Bernoulli **12**(2), 2006, 221–249

Selecting models with different spectral density matrix structures by the cross-validated log likelihood criterion

YASUMASA MATSUDA[1], YOSHIHIRO YAJIMA[2] and HOWELL TONG[3]

[1]*Faculty of Economics, Tohoku University, 27-1 Kawauchi, Aoba-ku, Sendai 980-8576, Japan. E-mail: matsuda@econ.tohoku.ac.jp*
[2]*Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: yajima@e.u-tokyo.ac.jp*
[3]*Department of Statistics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. E-mail: h.tong@lse.ac.uk*

We propose the cross-validated log likelihood (CVLL) criterion for selecting multivariate time series models with different forms of the spectral density matrix, which correspond to different constraints on the component time series such as mutual independence, separable correlation, time reversibility, graphical interaction and others. We obtain asymptotic properties of the CVLL, and demonstrate the empirical properties of the CVLL selection with both simulated and real data.

Keywords: conditional independence; consistency; graphical model; Kullback–Leibler divergence; model selection; multivariate time series; periodogram; spectral density matrix

1. Introduction

Let $\{X_{t,a} : t = \ldots, -1, 0, 1, \ldots ; a = 1, \ldots, r\}$ be an r-variate stationary time series and $f(\lambda)$, $-\pi \leqslant \lambda \leqslant \pi$, be its spectral density matrix (sdm). To analyse time series, parametric models are often fitted to $f(\lambda)$. With a parametric matrix-valued function $g_\theta(\lambda)$, a parametric model is described as

$$\mathcal{F}_\Theta := \{g_\theta(\lambda), \theta \in \Theta \subset \mathbb{R}^p\} \subset \mathcal{F}, \tag{1}$$

where Θ is a parameter space and $\mathcal{F}$ is the set of all possible $r \times r$ sdms. It follows that fitting a parametric model is equivalent to identifying a subset in $\mathcal{F}$. Autoregressive moving average (ARMA) models are typical examples of parametric models. Specifically, the sdms of these models consist of components which are rational functions of $\exp(i\lambda)$ and with parameters θ.

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 an appropriate model from several candidate parametric models. In particular, the order determination for ARMA processes by AIC has been investigated extensively in the literature. For discussions of the order selection by AIC see, for example, Brockwell and Davis (1991, Section 9.3), Shibata (1980), Hurvich and Tsai (1989).

Though parametric models (1) provide useful tools for linear time series analysis, we are often interested in aspects of the component time series which are of such a very different nature that a radically different approach is called for. For example, we may be interested in investigating the plausibility of mutual independence of the component time series without wishing to make as strong an assumption as a rational sdm. Other general aspects of the component time series that are often of interest include separable correlation, time reversibility and graphical interaction. For the above purpose, let G_{ab}, a, $b = 1, \ldots, r$, be a function from $\mathbb{R}^{r^2}$ to $\mathbb{R}$ and consider the subset $\mathcal{F}_G \subset \mathcal{F}$ prescribed by G in the following way:

$$\mathcal{F}_G := \{g(\lambda) := (g_{ab}(\lambda)) \in \mathcal{F} |$$
$$g_{ab}(\lambda) = G_{ab}(\theta,\, g_{cd}(\lambda),\, c,\, d = 1, \ldots, r),\, a,\, b = 1, \ldots, r, \text{ almost everywhere in } [-\pi, \pi]\}. \tag{2}$$

The subset $\mathcal{F}_G \subset \mathcal{F}$ is defined in a very different way from the parametric models (1) and is called an *sdm model* described by G_{ab}, a, $b = 1, \ldots, r$. In general G_{ab} need not depend on a parameter, although in some cases it may. We give examples of both situations in Section 4, which will show how $\mathcal{F}_G$ can easily accommodate the component-time-series-specific aspects mentioned above.

The purpose of this paper is to propose a selection criterion for sdm models corresponding to different G_{ab}. Many existing criteria developed for model (1) rely on penalizing model complexity by reference to 'the number of parameters'. However, as we have seen, for an sdm model the notion of a parameter need not be relevant, and even when it is there is usually no simple way to count the number of parameters. An alternative approach is cross-validation (CV), which requires no such counting. For examples, see 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 and Cheng and Tong (1992) for order determination of nonparametric autoregressive models. CV penalizes the complexity of a model by the *leave-one-out* approach instead of an explicit expression for model complexity.

In this paper, we adopt the cross-validated log quasi-likelihood – which we refer to for convenience as the cross-validated log likelihood (CVLL) – as a criterion for sdm model selection. The distinctive features are as follows. First, the sdm model having the smallest CVLL is asymptotically the one with the minimum mean integrated squared estimation error for $f(\lambda)$. Secondly, in contrast to the inconsistency of CV selection in parametric situations (Shao 1988, Kavalieris 1989), it should be emphasized that CV can provide consistent selection procedures outside the parametric context.

This paper is organized as follows. We introduce the CVLL criterion in Section 2, and show its asymptotic properties in Section 3. In Section 4 we give several examples of the sdm models. In Section 5 we apply the CVLL criterion to graphical modelling and prove the consistency of the CVLL selection for graphical models. In Section 6 we use simulation and real data to show empirical properties of the criterion. In Section 7 we give concluding remarks. In Sections 8 and 9 we finally present the proofs of lemmas and theorems, respectively.

2. The CVLL criterion

Throughout this paper, A_{ab} and A^{ab} are generic symbols for the (a, b)th element of matrices A and A^{-1} respectively, and A^{T} is the matrix transpose of A.

Let $\{\mathbf{X}_t = (X_{t,a}, a = 1, \ldots, r)^{\mathrm{T}}, -\infty < t < \infty\}$ be an r-dimensional stationary time series with the sdm $f(\lambda)$, $-\pi \leqslant \lambda \leqslant \pi$. We define $f(\lambda)$ outside the region $[-\pi, \pi]$ to have a period of 2π. Suppose $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are given. The discrete Fourier transform $W(\lambda)$ and the periodogram matrix $I(\lambda)$ of $\mathbf{X}_t$ are defined as follows:

$$W(\lambda) := (2\pi n)^{-1/2} \sum_{t=1}^{n} \mathbf{X}_t \exp(-\mathrm{i}\lambda t),$$

$$I(\lambda) := W(\lambda)\overline{W}(\lambda)^{\mathrm{T}}.$$

$I(\lambda)$ is defined periodically with the period 2π for $\lambda \notin [-\pi, \pi]$. Let $\lambda_j = 2\pi j/n$, $j = -[(n-1)/2], \ldots, -1, 0, 1, \ldots, [n/2]$, be the jth Fourier frequency and w_k, $k = -m/2, \ldots, m/2$, be a weight sequence. It is well known that the sdm $f(\lambda)$ is consistently estimated under suitable conditions by the smoothed periodogram

$$\hat{f}_{ab}(\lambda_j) := \left(\sum_{k=-m/2}^{m/2} w_k \right)^{-1} \sum_{k=-m/2}^{m/2} w_k I_{ab}(\lambda_{j+k}), \qquad a, b = 1, \ldots, r.$$

Under an sdm model described by G in (2), define the sdm $g(\lambda) = (g_{ab}(\lambda))$ and the estimator $\hat{g}(\lambda) = (\hat{g}_{ab}(\lambda))$, $a, b = 1, \ldots, r$, by

$$g_{ab}(\lambda) := G_{ab}(\theta, f_{cd}(\lambda), c, d = 1, \ldots, r) \tag{3}$$

and

$$\hat{g}_{ab}(\lambda_j) := G_{ab}(\hat{\theta}, \hat{f}_{cd}(\lambda_j), c, d = 1, \ldots, r), \tag{4}$$

respectively, where $\hat{\theta}$ is an estimator for θ. Let us define the CVLL for an sdm model described by G with the cross-validated version of (4).

Definition 1 *Cross-validated log likelihood. For an sdm model described by G, the CVLL is defined by*

$$\mathrm{CVLL}(G) := \sum_{j=1}^{[n/2]} \log\det\big(\hat{g}_{-j}(\lambda_j)\big) + \mathrm{tr}\Big(I(\lambda_j)\hat{g}_{-j}^{-1}(\lambda_j)\Big), \tag{5}$$

where $\hat{g}_{-j}(\lambda_j) = (\hat{g}_{ab,-j}(\lambda_j))$, $a, b = 1, \ldots, r$, *and*

$$\hat{g}_{ab,-j}(\lambda_j) := G_{ab}\Big(\hat{\theta}, \hat{f}_{cd,-j}(\lambda_j), c, d = 1, \ldots, r\Big),$$

$$\hat{f}_{cd,-j}(\lambda_j) := \left(\sum_{k=-m/2, k\neq 0}^{m/2} w_k\right)^{-1} \sum_{k=-m/2, k\neq 0}^{m/2} w_k I_{cd}(\lambda_{j+k}). \tag{6}$$

The sdm model which minimizes the CVLL over given candidates is the CVLL selected model. Not the usual spectral estimator (4) but the cross-validated version (6) must be put in the CVLL (5) to make the CVLL criterion work for sdm model selection, which is justified empirically and theoretically. See the proof of Theorem 2 in Section 9 for the theoretical justification, where cross-validation is required to prove that the third term of (26) is $o_p(nm^{-1})$. Hurvich (1985) and Beltrao and Bloomfield (1987) used the CVLL criterion to determine the optimal bandwidth for univariate kernel spectrum estimates. Hurvich's definition may be viewed as a univariate version of our CVLL, not for sdm model selection but for bandwidth selection.

