

---

# Bayesian Structure Learning for Stationary Time Series: Supplementary Material

---

**Alex Tank**

University of Washington  
alextank@uw.edu

**Nicholas J. Foti**

University of Washington  
nfoti@uw.edu

**Emily B. Fox**

University of Washington  
ebfox@uw.edu

## Abstract

In this supplement we provide more background on spectral analysis of time series and the complex normal and complex inverse Wishart distributions. The hyper complex inverse Wishart is then introduced in more detail and its marginal likelihood is derived. We also provide more detail about the periodogram smoothing plug in method referenced in Section 4 of the main text. Finally, we provide details about the global stock index data set and compare the underlying conditional independence graphs learned by our method with that of Songsiri et al. [1]. We also provide supplemental simulations showcasing the predictive performance of our approach in the frequency domain.

## 1 Spectral Analysis Background

Spectral analysis is an approach to analyzing stationary time series where the main object of interest is the *spectral density*. For many applications, the spectral density is often a more informative object about the underlying physical process than autoregressive coefficients or the lagged autocovariance. In particular, it allows one to read off which frequency components are prominent in a time series, and usually certain frequency bands have scientific relevance. The theoretical justification for caring about the spectral density arises from the classic *spectral representation theorem* [2]. Informally, this theorem states that under some conditions, any stationary vector stochastic process  $x(t) \in \mathbb{R}^p$  can be written as

$$x(t) = \int_{-\pi}^{\pi} e^{-it\lambda} dZ(\lambda) \quad (1)$$

where  $dZ(\lambda) \in \mathbb{C}^p$  is an *orthogonal increments process* such that  $E(Z(\lambda)Z(\lambda')^*) = S(\lambda)$ , where  $S(\lambda)$  is the spectral density matrix and  $E(Z(\lambda)Z(\lambda')^*) = 0$  for  $\lambda \neq \lambda'$ .

Intuitively, this theorem states that the amplitudes of certain frequencies in a stationary process are independent across frequencies and the within frequency covariance is given by the spectral density matrix. As noted in the main text, the spectral density matrix also arises as the Fourier transformation of the lagged auto covariance matrices of the process

$$S(\lambda) = \sum_{h=-\infty}^{\infty} \Gamma(h)e^{-i\lambda h} \quad (2)$$

where  $\Gamma(h) = E(x(t)x(t+h)^T)$ . The statistical task then becomes estimation of the spectral density matrix  $S(\lambda)$  from a finite observed time series  $x = [x(1), \dots, x(T)]$ . The most basic estimator is given by the periodogram, defined as:

$$P_k = d_k d_k^* \quad (3)$$

where  $d_k$  is the DFT of the observed time series at Fourier frequency  $\lambda_k$ , (Eq. (7) of the main text). While this estimator is asymptotically unbiased, it is not consistent since its variance does not go to zero. Instead, techniques that smooth across nearby frequencies provide consistent estimators of the spectral density, and are commonly used in practice [3]. Finally, the periodogram estimates at different frequencies,  $P_k$  and  $P'_k$ , are asymptotically uncorrelated, providing some intuition as to why the Whittle approximation decomposes into independent terms for each frequency [4]. More details on the spectral approach to analyzing stationary time series is provided in Brillinger, 2001 [4].

## 2 Smoothing the Periodogram for a Single Time Series via the Plug in Method

In this section we provide more details on the plug in method for smoothing the periodogram obtained from a single realization of a multivariate time series mentioned in Sec. 4 of the main text. First we provide some background on classical frequentist approaches to smoothing the periodogram to obtain consistent estimators of the spectral density (as  $T$  increases).

