate than (17), since M, which is the number of the cycle, can be made smaller in the FCA. Equation (17) is, however, more suited to the backward elimination mentioned above than the FCA, since a clique is difficult to find because of its geometrical property. This is the reason why we adopt the algorithm (17) instead of the FCA.

Remark 2. The CVLL criterion can also be applied to the graphical modelling for independent series in exactly the same way, since an independent series is considered to be a time series with a constant spectral density matrix.

6 Empirical studies

In this section, our interest is focused on the empirical properties of the CVLL criterion for spectral characterizations. We will use simulation, and the gold price series in five countries, to examine the properties.

First we apply the CVLL criterion to the characterization for independence of component series (9). Consider a three dimensional multivariate time series generated by

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.2 & x & 0.0 \\ x & -0.2 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \varepsilon_{t,3} \end{pmatrix}, \tag{18}$$

where $\{\varepsilon_{t,a}, a = 1, \ldots, 3\}$ is a sequence of normally distributed random vectors with mean 0 and variance matrix E_3. Note that $\{X_{t,1}, X_{t,2}\}$ and $\{X_{t,3}\}$ are independent for $x \neq 0$, and $\{X_{t,1}\}$,$\{X_{t,2}\}$ and $\{X_{t,3}\}$ are mutually independent for $x = 0$. We consider the following four characterization as candidates:

Case I: *no constraint* (8).
Case II : $M_1 = \{1\}, M_2 = \{2, 3\}$ in (9).
Case III: $M_1 = \{1, 2\}, M_2 = \{3\}$ in (9).
Case IV: $M_1 = \{1\}, M_2 = \{2\}, M_3 = \{3\}$ in (9).

All the cases except for Case II are in Category II for $x \neq 0$, and all the cases are in Category II for $x = 0$.

We selected the case which has the smallest CVLL for the simulated series by (18) 500 times. We show the empirical frequencies of the CVLL selection in Table 1 for $x = 0.0, 0.1$ and 0.2 when the sample sizes are 101, 201 and 401. Table 1 shows the consistency of the CVLL selection clearly. The CVLL tends to select the efficient characterizations: Case III for $x \neq 0$ and Case IV for $x = 0$ as the sample sizes get larger.

sample size	x	Case I	Case II	Case III	Case IV
	0.0	0.018	0.176	0.178	0.628
101	0.1	0.034	0.138	0.302	0.526
	0.2	0.090	0.074	0.682	0.154
	0.0	0.022	0.156	0.138	0.684
201	0.1	0.048	0.082	0.416	0.454
	0.2	0.086	0.006	0.858	0.050
	0.0	0.016	0.106	0.118	0.760
401	0.1	0.036	0.044	0.590	0.330
	0.2	0.044	0.000	0.952	0.004

Table 1: The empirical frequencies of selection by the CVLL criterion based on 500 replications. The Case which had the smallest CVLL was selected among Cases I-IV for time series (18).

Next we apply the CVLL criterion to graphical modelling using simulated series. Let us consider the following vector autoregressive model:

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \\ X_{t,4} \\ X_{t,5} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.0 & 0.3 & 0.0 & 0.3 \\ 0.3 & -0.2 & x & 0.0 & 0.0 \\ 0.2 & x & 0.3 & 0.0 & 0.0 \\ 0.2 & 0.3 & 0.0 & 0.3 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \\ X_{t-1,4} \\ X_{t-1,5} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \varepsilon_{t,3} \\ \varepsilon_{t,4} \\ \varepsilon_{t,5} \end{pmatrix}, \tag{19}$$

where $\{\varepsilon_{t,a}, a = 1, \ldots, 5\}$ is a sequence of normally distributed random vectors with mean 0 and variance matrix E_5. By Dahlhaus (2000, page 164), this process provides an example of a graph (V, E) such that

$$E^c = \begin{cases} \{(2,5), (3,4)\}, & \text{if } x \neq 0, \\ \{(2,5), (3,4), (2,3)\}, & \text{if } x = 0. \end{cases}$$

We consider the following four candidate graphs $(V, E_i), i = 1, \ldots, 4$:

Graph I: $E_1^c = \phi$, which is equivalent to *no constraint*.
Graph II: $E_2^c = \{(2,5)\}$.
Graph III: $E_3^c = \{(2,5), (3,4)\}$.
Graph IV: $E_4^c = \{(2,5), (3,4), (2,3)\}$.

Note that all the graphs except for Graph IV are in Category II for $x \neq 0$, and all the graphs are in Category II for $x = 0$.

We selected the graph which has the smallest CVLL from Graphs I-IV for simulated time series by the autoregressive model (19) 500 times. We show the frequencies of the CVLL selection in Table 2 when $x = 0.0, 0.1$ and 0.2 and sample sizes are 101,201 and 401. Table 2 supports the consistency of the CVLL selection for graphical modelling. The CVLL tends to select the efficient graphs: Graph III for $x \neq 0$ and Graph IV for $x = 0$, as the sample sizes increase.

Finally, we consider a seven-dimensional multivariate time series of pollutants, weather and daily hospital admissions of patients suffering from circulatory and respiratory problems. The pollutant and weather data are the daily average levels of sulphur dioxide (SO_2,

sample size	x	Graph I	Graph II	Graph III	Graph IV
	0.0	0.020	0.050	0.134	0.796
101	0.1	0.018	0.064	0.244	0.674
	0.2	0.022	0.074	0.602	0.302
	0.0	0.010	0.056	0.122	0.812
201	0.1	0.026	0.062	0.342	0.570
	0.2	0.058	0.076	0.786	0.080
	0.0	0.006	0.016	0.078	0.900
401	0.1	0.026	0.054	0.492	0.428
	0.2	0.018	0.070	0.912	0.000

Table 2: The empirical frequencies of selection by the CVLL criterion based on 500 replications. The Graph which had the smallest CVLL was selected among Graphs I-IV for time series (19).

μg m^{-3}), nitrogen dioxide (NO_2, μg m^{-3}), respirable suspended particulates (μg m^{-3}), ozone (O_3, μg m^{-3}), temperature (Celsius) and relative humidity (%), which were collected daily in Hong Kong from January 1st, 1994, to December 31st, 1995, and are shown in Figure 1. Taking the admissions series as the response variable and the other series as the explanatory variables, Xia *et al.* (2002) analyzed the data by a semiparametric regression model. Their method is motivated by the non-linearity expected to exist in some parts of the very complex weather-pollutant interaction.