3. Asymptotic properties of the CVLL

An sdm model described by G is correct for $\mathbf{X}_t$ if $f(\lambda) \in \mathcal{F}_G$; it is incorrect if $f(\lambda) \notin \mathcal{F}_G$. Following the notation of Shao (1993), we designate the latter models as category I, and the former as category II. Note that if an sdm model G is in category II, any model G_2 such that $\mathcal{F}_{G_2} \supset \mathcal{F}_G$ is in category II.

We evaluate the asymptotic behaviours of the CVLL for an sdm model described by G. They are shown to depend on the category to which G belongs by Theorems 1 and 2 under the following assumptions:

Assumption 1. $\{\mathbf{X}_t\}$ *is an r-dimensional Gaussian stationary process.*

Assumption 2. $f(\lambda)$ *is positive definite for* $-\pi \leqslant \lambda < \pi$.

Assumption 3. $f_{ab}(\lambda)$, $a, b = 1, \ldots, r$, $-\pi \leqslant \lambda < \pi$, *is twice continuously differentiable.*

Assumption 4. $m = O(n^{\beta})$, $1/2 < \beta < 3/4$, *and the weight function* w_k, $k = -m/2, \ldots, m/2$, *is given with a positive continuous even function* $u(x)$ on $[-1/2, 1/2]$ *by*

$$w_k = u\left(\frac{k}{m}\right), \qquad k = -m/2, \ldots, m/2.$$

Assumption 5. *When* G_{ab}, $a, b = 1, \ldots, r$, *depends on a parameter* θ, *there exists a* θ_0 *such that* $\hat{\theta} - \theta_0 = O_p(n^{1/2})$ *and satisfing*

$$f_{ab}(\lambda) = G_{ab}(\theta_0, f_{cd}(\lambda), c, d = 1, \ldots, r), \qquad a, b = 1, \ldots, r,$$

if G is in category II.

Theorem 1. *Under Assumptions 1–5, if* $G_{ab}(\theta, y_{cd}, c, d = 1, \ldots, r)$, $a, b = 1, \ldots, r$, *is continuously differentiable with respect to* θ *and* y_{cd},

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

where $g(\lambda)$ *is the sdm* (3) *under* G *in which* θ *is replaced with* θ_0 *and* $\mathrm{KL}(f, g)$ *is the Kullback–Leibler divergence between the two Gaussian stationary processes whose sdms are* $f(\lambda)$ *and* $g(\lambda)$, *that is,*

$$\mathrm{KL}(f, g) := \frac{1}{2\pi}\int_0^{\pi} \{\mathrm{tr}(f(\lambda)g^{-1}(\lambda) - E_r) - \log\det(f(\lambda)g^{-1}(\lambda))\}\mathrm{d}\lambda.$$

See Section 9 for the proof of Theorem 1. Theorem 1 implies that an sdm model whose $\mathrm{KL}(f, g)$ is smallest gives the smallest CVLL asymptotically. Theorem 1 is, however, powerless to distinguish among the sdm models in category II, since $\mathrm{KL}(f, g) = 0$ for all the models in category II. It is necessary to evaluate the terms of smaller order than n in probability when G is in category II, which is expressed by the asymptotic mean integrated squared error (AMISE) of G.

Definition 2 *Asymptotic mean integrated squared error. When G is in category II, the asymptotic mean integrated squared error of G is defined by*

$$\mathrm{AMISE}(G) := p\lim_{n\to\infty} \frac{m}{n} \sum_{j=1}^{[n/2]} \mathrm{tr}\{(\hat{g}(\lambda_j) - f(\lambda_j))f^{-1}(\lambda_j)\}^2,$$

where $\hat{g}(\lambda_j)$ *is the estimated sdm* (4) *under* G.

Theorem 2. *Under Assumptions 1–5, if G is in category II and* $G_{ab}(\theta, y_{cd}, c, d = 1, \ldots, r)$, $a, b = 1, \ldots, r$, *is three times continuously differentiable with respect to* θ *and* y_{cd}, *then*

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

and

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

where

$$C_u := \left(\int_{-1/2}^{1/2} u(x)\mathrm{d}x \right)^{-2} \int_{-1/2}^{1/2} u^2(x)\mathrm{d}x,$$

$$\mu_{\alpha\beta\gamma\nu}(\lambda) := \sum_{a,b=1}^{r} \sum_{e_1,e_2=1}^{r} h_{ae_1,\alpha\beta}(f(\lambda)) h_{be_2,\gamma\nu}(f(\lambda)) f^{e_1 b}(\lambda) f^{e_2 a}(\lambda),$$

$$h_{ae_1,\alpha\beta}(y) := \frac{\partial G_{ae_1}(\theta_0,\, y_{cd},\, c,\, d = 1, \ldots, r)}{\partial y_{\alpha\beta}}.$$

The proof of Theorem 2 is given in Section 9. Theorem 2 implies that the sdm model in category II which attains the smallest AMISE gives the smallest CVLL asymptotically. Corollary 1 summarizes the consequences of Theorems 1 and 2.

Corollary 1. *Let G_i, $i = 1, 2$, describe sdm models. Under Assumptions 1–5, and if G_i, $i = 1, 2$, is three times differentiable,*

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

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

2. *if both G_1 and G_2 are in category II and* $\mathrm{AMISE}(G_1) > \mathrm{AMISE}(G_2)$*, then*

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

Let us define consistency of CVLL selection here. Let S be a set of candidate sdm models from which the CVLL criterion selects.

Definition 3 *Consistency of CVLL selection for S. Let G_i, $i = 1, 2$, describe two sdm models in $S \cap \{$category II$\}$. CVLL selection is consistent for S, if $\mathcal{F}_{G_1} \supset \mathcal{F}_{G_2}$ implies* $\lim_{n\to\infty} P(\mathrm{CVLL}(G_1) > \mathrm{CVLL}(G_2)) = 1$.

It follows from Corollary 1 that the consistency for S is equivalent to the statement: $\mathcal{F}_{G_1} \supset \mathcal{F}_{G_2}$ implies $\mathrm{AMISE}(G_1) > \mathrm{AMISE}(G_2)$ for G_i, $i = 1, 2$ in $S \cap \{\text{category II}\}$. We shall show in Sections 4 and 5 that CVLL selection is consistent for some important specific cases: the decomposition into independent subseries and graphical modelling. We leave it to future studies to investigate if the consistency holds true for general sdm model selection, namely to show whether $\mathcal{F}_{G_1} \supset \mathcal{F}_{G_2}$ implies $\mathrm{AMISE}(G_1) > \mathrm{AMISE}(G_2)$ for any G_i, $i = 1, 2$, in category II.

Remark 1. The same weight function should be used for all the candidates to evaluate their CVLLs, since their AMISEs depend on weight functions. In choosing the bandwidth m, it seems plausible in practice to adopt the one which satisfies Assumption 4 and minimizes the CVLL.

4. Examples of sdm models

In this section, we show typical examples of sdm models and derive their AMISE values by (7). In the following, for a condition C, let I_C take the value 1 if C is satisfied and 0 otherwise. Let y be an $r \times r$ positive definite matrix. In this sections Assumptions 1–4 hold, and in addition Assumption 5 holds in Example 3.

Example 1 *No constraint.* Consider the sdm model described by

$$G_{ab}(y_{cd}, c, d = 1, \ldots, r) := y_{ab}, \qquad a, b = 1, \ldots, r. \tag{8}$$

In this case $\mathcal{F}_G$ reduces to $\mathcal{F}$, which means that (8) is always in category II. Since $h_{ae_1,\alpha\beta}(f(\lambda)) = I_{(a,e_1)=(\alpha,\beta)}$, we have $\mu_{\alpha\beta\gamma\nu}(\lambda) = f^{\beta\gamma}(\lambda)f^{\nu\alpha}(\lambda)$. Hence,

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

Example 2 *Mutual independence among subseries.* Let $M_1 \cup \ldots \cup M_p$ be a partition of the set $\{1, 2, \ldots, r\}$. Suppose that an r-dimensional time series is grouped into p independent subseries $\mathbf{X}_t = (X_{t,M_1}, \ldots, X_{t,M_p})^{\mathrm{T}}$ such that $\{X_{t,M_i}\}$ and $\{X_{t,M_j}\}$ are mutually independent for $i \neq j$. This is the sdm model described by

$$G_{ab}(y_{cd}, c, d = 1, \ldots, r) := \begin{cases} y_{ab}, & \text{if } a \in M_i \text{ and } b \in M_i \text{ for some } i, \\ 0, & \text{otherwise} \end{cases} \qquad a, b = 1, \ldots, r. \tag{9}$$

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

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

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 $f(\lambda)$ is block diagonal,

$$\begin{aligned}\text{AMISE}(G) &= \frac{C_u}{2\pi}\int_0^{\pi} \sum_{(\alpha,\beta)\in E} \sum_{(\gamma,\nu)\in E} f^{\beta\gamma}(\lambda)f^{\nu\alpha}(\lambda)f_{\alpha\nu}(\lambda)f_{\gamma\beta}(\lambda)\mathrm{d}\lambda \\ &= \frac{C_u \#\{E\}}{2} \\ &= \frac{C_u(\#\{M_1\}^2 + \ldots + \#\{M_p\}^2)}{2},\end{aligned} \tag{10}$$

where $\#$ denotes the cardinality of a set.

