

GRAPHICAL INTERACTION MODELS FOR TIME SERIES: PARAMETER ESTIMATION AND MODEL SELECTION

MICHAEL EICHLER

The University of Chicago

ABSTRACT. We present a parametric approach for graphical interaction modelling in multivariate stationary time series. In these models, the possible dependencies between the components of the process are represented by edges in an undirected graph. We consider vector autoregressive models and propose a parametrization in terms of inverse covariances, which are constrained to zero for missing edges. The parameter can be estimated by minimization of Whittle's log-likelihood, which leads to similar likelihood equation as for covariance selection models. We discuss the problem of model selection and prove asymptotic efficiency of AIC-like criteria.

1. INTRODUCTION

Graphical models have become an important tool for analyzing multivariate data. While the theory originally has been developed for variables that are sampled with independent replications, graphical models recently have been applied also to stationary multivariate time series (e.g. Brillinger 1996, Dahlhaus 2000, Eichler 1999, 2001, 2002, Dahlhaus and Eichler 2003).

A particularly simple graphical representation is provided by graphical interaction models, which visualize dependencies by undirected graphs. For time series, a similar approach has been proposed by Dahlhaus (2000) who introduced so-called partial correlation graphs, in which each component of the time series is represented by one vertex in the graph. Dahlhaus et al. (1997) suggested a nonparametric test for the presence of an edge in the partial correlation graph based on the maximum of the spectral coherence. The concept of partial correlation graphs has been used in many application from various fields (Dahlhaus et al. 1997, Timmer et al. 2000, Gather et al. 2002, Fried and Didelez 2003, Fried et al. 2003).

The main disadvantage of the current nonparametric approach to graphical modelling based on partial correlation graphs is the lack of a rigorous theory for identifying the best fitting graph. An alternative to the nonparametric approach is the fitting of parametric graphical models where the parameters are constrained with respect to undirected graphs. The problem of estimating the dependence structure of the process now becomes a problem of model selection where the best approximating model minimizes some chosen model distance such as the Kullback-Leibler information divergence.

2000 *Mathematics Subject Classification.* Primary 62M10; Secondary 62M15,62F30.

Key words and phrases. Graphical interaction models, multivariate stationary time series, Whittle likelihood, autoregressive models, inverse covariance, model selection.

In this paper, we propose graphical interaction models for stationary time series that are defined in terms of inverse covariances. In these models, the conditional independences encoded by an undirected graph G correspond to zero constraints on the parameters.

In Section 2, we define graphical interaction models in terms of inverse covariances and indicate their relation to graphical vector autoregressive models. In Section 3, we derive implicit equations for the Whittle estimators, which are similar to the equations for the maximum likelihood estimate in the case of ordinary Gaussian covariance selection models. In Section 4 we investigate the asymptotic behaviour of the Kullback-Leibler information and show that it can be approximated by a deterministic function. This is exploited in Section 5 to derive the asymptotic efficiency of the proposed AIC-like model selection criterion.

2. GRAPHICAL INTERACTION MODELS AND VECTOR AUTOREGRESSIONS

Let $X = \{X(t), t \in \mathbb{Z}\}$ be a multivariate stationary Gaussian process of dimension d defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We suppose that X is purely non-deterministic and has mean zero. Furthermore, denoting the spectral matrix of X at frequency λ by $f(\lambda)$, we assume there exists a real constant $c \geq 1$ such that

$$c^{-1} \mathbb{1}_d \leq f(\lambda) \leq c \mathbb{1}_d \quad \text{for all } \lambda \in [-\pi, \pi]. \quad (2.1)$$

Here, $\mathbb{1}_d$ is the identity matrix and, for matrices A and B , we write $A \leq B$ if $B - A$ is nonnegative definite. Under these assumptions, X has a mean-square convergent autoregressive representation

$$X(t) = \sum_{u=1}^{\infty} A(u) X(t-u) + \varepsilon(t), \quad (2.2)$$

where $\{\varepsilon(t), t \in \mathbb{Z}\}$ is a Gaussian white noise process with non-singular covariance matrix Σ .

In this paper, we consider undirected graphs to describe the dependence structure of the process X . Let $V = \{1, \dots, d\}$ be the set of indices of X . Then a *graph* G over V is given by an ordered pair (V, E) where the elements on V represent the *vertices* or *nodes* of the graph and E is a collection of *edges* e denoted as $a - b$ for distinct nodes a, b in V .

For a process X , we say that two components X_a and X_b are *conditionally independent* given $X_{V_{ab}}$, where $V_{ab} = V \setminus \{a, b\}$, if the corresponding σ -algebras generated by X_a , X_b , and $X_{V_{ab}}$ satisfy

$$\sigma\{X_a(t), t \in \mathbb{Z}\} \perp\!\!\!\perp \sigma\{X_b(t), t \in \mathbb{Z}\} \mid \sigma\{X_{V_{ab}}(t), t \in \mathbb{Z}\}.$$

In this case, we write $X_a \perp\!\!\!\perp X_b \mid X_{V_{ab}}$. If the process X is Gaussian, the conditional independence of X_a and X_b given $X_{V_{ab}}$ can be expressed in terms of the residual processes of X_a and X_b after removing the linear effects of $X_{V_{ab}}$. More precisely, let

$$\varepsilon_{a|V_{ab}}(t) = X_a(t) - \mathbb{E}(X_a(t) \mid X_{V_{ab}}(s), s \in \mathbb{Z}) \quad (2.3)$$

and

$$\varepsilon_{b|V_{ab}}(t) = X_b(t) - \mathbb{E}(X_b(t) \mid X_{V_{ab}}(s), s \in \mathbb{Z}) \quad (2.4)$$

for all $t \in \mathbb{Z}$. Then X_a and X_b are conditionally independent given $X_{V_{ab}}$ if and only if $\varepsilon_{a|V_{ab}}(t)$ and $\varepsilon_{b|V_{ab}}(s)$ are independent for all $t, s \in \mathbb{Z}$.

The dependence structure of a process X can be graphically represented by an undirected graph that encodes the pairwise conditional independences for the process.

Definition 2.1 (Conditional independence graph). Let X be a stationary process of dimension d . The *conditional independence graph* associated with X is a graph $G = (V, E)$ with vertex set $V = \{1, \dots, d\}$ and edge set E such that

$$a - b \notin E \Leftrightarrow X_a \perp\!\!\!\perp X_b \mid X_{V_{ab}}$$

for all pairs $a, b \in V$.

We note, that under additional assumptions more general conditional independence relations can be derived from the conditional independence graph G . Such properties that allow to associate certain statements about the graph G with corresponding conditional independence statements about the variables in X are usually called Markov properties with respect to G . For instance, if X is a Gaussian process and condition (2.1) holds, then X satisfies also the so-called global Markov property with respect to G . For details, we refer to Dahlhaus (2000).

Inference about conditional independence graphs for time series commonly is based in the frequency domain (e.g. Dahlhaus 2000, Dahlhaus et al. 1997, Fried and Didelez 2003). Here, the dependence between components X_a and X_b given $X_{V_{ab}}$ can be described by the partial cross-spectrum of X_a and X_b given $X_{V_{ab}}$,

$$f_{ab|V_{ab}}(\lambda) = f_{\varepsilon_a|V_{ab}\varepsilon_b|V_{ab}}(\lambda),$$

or, equivalently, by the partial spectral coherence of X_a and X_b given $X_{V_{ab}}$,

$$R_{ab|V_{ab}}(\lambda) = \frac{f_{ab|V_{ab}}(\lambda)}{(f_{aa|V_{ab}}(\lambda)f_{bb|V_{ab}}(\lambda))^{1/2}} = \frac{f_{\varepsilon_a|V_{ab}\varepsilon_b|V_{ab}}(\lambda)}{(f_{\varepsilon_a|V_{ab}\varepsilon_a|V_{ab}}(\lambda)f_{\varepsilon_b|V_{ab}\varepsilon_b|V_{ab}}(\lambda))^{1/2}},$$

where $\varepsilon_a|V_{ab}$ and $\varepsilon_b|V_{ab}$ are the residual processes given by (2.3) and (2.4), respectively. It follows that

$$X_a \perp\!\!\!\perp X_b \mid X_{V \setminus \{a,b\}} \Leftrightarrow R_{ab|V_{ab}}(\lambda) = 0 \text{ for all } \lambda \in [-\pi, \pi]. \quad (2.5)$$

For random vectors, the partial correlations can be obtained from the inverse of the covariance matrix. Dahlhaus (2000) showed that a similar relationship holds between the partial spectral coherences and the inverse of the spectral matrix.

Lemma 2.2. *Suppose that X is a vector-valued stationary process such that condition (2.1) holds. Then, if $g(\lambda) = f(\lambda)^{-1}$ denotes the inverse spectral matrix, we have*

$$R_{ab|V_{ab}}(\lambda) = -\frac{g_{ab}(\lambda)}{\sqrt{g_{aa}(\lambda)g_{bb}(\lambda)}}.$$

Proof. The lemma has been proved in Dahlhaus (2000), Theorem 2.4. \square

This relation not only provides an efficient method for computing the partial spectral coherences of a process X , but also allows a new time domain based characterization of conditional independences and thus of the absent edges in the conditional independence graph. For this, let $R = (R(u-v))_{u,v \in \mathbb{Z}}$ with $R(u) = \mathbb{E}(X(t)X(t+u)')$

be the infinite dimensional covariance matrix of X and let $R^{(i)} = R^{-1}$ be the inverse of R . Then $R^{(i)}$ is related to the Fourier transform of the inverse spectral matrix by

$$R^{(i)}(u) = \frac{1}{4\pi} \int_{\Pi} f(\lambda)^{-1} \exp(i\lambda u) d\lambda \quad (2.6)$$

for all $u \in \mathbb{Z}$ (e.g. Shaman 1975, 1976).

Proposition 2.3. *Let X be a multivariate stationary process such that condition (2.1) holds. Then*

$$X_a \perp\!\!\!\perp X_b \mid X_{V \setminus \{a,b\}} \Leftrightarrow R_{ab}^{(i)}(u) = 0 \text{ for all } u \in \mathbb{Z}.$$

Proof. The relation follows directly from Lemma 2.2 and (2.6). \square

The proposition suggests to parametrize graphical interaction models directly by inverse covariances thus making use of the zero constraints imposed by the absence of edges in the graph.

Definition 2.4 (Graphical interaction model). Let X be a multivariate stationary Gaussian process satisfying (2.1). Furthermore, let $R^{(i)}(u)$, $u \in \mathbb{Z}$, be the inverse covariances of X . We say that X belongs to the graphical interaction model of order p associated with graph G if $R^{(i)}(u) = 0$ for all $|u| > p$ and

$$a - b \notin E \Rightarrow R_{ab}^{(i)}(u) = R_{ba}^{(i)}(u) = 0 \text{ for all } u \in \mathbb{Z}.$$

Alternatively, the autoregressive representation (2.2) suggests to model the process X by vector autoregressive models of order p , that is,

$$X(t) = \sum_{u=1}^p A(u) X(t-u) + \varepsilon(t)$$

and $\text{var}(\varepsilon(t)) = \Sigma$. If we further assume that the process is causal, the spectral density matrix $f(\lambda)$ exists and satisfies condition (2.1). Thus the inverse spectral matrix $f^{-1}(\lambda)$ also exists and is given by

$$f^{-1}(\lambda) = 2\pi A(e^{i\lambda})' K A(e^{-i\lambda}), \quad (2.7)$$

where $K = \Sigma^{-1}$ and $A(z) = \mathbb{1}_d - A(1)z - \dots - A(p)z^p$ is the characteristic polynomial of the process (e.g. Dahlhaus 2000). From (2.5) and Lemma 2.2, it follows that X_a and X_b are conditionally independent given $X_{V_{ab}}$ if and only if

$$\sum_{k,l=1}^d \sum_{u,v=0}^p K_{kl} A_{ki}(u) A_{lj}(v) \exp(i\lambda(v-u)) = 0.$$

for all $\lambda \in [-\pi, \pi]$, which yields the following $2p+1$ restrictions on the parameters

$$\sum_{k,l=1}^d \sum_{u=0}^p K_{kl} A_{ki}(u) A_{lj}(u+h) = 0, \quad h = -p, \dots, p,$$

where $A(0) = -\mathbb{1}_d$ and $A(u) = 0$ if $u < 0$ or $u > p$. It is clear from these expressions that graphical modelling with constraints on the autoregressive parameters would be very difficult.