Here , we explore the extent to which our simple linear-based technique can yield useful results even in a complex situation. We applied graphical models to the data. All the series were detrended by extracting the 15 days moving averages. For the hospital admissions series, the trend component caused by the day of the week effect (Xia *et al.*, 2002, page 378) was removed by a simple regression method using dummy variables. For the seven-dimensional detrended time series, we estimated the spectral coherency $|f_{ab}(\lambda_j)|/\sqrt{f_{aa}(\lambda_j)f_{bb}(\lambda_j)},\ a,b=1,\dots,7$, and the spectral partial coherency $|f^{ab}(\lambda_j)|/\sqrt{f^{aa}(\lambda_j)f^{bb}(\lambda_j)},\ a,b=1,\dots,7$, which are shown in Figure 2. The alignment technique for the coherency estimation (e.g., Priestley (1981, Section 9.5.4)) was used to increase the estimation accuracy, though it was not used for the partial coherency estimation, since the technique does not preserve the positive semi-definiteness of the estimated density matrix required for the matrix inversion. In Figure 3, we show the partial correlation graph estimated with the backward elimination mentioned at the end of Section 5.

Figures 2 and 3 suggest the following features.

1. The number of the hospital admissions (HA) is conditionally independent of SO_2, humidity and particulates, while it is conditionally dependent on NO_2, O_3 and temperature especially at low frequencies, which suggests a long range dependence between HA and the latter pollutant variables. Xia and Tong (2004) used a weighted average of the past levels of the pollutant variables up to time period L(they took $L = 365$) for the explanatory variables. Our observation supports the validity of their cumulative model.

2. Temperature is the most influential among the three variables dependent conditionally on HA, which reinforces the observation of Xia *et al.* (2002) that weather has an important role to play for HA.

3. Humidity is conditionally dependent strongly on NO_2, O_3 and particulates. Humidity has

something to do with the process of boosting the impact of the above pollutants under certain conditions, and is therefore considered to be an indirect but principal factor for HA.

4. Figure 3 supports the fact that NO_2 plays an important role in the process of O_3 generation (Dahlhaus, 2000,page 168).

Discussion

This paper proposes a procedure for characterization selection by the CVLL criterion. Considering undirected graphs to be an example of characterizations, we provide a consistent method for graphical modelling of multivariate time series. Interesting extensions of graphical modelling are to the following two cases. One is directed graphs which can detect causal relationships among variables. Undirected graphs only describe mutual relations which cannot make clear which variables are the cause and which the effect. The extension to directed graphical modelling is a challenging problem. The other is nonlinear undirected graphical modelling. Our approach considers only linear relationships among variables, and is powerless to detect nonlinear relationships, e.g., the nonlinear dependency between the hospital admissions and the pollutant variables identified by Xia *et al.*(2002). We leave the interesting extensions of the graphical modelling to future studies.

Acknowledgement

This manuscript was prepared while the first author was visiting the London School of Economics and Political Science as a research fellow, financed by the Japanese Ministry of Education, Culture, Sports, Science and Technology. We thank the Hong Kong Research Grant Council (Grant # HKU 1049/03P) and the University of Hong Kong (Distinguished Research Achievement Award) for partial supports to the third author. We thank Dr. C. M. Wong of the University of Hong Kong for making the data available to us.

APPENDIX

Proofs

Proof of Theorem 1. For notational simplicity, put $W_j = W(\lambda_j),\ I_j = I(\lambda_j),\ f_j = f(\lambda_j),\ g_j = g(\lambda_j),\ \hat{f}_{-j,j} = \hat{f}_{-j}(\lambda_j),\ \hat{g}_{-j,j} = \hat{g}_{-j}(\lambda_j)$.

$$\begin{aligned}\mathrm{CVLL}(G) &= \sum_{j=1}^{[n/2]} [\log\det f_j + r] + \sum_{j=1}^{[n/2]} \left[\mathrm{tr}\left(f_j g_j^{-1}\right) - \log\det\left(f_j g_j^{-1}\right) - r\right] \\ &\quad + \sum_{j=1}^{[n/2]} \left[\log\det\left(\hat{g}_{-j,j} g_j^{-1}\right) + \mathrm{tr}\left(\hat{g}_{-j,j}^{-1} - g_j^{-1}\right) f_j \right. \\ &\quad \left. + \mathrm{tr}\left(I_j - f_j\right)\left(\hat{g}_{-j,j}^{-1} - g_j^{-1}\right) + \mathrm{tr}\left(I_j - f_j\right) g_j^{-1}\right]. \end{aligned} \tag{20}$$

We will show that each of the last four terms is $o_p(n)$. Then Theorem 1 is proved since the second term is $\frac{n}{2\pi}\mathrm{KL}(f,g) + o(n)$.

Put $D_j = \left(\hat{g}_{-j,j} - g_j\right) g_j^{-1}$. Since $\max_j |D_{j,ab}| = o_p(n^{-1/4})$ for $a, b = 1, \ldots, r$ by Proposition 1 of Matsuda and Yajima (2004),

$$\begin{aligned}\sum_j \log\det\left(\hat{g}_{-j,j} g_j^{-1}\right) &= \sum_j \log\det(D_j + E) \\ &= O\left(\sum_j |\mathrm{tr}(D_j)|\right) \\ &= o_p(n).\end{aligned}$$

In the same way,

$$\begin{aligned}\sum_j \mathrm{tr}\left(\hat{g}_{-j,j}^{-1} - g_j^{-1}\right) f_j &= \sum_j \mathrm{tr}\left(\hat{g}_{-j,j}^{-1}\left(g_j - \hat{g}_{-j,j}\right) g_j^{-1} f_j\right) \\ &= o_p(n).\end{aligned}$$

By Lemma 1 of Yajima and Matsuda (2004), $\sum_j \mathrm{tr}(I_j - f_j)g_j^{-1} = \sum_j \mathrm{tr}(I_j - E(I_j))g_j^{-1} + O(\log n)$, and

$$\begin{aligned}E\left(\sum_j \mathrm{tr}(I_j - E(I_j))g_j^{-1}\right)^2 &= \sum_{a,b,c,d}\sum_{j,k} E\left(I_{j,ab} - E(I_{j,ab})\right)\left(I_{k,cd} - E(I_{k,cd})\right) g_j^{ba} g_k^{dc} \\ &= \sum_{a,b,c,d}\sum_j f_{j,ad} f_{j,cb} g_j^{ba} g_j^{dc} + O(\log n) \\ &= O(n).\end{aligned}$$