Let S be the set of sdm models which decompose the original time series into all possible independent subseries. It follows from (10) and Theorem 2 that CVLL selection is consistent for S in the sense of Definition 3.

Example 3 *Separable correlations.* A separable correlation is defined to satisfy

$$\mathrm{cov}(X_{t,a}, X_{s,b}) = \sigma_{ab}\rho(t-s), \qquad a, b = 1, \ldots, r,$$

for all t, s, where (σ_{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; Matsuda and Yajima 2004). A separable model includes a mutually independent component model with an identical autocorrelation as the special case when $\sigma_{\alpha\gamma} = 0$ for $\alpha \neq \gamma$, which is often useful for longitudinal data analysis (Diggle *et al.* 1994, Chapter 4).

By 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 non-negative integrable function on $[-\pi, \pi]$. Hence the separable model is the sdm model described by

$$G_{ab}(\sigma, y_{cd}, c, d = 1, \ldots, r) := \frac{\sigma_{ab}}{r}\sum_{i=1}^{r}\frac{y_{ii}}{\sigma_{ii}}, \qquad a, b = 1, \ldots, r, \tag{12}$$

where $\sigma = (\sigma_{ab})$, $a, b = 1, \ldots, r$, is the parameter vector. This model is an example of the case where G depends on a parameter. Noting that

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

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

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

which reduces to $C_u/2$ in the special case where $\sigma_{\alpha\gamma} = 0$ for $\alpha \neq \gamma$. It follows that CVLL selection is consistent for the set of the sdm models obtained from (8) and the separable models in the sense of Definition 3. See Remark 2 below for a practical application.

Example 4 *Time reversibility.* An r-dimensional time series $\{\mathbf{X}_t\}$ is said to be time-reversible if, for any $k = 1, 2, \ldots$ and any k-tuple $t_1 < \ldots < t_k$, the joint distribution of $(\mathbf{X}_{t_1}, \ldots, \mathbf{X}_{t_k})$ is the same as that of $(\mathbf{X}_{-t_1}, \ldots, \mathbf{X}_{-t_k})$. See Chan *et al.* (2005) for a recent discussion of time reversibility for multivariate time series. The necessary and sufficient condition of time reversibility for Gaussian time series is that the spectral density matrix is real-valued. Hence time reversibility is identified with the sdm model described by

$$G_{ab}(y_{cd},\, c,\, d = 1,\, \ldots,\, r) := \frac{y_{ab} + y_{ba}}{2}, \qquad a,\, b = 1,\, \ldots,\, r.$$

Noting that

$$h_{ae_1,\alpha\beta}(f(\lambda)) = \frac{I_{a=\alpha} I_{e_1=\beta} + I_{a=\beta} I_{e_1=\alpha}}{2},$$

$$\mu_{\alpha\beta\gamma\nu}(\lambda) = \frac{f^{\beta\gamma}(\lambda) f^{\nu\alpha}(\lambda) + f^{\beta\nu}(\lambda) f^{\gamma\alpha}(\lambda) + f^{\alpha\gamma}(\lambda) f^{\nu\beta}(\lambda) + f^{\alpha\nu}(\lambda) f^{\gamma\beta}(\lambda)}{4},$$

we have

$$\mathrm{AMISE}(G) = \frac{C_u(r^2 + r)}{4}.$$

It follows that CVLL selection is consistent for the set of the sdm models obtained from (8) and the time reversibility in the sense of Definition 3. See Remark 2 below for a practical application.

Example 5 *Graphical models.* Graphical models were originally defined for a random vector and were estimated based on its independent realizations. See Lauritzen (1996) for the basic notation and definitions of graphical models. Dahlhaus (2000) extended the concept of undirected conditional independence graphs to multivariate time series. The key idea is to regard a graphical model for multivariate time series as a kind of sdm model, which makes it possible to apply the CVLL criterion to graphical modelling. In the next section, we discuss this application specifically.

Remark 2. The CVLL criterion can be applied to the following testing problem. For an sdm model described by G,

$$H_0 : G \text{ is in category II} \quad \text{versus} \quad H_1 : G \text{ is in category I}.$$

Accept H_0 if the CVLL of G is smaller than that of (8), and accept H_1 otherwise. Then the testing procedure is consistent if AMISE(G) is smaller than $C_u r^2/2$.

Remark 3. The expression for an sdm model is not always unique. It is desirable to adopt the one having the smaller AMISE if possible, which makes it easier to distinguish it from other redundant sdm models. For example, the separable correlations can be expressed in another way:

$$G_{2,ab}(\sigma,\, y_{cd},\, c,\, d = 1,\, \ldots,\, r) := \sigma_{ab} \frac{y_{11}}{\sigma_{11}}, \qquad a,\, b = 1,\, \ldots,\, r, \tag{13}$$

whose AMISE is equal to $C_u r/2$, which is larger than that of (12). Hence expression (12) is preferable to (13).

5. Application to graphical models

We will show in this section that the CVLL criterion is an effective tool to identify the undirected graph for multivariate Gaussian time series $\{X_{t,a},\ a = 1, \ldots, r\}$. Set $X_a = \{X_{t,a},\ -\infty < t < \infty\}$, $Y_{t,ab} = \{X_{t,j},\ j \neq a, b\}$ and $Y_{ab} = \{Y_{t,ab},\ -\infty < t < \infty\}$. The conditional independence between X_a and X_b given Y_{ab} is defined by

$$X_a \perp X_b | \mathrm{Y}_{ab} \Leftrightarrow \mathrm{cov}(\varepsilon_{a|\{a,b\}^c}(s), \varepsilon_{b|\{a,b\}^c}(t)) = 0, \qquad \text{for all } s, t \in Z, \tag{14}$$

where

$$\varepsilon_{a|\{a,b\}^c}(t) = X_{t,a} - \mu_a^{\mathrm{opt}} - \sum_{u-\infty}^{\infty} \mathrm{d}_a^{\mathrm{opt}}(t-u) Y_{u,ab},$$

which minimizes

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

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

Definition 4 *Partial correlation graph. Let $V = \{1, \ldots, r\}$ be the set of vertices and $E(\subset V \times V)$ be a set of edges. Let $(a, b) \notin E$ if and only if $X_a \perp X_b | \mathrm{Y}_{ab}$. Then (V, E) is called a partial correlation graph for time series.*

Dahlhaus (2000, Theorem 2.4) proved that (14) is equivalent to

$$f^{ab}(\lambda) \equiv 0, \qquad -\pi \leqslant \lambda < \pi. \tag{15}$$

It follows that the missing edges in the partial correlation graph can uniquely be identified from the zeros in the inverse of the sdm $f(\lambda)$. With a suitable estimator $\hat{f}(\lambda)$, Dahlhaus (2000: 171) used the test statistic

$$\sup_{\lambda} \frac{|\hat{f}^{ab}(\lambda)|^2}{\hat{f}^{aa}(\lambda) \hat{f}^{bb}(\lambda)}$$

to detect the zeros in his empirical studies. Though the statistic is intuitively appealing and easy to construct, it is difficult to determine the null distribution when (15) is true, as he mentioned.

Instead we apply the CVLL criterion to graphical model selection. Let (V, E) be a graph for $\mathbf{X}_t$ with the sdm $f(\lambda)$. For the graph (V, E), consider the sdm $g(\lambda)$ which satisfies

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

The unique existence of the sdm $g(\lambda)$ is guaranteed in Lemma 7. Let us define the function G_{ab} by

$$G_{ab}(f_{cd}(\lambda), c, d = 1, \ldots, r) := g_{ab}(\lambda), \qquad a, b = 1, \ldots, r. \tag{17}$$

It follows from (15) that the graph (V, E) is identified with the sdm model described by the function G_{ab}. We derive the AMISE of a graphical model by applying the implicit function theorem to (16) in (7).

Theorem 3. *Let G describe a graphical model (V, E). Then under Assumptions 1–4,*

$$\mathrm{AMISE}(G) = \frac{C_u(r^2 - 2M_E)}{2},$$

where $M_E := \#\{(a, b)|(a, b) \notin E, a < b\}$, which is the number of conditional independent pairs of (V, E).