On the other hand, it is well known that for a VAR(p) process the inverse covariances $R^{(i)}(u)$ vanish for all $|u| > p$ (e.g. Bhansali 1980, Battaglia 1984). Because of

the uniqueness of the factorization in (2.7) (cf. Masani 1966), a VAR(p) process is also determined by the set of inverse covariances

$$\theta = (\text{vech}(R^{(i)}(0))', \text{vec}(R^{(i)}(1))', \dots, \text{vec}(R^{(i)}(p))')',$$

where as usual the vec stacks the columns of the matrix and vech stacks only the elements contained in the lower triangular submatrix. Thus a process X belongs to a graphical interaction model of order p associated with graph G if and only if it belongs to a VAR(p) model that satisfies the pairwise independence relations encoded by the graph G . Therefore we also say that X belongs to a *graphical vector autoregressive model* of order p associated with graph G , which we denote by VAR(p, G).

3. MODEL FITTING

A fundamental, information theoretic measure for the separation or distance between two probability distributions is the Kullback-Leibler information (Kullback and Leibler 1951), which gives the mean information per observation for the discrimination between the true and a fitted distribution. Let $X(1), \dots, X(T)$ be observations from a multivariate Gaussian stationary process specified by some infinite parameter θ_0 . Then for density functions p_{θ_0} and p_{θ} and spectral matrices f_{θ_0} and f_{θ} the Kullback-Leibler information between the process and a fitted model specified by the parameter θ is given by

$$\begin{aligned} I(\theta, \theta_0) &= \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\theta_0} \left(\frac{p_{\theta}(X_1, \dots, X_T)}{p_{\theta_0}(X_1, \dots, X_T)} \right) \\ &= \frac{1}{4\pi} \int_{\Pi} \left\{ \log \left(\frac{\det f_{\theta}(\lambda)}{\det f_{\theta_0}(\lambda)} \right) + \text{tr}[f_{\theta_0}(\lambda) f_{\theta}^{-1}(\lambda) - \mathbb{1}_d] \right\} d\lambda \end{aligned}$$

(cf Parzen 1983). Minimization of $I(\theta, \theta_0)$ with respect to θ is equivalent to minimizing

$$\mathcal{L}(\theta) = \frac{1}{4\pi} \int_{\Pi} \left(\log \det f_{\theta}(\lambda) + \text{tr}[f_{\theta_0}(\lambda) f_{\theta}^{-1}(\lambda)] \right) d\lambda.$$

In the following we assume that X is a vector autoregressive process of infinite order to which we will fit graphical vector autoregressions of finite order p . Allowing the order to diverge to infinity for increasing sample size, this implies that asymptotically the process can be fitted by the correct model which is crucial in our investigation of the asymptotic properties of the Kullback-Leibler information.

Assumption 3.1. X is a multivariate stationary Gaussian process defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that the following conditions hold.

- (i) The spectral matrix $f(\lambda)$ of $\{X(t)\}$ exists and satisfies the boundedness condition

$$a_1 \mathbb{1}_d \leq g(\lambda) \leq a_2 \mathbb{1}_d \quad \forall \lambda \in [-\pi, \pi]$$

for constants a_1 and a_2 such that $0 < a_1 \leq a_2 < \infty$.

- (ii) There exists $\beta > 1$ such that the covariances $R(u)$ of $\{X(t)\}$ satisfy

$$\sum_{u \in \mathbb{Z}} |u|^{\beta} \|R(u)\| < \infty.$$

- (iii) $\{X(t)\}$ has conditional independence graph $G_0 = (V, E_0)$.

Under these assumptions, X has an autoregressive representation (2.2). As in previous section, we parametrize graphical vector autoregressive models by the inverse covariances $R_{ij}^{(i)}(u)$. Thus we get infinite dimensional parameter vectors

$$\theta = (\text{vech}(R^{(i)}(0))', \text{vec}(R^{(i)}(1))', \text{vec}(R^{(i)}(2))', \dots)'$$

In the following, we denote the spectral matrices, covariances, and inverse covariances specified by the parameter θ by $f_\theta(\lambda)$, $R_\theta(u)$, and $R_\theta^{(i)}(u)$, respectively.

Assumption 3.2. Θ is a subset of $\ell^2(\mathbb{R})$ such that the following conditions hold.

(i) The spectral matrices f_θ satisfy for all $\theta \in \Theta$ the boundedness condition

$$b_1 \mathbb{1}_d \leq f_\theta(\lambda) \leq b_2 \mathbb{1}_d \quad \forall \lambda \in [-\pi, \pi]$$

for constants b_1 and b_2 such that $0 < b_1 \leq b_2 < \infty$.

(ii) There exists a constant $C > 0$ such that the covariances $R_\theta(u)$ satisfy

$$\sum_{u \in \mathbb{Z}} |u|^\beta \|R_\theta(u)\| < C$$

for all $\theta \in \Theta(p, G)$, where β is the same as in Assumption 3.1.

(iii) There exists θ_0 in Θ such that $f_{\theta_0}(\lambda) = f(\lambda)$ for all $\lambda \in [-\pi, \pi]$ and θ_0 belongs to the interior of Θ .

Next, let \mathcal{G} denote the set of all graphs $G = (V, E)$ such that $V = \{1, \dots, d\}$ and $E \subseteq \{a - b \mid a, b \in V, a \neq b\}$. For $p \in \mathbb{N}$ and $G \in \mathcal{G}$, the VAR(p, G) model is now given by the parameter space

$$\Theta(p, G) = \{\theta \in \Theta \mid R_{ab, \theta}^{(i)}(u) = 0 \text{ if } a - b \notin E \text{ or } |u| > p\}.$$

Let $I_{p, G}$ denote the set of indices for which $\Theta(p, G)$ is not constrained to zero and $\pi_{p, G}$ the projection of $\ell^2(\mathbb{R})$ onto the subspace spanned by $\Theta(p, G)$.

Minimization of the Kullback-Leibler information $I(\theta, \theta_0)$, or equivalently $\mathcal{L}(\theta)$, with respect to $\theta \in \Theta(p, G)$ yields the best VAR(p, G) approximation of $\{X(t)\}$, which we denote by the parameter

$$\theta_0(p, G) = \underset{\theta \in \Theta(p, G)}{\text{argmin}} \mathcal{L}(\theta).$$

We require that $\theta_0(p, G)$ exists and is uniquely defined.

Assumption 3.3. The best approximation $\theta_0(p, G)$ in $\Theta(p, G)$ with respect to the Kullback-Leibler information $I(\theta, \theta_0)$ is unique and belongs to the relative interior of $\Theta(p, G)$ with respect to the seminorm $\|\cdot\|_{\pi_{p, G}}$.

Using matrix calculus (eg Harville 1997), we obtain the following derivatives of $\mathcal{L}(\theta)$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_k} = \frac{1}{4\pi} \int_{\Pi} \text{tr} \left[(f_{\theta_0}(\lambda) - f_\theta(\lambda)) \frac{\partial f_\theta^{-1}(\lambda)}{\partial \theta_k} \right] d\lambda. \quad (3.1)$$

Since the inverse spectral matrix is linear in the parameters we get an explicit formula for its derivatives. Let θ_k correspond to $R_{ab}^{(i)}(u)$. Then

$$\frac{\partial f_{ij, \theta}^{-1}(\lambda)}{\partial \theta_k} = \begin{cases} 2\pi \delta_{ia} \delta_{ja} & \text{if } a = b \text{ and } u = 0 \\ 2\pi [\delta_{ia} \delta_{jb} \exp(-i\lambda u) + \delta_{ib} \delta_{ja} \exp(i\lambda u)] & \text{else} \end{cases}.$$

Substituted into (3.1) we therefore get

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_k} = 0 \Leftrightarrow \int_{\Pi} (f_{\theta_0, ab}(\lambda) - f_{ab, \theta}(\lambda)) \exp(i\lambda u) d\lambda = 0. \quad (3.2)$$

This leads to the following set of equations, which characterize the best VAR(p, G) approximation $\theta_0(p, G)$,

$$\begin{aligned} R_{ij, \theta_0(p, G)}(u) &= R_{ij, \theta_0}(u) & \forall i - j \in E & \quad \forall u \in \{-p, \dots, p\} \\ R_{ij, \theta_0(p, G)}^{(i)}(u) &= 0 & \forall i - j \notin E & \quad \forall u \in \{-p, \dots, p\} \end{aligned} \quad (3.3)$$

and additionally $R_{\theta_0(p, G)}^{(i)}(u) = 0$ for all $|u| > p$.

In the following we will also need the second and third derivatives of $\mathcal{L}(\theta)$. Using results on matrix differentiation (eg Harville 1997) and by linearity of $f_{\theta}^{-1}(\lambda)$ in the parameters we obtain for the second derivatives

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j} &= \frac{1}{4\pi} \int_{\Pi} \text{tr} \left[(f_{\theta_0}(\lambda) - f_{\theta}(\lambda)) \frac{\partial^2 f_{\theta}^{-1}(\lambda)}{\partial \theta_i \partial \theta_j} \right] d\lambda - \int_{\Pi} \text{tr} \left[\frac{\partial f_{\theta}(\lambda)}{\partial \theta_i} \frac{\partial f_{\theta}^{-1}(\lambda)}{\partial \theta_j} \right] d\lambda \\ &= \frac{1}{4\pi} \int_{\Pi} \text{tr} \left[f_{\theta}(\lambda) \frac{\partial f_{\theta}^{-1}(\lambda)}{\partial \theta_i} f_{\theta}(\lambda) \frac{\partial f_{\theta}^{-1}(\lambda)}{\partial \theta_j} \right] d\lambda, \end{aligned}$$

and similarly for the third derivatives

$$\frac{\partial^3 \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} = \frac{1}{2\pi} \int_{\Pi} \text{tr} \left[f_{\theta}(\lambda) \frac{\partial f_{\theta}^{-1}(\lambda)}{\partial \theta_i} f_{\theta}(\lambda) \frac{\partial f_{\theta}^{-1}(\lambda)}{\partial \theta_j} f_{\theta}(\lambda) \frac{\partial f_{\theta}^{-1}(\lambda)}{\partial \theta_k} \right] d\lambda.$$

We will denote the vector of first derivatives by

$$\nabla \mathcal{L}(\theta) = \left(\frac{\partial \mathcal{L}(\theta)}{\partial \theta_i} \right)_{i \in \mathbb{N}},$$

and the matrix of second derivatives by

$$\nabla^2 \mathcal{L}(\theta) = \left(\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j} \right)_{i, j \in \mathbb{N}}.$$

The linearity of $f_{\theta}^{-1}(\lambda)$ in θ also implies that

$$\theta_1' \nabla^2 \mathcal{L}(\theta) \theta_2 = \frac{1}{4\pi} \int_{\Pi} \text{tr} [f_{\theta}(\lambda) f_{\theta_1}^{-1}(\lambda) f_{\theta}(\lambda) f_{\theta_2}^{-1}(\lambda)] d\lambda. \quad (3.4)$$

In practice, model distances such as the Kullback-Leibler information need to be estimated as they depend on the unknown parameter θ_0 . Akaike (1973) pointed out that the Kullback-Leibler information is related to the method of maximum likelihood. Therefore, given observations $X(1), \dots, X(T)$ from the process X , minimum distance estimates can be obtained by maximizing the Gaussian likelihood function or, equivalently, minimizing the $-1/T \log$ likelihood function

$$\mathcal{L}_T^*(\theta) = \frac{1}{2} \log(2\pi) + \frac{1}{2T} \log \det R_{\theta, T} + \frac{1}{2T} X_T' R_{\theta, T}^{-1} X_T, \quad (3.5)$$

where $R_{\theta, T} = (R_{\theta}(u - v))_{u, v=1, \dots, T}$. A more favourable choice for fitting graphical autoregressive models is the likelihood approximation suggested by Whittle (1953,