Hence $\sum_j \mathrm{tr}(I_j - f_j)g_j^{-1} = O_p(n^{1/2})$. Finally, by Proposition 1 of Matsuda and Yajima (2004),

$$\begin{aligned}\left|\sum_j \mathrm{tr}\left(I_j - f_j\right)\left(\hat{g}_{-j,j}^{-1} - g_j^{-1}\right)\right| &\leq \sum_{a,b} \max_j \left|\hat{g}_{-j,j}^{ba} - g_j^{ba}\right| \sum_j \left|I_{j,ab} - f_{j,ab}\right| \\ &= o_p(n^{-1/4})O_p(n) \\ &= o_p(n).\end{aligned}$$

Proof of Theorem 2. Since $f_j = g_j$ for all $j = 1, \ldots, [n/2]$, the second term of (20) is 0 and we have

$$\begin{aligned}\mathrm{CVLL}(G) &= \sum_{j=1}^{[n/2]}\left[\log\det f_j + r + \mathrm{tr}\left(I_j - f_j\right) f_j^{-1}\right] \\ &\quad + \sum_{j=1}^{[n/2]}\left[\log\det\left(\hat{g}_{-j,j} f_j^{-1}\right) + \mathrm{tr}\left(\hat{g}_{-j,j}^{-1} - f_j^{-1}\right) f_j\right] \\ &\quad + \sum_{j=1}^{[n/2]} \mathrm{tr}\left(I_j - f_j\right)\left(\hat{g}_{-j,j}^{-1} - f_j^{-1}\right).\end{aligned} \tag{21}$$

First we show that the third term is $o_p(n/m)$. For simplicity, we give a proof for $r = 1$ when θ is one dimensional, but the general case is proved essentially in the same way. By applying the Taylor's expansion,

$$\begin{aligned}\hat{g}_{-j,j}^{-1} - f_j^{-1} &= G^{-1}(\hat{\theta}, \hat{f}_{-j,j}) - G^{-1}(\theta_0, f_j) \\ &= \sum_{i+k=1, i\geq 0, k\geq 0}^{2} \frac{\partial^{i+k} G^{-1}(\theta_0, f_j)}{\partial\theta^i \partial y^k}(\hat{\theta} - \theta_0)^i (\hat{f}_{-j,j} - f_j)^k \\ &\quad + \sum_{i+k=3, i\geq 0, k\geq 0} \frac{\partial^{i+k} G^{-1}(\theta^*, y^*)}{\partial\theta^i \partial y^k}(\hat{\theta} - \theta_0)^i (\hat{f}_{-j,j} - f_j)^k, \\ &= \sum_{i+k=1, i\geq 0, k\geq 0}^{2} M_{ik,j}(\hat{\theta} - \theta_0)^i (\hat{f}_{-j,j} - f_j)^k + \sum_{i+k=3, i\geq 0, k\geq 0} N_{ik,j}(\hat{\theta} - \theta_0)^i (\hat{f}_{-j,j} - f_j)^k, \text{ say,}\end{aligned}$$

where y^* and θ^* are mean values of $(f_j, \hat{f}_{-j,j})$ and $(\theta_0, \hat{\theta})$, respectively. Note that $N_{ik,j}$ is stochastic, though $M_{ik,j}$ is non-random. We will show that

$$\sum_{j=1}^{[n/2]} \left(I_j - f_j\right) M_{10,j}(\hat{\theta} - \theta_0) = o_p(n/m),$$

$$\sum_{j=1}^{[n/2]} (I_j - f_j)\, M_{01,j}(\hat{f}_{-j,j} - f_j) = o_p(n/m),$$
$$\sum_{j=1}^{[n/2]} (I_j - f_j)\, N_{03,j}(\hat{f}_{-j,j} - f_j)^3 = o_p(n/m),$$

which complete the proof for the third term of (21) being $o_p(n/m)$, since the three terms dominate the others. First,

$$\begin{aligned}\left|\sum_j (I_j - f_j) M_{10,j}(\hat{\theta} - \theta_0)\right| &= \left|\hat{\theta} - \theta_0\right| \left|\sum_j (I_j - f_j) M_{10,j}\right| \\ &= O_p\left(n^{-1/2}\right) O_p\left(n^{1/2}\right) = O_p(1).\end{aligned}$$

Next, by Lemma 2 of Yajima and Matsuda(2004), we have

$$\sum_j (I_j - f_j) M_{01,j}(\hat{f}_{-j,j} - f_j) = \sum_j (I_j - E(I_j)) M_{01,j}(\hat{f}_{-j,j} - E(\hat{f}_{-j,j})) + O_p(m^2/n),$$

and by Lemma 1 of Yajima and Matsuda (2004) we have

$$\begin{aligned}&E\left\{\sum_j (I_j - E(I_j)) L_{1,j}(\hat{f}_{-j,j} - E(\hat{f}_{-j,j}))\right\}^2 \\ =\ & \frac{1}{m^2}\sum_{j_1,j_2}\sum_{k_1,k_2\neq 0} M_{01,j_1} M_{01,j_2} E\left[(I_{j_1} - E(I_{j_1}))(I_{j_2} - E(I_{j_2}))(I_{j_1+k_1} - E(I_{j_1+k_1}))(I_{j_2+k_2} - E(I_{j_2+k_2}))\right] \\ =\ & \frac{1}{m^2}\sum_{j_1,j_2}\sum_{k_1,k_2\neq 0} M_{01,j_1} M_{01,j_2} \left(\mathrm{cum}(I_{j_1}, I_{j_2})\mathrm{cum}(I_{j_1+k_1}, I_{j_2+k_2})\right. \\ & \left. +\mathrm{cum}(I_{j_1}, I_{j_1+k_1})\mathrm{cum}(I_{j_2}, I_{j_2+k_2}) + \mathrm{cum}(I_{j_1}, I_{j_2+k_2})\mathrm{cum}(I_{j_2}, I_{j_1+k_1})\right) \\ =\ & O\left(1/m^2 nm\right) + O\left(1/m^2 n^2 m^2 \log^2 / n^2\right) + O\left(1/m^2 nm\right) = O\left(n/m\right).\end{aligned}$$