The proof of Theorem 3 is given in Section 9. Let S be the class of $2^{{}_rC_2}$ graphical models $\{(V, E)|M_E = 0, 1, \ldots, {}_rC_2\}$. It follows from Theorems 2 and 3 that CVLL selection is consistent for S in the sense of Definition 3.

In the following, we show a practical way to perform graphical model selection with the CVLL. The explicit form of G defined by (17), which is necessary to evaluate the CVLL by (5), cannot be given for $M_E > 1$. Hence we use the recursion introduced by Wermuth and Scheidt (1977) as an alternative way. For a positive definite matrix y, let us show the recursion to obtain the matrix g which satisfies

$$\begin{aligned} &g_{ab} = y_{ab}, &&(a, b) \in E,\\ &g^{ab} = 0, &&(a, b) \notin E. \end{aligned} \tag{18}$$

Let $F_i = \{(a_i, b_i), (b_i, a_i)\}$, $i = 0, 1, \ldots, M_E - 1$, be the elements of $\{\{(a, b), (b, a)\}|(a, b) \notin E\}$, and put $n' = n \pmod{M_E}$. Set $g_0 = y$ and calculate g_n, $n = 1, 2, \ldots,$ recursively by

$$g_{n,ab} = \begin{cases} g_{n-1,ab} + \dfrac{g_{n-1}^{ab}}{g_{n-1}^{aa} g_{n-1}^{bb} - g_{n-1}^{ab} g_{n-1}^{ba}}, & \text{for } (a, b) \in F_{n'},\\[2ex] g_{n-1,ab}, & \text{for } (a, b) \notin F_{n'}. \end{cases} \tag{19}$$

Then it follows from Proposition 3 of Speed and Kiiveri (1986) that $g := \lim_{n\to\infty} g_n$ satisfies (18).

The graph with the smallest CVLL among the $2^{{}_rC_2}$ candidate graphs is the CVLL selected graph. However, it is difficult to perform the search in practice, especially for large r. This and the recursion (19) lead to a backward stepwise selection procedure. Let G_k describe the graphical models (V, E_k), $k = 0, 1, \ldots,$ in the following procedure.

0. Put $k = 0$, set (V, E_0) as the graph with no conditional independent pairs and calculate $\mathrm{CVLL}(G_0)$.
1. Calculate the CVLLs of the candidates (V, E_{k+1}^i), $i = 1, \ldots, {}_rC_2 - k$, via (19), each of which has one more conditional independent pair than the graph (V, E_k).
2. Select the best graph (V, E_{k+1}) minimizing the CVLL among the candidates.
3. If $\mathrm{CVLL}(G_{k+1}) < \mathrm{CVLL}(G_k)$, set $k = k + 1$ and return to step 1. Otherwise, stop the procedure and take (V, E_k) as the selected model.

Remark 4. Speed and Kiiveri (1986, pp. 146–147) showed that the recursion (19) was a special case of the first cyclic algorithm (FCA). In the FCA, they set F_i, $i = 0, 1, \ldots, M_E - 1$, as the off-diagonal elements of the cliques of the complimentary graph $(V, \tilde{E})$ instead of the conditional independent pairs. The FCA is more effective and accurate than (19), since M_E, which is the number of the cycle, can be made smaller in the FCA. Equation (19) is, however, more suitable for practical use, since it is hard to find a clique in computer programming.

6. Empirical studies

In this section, our interest is focused on the empirical properties of CVLL selection. We shall conduct computational simulation and analyse hospital admissions data in Hong Kong to examine the properties. Throughout this section, we use the weight function $w_k = \cos(\pi k/m)$, $k = -m/2, \ldots, m/2$, with the bandwidth m which was set to minimize the CVLL.

First we apply the CVLL criterion to sdm model (9) which stipulates independence of component time series. Consider a three-variate time series generated by

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} = \begin{pmatrix} 0.2 & x & 0.0 \\ x & -0.2 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \varepsilon_{t,3} \end{pmatrix}, \tag{20}$$

where $\{\varepsilon_{t,a}, a = 1, 2, 3\}$ 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 sdm models as candidates:

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

For $x \neq 0$, cases II and IV are in category I and cases I and III are in category II. For $x = 0$, all the cases are in category II.

We select the case which has the smallest CVLL 1000 times. Table 1 shows the empirical frequencies of the CVLL selection for $x = 0.0$, 0.1 and 0.2 when the sample sizes are 101, 201 and 401. Table 1 clearly lends support to the consistency of the CVLL selection. The CVLL tends to select the optimal model – case III for $x \neq 0$ and case IV for $x = 0$ – more frequently as the sample size increases.

Table 1. Empirical frequencies of selection by the CVLL criterion based on 1000 replications. The case which has the smallest CVLL is selected from cases I–IV for time series (20)

Sample size	x	Case I	Case I	Case III	Case IV
101	0.0	0.019	0.146	0.162	0.673
	0.1	0.052	0.127	0.320	0.501
	0.2	0.103	0.056	0.664	0.177
201	0.0	0.021	0.144	0.139	0.69
	0.1	0.053	0.082	0.461	0.404
	0.2	0.089	0.012	0.867	0.032
401	0.0	0.015	0.117	0.108	0.760
	0.1	0.031	0.056	0.639	0.274
	0.2	0.062	0.000	0.938	0.000

Next we apply the CVLL criterion to graphical modelling. Consider the following vector autoregressive model:

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \\ X_{t,4} \\ X_{t,5} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.0 & 0.3 & 0.0 & 0.3 \\ 0.3 & -0.2 & x & 0.0 & 0.0 \\ 0.2 & x & 0.3 & 0.0 & 0.0 \\ 0.2 & 0.3 & 0.0 & 0.3 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \\ X_{t-1,3} \\ X_{t-1,4} \\ X_{t-1,5} \end{pmatrix} \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \varepsilon_{t,3} \\ \varepsilon_{t,4} \\ \varepsilon_{t,5} \end{pmatrix}, \tag{21}$$

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: 164), this process provides an example of a graph (V, E) such that

$$\{(a, b) | (a, b) \notin E,\ a < b\} = \begin{cases} (2, 5), (3, 4)\}, & \text{if } x \neq 0, \\ \{(2, 5), (3, 4), (2, 3)\}, & \text{if } x = 0. \end{cases}$$

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

I. $\{(a, b) | (a, b) \notin E_1,\ a < b\} = \varnothing$, which is equivalent to no constraint.
II. $\{(a, b) | (a, b) \notin E_2,\ a < b\} = \{(2, 5)\}$.
III. $\{(a, b) | (a, b) \notin E_3,\ a < b\} = \{(2, 5), (3, 4)\}$.
IV. $\{(a, b) | (a, b) \notin E_4,\ a < b\} = \{(2, 5), (3, 4), (2, 3)\}$.

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

We select the graph which has the smallest CVLL 1000 times. Table 2 shows the empirical frequencies of the CVLL selection when $x = 0.0$, 0.1 and 0.2 and sample sizes are 101, 201 and 401. Table 2 lends support to the consistency of the CVLL selection for graphical modelling. The CVLL tends to select the optimal graph – graph III for $x \neq 0$ and graph IV for $x = 0$ – more frequently as the sample size increases.

Finally, we consider a seven-variate time series of pollutants, weather and daily hospital admissions of patients suffering from circulatory and respiratory problems. The pollutant

Table 2. The empirical frequencies of selection by the CVLL criterion based on 1000 replications. The graph which has the smallest CVLL is selected from graphs I–IV for time series (21)

Sample size	x	Graph I	Graph II	Graph III	Graph IV
101	0.0	0.025	0.037	0.134	0.804
	0.1	0.024	0.074	0.271	0.631
	0.2	0.035	0.099	0.638	0.228
201	0.0	0.033	0.039	0.119	0.809
	0.1	0.033	0.089	0.383	0.495
	0.2	0.049	0.099	0.801	0.051
401	0.0	0.010	0.036	0.109	0.845
	0.1	0.021	0.064	0.578	0.337
	0.2	0.025	0.095	0.878	0.002

and weather data are the daily average levels of sulphur dioxide (SO_2, $\mu g\,m^{-3}$), nitrogen dioxide (NO_2, $\mu g\,m^{-3}$), respirable suspended particulates ($\mu g\,m^{-3}$), ozone (O_3, $\mu g\,m^{-3}$), temperature (degrees Celsius) and relative humidity (%), which were collected daily in Hong Kong from 1 January 1994 to 31 December 1995 (Figure 1). Taking the admissions series as the response variable and the other series as the explanatory variables, Xia *et al.* (2002) analysed the data using a semiparametric regression model. Their method is motivated by the nonlinearity 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 apply graphical models to the data. All the series are detrended by extracting the 15-day moving averages. For the hospital admissions series, the trend component caused by the day-of-the-week effect (Xia *et al.* 2002: 378) is removed by a simple regression method using dummy variables. For the seven-variate detrended time series, we estimate the spectral coherency $|f_{ab}(\lambda_j)|^2/\{f_{aa}(\lambda_j)f_{bb}(\lambda_j)\}$, $a, b = 1, \ldots, 7$, and the spectral partial coherency $|f^{ab}(\lambda_j)|^2/\{f^{aa}(\lambda_j)f^{bb}(\lambda_j)\}$, $a, b = 1, \ldots, 7$, as shown in Figure 2. The alignment technique for the coherency estimation (see Priestley 1981, Section 9.5.4) is used to increase the estimation accuracy, though it is 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 by the backward stepwise selection mentioned in Section 5.