1954). Approximating the matrix $R_{\theta,T}^{-1}$ by the corresponding matrix of inverse covariances (cf. Shaman 1975, 1976) together with the Szegö identity (cf. Grenander and Szegö 1958) leads to the Whittle likelihood

$$\mathcal{L}_T(\theta) = \frac{1}{4\pi} \int_{\Pi} \left(\log \det f_{\theta}(\lambda) + \text{tr} [I^{(T)}(\lambda) f_{\theta}(\lambda)^{-1}] \right) d\lambda,$$

which estimates $\mathcal{L}(\theta)$ consistently. Thus we get as a minimum distance estimate the Whittle estimate

$$\hat{\theta}_T(p, G) = \underset{\theta \in \Theta(p, G)}{\text{argmin}} \mathcal{L}_T(\theta).$$

The first derivative of the Whittle likelihood is

$$\frac{\partial \mathcal{L}_T(\theta)}{\partial \theta_i} = \frac{1}{4\pi} \int_{\Pi} \text{tr} \left[(I^{(T)}(\lambda) - f_{\theta}(\lambda)) \frac{\partial f_{\theta}^{-1}(\lambda)}{\partial \theta_i} \right] d\lambda. \quad (3.6)$$

Since $f_{\theta}^{-1}(\lambda)$ is linear in θ , the data dependent term vanishes in the second derivative and we find for all $\theta \in \Theta$

$$\nabla^2 \mathcal{L}_T(\theta) = \nabla^2 \mathcal{L}(\theta). \quad (3.7)$$

Consequently, also the third derivatives of $\mathcal{L}_T(\theta)$ and $\mathcal{L}(\theta)$ are equal. Setting the first derivative to zero leads to the following characterization of the Whittle estimates in the VAR(p, G) model.

Theorem 3.4. *Suppose that Assumptions 3.1 and 3.2 hold. Then the Whittle-estimate $\hat{\theta}_T(p, G)$ in the graphical autoregressive model VAR(p, G) is given by the equations*

$$\begin{aligned} R_{ij, \hat{\theta}_T(p, G)}(u) &= \hat{R}_{ij}(u) & \forall i - j \in E \ \forall u \in \{-p, \dots, p\}, \\ R_{ij, \hat{\theta}_T(p, G)}^{(i)}(u) &= 0 & \forall i - j \notin E \ \forall u \in \{-p, \dots, p\}, \end{aligned}$$

and $R_{\hat{\theta}_T(p, G)}^{(i)}(u) = 0$ for all $|u| > p$, where $\hat{R}_{ij}(u)$ is defined as

$$\hat{R}_{ij}(u) = \int_{\Pi} I_{ij}^{(T)}(\lambda) \exp(i\lambda u) d\lambda.$$

Proof. The result follows from the arguments leading to (3.3) applied to the first derivative in (3.6). \square

These equations are similar to the equations for the maximum likelihood estimates in ordinary Gaussian graphical models (cf Lauritzen 1996). More precisely, these are the restrictions for a Gaussian graphical model in which the set of vertices consists of the entire process X . This, however, is not surprising by the way the Whittle likelihood approximates the likelihood function in (3.5), as the Whittle likelihood mainly neglects edge effects due to observing only a finite horizon by substituting asymptotic approximations for the finite sample quantities $\det R_{\theta, T}$ and $R_{\theta, T}^{-1}$.

The asymptotic properties of the Whittle estimate in general are well known (eg Dzhaparidze and Yaglom 1983). For example, we have the following central limit theorem.

Theorem 3.5. *Under Assumptions 3.1 to 3.3 we have*

$$\sqrt{T}(\hat{\theta}_T(p, G) - \theta_0(p, G)) \xrightarrow{D} \mathcal{N}(0, c_h \Gamma(p, G)^{-1} \Gamma_0(p, G) \Gamma(p, G)^{-1})$$

where $c_h = H_4/H_2^2$, $\Gamma_0(p, G) = \pi_{p,G} \nabla^2 \mathcal{L}(\theta_0) \pi_{p,G}$ and $\Gamma(p, G) = \pi_{p,G} \nabla^2 \mathcal{L}(\theta_0(p, G)) \pi_{p,G}$ with $\Gamma(p, G)^{-1} = \pi_{p,G} \Gamma(p, G)^{-} \pi_{p,G}$ for any generalized inverse $\Gamma(p, G)^{-}$.

Proof. see Dzhaparidze and Yaglom (eg 1983, Section 5.6). \square

From the Whittle estimate $\hat{\theta}_T(p, G)$ we can finally compute estimates for the parameters A_1, \dots, A_p and Σ in (??).

- (a) From the estimates $R_{\hat{\theta}_T(p, G)}^{(i)}(u)$ for the inverse covariances we can obtain the covariances $R_{\hat{\theta}_T(p, G)}(u)$ via computation of $f_{\hat{\theta}_T(p, G)}^{-1}$ and $f_{\hat{\theta}_T(p, G)}$. Then estimates for the matrices A_1, \dots, A_p and Σ can be determined by solving the Yule-Walker equations

$$\sum_{u=0}^p A_u R_{\hat{\theta}_T(p, G)}(u - v) = \delta_{v0} \Sigma, \quad v = 0, \dots, p,$$

where $A_0 = -\mathbb{1}_d$.

- (b) The parameters A_1, \dots, A_p and Σ are related to the inverse covariances by the equation system

$$R_{\hat{\theta}_T(p, G)}^{(i)}(v) = \sum_{u=0}^{p-v} A_u \Sigma A_{u+v},$$

where again $A(0) = -\mathbb{1}_d$. This problem is equivalent to the estimation of moving average parameters from the covariances of a process. An iterative algorithm for solving such an equation system has been suggested e.g. by Tunnicliffe Wilson (1972).

We are now interested in the VAR(p, G) model, where $G \in \mathcal{G}$ and p is selected from a given range $1 \leq p \leq P_T$ with $P_T \leq T$, which minimizes the Kullback-Leibler information $I(\hat{\theta}_T(p, G), \theta_0)$ between the fitted model and the true process. A model selection (\hat{p}_T, \hat{G}_T) which has this optimality property at least asymptotically is called asymptotically efficient.

Definition 3.6 (Asymptotically efficient model selection). A selection of models $(\hat{p}_T, \hat{G}_T)_{T \in \mathbb{N}}$ with $1 \leq p \leq P_T$ and $G \in \mathcal{G}$ is called asymptotically efficient if

$$\frac{I(\hat{\theta}_T(\hat{p}_T, \hat{G}_T), \theta_0)}{\min_{1 \leq p \leq P_T} \min_{G \in \mathcal{G}} I(\hat{\theta}_T(p, G), \theta_0)} \xrightarrow{P} 1.$$

The derivation of model selection criteria with this optimality property is based upon an approximation of the Kullback-Leibler information by some deterministic function. For this, it is necessary that asymptotically the process $\{X(t)\}$ can be fitted by the correct model, which implies that P_T must diverge to infinity. On the other hand the approximation holds only if the stochastic variation in $I(\hat{\theta}_T(p, G), \theta_0)$ due to the estimate $\hat{\theta}_T(p, G)$ vanishes asymptotically for all $p \leq P_T$. More precisely, we require that the following conditions hold.

Assumption 3.7. $\{P_T\}_{T \in \mathbb{N}}$ is an integer-valued sequence such that $P_T \rightarrow \infty$ and $P_T^{\frac{5}{2}} \log(T)^2 / T \rightarrow 0$ as $T \rightarrow \infty$.

4. ASYMPTOTICAL EFFICIENCY OF A MODEL SELECTION

In this section, we investigate the asymptotic behaviour of the Kullback-Leibler information $I(\hat{\theta}_T(p, G), \theta_0)$ between a fitted and the true model. In particular, we derive an asymptotic lower bound for $I(\hat{\theta}_T(p, G), \theta_0)$, which is important for establishing the asymptotic efficiency of model selection criteria.

Taniguchi (1980) discussed the asymptotics of the final prediction error for fitting spectral models to univariate time series. Although the line of proof is similar, the conditions of Taniguchi do not hold for the parametrization by inverse covariances (also in the univariate case). In the next two lemmas we show that weaker statements about the second and third derivatives and the convergence of $|f_{\theta_0(p, G)}(\lambda) - f_{\theta_0}(\lambda)|$ can be derived under the Assumptions 3.1 to 3.3.

Lemma 4.1. *Suppose that Assumptions 3.1 and 3.2 hold. Then*

(i) *there exist constants c_1 and c_2 such that*

$$\|\nabla^2 \mathcal{L}(\theta)\| \leq c_1 < \infty \quad \text{and} \quad \|\nabla^2 \mathcal{L}(\theta)\|_{\inf} \geq c_2 > 0 \quad (4.1)$$

uniformly in $\theta \in \Theta$;

(ii) *for all $\eta \in \ell^2(\mathbb{R})$ and all $\zeta \in \pi_{p, G}(\mathbb{R}^\infty)$*

$$\left| \sum_{i, j, k=1}^{\infty} \frac{\partial^3 \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \eta_i \eta_j \zeta_k \right| \leq C \sqrt{k(p, G)} \|\eta\|^2 \|\zeta\|$$

uniformly in $\theta \in \Theta$, $1 \leq p \leq P_T$, and $G \in \mathcal{G}$. For $\zeta \in \ell^1(\mathbb{R})$ the term is bounded by $C \|\eta\|^2 \|\zeta\|_1$.

Lemma 4.2. *Suppose that Assumptions 3.1 to 3.3 hold. Then for all graphs G such that $G_0 \subseteq G$ we have*

$$\int_{\Pi} \|f_{\theta_0(p, G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)\|_2^2 d\lambda = O(p^{-(2\beta+1)}) \quad (4.2)$$

and

$$\int_{\Pi} \|f_{\theta_0(p, G)}(\lambda) - f_{\theta_0}(\lambda)\|_2^2 d\lambda = O(p^{-(2\beta+1)}), \quad (4.3)$$

where $\|A\|_2 = (\text{tr}(A^*A))^{1/2}$.

The methods in this section are based on Taylor expansions of the Kullback-Leibler information. For this we need the following lemma, which states the uniform consistency of the Whittle estimates.

Lemma 4.3. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then we have*

$$\max_{1 \leq p \leq P_T} \|\hat{\theta}_T(p, G) - \theta_0(p, G)\| = o_P(1).$$

The Kullback-Leibler information can now be studied by the help of Taylor expansions. We first note that $I(\hat{\theta}_T(p, G), \theta_0)$ can be written as

$$I(\hat{\theta}_T(p, G), \theta_0) = [\mathcal{L}(\hat{\theta}_T(p, G)) - \mathcal{L}(\theta_0(p, G))] + I(\theta_0(p, G), \theta_0)$$

and with a Taylor expansion of $\mathcal{L}(\hat{\theta}_T(p, G))$ about $\theta_0(p, G)$

$$\begin{aligned} &= \frac{1}{2} \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|_{\nabla^2 \mathcal{L}(\theta_0(p, G))}^2 + I(\theta_0(p, G), \theta_0) \\ &\quad + O_P\left(p^{\frac{1}{2}} \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|^3\right), \end{aligned} \quad (4.4)$$

where the first order term vanishes because of $\pi_{p, G} \nabla \mathcal{L}(\theta_0(p, G)) = 0$. In this decomposition of $I(\hat{\theta}_T(p, G), \theta_0)$ the first term represents the variance due to estimating $\theta_0(p, G)$ by $\hat{\theta}_T(p, G)$ while the second term can be approximated by $\frac{1}{2} \left\| \theta_0(p, G) - \theta_0 \right\|_{\nabla^2 \mathcal{L}(\theta_0)}^2$ and thus represents the bias due to fitting an incorrect model.