It follows that $\sum_j (I_j - f_j) L_{1,j}(\hat{f}_{-j,j} - f_j) = O_p\left(\sqrt{\frac{n}{m}}\right) + O_p\left(\frac{m^2}{n}\right) = o_p(\frac{n}{m})$. Finally, by Proposition 1 of Matsuda and Yajima (2004), we have

$$\begin{aligned}\left|\sum_j (I_j - f_j) N_{03,j}(\hat{f}_{-j,j} - f_j)^3\right| &\leq \max_j |N_{03,j}| \sqrt{\sum_j (I_j - f_j)^2}\sqrt{\sum_j (\hat{f}_{-j,j} - f_j)^6} \\ &= \sqrt{O_p\left(n\right)}\sqrt{o_p\left(nn^{-3/2}\right)} = o_p(n^{1/4}).\end{aligned}$$

Put $\delta_j = (\hat{g}_{-j,j} - f_j)\, f_j^{-1}$. Then the second term of (21) is

$$\begin{aligned}&\sum_j \log\det\left(\delta_j + E\right) - \mathrm{tr}\left[\delta_j(\delta_j + E)^{-1}\right] \\ =\ & \sum_j \mathrm{tr}\left(\delta_j - \frac{1}{2}\delta_j^2\right) - \mathrm{tr}\left(\delta_j(E - \delta_j)\right) + O\left(\sum_j \max_{a,b}\left|\delta_{j,ab}^3\right|\right) \\ =\ & \frac{1}{2}\sum_j \mathrm{tr}\left(\delta_j^2\right) + O_p\left(n^{1/4}\right).\end{aligned}$$

By the Taylor's expansion, we have

$$\sum_j \mathrm{tr}(\delta_j^2)$$

$$
\begin{aligned}
&= \sum_j \sum_{a,b} \delta_{j,ab}\delta_{j,ba} \\
&= \sum_j \sum_{a,b} \sum_{e_1,e_2} \left(\hat{g}_{-j,j,ae_1} - f_{j,ae_1}\right)\left(\hat{g}_{-j,j,be_2} - f_{j,be_2}\right) f_j^{e_1 b} f_j^{e_2 a} \\
&= \sum_j \sum_{a,b} \sum_{e_1,e_2} \left(\frac{\partial G_{ae_1}(\theta_0, f_j)}{\partial \theta}(\hat{\theta}-\theta_0) + \sum_{\alpha,\beta} \frac{\partial G_{ae_1}(\theta_0, f_j)}{\partial y_{\alpha\beta}}(\hat{f}_{-j,j,\alpha\beta} - f_{j,\alpha\beta}) \right) \\
&\quad \times \left(\frac{\partial G_{be_2}(\theta_0, f_j)}{\partial \theta}(\hat{\theta}-\theta_0) + \sum_{\gamma,\nu} \frac{\partial G_{be_2}(\theta_0, f_j)}{\partial y_{\gamma\nu}}(\hat{f}_{-j,j,\gamma\nu} - f_{j,\gamma\nu}) \right) f_j^{e_1 b} f_j^{e_2 a} + O\left(\sum_j \max_{a,b} \left| \hat{f}_{-j,j,ab} - f_j \right|^3 \right) \\
&= \sum_j \sum_{a,b} \sum_{e_1,e_2} \sum_{\alpha,\beta} \sum_{\gamma,\nu} h_{ae_1,\alpha\beta,j} h_{be_2,\gamma\nu,j} \left(\hat{f}_{-j,j,\alpha\beta} - f_{j,\alpha\beta} \right) \left(\hat{f}_{-j,j,\gamma\nu} - f_{j,\gamma\nu} \right) f_j^{e_1 b} f_j^{e_2 a} + o_p(n^{1/4}) \\
&= \sum_j \sum_{a,b} \sum_{e_1,e_2} \sum_{\alpha,\beta} \sum_{\gamma,\nu} h_{ae_1,\alpha\beta,j} h_{be_2,\gamma\nu,j} \left(\hat{f}_{-j,j,\alpha\beta} - E(\hat{f}_{-j,j,\alpha\beta}) \right) \left(\hat{f}_{-j,j,\gamma\nu} - E(\hat{f}_{-j,j,\gamma\nu}) \right) f_j^{e_1 b} f_j^{e_2 a} \\
&\quad + o_p\left(m^2 n^{-5/4}\right) + o_p(n^{1/4}),
\end{aligned}
$$

where we have used Proposition 1 of Matsuda and Yajima (2004) for the fourth equality and Lemma 2 of Yajima and Matsuda (2004) for the fifth equality. Let us define

$$
\begin{aligned}
\mu_{\alpha\beta\gamma\nu,j} &:= \sum_{a,b} \sum_{e_1,e_2} h_{ae_1,\alpha\beta,j} h_{be_2,\gamma\nu,j} f_j^{e_1 b} f_j^{e_2 a} \\
y_{\alpha\beta,j} &:= \hat{f}_{-j,j,\alpha\beta} - E(\hat{f}_{-j,j,\alpha\beta}).
\end{aligned}
$$

Then it will complete the proof of Theorem 2 to prove that

$$
E\left(\sum_j \sum_{\alpha,\beta,\gamma,\nu} \mu_{\alpha\beta\gamma\nu,j} y_{\alpha\beta,j} y_{\gamma\nu,j} \right) = \frac{n}{m}\mathrm{AMISE}(G) + o(n/m), \tag{22}
$$

$$
\mathrm{var}\left(\sum_j \sum_{\alpha,\beta,\gamma,\nu} \mu_{\alpha\beta\gamma\nu,j} y_{\alpha\beta,j} y_{\gamma\nu,j} \right) = O\left(n/m\right). \tag{23}
$$

By Lemma 1 of Yajima and Matsuda (2004), we have

$$
\begin{aligned}
E(y_{\alpha\beta,j} y_{\gamma\nu,j}) &= \frac{1}{m^2} \sum_{k,l \neq 0} E(I_{\alpha\beta,j+k} I_{\gamma\nu,j+l}) - E(I_{\alpha\beta,j+k}) E(I_{\gamma\nu,j+l}) \\
&= \frac{1}{m^2} \sum_{k,l \neq 0} E\left(W_{\alpha,j+k} \bar{W}_{\beta,j+k} W_{\gamma,j+l} \bar{W}_{\nu,j+l} \right) - E(I_{\alpha\beta,j+k}) E(I_{\gamma\nu,j+l}) \\
&= \frac{1}{m^2} \sum_{k \neq 0} f_{\alpha\nu,j+k} f_{\gamma\beta,j+k} + O\left(\log n / n\right) \\
&= \frac{1}{m} f_{\alpha\nu,j} f_{\gamma\beta,j} + O\left(m/n^2\right) + O(\log n/n).
\end{aligned}
$$