Figures 2 and 3 suggest the following features:

1. The number of 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 (2005) used a weighted average of the past levels of the pollutant variables up to time point L (they took $L = 365$) for the explanatory variables. Our observation lends support to the findings based on their cumulative model.

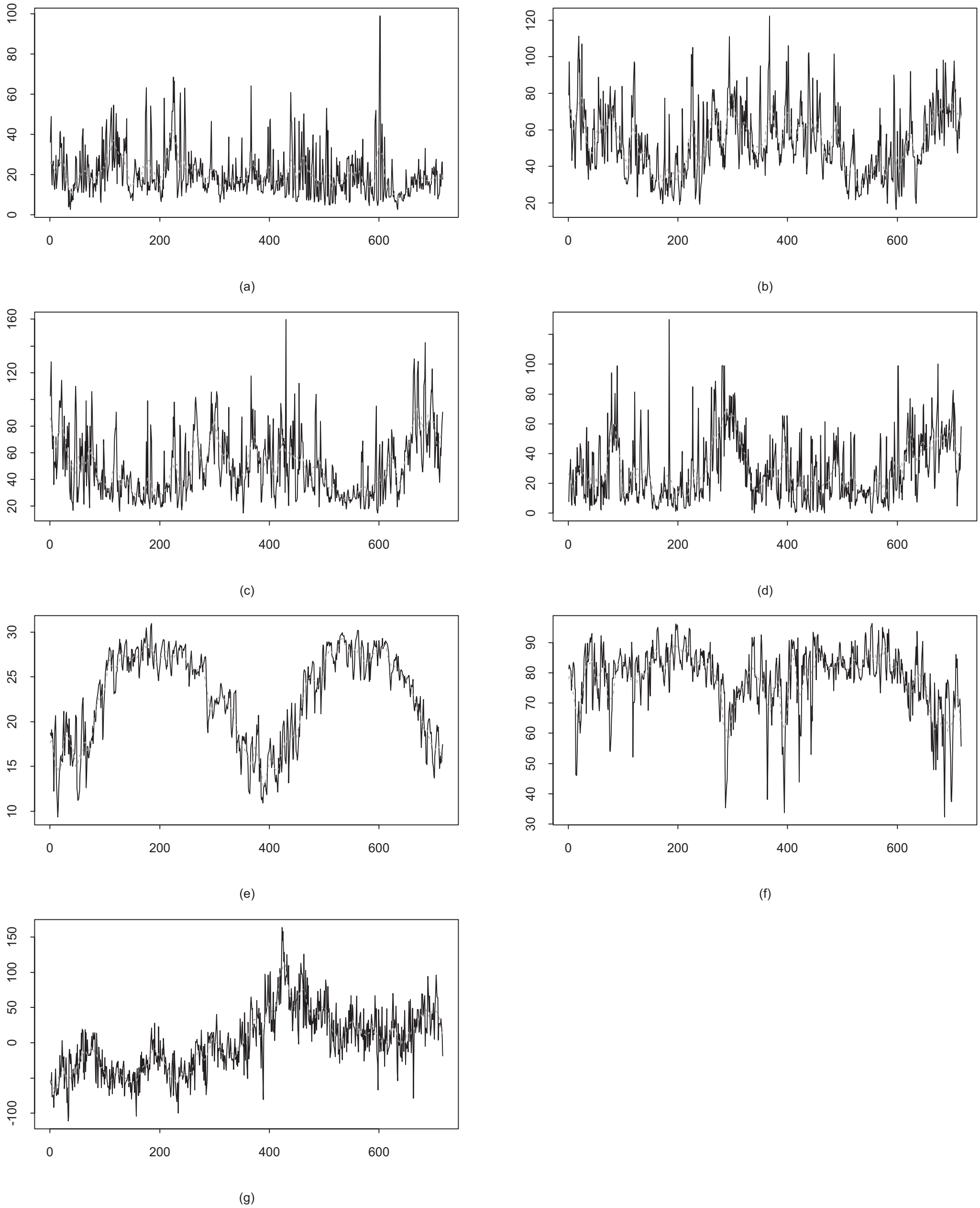

Figure 1. Average levels of (a) sulphur dioxide, (b) nitrogen dioxide, (c) respirable suspended particulates, (d) ozone, (e) temperature, (f) humidity, and (g) total number of daily hospital admissions of circulatory and respiratory patients, from 1 January 1994 to 31 December 1995, with trends estimated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

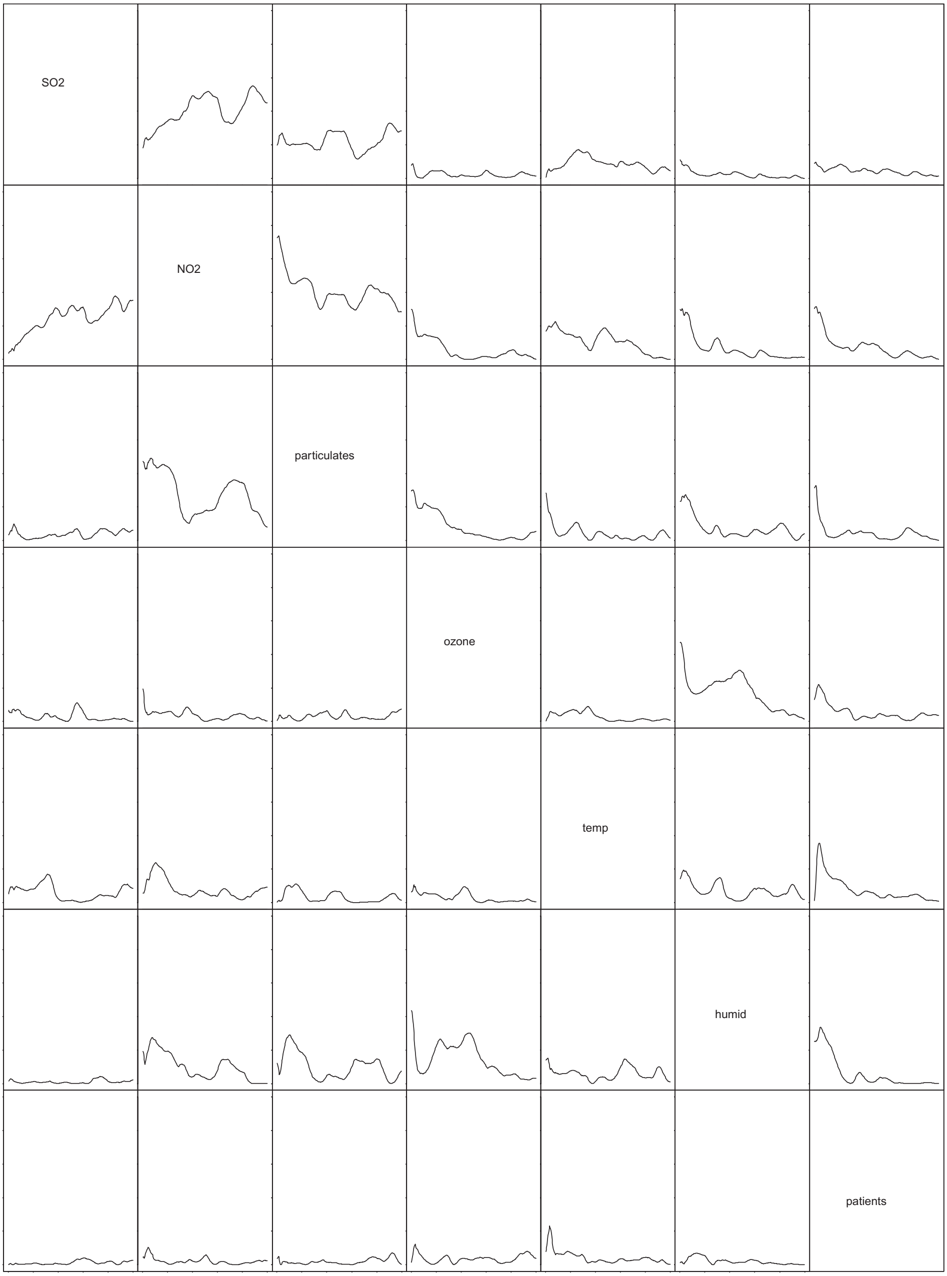

Figure 2. Spectral coherence (above diagonal) and partial spectral coherence (below diagonal) for the detrended time series of (a) SO_2, (b) NO_2, (c) particulates, (d) O_3, (e) temperature, (f) humidity and (g) hospital admissions.