When the spectral density itself is the primary object of interest, a common frequentist method is to smooth the periodogram to obtain a consistent estimator of the spectral density:

$$\hat{S}(\lambda_k) = \sum_{|j| < m} W_T(j) P_{k+j} \quad (4)$$

where  $P_k$  is the periodogram at frequency  $\lambda_k$  as introduced in the main text and  $W_T(j) \geq 0$ ,  $\sum_{|j| < m} W_T(j) = 1$  are some smoothing weights for a length  $T$  series and  $m$  is the smoothing window. This approach was used in the frequentist graph estimation frameworks in [5, 6, 7]. To ensure consistency as  $T \rightarrow \infty$  we must have  $m \rightarrow \infty$ ,  $\frac{m}{n} \rightarrow 0$ , and  $\sum_{|j| < m} W_T(j)^2 \rightarrow 0$  [2]. The asymptotic variance of  $\hat{S}_k$  scales as  $\sum_{|j| \leq m} W_T^2(j)$ , implying that the asymptotic effective sample size for a smoothed estimate of this form is  $(\sum_{|j| \leq m} W_T^2(j))^{-1}$  [2]. The Daniell smoother corresponds to taking  $W_T(j) = \frac{1}{2m+1}$  and has an intuitive (effective) sample size of  $2m+1$ , the size of the smoothing window. Intuitively, this holds asymptotically since as  $T \rightarrow \infty$  the sample periodograms become independent at different frequencies implying a sample size of  $2m+1$ , the number of (asymptotically) independent samples.

Inspired by the use of this smoothing technique in previous TGM procedures [5, 6, 7] we develop a similar procedure tailored to our objective function in Eq. (14). We plug in a smoothed estimate of the spectral density matrix, scaled by the asymptotic effective degrees of freedom, for the periodogram,  $P_k$ , in Eq. (14). Specifically, we set  $W_k^* = W_k + (\sum_{|j| \leq m} W_T^2(j))^{-1} \hat{S}_k$ . The degrees of freedom parameter  $\delta_k^*$  is similarly updated by adding the effective sample size of the smoother to the prior degrees of freedom:  $\delta_k^* = \delta_k + (\sum_{|j| \leq m} W_T^2(j))^{-1}$ . If we use the Daniell smoother outlined above the updates become  $W_k^* = W_k + \sum_{|j| \leq m} P_{k+j}$  and  $\delta^* = \delta_k + 2m + 1$ . In practice we set  $m = \lfloor \frac{\sqrt{T}}{2} \rfloor$  to ensure that the conditions for consistency of  $\hat{S}_k$  are met.

### 3 The Complex Normal and Complex Inverse Wishart Distributions

The complex normal distribution is a generalization of the multivariate normal distribution to the complex domain. Let  $Z \in \mathbb{C}^p$  be a complex random variable.  $Z$  is distributed as a complex normal distribution,  $\mathcal{N}_c(0, \Sigma)$ , with zero mean and complex Hermitian positive definite covariance matrix  $\Sigma \in \mathbb{C}^{p \times p}$  if it has density given by

$$p(z) = \frac{1}{\pi^p |\Sigma|} e^{-z^* \Sigma^{-1} z}, \quad (5)$$

where  $z^* = \bar{z}^T$  denotes the conjugate transpose of  $z$ . If  $Z \sim \mathcal{N}_c(0, \Sigma)$  then the distribution over  $Z$  can be represented equivalently as a joint distribution over the real and

imaginary elements of  $Z = X + iY$ ,  $X, Y \in \mathbb{R}^p$

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(0, \begin{bmatrix} \text{Re}\Sigma & -\text{Im}\Sigma \\ \text{Im}\Sigma & \text{Re}\Sigma \end{bmatrix}), \quad (6)$$

where  $\text{Re}\Sigma$  and  $\text{Im}\Sigma$  indicate the real and imaginary components of  $\Sigma$ , respectively. Thus we see that the real and imaginary components are independent iff  $\text{Im}\Sigma = 0$ . As in the non-complex case, the marginal likelihood of  $X_A$  for some subset of nodes  $A \subseteq \{1, \dots, p\}$ , is given by  $X_A \sim \mathcal{N}_c(0, \Sigma_A)$ , where  $\Sigma_A$  is the matrix formed by selecting the rows and columns of  $\Sigma$  in  $A$ .