The last equality is given by the Taylor's expansion. It follows from Exercise 1.7.14 of Brillinger (1981) that the left hand side of (22) is

$$
\begin{aligned}
&\sum_{\alpha,\beta,\gamma,\nu} \sum_j \mu_{\alpha\beta\gamma\nu,j} \frac{1}{m} f_{\alpha\nu,j} f_{\gamma\beta,j} + O(\log n) \\
&= \frac{n}{m} \frac{1}{2\pi} \int_0^\pi \sum_{\alpha,\beta,\gamma,\nu} \mu_{\alpha\beta\gamma\nu}(\lambda) f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda) d\lambda + o(n/m).
\end{aligned}
$$

We can put $\mu_{\alpha\beta\gamma\nu,j} = 1$ without loss of generality. By Lemmas 3, 4 of Yajima and Matsuda (2004), the left hand side of (23) is evaluated as

$$
\begin{aligned}
\mathrm{var}\left(\sum_j y_{\alpha\beta,j} y_{\gamma\nu,j}\right) &= \mathrm{cov}\left(\sum_j y_{\alpha\beta,j} y_{\gamma\nu,j}, \sum_k \bar{y}_{\alpha\beta,k} \bar{y}_{\gamma\nu,k}\right) \\
&= \sum_{j,k} [\mathrm{cum}(y_{\alpha\beta,j}, y_{\gamma\nu,j}, \bar{y}_{\alpha\beta,k}, \bar{y}_{\gamma\nu,k}) \\
&\quad + \mathrm{cum}(y_{\alpha\beta,j}, \bar{y}_{\alpha\beta,k}) \mathrm{cum}(y_{\gamma\nu,j}, \bar{y}_{\gamma\nu,k}) + \mathrm{cum}(y_{\alpha\beta,j}, y_{\gamma\nu,k}) \mathrm{cum}(\bar{y}_{\alpha\beta,k}, \bar{y}_{\gamma\nu,j})] \\
&= O(n^2 m^{-3}) + O(nmm^{-2}) + O(n^2 \log^2 n / n^2) = O(n/m),
\end{aligned}
$$

which proves (23).

Proof of Theorem 3. We derive the AMISE(G) according to (7). Since

$$
\sum_{e_1=1}^{r} h_{ae_1,\alpha\beta}(\lambda) f^{e_1 b}(\lambda) = I_{a=\alpha,(\alpha,\beta)\in E} f^{\beta b}(\lambda) + \sum_{e_1=1}^{r} I_{(a,e_1)\notin E} h_{ae_1,\alpha\beta}(\lambda) f^{e_1 b}(\lambda),
$$

the integrand of AMISE(G) is

$$
\begin{aligned}
&\left\{ \sum_{(\alpha,\beta)\in E,(\gamma,\nu)\in E} f^{\beta\gamma}(\lambda) f^{\nu\alpha}(\lambda) + \sum_{(\alpha,\beta)\in E} \sum_{\gamma,\nu=1}^{r} \sum_{(b,e_2)\notin E} h_{be_2,\gamma\nu}(\lambda) f^{\beta b}(\lambda) f^{e_2\alpha}(\lambda) \right. \\
&\quad + \sum_{(\gamma,\nu)\in E} \sum_{\alpha,\beta=1}^{r} \sum_{(a,e_1)\notin E} h_{ae_1,\alpha\beta}(\lambda) f^{e_1\gamma}(\lambda) f^{\nu a}(\lambda) \\
&\quad \left. + \sum_{\alpha,\beta,\gamma,\nu=1}^{r} \sum_{(a,e_1)\notin E,(b,e_2)\notin E} h_{ae_1,\alpha\beta}(\lambda) h_{be_2,\gamma\nu}(\lambda) f^{e_1 b}(\lambda) f^{e_2 a}(\lambda) \right\} f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda) \\
&= \eta_1(\lambda) + \eta_2(\lambda) + \eta_3(\lambda) + \eta_4(\lambda), \quad \text{say.}
\end{aligned}
$$

Then

$$
\eta_1(\lambda) = r^2 - 4\sharp\{E^c\} + \sum_{(\alpha,\beta)\notin E,(\gamma,\nu)\notin E} f^{\beta\gamma}(\lambda) f^{\nu\alpha}(\lambda) f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda).
$$

By Lemma 2(i), we have

$$
\begin{aligned}
\eta_2(\lambda) &= \sum_{(\alpha,\beta)\in E} \sum_{(b,e_2)\notin E} f^{\beta b}(\lambda) f^{e_2\alpha}(\lambda) f_{\alpha e_2}(\lambda) f_{b\beta}(\lambda) \\
&= 2\sharp\{E^c\} - \sum_{(\alpha,\beta)\notin E} \sum_{(b,e_2)\notin E} f^{\beta b}(\lambda) f^{e_2\alpha}(\lambda) f_{\alpha e_2}(\lambda) f_{b\beta}(\lambda).
\end{aligned}
$$

In the same way,

$$
\eta_3(\lambda) = 2\sharp\{E^c\} - \sum_{(\gamma,\nu)\notin E} \sum_{(a,e_1)\notin E} f^{e_1\gamma}(\lambda) f^{\nu a}(\lambda) f_{a\nu}(\lambda) f_{\gamma e_1}(\lambda).
$$

By Lemma 2(i) and (ii), we have

$$
\begin{aligned}
\eta_4(\lambda) &= \sum_{(a,e_1)\notin E,(b,e_2)\notin E} \sum_{(\gamma,\nu)\in E} h_{be_2,\gamma\nu}(\lambda) f^{e_1 b}(\lambda) f^{e_2 a}(\lambda) \sum_{\alpha,\beta=1}^{r} h_{ae_1,\alpha\beta}(\lambda) f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda) \\
&= \sum_{(a,e_1)\notin E,(b,e_2)\notin E} f^{e_1 b}(\lambda) f^{e_2 a}(\lambda) \sum_{(\gamma,\nu)\in E} h_{be_2,\gamma\nu}(\lambda) f_{a\nu}(\lambda) f_{\gamma e_1}(\lambda) \\
&= \sum_{(a,e_1)\notin E,(b,e_2)\notin E} f^{e_1 b}(\lambda) f^{e_2 a}(\lambda) \left(f_{ae_2}(\lambda) f_{be_1}(\lambda) + x_\infty(a, e_1, b, e_2, \lambda)\right) \\
&= \sum_{(a,e_1)\notin E,(b,e_2)\notin E} f^{e_1 b}(\lambda) f^{e_2 a}(\lambda) f_{ae_2}(\lambda) f_{be_1}(\lambda) - \sum_{(a,e_1)\notin E} 1 \\
&= \sum_{(a,e_1)\notin E,(b,e_2)\notin E} f^{e_1 b}(\lambda) f^{e_2 a}(\lambda) f_{ae_2}(\lambda) f_{be_1}(\lambda) - 2\sharp\{E^c\},
\end{aligned}
$$

which completes the proof of Theorem 3.