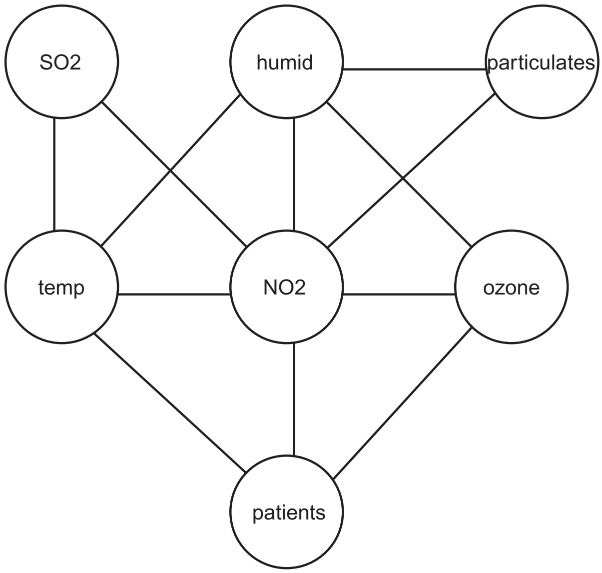

Figure 3. Estimated partial correlation graph for the detrended time series of Figure 1.

2. Temperature is the most influential of 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 shows its strong dependence on NO_2, O_3 and particulates. Humidity is a factor which may boost the impact of the pollutants under certain conditions, and is therefore considered to have an indirect but principal effect for HA.
4. Figure 3 supports the fact that NO_2 plays an important role in the process of O_3 generation (Dahlhaus 2000: 168).

7. Concluding remarks

This paper proposes the CVLL criterion for sdm model selection for time series. Regarding a graphical model for time series as an example of sdm models, we provide a consistent method for graphical modelling by the CVLL criterion. There are interesting extensions of graphical modelling 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, such as the nonlinear dependency between the hospital admissions and the pollutant variables identified by Xia *et al.* (2002). We leave these interesting extensions of graphical modelling to future studies.

8. Lemmas

We require a number of lemmas. The proofs of Lemmas 1–4 are given by Yajima and Matsuda (2003, Lemmas 1–4).

Lemma 1. *Under Assumptions 1, 3 and 4,*

$$\mathrm{E}(W_a(\lambda_j)\overline{W}_b(\lambda_j)) - f_{ab}(\lambda_j) = O(n^{-1}\log n), \qquad -m/2+1 \leqslant j \leqslant [n/2]+m/2,$$
$$\mathrm{E}(W_a(\lambda_j)W_b(\lambda_k)) = O(n^{-1}\log n), \qquad -m/2+1 \leqslant j \leqslant k \leqslant [n/2]+m/2,\ j+k \neq 0,\ n,$$
$$\mathrm{E}(W_a(\lambda_j)W_b(\lambda_k)) - f_{ab}(\lambda_j) = O(n^{-1}\log n), \qquad -m/2+1 \leqslant j \leqslant k \leqslant [n/2]+m/2,\ j+k = 0,\ n,$$
$$\mathrm{E}(W_a(\lambda_j)\overline{W}_b(\lambda_k)) = O(n^{-1}\log n), \qquad -m/2+1 \leqslant j < k \leqslant [n/2]+m/2,$$

uniformly in j and k for $a, b = 1, 2, \ldots, r$.

Lemma 2. *Under Assumptions 1, 3 and 4,*

$$\mathrm{E}(\hat{f}_{ab}(\lambda_j)) - f_{ab}(\lambda_j) = O(m^2 n^{-2}),$$

uniformly in $j = 1, 2, \ldots, [n/2]$, *for* $a, b = 1, 2, \ldots, r$.

Lemma 3. *Under Assumptions 1, 3 and 4,*

$$\mathrm{cov}(\hat{f}_{ab}(\lambda_j), \hat{f}_{cd}(\lambda_k)) = \begin{cases} O(m^{-1}), & \text{if } |j-k| \leqslant m, \\ O(n^{-2}\log^2 n), & \text{if } |j-k| > m, \end{cases}$$

uniformly in $j, k = 1, 2, \ldots, [n/2]$, *for* $a, b, c, d = 1, 2, \ldots, r$.

Lemma 4. *Under Assumptions 1, 3 and 4,*

$$\mathrm{cum}(\hat{f}_{a_1 b_1}(\lambda_{j_1}), \hat{f}_{a_2 b_2}(\lambda_{j_2}), \ldots, \hat{f}_{a_k b_k}(\lambda_{j_k})) = O(m^{1-k}),$$

uniformly in $1 \leqslant j_i \leqslant [n/2]$ *for any* $k \geqslant 3$ *and* $1 \leqslant a_i, b_i \leqslant r$, $i = 1, \ldots, k$.

Lemma 5. *For an invertible* $r \times r$ *matrix* A,

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

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

Proof. The result is derived simply by componentwise calculaton using rule 9 of Lütkepohl (1991: 471). $\square$

Lemma 6. *Let* $U = (U_1, \ldots, U_p)^{\mathrm{T}}$ *and* $V = (V_1, \ldots, V_q)^{\mathrm{T}}$ *be mutually independent random vectors whose covariance matrices are positive definite. Then the covariance matrix of the* pq*-dimensional random vector* $(U_i V_j)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, *is positive definite.*

Proof. The covariance matrix of $(U_i V_j)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, is the Kronecker product of the covariance matrices of U and V. Hence the assertion follows from Lütkepohl (1991, A.11 (6)). $\square$

Let (V, E) be a graphical model for r-dimensional time series with $M_E := \#\{(a, b)|(a, b) \notin E, a < b\}$, and let (a_i, b_i), $i = 1, \ldots, 2M_E$, and (p_i, q_i), $i = 1, \ldots, r^2 - 2M_E$, be the elements of the sets $\{(a, b)|(a, b) \notin E\}$ and $\{(p, q)|(p, q) \in E\}$, respectively.

Lemma 7. *Let* y *be a positive definite* $r \times r$ *matrix with the property that* $y^{ab} = 0$ *for* $(a, b) \notin E$. *Then there exist holomorphic functions* $G_{a_i b_i}$, $i = 1, \ldots, 2M_E$, *on a neighbourhood of* $(y_{p_j q_j})$, $j = 1, \ldots, r^2 - 2M_E$, *such that*

$$y_{a_i b_i} = G_{a_i b_i}(y_{p_j q_j}, j = 1, \ldots, r^2 - 2M_E), \qquad i = 1, \ldots, 2M_E. \tag{22}$$

Define the derivatives of $G_{a_i b_i}$ *as* $h_{a_i b_i, p_j q_j}(y) := \partial G_{a_i b_i}(y)/\partial y_{p_j q_j}$, $i = 1, \ldots, 2M_E$, $j = 1, \ldots, r^2 - 2M_E$. *Then*

$$A(y)H(y) + B(y) = O_{2M_E, r^2 - 2M_E}, \tag{23}$$

where $A(y)$, $H(y)$ *and* $B(y)$ *are* $2M_E \times 2M_E$, $2M_E \times (r^2 - 2M_E)$ *and* $2M_E \times (r^2 - 2M_E)$ *matrices, respectively, given by*

$$\begin{aligned}
A(y) &:= -y^{a_i a_j} y^{b_j b_i}), && i, j = 1, \ldots, 2M_E, \\
H(y) &:= (h_{a_i b_i, p_j q_j}(y)), && i = 1, \ldots, 2M_E, j = 1, \ldots, r^2 - 2M_E, \\
B(y) &:= (-y^{a_i p_j} y^{q_j b_i}), && i = 1, \ldots, 2M_E, j = 1, \ldots, r^2 - 2M_E.
\end{aligned}$$

Proof. y satisfies the following $2M_E$ restrictions:

$$y^{a_i b_i} = 0, \qquad i = 1, \ldots, 2M_E.$$

The Jacobian for the $2M_E$ equations is, by Lemma 5, $|A(y)|$. Note that $-A(y)$ is equal to the covariance matrix of a random vector $c := (X_{a_1} \cdot Y_{b_1}, \ldots, X_{a_{2M_E}} \cdot Y_{b_{2M_E}})^{\mathrm{T}}$, where $(X_{a_i}, i = 1, \ldots, 2M_E)$ and $(Y_{b_i}, i = 1, \ldots, 2M_E)$ are zero-mean and mutually independent random vectors with $\mathrm{E}(X_{a_i} \overline{X}_{a_j}) = y^{a_i a_j}$, $\mathrm{E}(Y_{b_i} \overline{Y}_{b_j}) = y^{b_j b_i}$, for $i, j = 1, \ldots, 2M_E$. Let $a_{i_1}, \ldots, a_{i_p}$ be the largest subset of $a_1, \ldots, a_{2M_E}$ with distinct elements $a_{i_m} \neq a_{i_n}$, $(m \neq n)$. Define $b_{i_1}, \ldots, b_{i_q}$ in the same way. Then the covariance matrices $U = (X_{a_{i_1}}, \ldots, X_{a_{i_p}})^{\mathrm{T}}$ and $V = (Y_{b_{i_1}}, \ldots, Y_{b_{i_q}})^{\mathrm{T}}$ are positive definite. Noting that c is a subvector of $(U_i V_j)$,