The conjugate prior distribution for  $\Sigma$  is given by the complex inverse Wishart,  $\Sigma \sim IW_c(\delta, W)$ , with degrees of freedom parameter  $\delta > 0$  and centering matrix  $W \in \mathbb{C}^{p \times p}$ , Hermitian positive definite. Its density is given by

$$p(\Sigma|W, \delta) = B(W, \delta) |\Sigma|^{-(\delta+2p)} e^{-\text{tr}W\Sigma^{-1}} \quad (7)$$

with normalization constant

$$B(W, \delta) = \frac{|W|^{\delta+p}}{\pi^{\frac{p(p-1)}{2}} \prod_{j=1}^p (\delta + p - j)!}.$$

Note that we have used an alternative parameterization of the inverse Wishart distribution commonly used in the graphical modeling literature [8]. The marginal distribution of  $\Sigma_A$  where  $A \subseteq \{1, \dots, p\}$  is given by  $\Sigma_A \sim IW_c(\delta, W_A)$ .

### 4 Marginal Likelihood for the Hyper Complex Inverse Wishart

We define the hyper-complex inverse Wishart distribution for a graph  $G = \{V, E\}$  in the main paper as the restriction of the complex inverse Wishart distribution to  $\Sigma \in \mathbb{C}^{p \times p}$  with a zero pattern in  $\Sigma^{-1}$  specified by  $G$ . Its density is given by:

$$p(\Sigma|\delta, W, G) = \mathbf{1}_{\Sigma \in M^+(G)} h(W, \delta, G) |\Sigma|^{-(\delta+2p)} e^{-\text{tr}W\Sigma^{-1}} \quad (8)$$

where  $h(W, \delta, G)$  is a normalization constant and  $M^+(G)$  is the set of positive definite matrices with zeros in their inverse that obey the conditional independence properties of  $G$ .

Due to the fact that the complex inverse Wishart distribution is conjugate to the complex normal distribution for an unrestricted  $\Sigma$ , by Proposition 5.1 in [9] it follows that the hyper complex inverse Wishart distribution is a strong hyper-Markov distribution. It follows that for decomposable  $G$  the complex hyper inverse Wishart density can be written in terms of the cliques,  $\mathcal{C}$ , and separators,  $\mathcal{S}$ , of  $G$ :

$$p(\Sigma|\delta, W, G) = \mathbf{1}_{\Sigma \in M^+(G)} \frac{\prod_{C \in \mathcal{C}} p(\Sigma_C | W_C, \delta)}{\prod_{S \in \mathcal{S}} p(\Sigma_S | W_S, \delta)} \quad (9)$$

where  $p(\Sigma_C|W_C, \delta)$  is the unrestricted complex inverse Wishart density for  $\Sigma_C$ . This decomposition implies that the normalization constant for Equation (8) is also given by the ratio of complex inverse Wishart normalization constants for cliques and separators

$$h(W, \delta, G) = \frac{\prod_{C \in \mathcal{C}} B(W_C, \delta)}{\prod_{S \in \mathcal{S}} B(W_S, \delta)}. \quad (10)$$

If  $Z_1, \dots, Z_N \stackrel{i.i.d.}{\sim} \mathcal{N}_c(0, \Sigma)$ , then the joint distribution of  $Z_1, \dots, Z_N$ , and  $\Sigma$  can be written as:

$$p(z_1, \dots, z_N, \Sigma | G, W, \delta) \propto \quad (11)$$

$$\mathbf{1}_{\Sigma \in M^+(G)} \frac{h(W, \delta, G)}{\pi^{Np}} |\Sigma|^{-(\delta+N+2p)} e^{-\text{tr}(W + \sum_{i=1}^N z_i z_i^*) \Sigma^{-1}}. \quad (12)$$

We note that the part dependent on  $\Sigma$  is the kernel for a  $HIW_c(W + \sum_{i=1}^N z_i z_i^*, \delta + N, G)$  distribution, it follows that the marginal distribution of  $Z_1, \dots, Z_N | G, W, \delta$  is given by the ratio of prior and posterior normalization constants of the complex hyper inverse Wishart distribution times a likelihood constant:

$$p(z_1, \dots, z_n | G, W, \delta) = \frac{h(W, \delta, G)}{\pi^{Np} h(W + \sum_{i=1}^N z_i z_i^*, \delta + N, G)}. \quad (13)$$