Lemmas

Lemma 1. *For an invertible $n \times n$ matrix A, we have*

$$\frac{\partial A^{pq}}{\partial A_{\alpha\beta}} = -A^{p\alpha}A^{\beta q},$$

for $p, q, \alpha, \beta = 1, \ldots, n$.

Proof of Lemma 1. It is derived simply by calculating componentwise the rule 9 of Lütkepohl (1991, page 471).

Lemma 2. *Let G_n be the characterization defined by $G_n(\mathrm{vec}(f(\lambda))) := g_n(\lambda)$ in (17) and $h_{n,be_2,\gamma\nu}(\lambda) = \frac{\partial G_{n,be_2}(y)}{\partial y_{\gamma\nu}}\Big|_{y=\mathrm{vec}(f(\lambda))}$. Put*

$$x_n(\alpha,\beta,b,e_2,\lambda) \quad := \quad \sum_{\gamma,\nu=1}^{r} h_{n,be_2,\gamma\nu}(\lambda) f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda) - f_{\alpha e_2}(\lambda) f_{b\beta}(\lambda),$$

for $(b, e_2) \notin E$, and define $x_\infty(\alpha,\beta,p,q,\lambda)$ as the limiting value of $x_n(\alpha,\beta,p,q,\lambda)$ as n tends to ∞.

(i) If $(\alpha, \beta) \in E$,

$$x_n(\alpha,\beta,b,e_2,\lambda) \quad = \quad 0, \quad n = 1, 2, \ldots.$$

(ii) If $(\alpha, \beta) \notin E$,

$$\sum_{(p,q)\notin E} x_\infty(\alpha,\beta,p,q,\lambda) f^{bp}(\lambda) f^{qe_2}(\lambda) + I_{b=\beta} I_{\alpha=e_2} \quad = \quad 0,$$

for $(b, e_2) \notin E$.

Proof of (i). Let $n' = n(\mathrm{mod}\ M)$. It is clear from (17) that for$(b, e_2) \notin F_{n'}$

$$x_n(\alpha,\beta,b,e_2,\lambda) = x_{n-1}(\alpha,\beta,b,e_2,\lambda), \tag{24}$$

since

$$h_{n,be_2,\gamma\nu}(\lambda) \quad = \quad h_{n-1,be_2,\gamma\nu}(\lambda).$$

For $(b, e_2) \in F_{n'}$, by applying Lemma 1 to (17) we have

$$\begin{aligned} h_{n,be_2,\gamma\nu}(\lambda) &= h_{n-1,be_2,\gamma\nu}(\lambda) - \sum_{p,q=1}^{r} \frac{f^{bp}(\lambda) f^{qe_2}(\lambda)}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} h_{n-1,pq,\gamma\nu}(\lambda) \\ &= h_{n-1,be_2,\gamma\nu}(\lambda) - \sum_{p,q\notin E} \frac{f^{bp}(\lambda) f^{qe_2}(\lambda)}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} h_{n-1,pq,\gamma\nu}(\lambda) - \frac{f^{b\gamma}(\lambda) f^{\nu e_2}(\lambda)}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} I_{\{(\gamma,\nu)\in E\}}. \end{aligned} \tag{25}$$

Hence by putting (25) into the definition of x_n, we have for $(b, e_2) \in F_{n'}$,

$$\begin{aligned} & x_n(\alpha,\beta,b,e_2,\lambda) \\ = \; & x_{n-1}(\alpha,\beta,b,e_2,\lambda) - \sum_{(p,q)\notin E} \sum_{\gamma,\nu=1}^{r} \frac{f^{bp}(\lambda) f^{qe_2}(\lambda) h_{n-1,pq,\gamma\nu}(\lambda)}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda) \\ & - \sum_{(\gamma,\nu)\in E} \frac{f^{b\gamma}(\lambda) f^{\nu e_2}(\lambda) f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda)}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} \\ = \; & x_{n-1}(\alpha,\beta,b,e_2,\lambda) - \sum_{(p,q)\notin E} \frac{f^{bp}(\lambda) f^{qe_2}(\lambda)(f_{\alpha q}(\lambda) f_{p\beta}(\lambda) + x_{n-1}(\alpha,\beta,p,q,\lambda))}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} \\ & - \sum_{(\gamma,\nu)\in E} \frac{f^{b\gamma}(\lambda) f^{\nu e_2}(\lambda) f_{\alpha\nu}(\lambda) f_{\gamma\beta}(\lambda)}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} \\ = \; & x_{n-1}(\alpha,\beta,b,e_2,\lambda) - \frac{I_{b=\beta} I_{\alpha=e_2}}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} - \sum_{(p,q)\notin E} \frac{f^{bp}(\lambda) f^{qe_2}(\lambda)}{f^{bb}(\lambda) f^{e_2e_2}(\lambda)} x_{n-1}(\alpha,\beta,p,q,\lambda). \end{aligned} \tag{26}$$

We use induction to prove (i). $x_1(\alpha,\beta,b,e_2,\lambda)=0$ for $(b,e_2)\notin E$, since by Lemma 1

$$h_{1,be_2,\gamma\nu}(\lambda)=\begin{cases} I_{b=\gamma}I_{e_2=\nu}-\frac{f^{b\gamma}(\lambda)f^{\nu e_2}(\lambda)}{f^{bb}(\lambda)f^{e_2e_2}(\lambda)}, & \text{if } (b,e_2)\in F_1,\\ I_{b=\gamma}I_{e_2=\nu}, & \text{otherwise .}\end{cases} \tag{27}$$

Suppose $x_{n-1}(\alpha,\beta,b,e_2,\lambda)=0$ for $(b,e_2)\notin E$. Then $x_n(\alpha,\beta,b,e_2,\lambda)=0$ by the recursion (24) and (26).