$i = 1, \dots, p$, $j = 1, \dots, q$, and applying Lemma 6 to the U, V, we have $|-A(y)| > 0$. The assertion follows from the implicit function theorem. □

Lemma 8. *For a positive definite $r \times r$ matrix y with the property that $y^{ab} = 0$ for $(a, b) \notin E$, define*

$$x(\alpha, \beta, a, b, y) := \sum_{(p,q)\in E} h_{ab,pq}(y) y_{\alpha q} y_{p\beta} - y_{\alpha b} y_{a\beta}.$$

(i) *If $(\alpha, \beta) \in E$, then for $(a, b) \notin E$,*

$$x(\alpha, \beta, a, b, y) = 0.$$

(ii) *If $(\alpha, \beta) \notin E$, then for $(c, d) \notin E$,*

$$\sum_{(a,b)\notin E} x(\alpha, \beta, a, b, y) y^{ca} y^{bd} + I_{c=\beta} I_{d=\alpha} = 0.$$

Proof. (i) Define the $(r^2 - 2M_E) \times 1$ vector

$$u(y) := (y_{\alpha q_i} y_{p_i \beta}), \qquad i = 1, \dots, r^2 - 2M_E$$

and $2M_E \times 1$ vector

$$v(y) := (y_{\alpha b_i} y_{a_i \beta}), \qquad i = 1, \dots, 2M_E.$$

Multiplying (23) on the right by $u(y)$, we have

$$A(y)H(y)u(y) + B(y)u(y) = O_{2M_E,1}.$$

Hence

$$\begin{aligned} A(y)(H(y)u(y) - v(y)) &= -(B(y)u(y) + A(y)v(y)) \\ &= O_{2M_E,1}, \end{aligned}$$

since the ith component of the right-hand side in the first equation is equal to

$$\begin{aligned} \sum_{(p,q)\in E} y^{a_i p} y^{q b_i} y_{\alpha q} y_{p\beta} + \sum_{(p,q)\notin E} y^{a_i p} y^{q b_i} y_{\alpha q} y_{p\beta} &= \sum_{p,q=1}^{r} y^{a_i p} y^{q b_i} y_{\alpha q} y_{p\beta} \\ &= I_{\beta=a_i} I_{\alpha=b_i} \qquad (24) \\ &= 0. \end{aligned}$$

(ii). Let us define the $2M_E \times 1$ vectors

$$X(\alpha, \beta) := H(y)u(y) - v(y),$$

$$J(\alpha, \beta) := (I_{\beta=a_i} I_{\alpha=b_i}), \qquad i = 1, \dots, 2M_E.$$

Then from (24),

$$X(\alpha, \beta) - A(y)^{-1}J(\alpha, \beta) = O_{2M_E,1}.$$

Multiplying the above expression on the left by the $1 \times 2M_E$ vector $(y^{ca_i} y^{b_i d})$, $i = 1, \ldots, 2M_E$, we obtain the result. □

9. Proofs of theorems

Throughout this section, assume that $\sum_{k=-m/2}^{m/2} w_k = 1$ without loss of generality.

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

$$\begin{aligned}
\mathrm{CVLL}(G) = &\sum_{j=1}^{[n/2]} (\log\det f_j + r) + \sum_{j=1}^{[n/2]} \{\mathrm{tr}(f_j g_j^{-1}) - \log\det(f_j g_j^{-1}) - r\} \\
&+ \sum_{j=1}^{[n/2]} \Big\{ \log\det\Big(\hat{g}_{-j,j} g_j^{-1}\Big) + \mathrm{tr}\Big(\hat{g}_{-j,j}^{-1} - g_j^{-1}\Big) f_j \\
&+ \ \mathrm{tr}\big(I_j - f_j\big)\Big(\hat{g}_{-j,j}^{-1} - g_j^{-1}\Big) + \mathrm{tr}\big(I_j - f_j\big) g_j^{-1} \Big\}.
\end{aligned} \tag{25}$$

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

Put $D_j = (\hat{g}_{-j,j} - g_j) g_j^{-1}$. 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),

$$\begin{aligned}
\sum_{j=1}^{[n/2]} \log\det\Big(\hat{g}_{-j,j} g_j^{-1}\Big) &= \sum_j \log\det(D_j + E) \\
&= O\Bigg(\sum_j |\mathrm{tr}(D_j)|\Bigg) \\
&= o_p(n).
\end{aligned}$$

In the same way,

$$\sum_{j=1}^{[n/2]} \mathrm{tr}\left(\hat{g}_{-j,j}^{-1} - g_j^{-1}\right) f_j = \sum_j \mathrm{tr}\left(\hat{g}_{-j,j}^{-1}\left(g_j - \hat{g}_{-j,j}\right) g_j^{-1} f_j\right)$$
$$= o_p(n).$$

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

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

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),

$$\begin{aligned}
\left|\sum_{j=1}^{[n/2]} \mathrm{tr}\left(I_j - f_j\right)\left(\hat{g}_{-j,j}^{-1} - g_j^{-1}\right)\right| &\leqslant \sum_{a,b=1}^{r} \max_j |\hat{g}_{-j,j}^{ba} - g_j^{ba}| \sum_j |I_{j,ab} - f_{j,ab}| \\
&= o_p(n^{-1/4}) O_p(n) \\
&= o_p(n).
\end{aligned}$$

□

Proof of Theorem 2. Since $f_j = g_j$ for all $j = 1, \ldots, [n/2]$, the second term of (25) is 0 and we have

$$\begin{aligned}
\mathrm{CVLL}(G) = &\sum_{j=1}^{[n/2]} \{\log\det f_j + r + \mathrm{tr}(I_j - f_j)f_j^{-1}\} \\
&+ \sum_{j=1}^{[n/2]} \left\{\log\det\left(\hat{g}_{-j,j} f_j^{-1}\right) + \mathrm{tr}\left(\hat{g}_{-j,j}^{-1} - f_j^{-1}\right) f_j\right\} \qquad (26) \\
&+ \sum_{j=1}^{[n/2]} \mathrm{tr}\left(I_j - f_j\right)\left(\hat{g}_{-j,j}^{-1} - f_j^{-1}\right).
\end{aligned}$$

First we shall show that the third term is $o_p(nm^{-1})$. For simplicity, we give a proof for $r = 1$ and one-dimensional θ, but the general case is proved in essentially the same way. By applying Taylor's expansion,

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

say, 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 random, while $M_{ik,j}$ is a constant. It suffices to show that

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

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

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

since these terms dominate the others. First,

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

Next, by Lemmas 1 and 2,

$$
\sum_{j=1}^{[n/2]} (I_j - f_j) M_{01,j} (\hat{f}_{-j,j} - f_j) = \sum_{j=1}^{[n/2]} (I_j - \mathrm{E}(I_j)) M_{01,j} (\hat{f}_{-j,j} - \mathrm{E}(\hat{f}_{-j,j})) + O_p(m^2 n^{-3/2}),
$$

and by Theorem 2.3.2 of Brillinger (1981),

$$\mathrm{E}\left\{\sum_{j=1}^{[n/2]}(I_j-\mathrm{E}(I_j))M_{01,j}(\hat{f}_{-j,j}-\mathrm{E}(\hat{f}_{-j,j}))\right\}^2$$

$$=\sum_{j_1,j_2=1}^{[n/2]}\sum_{k_1,k_2=-m/2,k_1,k_2\neq 0}^{m/2}M_{01,j_1}M_{01,j_2}w_{k_1}w_{k_2}$$

$$\times\,\mathrm{E}\{(I_{j_1}-\mathrm{E}(I_{j_1}))(I_{j_2}-\mathrm{E}(I_{j_2}))(I_{j_1+k_1}-\mathrm{E}(I_{j_1+k_1}))(I_{j_2+k_2}-\mathrm{E}(I_{j_2+k_2}))\}$$

$$=\sum_{j_1,j_2}\sum_{k_1,k_2\neq 0}M_{01,j_1}M_{01,j_2}w_{k_1}w_{k_2}\{\mathrm{cum}(I_{j_1},\,I_{j_2})\mathrm{cum}(I_{j_1+k_1},\,I_{j_2+k_2})$$

$$+\,\mathrm{cum}(I_{j_1},\,I_{j_1+k_1})\mathrm{cum}(I_{j_2},\,I_{j_2+k_2})+\mathrm{cum}(I_{j_1},\,I_{j_2+k_2})\mathrm{cum}(I_{j_2},\,I_{j_1+k_1})$$