In the following we investigate the asymptotic behaviour of the variance term. Taylor expansion for the first derivative of $\mathcal{L}_T(\hat{\theta}_T(p, G))$ yields

$$\nabla \mathcal{L}_T(\hat{\theta}_T(p, G)) - \nabla \mathcal{L}_T(\theta_0(p, G)) = \nabla^2 \mathcal{L}_T(\theta_0(p, G))(\hat{\theta}_T(p, G) - \theta_0(p, G)) + Z(p, G).$$

where

$$Z_i(p, G) = \sum_{j, k \in I(p, G)} \frac{\partial^3 \mathcal{L}_T(\bar{\theta}_T)}{\partial \theta_i \partial \theta_j \partial \theta_k} (\hat{\theta}_T(p, G)_j - \theta_0(p, G)_j) (\hat{\theta}_T(p, G)_k - \theta_0(p, G)_k)$$

and $\bar{\theta}_T = \theta_0(p, G) + \xi(\hat{\theta}_T(p, G) - \theta_0(p, G))$ for some $\xi \in [0, 1]$. By the definition of $\hat{\theta}_T(p, G)$ and $\theta_0(p, G)$ we have $\pi_{p, G}(\nabla \mathcal{L}_T(\hat{\theta}_T(p, G))) = \pi_{p, G}(\nabla \mathcal{L}(\theta_0(p, G))) = 0$. Noting that by (3.7) the second derivative of $\mathcal{L}_T(\theta)$ can be replaced by that of $\mathcal{L}(\theta)$, we get

$$\pi_{p, G} \left(\nabla \mathcal{L}_T(\hat{\theta}_T(p, G)) - \nabla \mathcal{L}_T(\theta_0(p, G)) \right) = \Gamma(p, G)(\hat{\theta}_T(p, G) - \theta_0(p, G)) + \pi_{p, G}(Z(p, G)),$$

where $\Gamma(p, G) = \pi_{p, G} \nabla^2 \mathcal{L}(\theta_0(p, G)) \pi_{p, G}$ as in Theorem 3.5. This leads to the equation

$$\begin{aligned} &\left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \\ &= \left[\nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right]' \left[(\hat{\theta}_T(p, G) - \theta_0(p, G)) + \Gamma(p, G)^{-1} Z(p, G) \right] \\ &= \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|_{\Gamma(p, G)}^2 + 2Z(p, G)'(\hat{\theta}_T(p, G) - \theta_0(p, G)) + \left\| Z(p, G) \right\|_{\Gamma(p, G)^{-1}}^2. \end{aligned}$$

Noting that in the first term $\Gamma(p, G)$ can be replaced by $\nabla^2 \mathcal{L}(\theta_0(p, G))$, we finally have by Lemma 4.1

$$\begin{aligned} &\left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|_{\nabla^2 \mathcal{L}(\theta_0(p, G))}^2 \\ &= \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 + O_P\left(p^{\frac{1}{2}} \|\tilde{\theta}\|^3 + p \|\tilde{\theta}\|^4\right) \end{aligned} \quad (4.5)$$

with the abbreviation $\tilde{\theta} = \hat{\theta}_T(p, G) - \theta_0(p, G)$. In the next lemmas we prove that the moments of the first term on the right side can be approximated by the moments of a χ^2 -distribution up to the fourth order.

Lemma 4.4. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then*

$$\begin{aligned} &\mathbb{E} \left[T \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \right] \\ &= c_h \text{tr} \left[\Gamma(p, G)^{-1} \Gamma_0(p, G) \right] + O\left(\frac{p^{\frac{5}{2}} \log(T)}{T}\right) \end{aligned}$$

If the graph G contains the true graph G_0 , we further have

$$\text{tr}[\Gamma(p, G)^{-1}\Gamma_0(p, G)] = k(p, G) + O(p^{1-\beta}).$$

Lemma 4.5. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then if $G_0 \subseteq G$*

$$\begin{aligned} \mathbb{E} \left[\frac{T}{c_h} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 - k(p, G) \right]^4 \\ = 48k(p, G) + 12k(p, G)^2 + O(p^{4-\beta}) + O\left(\frac{p^{\frac{11}{2}} \log(T)^2}{T}\right), \end{aligned}$$

otherwise if $G_0 \not\subseteq G$

$$\mathbb{E} \left[\frac{T}{c_h} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \right]^2 = O(p^2).$$

It follows now from (4.5) and Lemma 4.1 that

$$\left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|^2 \leq C \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|_{\nabla^2 \mathcal{L}(\theta_0(p, G))}^2 = O_P\left(\frac{k(p, G)}{T}\right).$$

This implies that equation (4.4) for the Kullback-Leibler information can be rewritten

$$I(\hat{\theta}_T(p, G), \theta_0) = I(\theta_0(p, G), \theta_0) + \frac{1}{2} \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|_{\nabla^2 \mathcal{L}(\theta_0(p, G))}^2 + o_P\left(\frac{k(p, G)}{T}\right).$$

Further Lemma 4.5 suggests that $I(\hat{\theta}_T(p, G), \theta_0)$ can be approximated by the following function

$$L_T(p, G) = \frac{k(p, G)c_h}{2T} + I(\theta_0(p, G), \theta_0).$$

The next results show that this approximation holds uniformly for all $1 \leq p \leq P_T$, which allows us to reformulate asymptotic efficiency in terms of the sequence (p_T^*, G_T^*) which minimizes $L_T(p, G)$.

Definition 4.6. (p_T^*, G_T^*) is the sequence of models which attains the minimum of $L_T(p, G)$,

$$(p_T^*, G_T^*) = \underset{1 \leq p \leq P_T, G \in \mathcal{G}}{\text{argmin}} L_T(p, G)$$

for all $T \in \mathbb{N}$.

Under Assumption 3.7 $L_T(P_T, G_0)$ and therefore also $L_T(p_T^*, G_T^*)$ converge to zero as $T \rightarrow \infty$. But the second term of $L_T(p, G)$ does not vanish if the approximating model is wrongly specified, which is the case for finite p or if G does not contain the true graph G_0 . Thus it follows that p_T^* diverges to infinity as $T \rightarrow \infty$ and $G_0 \subseteq G_T^*$ for almost all $T \in \mathbb{N}$.

Lemma 4.7. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then*

$$\max_{1 \leq p \leq P_T} \max_{G \in \mathcal{G}} \left| \frac{T \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 - k(p, G)c_h}{TL_T(p, G)} \right| = o_P(1).$$

Theorem 4.8. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then*

$$\max_{1 \leq p \leq P_T} \max_{G \in \mathcal{G}} \left| \frac{I(\hat{\theta}_T(p, G), \theta_0)}{L_T(p, G)} - 1 \right| = o_P(1).$$

Proof of Theorem 4.8. Let G be fixed. We first note that equation (4.5) together with Lemma 4.7 implies that

$$\max_{1 \leq p \leq P_T} \|\hat{\theta}_T(p, G) - \theta_0(p, G)\|^2 = O_P\left(\frac{P_T}{T}\right).$$

Since further $p/TL_T(p, G) \leq C$ uniformly in $p \in \mathbb{N}$, it follows from (4.4) and (4.5)

$$\begin{aligned} & \max_{1 \leq p \leq P_T} \left| \frac{I(\hat{\theta}_T(p, G), \theta_0) - L_T(p, G)}{L_T(p, G)} \right| \\ &= \max_{1 \leq p \leq P_T} \left| \frac{T \|\hat{\theta}_T(p, G) - \theta_0(p, G)\|_{\nabla^2 \mathcal{L}(\theta_0(p, G))}^2 - k(p, G)c_h}{2TL_T(p, G)} \right| + O_P\left(\frac{P_T}{\sqrt{T}}\right) \\ &\leq \max_{1 \leq p \leq P_T} \left| \frac{T \|\nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G))\|_{\Gamma(p, G)^{-1}} - k(p, G)c_h}{2TL_T(p, G)} \right| + o_P(1), \end{aligned}$$

from which the assertion follows by Lemma 4.7. \square

The uniform approximation of the Kullback-Leibler information by $L_T(p, G)$ leads now to an asymptotic lower bound for $I(\hat{\theta}_T, \theta_0)$ and a new characterization of the asymptotic efficiency.

Theorem 4.9. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. If $(\hat{p}_T, \hat{G}_T)_{T \in \mathbb{N}}$ is a random sequence such that $1 \leq \hat{p}_T \leq P_T$ and $\hat{G}_T \in \mathcal{G}$, then we have for all $\varepsilon > 0$*

$$\lim_{T \rightarrow \infty} \mathbb{P}\left(\frac{I(\hat{\theta}_T(\hat{p}_T, \hat{G}_T), \theta_0)}{L_T(\hat{p}_T^*, \hat{G}_T^*)} \geq 1 - \varepsilon\right) = 1.$$

Proof. The result is a direct consequence of the inequality

$$\frac{I(\hat{\theta}_T(\hat{p}_T, \hat{G}_T), \theta_0)}{L_T(\hat{p}_T^*, \hat{G}_T^*)} \geq \frac{I(\hat{\theta}_T(\hat{p}_T, \hat{G}_T), \theta_0)}{L_T(\hat{p}_T, \hat{G}_T)}$$

and Theorem 4.8. \square

Corollary 4.10. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then a random sequence $(\hat{p}_T, \hat{G}_T)_{T \in \mathbb{N}}$ such that $1 \leq \hat{p}_T \leq P_T$ and $\hat{G}_T \in \mathcal{G}$ is asymptotically efficient if and only if it satisfies*

$$\frac{I(\hat{\theta}_T(\hat{p}_T, \hat{G}_T), \theta_0)}{L_T(\hat{p}_T^*, \hat{G}_T^*)} \xrightarrow{P} 1.$$

Proof. Let $(p_T^\circ, G_T^\circ)_{T \in \mathbb{N}}$ be the (random) sequence which minimizes $I(\hat{\theta}_T(p, G), \theta_0)$. Since accordingly $I(\hat{\theta}_T(p_T^\circ, G_T^\circ), \theta_0) \leq I(\hat{\theta}_T(\hat{p}_T^*, \hat{G}_T^*), \theta_0)$ we get

$$\mathbb{P}\left(\frac{I(\hat{\theta}_T(p_T^\circ, G_T^\circ), \theta_0)}{L_T(\hat{p}_T^*, \hat{G}_T^*)} \geq 1 - \varepsilon\right) \leq \mathbb{P}\left(\frac{I(\hat{\theta}_T(\hat{p}_T^*, \hat{G}_T^*), \theta_0)}{L_T(\hat{p}_T^*, \hat{G}_T^*)} \geq 1 - \varepsilon\right),$$

which converges to zero as $T \rightarrow \infty$. Therefore

$$\frac{I(\hat{\theta}_T(\hat{p}_T, \hat{G}_T), \theta_0)}{I(\hat{\theta}_T(p_T^\circ, G_T^\circ), \theta_0)} \xrightarrow{P} 1 \quad \Leftrightarrow \quad \frac{I(\hat{\theta}_T(\hat{p}_T, \hat{G}_T), \theta_0)}{L_T(\hat{p}_T^*, \hat{G}_T^*)} \xrightarrow{P} 1,$$

from which the assertion follows. \square

5. ASYMPTOTICALLY EFFICIENT MODEL SELECTION

Model distances in general depend naturally on the unknown distribution of the observations and therefore cannot be minimized for model selection. Empirical model distances such as the log likelihood function can be used for the estimation of parameters within each model, but typically do not provide good overall estimates of the theoretical model distance and therefore need to be corrected (eg Shibata 1997).