Proof of (ii). There exist integers k,l for n in (26) such that $n=kM+l,\ 0\le l\le M-1$. Then for $(b,e_2)\in F_l$ we have

$$\begin{aligned} x_{kM+l}(\alpha,\beta,b,e_2,\lambda) &= x_{kM+l-1}(\alpha,\beta,b,e_2,\lambda)-\frac{I_{b=\beta}I_{\alpha=e_2}}{f^{bb}(\lambda)f^{e_2e_2}(\lambda)}\\ &\quad-\sum_{(p,q)\notin E}\frac{f^{bp}(\lambda)f^{qe_2}(\lambda)}{f^{bb}(\lambda)f^{e_2e_2}(\lambda)}x_{kM+l-1}(\alpha,\beta,p,q,\lambda).\end{aligned}$$

Letting $k\to\infty$ for each $l=0,\dots,M-1$, we have the result for $(b,e_2)\notin E$.

References

[1] Akaike, H. (1974). A new look at statistical model identification. *Transaction on Automatic Control* **19**,716-723.

[2] Beltrao, K.I. and Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.***8**,21-38.

[3] Brillinger, D. R. (1981). *Time Series Analysis: Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

[4] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.

[5] Cheng, B. and Tong, H. (1992). On consistent nonparametric order determination and chaos. *J. Royal Stat. Soc. Ser. B.***54**, 427-449.

[6] Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika***51**, 151-172.

[7] Dunsmuir, W. (1979) A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed white noise. *Ann. Statist.* 7, 490-506.

[8] Härdle, W., Hall, P. and Marron, J.S.(1988). How far are automatically chosen regression smoothing parameters from their optimum ? *J. Am. Stat. Assoc.***83**, 86-95.

[9] Haslett, J. and Raftery, A.E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Appl. Statist.***38**, 1-50.

[10] Hosoya, Y. (1997) A limit theory for long-range dependence and statistical inference on related models. *Ann. Statist.* 25, 105-137.

[11] Hurvich, C.M. (1985). Data-driven choice of a spectrum estimate: Extending the Cross-Validation methods. *J. Am. Stat. Assoc.***80**, 933-940.

[12] Hurvich, C.M. and Tsai, C.L.(1989). Regression and time series model selection in small samples. *Biometrika* **76**,297-307.

[13] Kavalieris, L. (1989). The estimation of the order of an autoregression using recursive residuals and cross-validation. *J. Time Ser. Anal.***10**,271-281.

[14] Lauritzen, S.L. (1996). *Graphical Models*. Oxford University Press, Oxford.

[15] Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag.

[16] Martin, R.J.(1990). The use of time series models and methods in the analysis of agricultural fields trials. *Commun. Statist. Theory Methods* **19**,55-81.

[17] Matsuda, Y. and Yajima, Y. (2004). On testing for separable correlations of multivariate time series. *J. Time Ser. Anal.***25**,501-528.

[18] Priestley, M.B.(1988). *Spectral Analysis and Time Series*. Vols. 1 and 2. Academic Press, New York.

[19] Shao, J. (1993). Linear model selection by cross-validation. *J. Am. Stat. Assoc.***88**, 486-494.

[20] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.***8**, 147-164.

[21] Speed, T.P. and Kiiveri, H. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.***14**, 138-150.

[22] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Royal Stat. Soc. Ser. B.***36**, 111-147.

[23] Wermuth, N. and Scheidt, E. (1977). Fitting a covariance selection model to a matrix. Algorithm AS105. *Appl. Statist.***26**, 88-92.

[24] Xia, Y. and Tong, H. (2004). Cumulative effects of air pollution on public health, in preparation.

[25] Xia, Y., Tong, H., Li, W.K. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. *J. Royal Stat. Soc. Ser. B.***64**, 363-410.

[26] Yajima, Y. and Matsuda, Y. (2004). On nonparametric and semiparametric testing for multivariate time series, submitted.