#### 4.1 Marginal Whittle Likelihood

The model in the main text places independent  $HIW_c(W_k, \delta_k, G)$  priors on each spectral density matrix in the Whittle likelihood,  $S_k \sim HIW_c(W_k, \delta_k, G) \forall k \in [T-1]$ . Applying the above marginal likelihood result to each frequency component in the Whittle approximation shows that the marginal likelihood of the data given a graph, a set of centering matrices  $W_0, \dots, W_{T-1}$ , and degrees of freedom,  $\delta_0, \dots, \delta_{T-1}$ , for each frequency can be approximated by a product of the normalization constants across frequencies:

$$p(\mathbf{X}_{1:N} | G) \approx \pi^{-NTp} \prod_{k=0}^{T-1} \frac{h(W_k, \delta_k, G)}{h(W_k^*, \delta_k^*, G)}. \quad (14)$$

where  $W_k^* = W_k + P_k$  and  $\delta_k^* = \delta_k + N$ . Indeed, this derivation shows that our prior specification for spectral density matrices is conjugate to the entire Whittle likelihood.

## 5 Prediction Simulations

To further validate our approach we analyze predictive performance in the frequency domain on simulated VAR(1) data as described in Section 6 of the main text. We set the dimension to  $p = 10$  and time series length to  $T = 2500$ . We split the simulated series in half forming a training

set and test set and then learn a graph and smoothed periodogram on the test set of the time series. We use the Whittle marginal likelihood presented in Eq. (14) to compare predictions between four graphs: the learned graph in the frequency domain,  $\hat{G}_{\text{spectral}}$ , the learned iid graph that treats the time series as independent observations,  $\hat{G}_{\text{iid}}$ , the full graph,  $G_{\text{full}}$ , with all edges included, and the empty graph,  $G_{\text{empty}}$ , with no edges. For prediction, the prior centering matrix,  $W_k$ , is given by the Daniell smoothed periodogram on the training data with its respective degrees of freedom, and  $W_k^*$  is given by  $W_k^* = W_k + P_k^{\text{test}}$ , where  $P_k^{\text{test}}$  is the periodogram on the test data. Results are displayed in Table 1. We see that the learned spectral graph does significantly better than the other graphs at prediction.

## 6 Global Stock Indices

### 6.1 Learned Graph Comparison with Previous Methods

For comparison with the method of Songsiri et al. [1], we provide the CIG graphs learned on the international stock data set using both their autoregressive method and our nonparametric Bayesian method in Figure 1. Further details on the meaning of the edge weights can be found in [1]. Notice that both graphs capture similar structure, for instance the connections between the US, Canada, Australia, and Japan. Additionally, both graphs contain tight clusters containing European countries.

### 6.2 Stock Data

In Table 2, we list the stock indices that were used in the main paper. The data was downloaded from [globalfinancialdata.com](http://globalfinancialdata.com) for the dates June 3, 1997 to June 30, 1999.

## References

- [1] J. Songsiri and L. Vandenberghe. Topology selection in graphical models of autoregressive processes. *JMLR*, 11:2671–2705, 2010.
- [2] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, NY, 1991.
- [3] Donald B. Percival and Andrew T. Walden. *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.
- [4] D.R. Brillinger. *Time Series: Data Analysis an Theory*. Holden-Day, 2001.
- [5] A. Jung, G. Hannak, and N. Görtz. Graphical LASSO based model selection for time series. *ArXiv e-prints*, 2014.

	$\hat{G}_{\text{spectral}}$	$\hat{G}_{\text{iid}}$	$G_{\text{empty}}$	$G_{\text{full}}$
Loglik.	$-81623 \pm 205$	$-87712 \pm 216$	$-90040 \pm 211$	$-103964 \pm 304$

Table 1: Predictive marginal log-likelihood for simulated VAR data in the frequency domain on a held out test set using the marginal Whittle likelihood in Eq. (14) under four different graphs ( $\pm$  indicates 1 standard error across simulation replicates).