Akaike (1973) considered model selection by minimizing the expected Kullback-Leibler information, which leads to a simple bias correction term for the log likelihood function. In our situation, we have the following Taylor approximation for $\mathcal{L}(\theta)$ and $\mathcal{L}_T(\theta)$

$$\begin{aligned}\mathcal{L}(\hat{\theta}_T(p, G)) &\approx \mathcal{L}(\theta_0(p, G)) + \frac{1}{2} \|\hat{\theta}_T(p, G) - \theta_0(p, G)\|_{\Gamma(p, G)}^2, \\ \mathcal{L}_T(\hat{\theta}_T(p, G)) &\approx \mathcal{L}_T(\theta_0(p, G)) - \frac{1}{2} \|\hat{\theta}_T(p, G) - \theta_0(p, G)\|_{\Gamma(p, G)}^2.\end{aligned}$$

Taking expectations this leads to the following version of Akaike's AIC criterion

$$C_T(p, G) = \mathcal{L}_T(\hat{\theta}_T(p, G)) + c_h \frac{k(p, G)}{T}.$$

In the following we show that the minimizing sequence

$$(\hat{p}_T, \hat{G}_T) = \underset{1 \leq p \leq P_T, G \in \mathcal{G}}{\operatorname{argmin}} C_T(p, G)$$

is asymptotically efficient with respect to the Kullback-Leibler information. For this, we first rewrite the criterion as

$$\begin{aligned}C_T(p, G) &= L_T(p, G) + [\mathcal{L}_T(\theta_0(p, G)) - \mathcal{L}(\theta_0(p, G))] \\ &\quad + \left\{ \frac{k(p, G)c_h}{2T} + [\mathcal{L}_T(\hat{\theta}_T(p, G)) - \mathcal{L}_T(\theta_0(p, G))] \right\} + \mathcal{L}(\theta_0).\end{aligned}$$

$C_T(p, G)$ estimates the Kullback-Leibler information well if the second to fourth term on the right side are negligible compared with the first term $L_T(p, G)$.

Lemma 5.1. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then*

$$\max_{1 \leq p \leq P_T} \max_{G \in \mathcal{G}} \left| \frac{\frac{k(p, G)c_h}{2T} + [\mathcal{L}_T(\hat{\theta}_T(p, G)) - \mathcal{L}_T(\theta_0(p, G))]}{L_T(p, G)} \right| = o_P(1).$$

The lemma shows that the third term is uniformly negligible compared with $L_T(p, G)$. In contrast, the fourth term is constant and the second term is of order $O_P(\sqrt{p/T})$, which follows from the proof of Lemma 4.2. However, the behaviour of the minimum (\hat{p}_T, \hat{G}_T) depends only on the differences

$$C_T(p, G) - C_T(p_T^*, G_T^*).$$

Obviously, here the fourth term cancels and it is therefore sufficient to show that

$$[\mathcal{L}_T(\theta_0(p, G)) - \mathcal{L}(\theta_0(p, G))] - [\mathcal{L}_T(\theta_0(p_T^*, G_T^*)) - \mathcal{L}(\theta_0(p_T^*, G_T^*))]$$

is uniformly negligible compared with $L_T(p, G)$.

Lemma 5.2. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then the difference*

$$\max_{1 \leq p \leq P_T} \max_{G \in \mathcal{G}} \frac{[\mathcal{L}_T(\theta_0(p, G)) - \mathcal{L}(\theta_0(p, G))] - [\mathcal{L}_T(\theta_0(p_T^*, G_T^*)) - \mathcal{L}(\theta_0(p_T^*, G_T^*))]}{L_T(p, G)}$$

tends to zero in probability.

Theorem 5.3. *Suppose that Assumptions 3.1 to 3.3 and 3.7 hold. Then the sequence of $(\hat{p}_T, \hat{G}_T)_{T \in \mathbb{N}}$ is asymptotically efficient, that is*

$$\frac{\mathcal{L}(\hat{p}_T, \hat{G}_T)}{L_T(p_T^*, G_T^*)} \xrightarrow{P} 1.$$

Proof. In Lemmas 5.1 and 5.2 we have shown that

$$\frac{[C_T(p, G) - C_T(p_T^*, G_T^*)]}{L_T(p, G)} - \frac{[L_T(p, G) - L_T(p_T^*, G_T^*)]}{L_T(p, G)}$$

converges to zero in probability uniformly in $1 \leq p \leq P_T$. It then follows from the inequalities $C_T(\hat{p}_T, \hat{G}_T) \leq C_T(p_T^*, G_T^*)$ and $L_T(\hat{p}_T, \hat{G}_T) \geq L_T(p_T^*, G_T^*)$ that for all $\varepsilon > 0$

$$\lim_{T \rightarrow \infty} \mathbb{P} \left(\frac{L_T(p_T^*, G_T^*)}{L_T(\hat{p}_T, \hat{G}_T)} \geq 1 - \varepsilon \right) = 1.$$

The result now follows from Theorem 4.8 and Corollary 4.10. \square

APPENDIX A. L-FUNCTIONS AND DATA TAPERS

The use of data tapers often improves the small sample properties of spectral estimates (eg Dahlhaus 1990). For the discussion of asymptotic properties of frequency domain statistics based from tapered data, we need the following function, which has been introduced by Dahlhaus (1983). Let $L^{(T)} : \mathbb{R} \rightarrow \mathbb{R}$ be the periodic extension (with period 2π) of

$$L^{(T)}(\lambda) = \begin{cases} T, & |\lambda| \leq 1/T \\ \frac{1}{|\lambda|}, & 1/T < |\lambda| \leq \pi \end{cases}. \quad (\text{A.1})$$

The properties of these functions are summarized by the following lemma. The proofs are straightforward and can be found in Dahlhaus (1983, 1990).

Lemma A.1. *Let $L^{(T)}(\lambda)$ be defined as in (A.1), $\alpha, \beta, \gamma \in \mathbb{R}$ and $r, s \in \mathbb{N}$. We obtain with a constant C independent of T and S*

- (i) $L^{(T)}(\alpha)$ is monotonically increasing in $T \in \mathbb{R}^+$ and decreasing in $\alpha \in [0, \pi]$.
- (ii) $L^{(T)}(c\alpha) \leq c^{-1}L^{(T)}(\alpha)$ for all $c \in (0, 1]$.
- (iii) $\int_{\Pi} L^{(T)}(\alpha) d\alpha \leq C \log(T)$.
- (iv) $\int_{\Pi} L^{(T)}(\beta + \alpha) L^{(S)}(\gamma - \alpha) d\alpha \leq C \max\{\log(T), \log(S)\} L^{(\min\{T, S\})}(\beta + \gamma)$.
- (v) $\int_{\Pi} L^{(T)}(\alpha)^r d\alpha \leq CT^{r-1}$.
- (vi) $\int_{\Pi} L^{(T)}(\beta + \alpha)^r L^{(S)}(\gamma - \alpha)^r d\alpha \leq C \max\{T^{r-1}, S^{r-1}\} L^{(\min\{T, S\})}(\beta + \gamma)^r$.

Let $h^{(T)}$ be a data taper and let $H_k^{(T)}(\lambda)$ be its Fourier transform. Under the assumptions on the taper function, $H_k^{(T)}(\lambda)$ can be bounded by

$$|H_k^{(T)}(\lambda)| \leq C L^{(T)}(\lambda) \quad (\text{A.2})$$

with a constant $C \in \mathbb{R}$ independent of T and λ . Furthermore, let $\{\Phi_2^{(T)}\}_{T \in \mathbb{N}}$ be the sequence of function given by

$$\Phi_2^{(T)}(\lambda) = \frac{|H_2^{(T)}(\lambda)|^2}{2\pi H_4^{(T)}(0)}. \quad (\text{A.3})$$

By Lemma A.1 and the upper bound in (A.2), it can be shown that $\{\Phi_2^{(T)}\}_{T \in \mathbb{N}}$ is a Dirac sequence (eg Dahlhaus 1983).

By Theorem 4.3.2 of Brillinger (1981) we have

$$\begin{aligned} & \text{cum}\{d_{a_1}^{(T)}(\alpha_1), \dots, d_{a_k}^{(T)}(\alpha_k)\} \\ &= (2\pi)^{k-1} H^{(T)}(\alpha_1 + \dots + \alpha_k) f_{a_1 \dots a_k}(\alpha_1, \dots, \alpha_{k-1}) + O(1) \end{aligned} \quad (\text{A.4})$$

uniformly in $\alpha_1, \dots, \alpha_k \in \Pi$.

APPENDIX B. PROOFS AND AUXILIARY RESULTS

Proof of Lemma 4.1. (i) For $\eta \in \ell^2(\mathbb{R})$ we define matrix valued functions h_η on $[-\pi, \pi]$ by

$$h_\eta(\lambda) = \sum_{u \in \mathbb{Z}} R_\eta^{(i)}(u) \exp(-i\lambda u). \quad (\text{B.1})$$

These matrices are hermitian, but for general $\eta \in \ell^2(\mathbb{R})$ not necessarily positive definite. By (3.4) and Assumption 3.2 (i)

$$\begin{aligned} \eta' \nabla^2 \mathcal{L}(\theta) \eta &= \int_{\Pi} \text{tr} [f_\theta(\lambda) h_\eta(\lambda) f_\theta(\lambda) h_\eta(\lambda)] d\lambda \\ &\leq \int_{\Pi} \|f_\theta(\lambda)\|^2 \text{tr} [h_\eta(\lambda) h_\eta(\lambda)^*] d\lambda \\ &\leq \sup_{\lambda \in [-\pi, \pi]} \|f_\theta(\lambda)\|^2 \int_{\Pi} \|h_\eta(\lambda)\|_2^2 d\lambda \\ &= \sup_{\lambda \in [-\pi, \pi]} \|f_\theta(\lambda)\|^2 \sum_{u \in \mathbb{Z}} \|R_\eta^{(i)}(u)\|_2^2 \leq C \|\eta\|^2. \end{aligned}$$

This proves the first inequality in (4.1). The second inequality is proven similarly with $\|f_\theta(\lambda)\|_{\text{inf}}$ substituted for $\|f_\theta(\lambda)\|$.

(ii) Suppose that $\eta, \zeta \in \ell^2(\mathbb{R})$ and let $h_\eta(\lambda)$ and $h_\zeta(\lambda)$ be defined as in (B.1). Since $f_\theta^{-1}(\lambda)$ is linear in θ , we obtain

$$\begin{aligned} \left| \sum_{i,j,k \in \mathbb{N}} \frac{\partial^3 \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \eta_i \eta_j \zeta_k \right| &\leq \int_{\Pi} \left| \text{tr} [f_\theta(\lambda) h_\eta(\lambda) f_\theta(\lambda) h_\eta(\lambda) f_\theta(\lambda) h_\zeta(\lambda)] \right| d\lambda \\ &\leq \int_{\Pi} \|f_\theta(\lambda)\|^3 \|h_\zeta(\lambda)\| \text{tr} [h_\eta(\lambda) h_\eta(\lambda)^*] d\lambda \\ &\leq b_2^3 \sup_{\lambda \in [-\pi, \pi]} \|h_\zeta(\lambda)\| \int_{\Pi} \|h_\eta(\lambda)\|_2^2 d\lambda. \end{aligned}$$

Noting that for $\zeta \in \pi_{p,G}(\mathbb{R}^\infty) \subseteq \ell^2(\mathbb{R})$

$$\|h_\zeta(\lambda)\| \leq \|h_\zeta(\lambda)\|_1 \leq \sum_{u \in \mathbb{Z}} \|R_\zeta^{(i)}(u)\|_1 \leq C\|\zeta\|_1 \leq C\sqrt{k(p,G)}\|\zeta\|$$

uniformly for all $\lambda \in [-\pi, \pi]$ and $\int_{\Pi} \|h_\eta(\lambda)\|_2^2 d\lambda \leq 2\|\eta\|_2^2$, the second part of the lemma follows. \square

Proof of Lemma 4.3. If $G_0 \subseteq G$, then $\theta_0(p, G)$ converges to θ_0 as $p \rightarrow \infty$. Therefore we have the following Taylor expansion for $\mathcal{L}(\theta)$

$$\mathcal{L}(\theta_0(p, G)) - \mathcal{L}(\theta_0) = \nabla \mathcal{L}(\theta_0)'(\theta_0(p, G) - \theta_0) + \frac{1}{2} \|\theta_0(p, G) - \theta_0\|_{\nabla^2 \mathcal{L}(\bar{\theta})}^2.$$

where $\bar{\theta} = \theta_0 + \xi(\theta_0(p, G) - \theta_0)$ for some $\xi \in [0, 1]$. The first term is zero since θ_0 minimizes $\mathcal{L}(\theta)$. For the second term we obtain by (3.4) and Lemma 4.1 the lower bound

$$\begin{aligned} & \frac{1}{2} \int_{\Pi} \text{tr} \left[f_{\bar{\theta}}(\lambda) (f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) f_{\bar{\theta}}(\lambda) (f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) \right] d\lambda \\ & \geq \frac{1}{2} \int_{\Pi} \|f_{\bar{\theta}}(\lambda)\|_{\text{inf}} \text{tr} \left[(f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) f_{\bar{\theta}}(\lambda) (f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) \right] d\lambda \\ & \geq \frac{1}{2} \int_{\Pi} \|f_{\bar{\theta}}(\lambda)\|_{\text{inf}}^2 \text{tr} \left[(f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) (f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) \right] d\lambda \\ & \geq \frac{b_1^2}{2} \int_{\Pi} \|f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)\|_2^2 d\lambda. \end{aligned} \quad (\text{B.2})$$