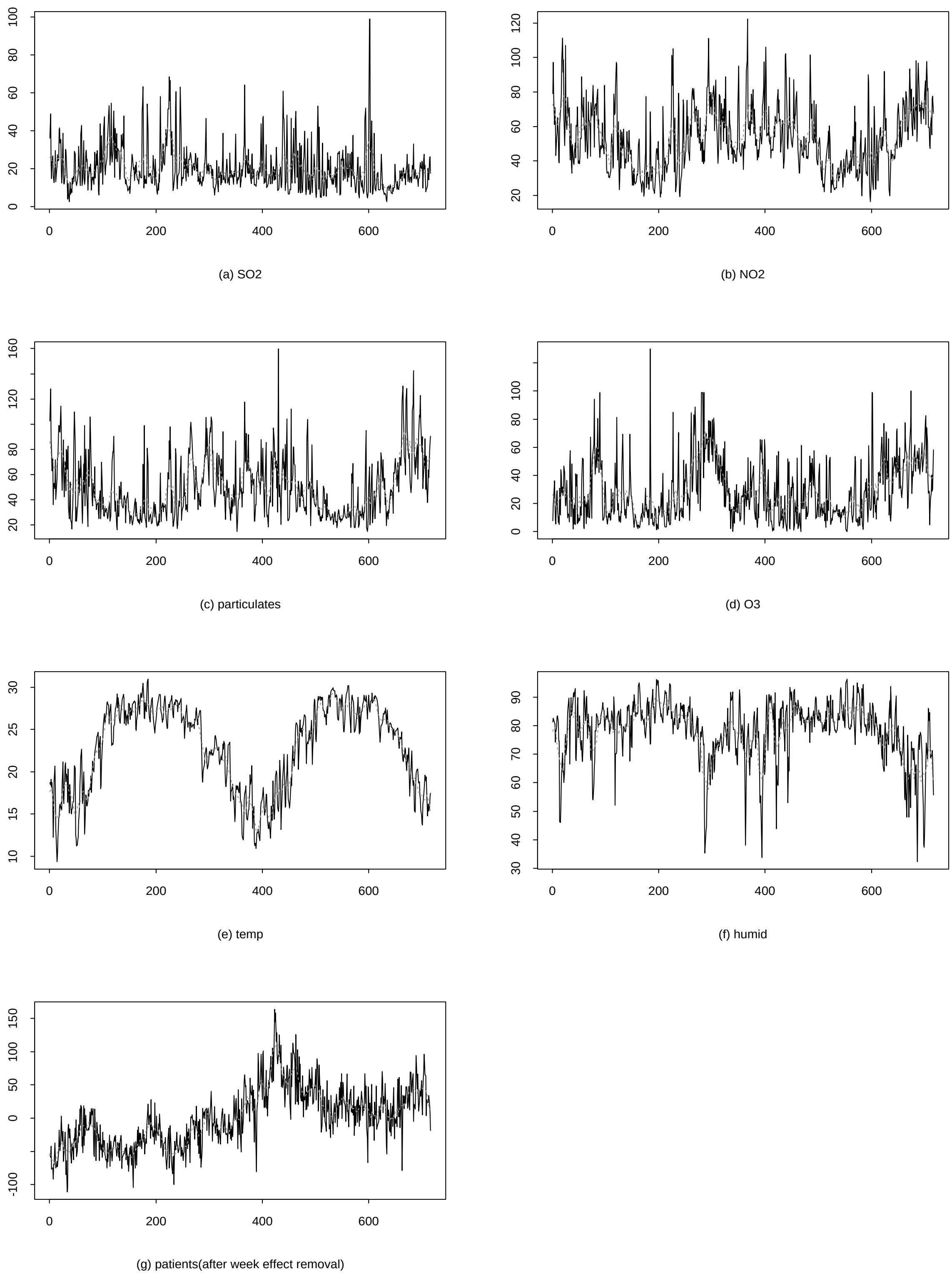

Figure 1: (a) Total number of daily hospital admissions of circulatory and respiratory patients and average levels of (b) sulphur dioxide, (c) nitrogen dioxide, (d) respirable suspended particulates, (e) ozone, (f) temperature and (g) humidity, from January 1st, 1994, to December 31st, 1995, with their trends estimated by 15 days moving averages.

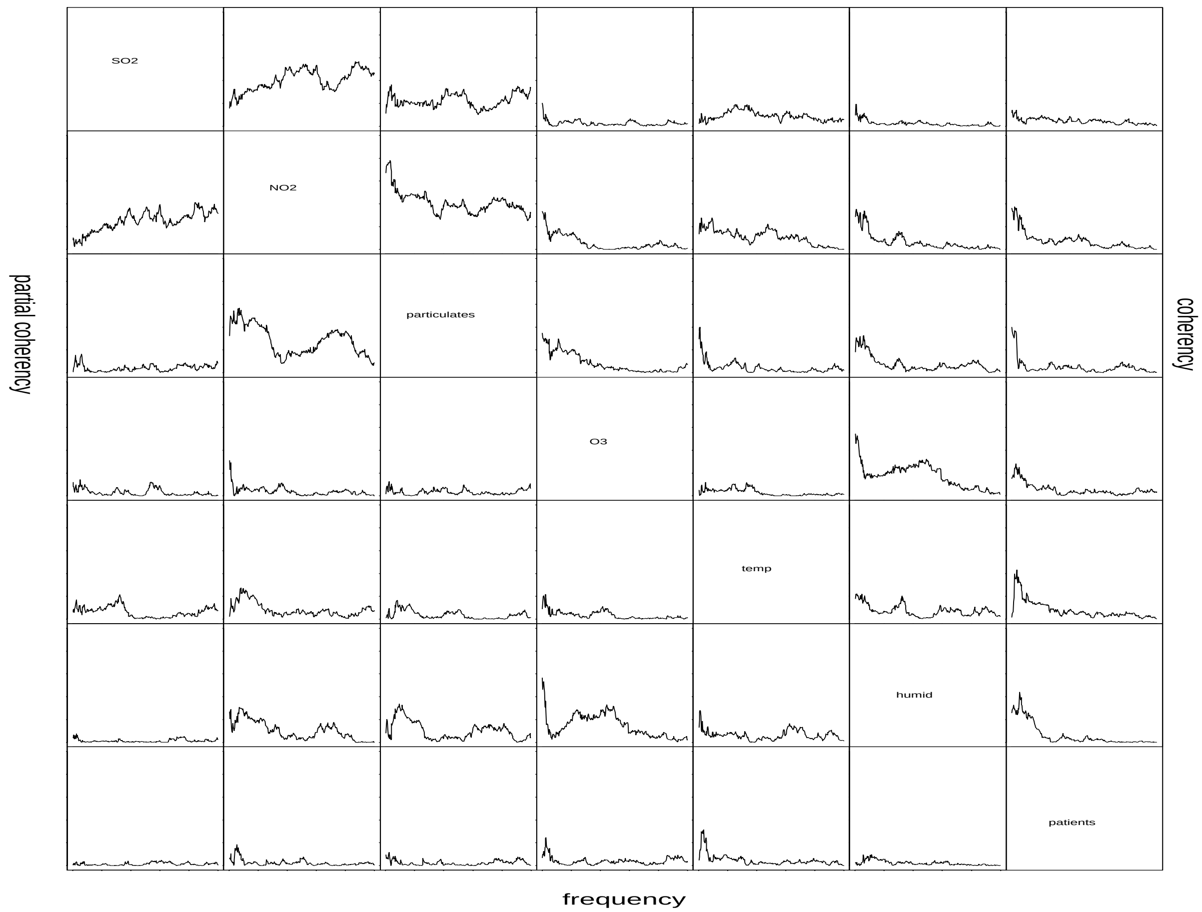

Figure 2: Spectral coherence (above diagonal) and partial spectral coherence (below diagonal) for the detrended time series of (a) SO_2, (b) NO_2, (c)particulates, (d) O_3, (e) temperature, (f) humidity and (g) hospital admissions.

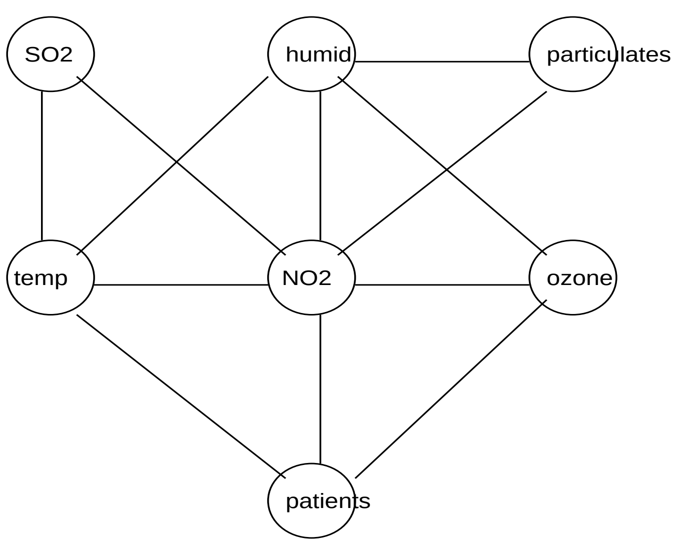

Figure 3: Estimated partial correlation graph for the detrended time series of (a) SO_2, (b) NO_2, (c)particulates, (d) O_3, (e) temperature, (f) humidity and (g) hospital admissions.