$$+\,\mathrm{cum}(I_{j_1},\,I_{j_2},\,I_{j_1+k_1},\,I_{j_2+k_2})\}$$

$$=O\big(m^{-2}nm\big)+O\big(m^{-2}n^2m^2(\log n)^4n^{-4}\big)+O\big(m^{-2}nm\big)+O\big(m^{-2}nm\big)=O\big(nm^{-1}\big),$$

where Lemma 1 is used in the last equality. It follows that

$$\sum_j(I_j-f_j)M_{01,j}(\hat{f}_{-j,j}-f_j)=O_p\left(n^{1/2}m^{-1/2}\right)+O_p\left(m^2n^{-3/2}\right)$$

$$=o_p(nm^{-1}).$$

Finally, by Proposition 1 of Matsuda and Yajima (2004),

$$\left|\sum_{j=1}^{[n/2]}(I_j-f_j)N_{03,j}(\hat{f}_{-j,j}-f_j)^3\right|\leqslant\max_j|N_{03,j}|\left\{\sum_j(I_j-f_j)^2\right\}^{1/2}\left\{\sum_j(\hat{f}_{-j,j}-f_j)^6\right\}^{1/2}$$

$$=\{O_p(n)\}^{1/2}\{o_p\left(nn^{-3/2}\right)\}^{1/2}=o_p(n^{1/4}).$$

Put $\delta_j=(\hat{g}_{-j,j}-f_j)f_j^{-1}$. Then the second term of (26) is

$$\sum_{j=1}^{[n/2]}\log\det(\delta_j+E)-\mathrm{tr}\{\delta_j(\delta_j+E)^{-1}\}$$

$$=\sum_j\mathrm{tr}\left(\delta_j-\frac{1}{2}\delta_j^2\right)-\mathrm{tr}\big(\delta_j(E-\delta_j)\big)+O\left(\sum_j\max_{a,b}|\delta^3_{j,ab}|\right)$$

$$=\frac{1}{2}\sum_j\mathrm{tr}\left(\delta_j^2\right)+o_p\left(n^{1/4}\right).$$

By Taylor's expansion,

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

where the fourth and fifth equalities follow from Proposition 1 of Matsuda and Yajima (2004) and Lemmas 1, 2, respectively. Let us define

$$
\mu_{\alpha\beta\gamma\nu}(\lambda_j) := \sum_{a,b=1}^{r} \sum_{e_1,e_2=1}^{r} h_{ae_1,\alpha\beta}(f_j) h_{be_2,\gamma\nu}(f_j) f_j^{e_1 b} f_j^{e_2 a},
$$

$$
y_{\alpha\beta,j} := \hat{f}_{-j,j,\alpha\beta} - \mathrm{E}(\hat{f}_{-j,j,\alpha\beta}).
$$

We shall prove that

$$\mathrm{E}\left(\sum_{j=1}^{[n/2]}\sum_{\alpha,\beta,\gamma,\nu=1}^{r}\mu_{\alpha\beta\gamma\nu}(\lambda_j)y_{\alpha\beta,j}y_{\gamma\nu,j}\right) = \frac{n}{m}\frac{C_u}{2\pi}\int_0^{\pi}\sum_{\alpha,\beta,\gamma,\nu=1}^{r}\mu_{\alpha\beta\gamma\nu}(\lambda)f_{\alpha\nu}(\lambda)f_{\gamma\beta}(\lambda)\mathrm{d}\lambda + o(nm^{-1}), \tag{27}$$

$$\mathrm{var}\left(\sum_{j=1}^{[n/2]}\sum_{\alpha,\beta,\gamma,\nu=1}^{r}\mu_{\alpha\beta\gamma\nu}(\lambda_j)y_{\alpha\beta,j}y_{\gamma\nu,j}\right) = O\left(nm^{-1}\right) \tag{28}$$

and thus complete the proof of Theorem 2.

By Lemma 1,

$$\begin{aligned}\mathrm{E}(y_{\alpha\beta,j}y_{\gamma\nu,j}) &= \sum_{k,l=-m/2,k,l\neq 0}^{m/2} w_k w_l\{\mathrm{E}(I_{\alpha\beta,j+k}I_{\gamma\nu,j+l}) - \mathrm{E}(I_{\alpha\beta,j+k})\mathrm{E}(I_{\gamma\nu,j+l})\} \\ &= \sum_{k=-m/2,k\neq 0}^{m/2} w_k^2 f_{\alpha\nu,j+k}f_{\gamma\beta,j+k} + O\left((nm)^{-1}\log n\right) \\ &= \left(\sum_k w_k^2\right)f_{\alpha\nu,j}f_{\gamma\beta,j} + O\left(mn^{-2}\right) + O((nm)^{-1}\log n).\end{aligned}$$

The last equality is given by Taylor's expansion. It follows from Exercise 1.7.14 of Brillinger (1981) that the left-hand side of (27) is

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

We can put $\mu_{\alpha\beta\gamma\nu}(\lambda_j) = 1$ without loss of generality. By Lemmas 3 and 4, the left-hand side of (28) is evaluated as

$$\begin{aligned}\mathrm{var}\left(\sum_j y_{\alpha\beta,j}y_{\gamma\nu,j}\right) &= \mathrm{cov}\left(\sum_j y_{\alpha\beta,j}y_{\gamma\nu,j},\ \sum_k \bar{y}_{\alpha\beta,k}\bar{y}_{\gamma\nu,k}\right) \\ &= \sum_{j,k}\{\mathrm{cum}(y_{\alpha\beta,j},\ y_{\gamma\nu,j},\ \bar{y}_{\alpha\beta,k},\ \bar{y}_{\gamma\nu,k}) \\ &\quad + \mathrm{cum}(y_{\alpha\beta,j},\ \bar{y}_{\alpha\beta,k})\mathrm{cum}(y_{\gamma\nu,j},\ \bar{y}_{\gamma\nu,k}) + \mathrm{cum}(y_{\alpha\beta,j},\ \bar{y}_{\gamma\nu,k})\mathrm{cum}(\bar{y}_{\alpha\beta,k},\ y_{\gamma\nu,j})\} \\ &= O(n^2m^{-3}) + 2\left(O(nmm^{-2}) + O(n^2(\log n)^2n^{-2})\right) = O(nm^{-1}),\end{aligned}$$

which proves (28). □

Proof of Theorem 3. We derive the AMISE of the graph (V, E) for a time series $\{X_{t,a},\ a = 1, \ldots, r\}$ applying (7).

Since $h_{ab,cd}(f(\lambda)) = 0$ for $(a, b) \notin E$, $(c, d) \notin E$ (see (22)), and

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

the integrand of AMISE(G) is

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

say. Then

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

By Lemma 8(i),

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

In the same way,

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

By Lemma 8(i) and (ii),

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

which completes the proof. □

Acknowledgements

This manuscript was prepared while Yasumasa Matsuda 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. Yoshihiro Yajima's research was partially supported by JSPS grant A(1)15200021. The authors are grateful to Dr. C.M. Wong of the University of Hong Kong for making the hospital admissions data available to us and wish to thank the editor and referee for comments that helped us significantly improve the paper.

References

Akaike, H. (1974) A new look at statistical model identification. *IEEE Trans. Automat. Control*, **19**, 716–723.

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

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

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

Chan, K.S., Ho, L.-H. and Tong, H. (2005) A note on time-reversibility of multivariate linear processes. Technical Report 344, Department of Statistics and Actuarial Research, University of Iowa, Iowa City.

Cheng, B. and Tong, H. (1992) On consistent nonparametric order determination and chaos. *J. Roy. Statist. Soc. Ser. B*, **54**, 427–449.

Dahlhaus, R. (2000) Graphical interaction models for multivariate time series. *Metrika*, **51**, 151–172.

Diggle, P.J., Liang, K. and Zeger S.L. (1994) *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

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.

Härdle, W., Hall, P. and Marron, J.S. (1988) How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.*, **83**, 86–95.

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.

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

Hurvich, C.M. (1985) Data-driven choice of a spectrum estimate: Extending the cross-validation methods. *J. Amer. Statist. Assoc.*, **80**, 933–940.

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

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

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

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

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.

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

Priestley, M.B. (1981) *Spectral Analysis and Time Series*, 2 vols. New York: Academic Press.

Shao, J. (1993) Linear model selection by cross-validation. *J. Amer. Statist. Assoc.*, **88**, 486–494.

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

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

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

Xia, Y. and Tong, H. (2005) Cumulative effects of air pollution on public health. *Statistics in Medicine*. To appear.

Xia, Y., Tong, H., Li, W.K. and Zhu, L. (2002) An adaptive estimation of dimension reduction space. *J. Roy. Statist. Soc. Ser. B*, **64**, 363–410.

Yajima, Y. and Matsuda, Y. (2003) On nonparametric and semiparametric testing for multivariate time series. Discussion Paper, CIRJE-F-253, Faculty of Economics, University of Tokyo.

Received December 2004 and revised August 2005