Similarly we get the upper bound

$$\mathcal{L}(\theta_0(p, G)) - \mathcal{L}(\theta_0) \leq \frac{b_2^2}{2} \int_{\Pi} \|f_{\theta_0(p,G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)\|_2^2 d\lambda. \quad (\text{B.3})$$

Now let G_S be the saturated graph with all edges included. Then by the equations in (3.3) the Fourier coefficients $R_{\theta_0(p,G_S)}(u)$ and $R_{\theta_0}(u)$ are equal for all $|u| \leq p$. Thus by Assumptions 3.1 and 3.2

$$\int_{\Pi} \|f_{\theta_0(p,G_S)}(\lambda) - f_{\theta_0}(\lambda)\|_2^2 d\lambda = \sum_{|u| > p} \|R_{\theta_0(p,G_S)}(u) - R_{\theta_0}(u)\|_2^2 = O(p^{-(2\beta+1)}). \quad (\text{B.4})$$

For the inverse spectral matrices the same holds since

$$\begin{aligned} \|f_{\theta_0(p,G_S)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)\|_2 &= \|f_{\theta_0(p,G_S)}^{-1}(\lambda) (f_{\theta_0}(\lambda) - f_{\theta_0(p,G_S)}(\lambda)) f_{\theta_0}^{-1}(\lambda)\|_2 \\ &\leq \|f_{\theta_0(p,G_S)}^{-1}(\lambda)\| \|f_{\theta_0}^{-1}(\lambda)\| \|f_{\theta_0(p,G_S)}(\lambda) - f_{\theta_0}(\lambda)\|_2, \end{aligned} \quad (\text{B.5})$$

where $\|f_{\theta_0(p,G_S)}(\lambda)^{-1}(\lambda)\|$ and $\|f_{\theta_0}^{-1}(\lambda)\|$ are bounded uniformly in λ by Assumptions 3.1 and 3.2. Now consider the inverse spectral matrix $f_{*}^{-1}(\lambda)$ with components

$$f_{ij,\theta^*}^{-1}(\lambda) = \begin{cases} f_{ij,\theta_0(p,G_S)}^{-1}(\lambda) & \text{if } (i, j) \in E \\ 0 & \text{if } (i, j) \notin E \end{cases}$$

where $\theta^* = \pi_{p,G}(\theta_0(p, G_S))$. Since $\theta_0(p, G_S) \rightarrow \theta_0$ also implies $\theta^* \rightarrow \theta_0$ and θ_0 belongs to the interior of Θ , there exists $p_0 \in \mathbb{N}$ such that $\theta^* \in \Theta(p, G) \subseteq \Theta$ for all $p \geq p_0$.

Since $G_0 \subseteq G$ it trivially holds that $|f_{ij,\theta^*}^{-1}(\lambda) - f_{ij,\theta_0}^{-1}(\lambda)| = 0$ for all $(i, j) \notin E$. Therefore we also have that

$$\int_{\Pi} \|f_{\theta^*}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)\|_2^2 d\lambda = O\left(p^{-(2\beta+1)}\right).$$

Since $\theta_0(p, G)$ minimizes $\mathcal{L}(\theta)$ for all $\theta \in \Theta(p, G)$, it then follows from (B.2) and (B.3) that

$$\begin{aligned} \int_{\Pi} \|f_{\theta_0(p, G)}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)\|_2^2 d\lambda &\leq C |\mathcal{L}(\theta_0(p, G)) - \mathcal{L}(\theta_0)| \\ &\leq C |\mathcal{L}(\theta^*) - \mathcal{L}(\theta_0)| \leq C \int_{\Pi} \|f_{\theta^*}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)\|_2^2 d\lambda, \end{aligned}$$

which together with (B.4) proves (4.2). The second part of the lemma then follows by the same argument used in (B.5) and the uniform boundedness of $\|f_{\theta_0(p, G_S)}^{-1}(\lambda)\|$ and $\|f_{\theta_0}(\lambda)\|$. \square

Proof of Lemma 4.3. We first consider the difference

$$\begin{aligned} |\mathcal{L}_T(\theta) - \mathcal{L}(\theta)| &= \left| \int_{\Pi} \text{tr}[(I^{(T)}(\lambda) - f_{\theta_0}(\lambda))f_{\theta}^{-1}(\lambda)] d\lambda \right| \\ &\leq \sum_{i,j=1}^d \sum_{|u| \leq p} \left| \int_{\Pi} (I_{ij}^{(T)}(\lambda) - f_{ij,\theta_0}(\lambda)) R_{ij,\theta}^{(i)}(u) \exp(i\lambda u) d\lambda \right| \end{aligned}$$

and further since $R_{ij,\theta}^{(i)}(u)$ does not depend on λ

$$\leq \max_{1 \leq i, j \leq d} \max_{|u| \leq p} |R_{ij,\theta}^{(i)}(u)| \sum_{i,j=1}^d \sum_{|u| \leq p} \left| \int_{\Pi} (I_{ij}^{(T)}(\lambda) - f_{ij,\theta_0}(\lambda)) \exp(i\lambda u) d\lambda \right|.$$

By Assumption 3.2 and the definition of $R_{ij,\theta}^{(i)}(u)$, the first factor can be bounded by $(4\pi^2 b_1)^{-1}$ uniformly over $\theta \in \Theta$. Further by Lemma B.1 each summand in the second factor has second moment of order $O(T^{-1})$ uniformly in $1 \leq u, v \leq P_T$. Thus

$$\mathbb{E} \left[\max_{1 \leq p \leq P_T} \sum_{|u| \leq p} \sum_{i,j=1}^d \left| \int_{\Pi} (I_{ij}^{(T)}(\lambda) - f_{ij,\theta_0}(\lambda)) \exp(i\lambda u) d\lambda \right|^2 \right] = O\left(\frac{P_T^2}{T}\right).$$

We therefore have

$$\max_{1 \leq p \leq P_T} \sup_{\theta \in \Theta(p, G)} |\mathcal{L}_T(\theta) - \mathcal{L}(\theta)| = O_P\left(\frac{P_T}{\sqrt{T}}\right),$$

which implies that

$$\mathcal{L}_T(\theta_0(p, G)) \xrightarrow{P} \mathcal{L}(\theta_0(p, G)) \quad \text{and} \quad |\mathcal{L}_T(\hat{\theta}_T(p, G)) - \mathcal{L}(\hat{\theta}_T(p, G))| \xrightarrow{P} 0$$

uniformly in $1 \leq p \leq P_T$. Together with the inequalities $\mathcal{L}_T(\hat{\theta}_T(p, G)) \leq \mathcal{L}_T(\theta_0(p, G))$ and $\mathcal{L}(\theta_0(p, G)) \leq \mathcal{L}(\hat{\theta}_T(p, G))$, this leads to the uniform convergence

$$\max_{1 \leq p \leq P_T} |\mathcal{L}(\hat{\theta}_T(p, G)) - \mathcal{L}(\theta_0(p, G))| = o_P(1).$$

Since the difference $|\mathcal{L}(\hat{\theta}_T(p, G)) - \mathcal{L}(\theta_0(p, G))|$ can be bounded from below by

$$\begin{aligned} & \mathcal{L}(\hat{\theta}_T(p, G)) - \mathcal{L}(\theta_0(p, G)) \\ &= \nabla \mathcal{L}(\theta_0(p, G))'(\hat{\theta}_T(p, G) - \theta_0(p, G)) + \frac{1}{2} \|\hat{\theta}_T(p, G) - \theta_0(p, G)\|_{\nabla^2 \mathcal{L}(\bar{\theta}_T)}^2 \\ &\geq \frac{1}{2} \|\hat{\theta}_T(p, G) - \theta_0(p, G)\|^2 \|\nabla^2 \mathcal{L}(\bar{\theta}_T)\|_{\text{inf}}, \end{aligned}$$

where $\bar{\theta}_T = \theta_0(p, G) + \xi(\hat{\theta}_T(p, G) - \theta_0(p, G))$ for some $\xi \in [0, 1]$, it follows that $\hat{\theta}_T(p, G)$ converges to $\theta_0(p, G)$ uniformly in $1 \leq p \leq P_T$. \square

Proof of Lemma 4.4. Noting that $\|\Gamma(p, G)^{-1}\| \leq C$ and $\|\Gamma(p, G)^{-1}\|_1 \leq Cp^{3/2}$, we obtain with Lemma B.1

$$\begin{aligned} & T\mathbb{E} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \\ &= \sum_{i_1, i_2 \in \mathbb{N}} T\mathbb{E} \left[\prod_{k=1}^2 \left(\frac{\partial \mathcal{L}_T(\theta_0(p, G))}{\partial \theta_{i_k}} - \frac{\partial \mathcal{L}_T(\theta_0(p, G))}{\partial \theta_{i_k}} \right) \right] \Gamma_{i_1 i_2}^{-1}(p, G) \\ &= \sum_{i_1, i_2 \in \mathbb{N}} \left[c_h \Gamma_{i_1 i_2, 0}(p, G) + O\left(\frac{p \log(T)}{T}\right) \right] \Gamma_{i_1 i_2}^{-1}(p, G) \\ &= c_h \text{tr}(\Gamma_0(p, G) \Gamma(p, G)^{-1}) + O\left(\frac{p^{\frac{5}{2}} \log(T)}{T}\right). \end{aligned}$$

where by the definition of $\Gamma(p, G)^{-1}$ all summands with $i_1, i_2 \notin I_{p, G}$ are zero, which yields p^2 nonzero summands.

For the second part we note that

$$\begin{aligned} & \left| \text{tr} \left[f_{\theta_0(p, G)}(\lambda) \frac{\partial f_{\theta_0(p, G)}^{-1}(\lambda)}{\partial \theta_i} f_{\theta_0(p, G)}(\lambda) \frac{\partial f_{\theta_0(p, G)}^{-1}(\lambda)}{\partial \theta_j} \right] - \text{tr} \left[f_{\theta_0}(\lambda) \frac{\partial f_{\theta_0}^{-1}(\lambda)}{\partial \theta_i} f_{\theta_0}(\lambda) \frac{\partial f_{\theta_0}^{-1}(\lambda)}{\partial \theta_j} \right] \right| \\ &\leq \left| \text{tr} \left[(f_{\theta_0(p, G)}(\lambda) - f_{\theta_0}(\lambda)) \frac{\partial f_{\theta_0(p, G)}^{-1}(\lambda)}{\partial \theta_i} f_{\theta_0(p, G)}(\lambda) \frac{\partial f_{\theta_0(p, G)}^{-1}(\lambda)}{\partial \theta_j} \right] \right| \\ &\quad + \left| \text{tr} \left[f_{\theta_0}(\lambda) \frac{\partial f_{\theta_0}^{-1}(\lambda)}{\partial \theta_i} (f_{\theta_0(p, G)}(\lambda) - f_{\theta_0}(\lambda)) \frac{\partial f_{\theta_0}^{-1}(\lambda)}{\partial \theta_j} \right] \right| \\ &\leq C \|f_{\theta_0(p, G)}(\lambda) - f_{\theta_0}(\lambda)\|_2 \left(\|f_{\theta_0(p, G)}(\lambda)\| + \|f_{\theta_0}(\lambda)\| \right) \end{aligned}$$