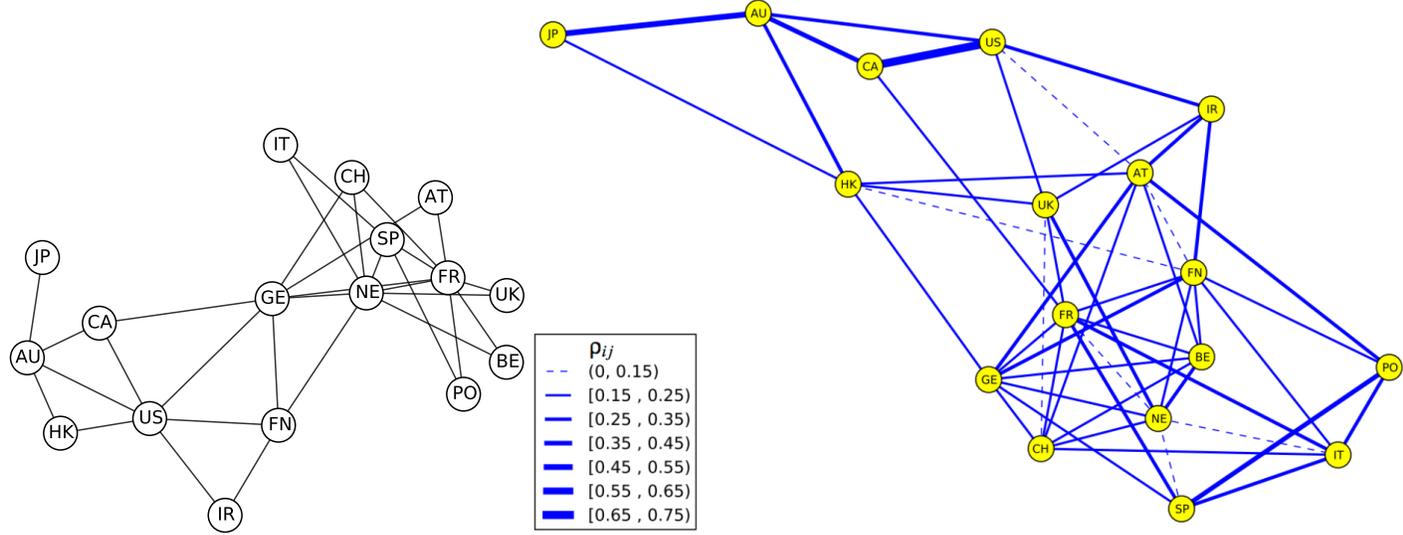


Figure 1: CIG graph learned on international stock data using (*left*) method presented in this paper and (*right*) the autoregressive method presented in Songsiri et al. [1] (Figure taken directly from [1]).

[6] F. R. Bach and M. I. Jordan. Learning graphical models for stationary time series. *IEEE Transactions on Signal Processing*, 52(8):2189–2199, 2004.

[7] R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51(2):157–172, 2000.

[8] P. Giudici and P. J. Green. Decomposable graphical Gaussian model determination. *Biometrika*, 86(4):785–801, 1999.

[9] A. P. Dawid and S. L. Lauritzen. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.*, 21(3):1272–1317, 1993.

Table 2: Stock index information.

Index Name	Ticker	Country	Country Code
Amsterdam Exchange Index	AEX	Netherlands	NE
All Ordinary Composite	AORD	Australia	AU
Austrian Traded Index	ATX	Austria	AT
BEL 20	BFX	Belgium	BE
CAC 40	FCHI	France	FR
FTSEMIB	FTMIB	Italy	IT
FTSE 100	FTSE	United Kingdom	UK
DAX 30	GDAX	Germany	GE
Toronto Stock Exchange 300	GSPTSE	Canada	CA
Hang Seng Composite	HSI	Hong Kong	HK
IBEX 35	IBEX	Spain	SP
Irish Stock Exchange Index	ISEQ	Ireland	IR
Nikkei 225	N225	Japan	JP
OMX Helsinki 25	OMXH25	Finland	FN
Portugal Stock Index	PSI20	Portugal	PO
S&P 500	SPX	United States	US
Swiss Market Index	SSMI	Switzerland	CH