It follows by Lemma 4.2 and the Cauchy-Schwarz inequality

$$\begin{aligned} & \text{tr} \left(\Gamma(p, G)^{-1} [\Gamma_0(p, G) - \Gamma(p, G)] \right) \\ &\leq C \|\Gamma(p, G)^{-1}\|_1 \left(\int_{\Pi} \|f_{\theta_0(p, G)}(\lambda) - f_{\theta_0}(\lambda)\|_2^2 d\lambda \right)^{\frac{1}{2}} = O(p^{1-\beta}), \end{aligned}$$

which completes the proof. \square

Proof of Lemma 4.5. First assume that $G_0 \subseteq G$. We evaluate for $m = 2, 3, 4$

$$\begin{aligned} & T^m \mathbb{E} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^{2m} \\ &= T^m \sum_{i_1, \dots, i_{2m}=1}^{\infty} \mathbb{E} \left[\prod_{k=1}^{2m} \left(\frac{\partial \mathcal{L}_T(\theta_0(p, G))}{\partial \theta_{i_k}} - \frac{\partial \mathcal{L}(\theta_0(p, G))}{\partial \theta_{i_k}} \right) \right] \prod_{k=1}^m \Gamma_{i_{2k-1} i_{2k}}^{-1}(p, G). \end{aligned} \quad (\text{B.6})$$

Setting

$$Y_{i_k} = \sqrt{\frac{T}{c_h}} \left(\frac{\partial \mathcal{L}_T(\theta_0(p, G))}{\partial \theta_{i_k}} - \frac{\partial \mathcal{L}(\theta_0(p, G))}{\partial \theta_{i_k}} \right)$$

we obtain from the product theorem for cumulants and Lemma B.2

$$\begin{aligned} \mathbb{E} \left[\prod_{k=1}^{2m} Y_{i_k} \right] &= \sum_{n=1}^{2m} \sum_{\pi_1, \dots, \pi_n} \prod_{r=1}^n \text{cum} \{Y_{i_{r,1}}, \dots, Y_{i_{r,n_r}}\} \\ &= \sum_{\substack{\pi_1, \dots, \pi_n \\ n_r=2}}^m \prod_{r=1}^m \text{cum} \{Y_{i_{r,1}}, Y_{i_{r,2}}\} + \sum_{\substack{i_{1,1}, i_{1,2}=1 \\ i_{1,1} \neq i_{1,2}}}^{2m} \text{cum} \{Y_{i_{1,1}}, Y_{i_{1,2}}\} O\left(\frac{\log(T)^2}{T}\right) \\ &\quad + O\left(\frac{\log(T)^4}{T^2}\right). \end{aligned} \quad (\text{B.7})$$

Lemma B.1 yields for the first term

$$\begin{aligned} & \sum_{\substack{\pi_1, \dots, \pi_n \\ n_r=2}}^m \prod_{r=1}^m \left[\Gamma_{i_{r,1} i_{r,2}, 0}(p, G) + O\left(\frac{p \log(T)}{T}\right) \right] \\ &= \sum_{\substack{\pi_1, \dots, \pi_n \\ n_r=2}}^m \prod_{r=1}^m \Gamma_{i_{r,1} i_{r,2}, 0}(p, G) + \sum_{\substack{\pi_1, \dots, \pi_n \\ n_r=2}}^m \sum_{k=1}^m \prod_{\substack{r=1 \\ r \neq k}}^m \Gamma_{i_{r,1} i_{r,2}, 0}(p, G) O\left(\frac{p \log(T)}{T}\right) + O\left(\frac{p^2 \log(T)^2}{T^2}\right). \end{aligned}$$

Substituting (B.7) into (B.6) we then obtain with $B_{p,G} = \Gamma_0(p, G) \Gamma(p, G)^{-1}$

$$\begin{aligned} & \mathbb{E} \left[\frac{T}{c_h} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \right]^4 \\ &= (\text{tr}(B_{p,G}))^4 + 12 \text{tr}(B_{p,G}^2) (\text{tr}(B_{p,G}))^2 + 12 (\text{tr}(B_{p,G}^2))^2 + 32 \text{tr}(B_{p,G}^3) \text{tr}(B_{p,G}) \\ &\quad + 48 \text{tr}(B_{p,G}^4) + O\left(\frac{p^{\frac{11}{2}} \log(T)^2}{T}\right) \end{aligned}$$

and further in the same way as in the proof of Lemma 4.4

$$= k(p, G)^4 + 12k(p, G)^3 + 44k(p, G)^2 + 48k(p, G) + O(p^{4-\beta}) + O\left(\frac{p^{\frac{11}{2}} \log(T)^2}{T}\right).$$

Similarly, we get

$$\begin{aligned} & \mathbb{E} \left[\frac{T}{c_h} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \right]^3 \\ &= k(p, G)^3 + 6k(p, G)^2 + 8k(p, G) + O(p^{3-\beta}) + O\left(\frac{p^{\frac{9}{2}} \log(T)^2}{T}\right) \end{aligned}$$

and

$$\begin{aligned}
 & \mathbb{E} \left[\frac{T}{c_h} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \right]^2 \\
 &= \text{tr}(\text{tr}(B_{p, G}))^2 + 2\text{tr}(B_{p, G}^2) + O\left(\frac{p^{\frac{7}{2}} \log(T)^2}{T}\right) \\
 &= k(p, G)^2 + 2k(p, G) + O(p^{2-\beta}) + O\left(\frac{p^{\frac{7}{2}} \log(T)^2}{T}\right).
 \end{aligned} \tag{B.8}$$

Together with Lemma B.1 this proves the first part of the lemma. The second part follows directly from equation (B.8). \square

Proof of Lemma 4.7. Since \mathcal{G} is finite, it is sufficient to prove the convergence for fixed $G \in \mathcal{G}$. First consider the case $G_0 \subseteq G$. We then get with Lemma 4.5

$$\begin{aligned}
 & \mathbb{E} \left[\max_{1 \leq p \leq P_T} \left| \frac{T \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 - k(p, G)}{T c_h L_T(p, G)} \right| \right]^4 \\
 & \leq \sum_{1 \leq p \leq P_T} \mathbb{E} \left[\frac{T \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 - k(p, G)}{T c_h L_T(p, G)} \right]^4 \\
 & \leq \sum_{1 \leq p \leq P_T} \left(\frac{48k(p, G) + 12k(p, G)^2 + O(p^{4-\beta})}{T^4 L_T(p, G)^4} \right) + O\left(\frac{P_T^{\frac{5}{2}} \log(T)^2}{T}\right) \\
 & \leq \sum_{1 \leq p \leq p_T^*} \left(\frac{Cp^2}{p_T^{*4}} + \frac{Cp^{4-\beta}}{p_T^{*4}} \right) + \sum_{p_T^* < p \leq P_T} \left(\frac{C}{p^2} + Cp^{-\beta} \right) + O\left(\frac{P_T^{\frac{5}{2}} \log(T)}{T}\right),
 \end{aligned}$$

which tends to zero by the assumptions on P_T and β and $p_T^* \rightarrow \infty$. On the other hand if G does not contain G_0 , $I(\theta_0(p, G), \theta_0)$ does not vanish for $p \rightarrow \infty$ and therefore $L_T(p, G)$ is bounded away from zero uniformly in $p \in \mathbb{N}$. It then follows from the second part of Lemma 4.5

$$\mathbb{E} \left[\max_{1 \leq p \leq P_T} \left\| \nabla \mathcal{L}_T(\theta_0(p, G)) - \nabla \mathcal{L}(\theta_0(p, G)) \right\|_{\Gamma(p, G)^{-1}}^2 \right]^2 = O\left(\frac{P_T^3}{T}\right),$$

which completes the proof. \square

Proof of Lemma 5.1. First, we obtain by a Taylor expansion of $\mathcal{L}_T(\theta_0(p, G))$ about $\hat{\theta}_T(p, G)$

$$\begin{aligned}
 & \mathcal{L}_T(\theta_0(p, G)) - \mathcal{L}_T(\hat{\theta}_T(p, G)) \\
 &= \nabla \mathcal{L}_T(\hat{\theta}_T(p, G))'(\theta_0(p, G) - \hat{\theta}_T(p, G)) \\
 & \quad + \frac{1}{2} \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|_{\nabla^2 \mathcal{L}_T(\hat{\theta}_T(p, G))}^2 + O_P\left(p^{\frac{1}{2}} \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|^3\right) \\
 &= \frac{1}{2} \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|_{\nabla^2 \mathcal{L}(\theta_0(p, G))}^2 + O_P\left(p^{\frac{1}{2}} \left\| \hat{\theta}_T(p, G) - \theta_0(p, G) \right\|^3\right),
 \end{aligned}$$

where we have used the identity in (3.7) and further a Taylor expansion for the second derivative together with Lemma 4.1 (ii) to get for $\theta \in \ell^2(\mathbb{R})$

$$\theta' \nabla^2 \mathcal{L}(\hat{\theta}_T(p, G)) \theta = \theta' \nabla^2 \mathcal{L}(\theta_0(p, G)) \theta + O\left(p^{\frac{1}{2}} \|\hat{\theta}_T(p, G) - \theta_0(p, G)\| \|\theta\|^2\right).$$

The result follows now as in the proof of Theorem 4.8 from (4.5) and Lemma 4.7. \square

Proof of Lemma 5.2. We have with the abbreviations $\theta_{p,G} = \theta_0(p, G)$ and $\theta_T^* = \theta_0(p_T^*, G_T^*)$

$$\begin{aligned} & [\mathcal{L}_T(\theta_{p,G}) - \mathcal{L}(\theta_{p,G})] - [\mathcal{L}_T(\theta_T^*) - \mathcal{L}(\theta_T^*)] \\ &= \frac{1}{4\pi} \int_{\Pi} \text{tr} \left[(f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_T^*}^{-1}(\lambda)) (I^{(T)}(\lambda) - g(\lambda)) \right] d\lambda \\ &= \frac{1}{4\pi} \int_{\Pi} \text{tr} \left([(f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) + (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda))] (I^{(T)}(\lambda) - f_{\theta_0}(\lambda)) \right) d\lambda. \end{aligned}$$

Noting that $L_T(p, G) \geq L_T(p_T^*, G_T^*)$, we therefore obtain

$$\begin{aligned} & \max_{1 \leq p \leq P_T} \max_{G \in \mathcal{G}} \left| \frac{[\mathcal{L}_T(\theta_{p,G}) - \mathcal{L}(\theta_{p,G})] - [\mathcal{L}_T(\theta_T^*) - \mathcal{L}(\theta_T^*)]}{L_T(p, G)} \right| \\ & \leq \max_{1 \leq p \leq P_T} \max_{G \in \mathcal{G}} \frac{C}{L_T(p, G)} \left| \int_{\Pi} \text{tr} \left[(f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) (I^{(T)}(\lambda) - f_{\theta_0}(\lambda)) \right] d\lambda \right| \\ & \quad + \frac{C}{L_T(p_T^*, G_T^*)} \left| \int_{\Pi} \text{tr} \left[(f_{\theta_T^*}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) (I^{(T)}(\lambda) - f_{\theta_0}(\lambda)) \right] d\lambda \right| \\ & \leq \max_{1 \leq p \leq P_T} \max_{G \in \mathcal{G}} \frac{C}{L_T(p, G)} \left| \int_{\Pi} \text{tr} \left[(f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) (I^{(T)}(\lambda) - f_{\theta_0}(\lambda)) \right] d\lambda \right|. \end{aligned}$$

Since \mathcal{G} is finite, it is now sufficient to show that for every $G \in \mathcal{G}$

$$\sum_{p=1}^{P_T} \frac{1}{L_T(p, G)^4} \mathbb{E} \left[\int_{\Pi} \text{tr} \left\{ [f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)] [I^{(T)}(\lambda) - f_{\theta_0}(\lambda)] \right\} d\lambda \right]^4 \quad (\text{B.9})$$

converges to zero. By the product theorem for cumulants, the assumption of normality, and Lemma B.1 we find for the mean

$$\begin{aligned} & \mathbb{E} \left[\int_{\Pi} \text{tr} \left[(f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) (I^{(T)}(\lambda) - f_{\theta_0}(\lambda)) \right] d\lambda \right]^4 \\ &= \left(\text{cum} \left\{ \int_{\Pi} \text{tr} \left[(f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) (I^{(T)}(\lambda) - f_{\theta_0}(\lambda)) \right] d\lambda, \right. \right. \\ & \quad \left. \left. \int_{\Pi} \text{tr} \left[(f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) (I^{(T)}(\lambda) - f_{\theta_0}(\lambda)) \right] d\lambda \right\} \right)^2 + O\left(\frac{\log(T)^3}{T^3}\right) \\ &= \frac{C}{T^2} \left[\int_{\Pi} \text{tr} \left[g(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) g(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) \right] d\lambda \right]^2 + O\left(\frac{\log(T)^3}{T^3}\right). \end{aligned}$$

Next, we derive a lower bound of similar form for $L_T(p, G)$. For this, we consider the Taylor expansion of $\mathcal{L}(\theta_{p,G})$ about θ_0

$$\mathcal{L}(\theta_{p,G}) = \mathcal{L}(\theta_0) + \frac{1}{2} \|\theta_{p,G} - \theta_0\|_{\nabla^2 \mathcal{L}(\theta_0)}^2 + O(\|\theta_{p,G} - \theta_0\|_1 \|\theta_{p,G} - \theta_0\|^2),$$

where the first term vanishes since θ_0 minimizes $\mathcal{L}(\theta)$. To show that the remainder is of smaller order than the quadratic term, we note that

$$\|\theta_{p,G} - \theta_0\|_1 \leq \sum_{|u| \leq p} \int_{\Pi} \|f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}(\lambda)\|_1 d\lambda + \sum_{|u| > p} \|R_{\theta_0}^{(i)}(u)\|_1$$

By Lemma 4.2 the first term is of order $O(p^{\frac{1}{2}-\beta})$, while the second term vanishes for $p \rightarrow \infty$ since $R_{\theta_0}^{(i)}(u)$ is absolutely summable. Therefore $\|\theta_{p,G} - \theta_0\|_1$ converges to zero. $L_T(p, G)$ can now be rewritten as

$$\begin{aligned} L_T(p, G) &= \frac{k(p, G)\sigma_h}{2T} + \frac{1}{2} \|\theta_{p,G} - \theta_0\|_{\nabla^2 \mathcal{L}(\theta_0)}^2 + o(\|\theta_{p,G} - \theta_0\|^2) \\ &\geq \frac{1}{8\pi} \int_{\Pi} \text{tr} \left[f_{\theta_0}(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) f_{\theta_0}(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) \right] d\lambda \\ &\quad + o(\|\theta_{p,G} - \theta_0\|^2). \end{aligned}$$

and thus

$$\frac{1}{L_T(p, G)^2} \left[\int_{\Pi} \text{tr} \left[f_{\theta_0}(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) f_{\theta_0}(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) \right] d\lambda \right]^2 \leq C$$

uniformly in $1 \leq p \leq P_T$. Therefore we obtain for (B.9)

$$\begin{aligned} &\sum_{p=1}^{P_T} \frac{C}{T^2 L_T(p, G)^4} \left[\int_{\Pi} \text{tr} \left[f_{\theta_0}(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) f_{\theta_0}(\lambda) (f_{\theta_{p,G}}^{-1}(\lambda) - f_{\theta_0}^{-1}(\lambda)) \right] d\lambda \right]^2 \\ &\leq \sum_{p=1}^{P_T} \frac{C}{T^2 L_T(p, G)^2} \leq \sum_{p=1}^{p_T^*} \frac{Cp}{p_T^{*2}} + \sum_{p=p_T^*+1}^{P_T} \frac{C}{p^2}. \end{aligned}$$

Since p_T^* diverges to infinity, both summands tend to zero as $T \rightarrow \infty$ and the proof is complete. \square

Lemma B.1. *Suppose that Assumptions 3.1, 3.2 and 3.7 hold. Then we have with $f(\lambda) = f_{\theta_0}(\lambda)$*

$$\begin{aligned} &T\mathbb{E} \left\{ \int_{\Pi^2} (I_{ij}^{(T)}(\lambda) - f_{ij}(\lambda)) (I_{kl}^{(T)}(\lambda) - f_{kl}(\lambda)) \exp(i\lambda u + i\mu v) d\lambda d\mu \right\} \\ &= \frac{2\pi H_4}{H_2^2} \int_{\Pi} \left[f_{ik}(\lambda) f_{lj}(\lambda) \exp(i\lambda(u-v)) + f_{il}(\lambda) f_{kj}(\lambda) \exp(i\lambda(u+v)) \right] d\lambda \\ &\quad + O\left(\frac{p \log(T)}{T}\right) \end{aligned}$$

uniformly in $|u|, |v| \leq p$.

Proof. By standard arguments we find that the mean above is equal to

$$\begin{aligned} &\frac{2\pi H_4}{H_2^2} \int_{\Pi^2} \left[f_{ik}(\lambda) f_{lj}(\lambda) \Phi_2^{(T)}(\lambda + \mu) + f_{il}(\lambda) f_{kj}(\lambda) \Phi_2^{(T)}(\lambda - \mu) \right] \exp(i\lambda u + i\mu v) d\lambda d\mu \\ &\quad + O\left(\frac{\log(T)}{T}\right), \end{aligned}$$

where $\Phi_2^{(T)}$ is defined as in (A.3) and the error term is uniformly bounded in $|u|, |v| \leq p$. Noting that $\exp(i\lambda u)$ is Lipschitz continuous in λ with constant $C|u|$, we have

$$\int_{\Pi} |\exp(iv(\lambda + \mu)) - \exp(iv\lambda)| \Phi_2^{(T)}(\mu) d\mu \leq \frac{C}{T} \int_{\Pi} |v\mu| L^{(T)}(\mu) d\mu \leq \frac{C|v| \log(T)}{T},$$

which proves the lemma. \square

Lemma B.2. *Under Assumptions 3.1 and 3.2 we have*

$$\left| \int_{\Pi} \text{cum} \{I_{ij}^{(T)}(\lambda) - f_{ij, \theta_0}(\lambda)\} \exp(-i\lambda u) d\lambda \right| = O\left(\frac{1}{T}\right)$$

and for $k \leq 16$

$$\left| \int_{\Pi^k} \text{cum} \{I_{i_1 j_1}^{(T)}(\lambda_1), \dots, I_{i_k j_k}^{(T)}(\lambda_k)\} \exp\left(-i \sum_{l=1}^k \lambda_l u_l\right) d\lambda_1 \cdots d\lambda_k \right| = O\left(\frac{\log(T)^{k-2}}{T^{k-1}}\right)$$

uniformly in $u, u_1, \dots, u_k \in \mathbb{Z}$.

Proof. The first part follows directly from $|\mathbb{E}(I_{ij}^{(T)}(\lambda)) - f_{ij, \theta_0}(\lambda)| = O(T^{-1})$ uniformly in $\lambda \in [-\pi, \pi]$. For the second part we note that by the normality assumption

$$\text{cum}\{d_{i_1}^{(T)}(\lambda_1), \dots, d_{i_k}^{(T)}(\lambda_k)\} = 0$$

if $k \geq 3$. Therefore it follows from the product theorem for cumulants and (A.4) that

$$\begin{aligned} & \int_{\Pi^k} |\text{cum}\{I_{i_1 j_1}^{(T)}(\lambda_1), \dots, I_{i_k j_k}^{(T)}(\lambda_k)\}| d\lambda_1 \cdots d\lambda_k \\ & \leq \frac{C}{T^k} \sum_{i.p.} \int_{\Pi^k} \prod_{r=1}^k L^{(T)}(\bar{\gamma}_r) d\lambda_1 \cdots d\lambda_k \leq \frac{C \log(T)^{k-2}}{T^{k-1}}, \end{aligned}$$

where we have used the notation from Appendix A. \square

ACKNOWLEDGMENTS

This paper is part of the author's doctoral thesis at the University of Heidelberg. The author would like to thank his advisor Professor R. Dahlhaus for his support.

REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csáki (eds), *2nd international Symposium on Information Theory*, Akadémia Kiado, Budapest. reprinted in: S. Kotz and N.L. Johnson, eds., *Breakthroughs in Statistics, Vol. I*, Springer, New York, 1994.
- Battaglia, F. (1984). Inverse covariances of a multivariate time series. *Metron* **2**(3-4), 117–129.
- Bhansali, R. J. (1980). Autoregressive and window estimates of the inverse correlation function. *Biometrika* **67**, 551–561.
- Brillinger, D. R. (1996). Remarks concerning graphical models for time series and point processes. *Revista de Econometria* **16**, 1–23.
- Dahlhaus, R. (1983). Spectral analysis with tapered data. *Journal of Time Series Analysis* **4**.

- Dahlhaus, R. (1990). Nonparametric high resolution spectral estimation. *Probability Theory and Related Fields* **85**, 147–180.
- Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157–172.
- Dahlhaus, R. and Eichler, M. (2003). Causality and graphical models in time series analysis. In P. Green, N. Hjort and S. Richardson (eds), *Highly structured stochastic systems*, University Press, Oxford.
- Dahlhaus, R., Eichler, M. and Sandkühler, J. (1997). Identification of synaptic connections in neural ensembles by graphical models. *Journal of Neuroscience Methods* **77**, 93–107.
- Dzhaparidze, K. O. and Yaglom, A. M. (1983). Spectrum parameter estimation in time series analysis. In P. R. Krishnaiah (ed.), *Developments in Statistics*, Vol. 4, Academic Press, New York.
- Eichler, M. (1999). *Graphical Models in Time Series Analysis*. Doctoral thesis, Universität Heidelberg.
- Eichler, M. (2001). Graphical modelling of time series. *Technical report*, The University of Chicago.
- Eichler, M. (2002). Granger-causality and path diagrams for multivariate time series. *Technical report*, The University of Chicago.
- Fried, R. and Didelez, V. (2003). Decomposability and selection of graphical models for time series, *biometrika*, 2003. *Biometrika* **90**, 251–267.
- Fried, R., Didelez, V. and Lanius, V. (2003). Latent variable analysis and partial correlation graphs for multivariate time series. *Technical report*, Fachbereich Statistik, Universität Dortmund.
- Gather, U., Imhoff, M. and Fried, R. (2002). Graphical models for multivariate time series from intensive care monitoring statistics in medicine, 21, 2685-2701 (2002). *Statistics in Medicine* **21**, 2685–2701.
- Grenander, U. and Szegö, G. (1958). *Toeplitz forms and their applications*. University of California Press, Berkeley.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22**, 79–86.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.
- Masani, P. (1966). Recent trends in multivariate prediction theory. In P. R. Krishnaiah (ed.), *Multivariate Analysis*, Academic Press, pp. 351–382.
- Parzen, E. (1983). Autoregressive spectral estimation. In D. R. Brillinger and P. R. Krishnaiah (eds), *Handbook of statistics*, Vol. 3, North Holland, Amsterdam.
- Shaman, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Annals of Statistics* **3**, 532–538.
- Shaman, P. (1976). Approximations for stationary covariance matrices and their inverses with the application to ARMA models. *Annals of Statistics* **4**, 292–301.
- Shibata, R. (1997). Bootstrap estimate of kullback-leibler information for model selection. *Statistica Sinica* **7**, 375–394.
- Taniguchi, M. (1980). On selection of the order of the spectral density model for a stationary process. *Ann. Inst. Statist. Math.* **32**, 401–419.
- Timmer, J., Lauk, M., Köster, B., Hellwig, B., Häußler, S., Guschlbauer, B., Radt, V., Eichler, M., Deuschl, G. and Lücking, C. H. (2000). Cross-spectral analysis of tremor time series. *Int. J. of Bifurcation and Chaos* **10**, 2595–2610.
- Tunncliffe Wilson, G. (1972). The factorization of matricial spectral densities. *SIAM J. Appl. Math.* **23**, 420–426.

DEPARTMENT OF STATISTICS, THE UNIVERSITY OF CHICAGO, 5734 S. UNIVERSITY AVENUE,
CHICAGO, IL 60637, U.S.A.

E-mail address: `eichler@galton.uchicago.edu/eichler@statlab.uni-heidelberg.de`