Prediction error matrix under model misspecification for multivariate harmonic time series regression models

Rinka Sagawa

Received January 5 2026; revised March 28 2026

Abstract. We consider the one-step-ahead prediction error matrix under misspecification caused by overfitting periodic components in the multivariate harmonic regression models. We show that the prediction error matrix obtained from the multivariate harmonic regression model admits a decomposition into the prediction error matrix of an autoregressive model and a periodic component characterized by the spectral density matrix. This decomposition provides a theoretical characterization of the effect of the model misspecification on prediction accuracy. In the data analysis, we apply the proposed decomposition with estimated parameters to the temperature data, leading to a basis for constructing an information criterion for selecting periodic components.

1 Introduction In time series analysis, periodicity is one of the fundamental characteristics of time series. Many real-world time series exhibit periodic patterns, including temperature data, electricity demand, and brainwave signals. Identifying and modeling such periodic behavior play an important role in understanding the underlying mechanisms of the observed phenomena. Therefore, estimating periodic components is an important problem in time series analysis.

Among time series models, the harmonic regression model is one of the most widely used models for representing periodic structures in time series (e.g. Hannan (1973), Brillinger (2001)). For example, the d-dimensional harmonic regression model is given by

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \left(\sum_{k=1}^{r_0} \boldsymbol{\alpha}_k \cos(t\theta_k) + \boldsymbol{\beta}_k \sin(t\theta_k) \right) + \boldsymbol{X}_t, \tag{1.1}$$

where $\boldsymbol{Y}_t \in \mathbb{R}^d$ is the observed vector and $\boldsymbol{\mu}$ is the mean vector, and $\boldsymbol{X}_t$ is a zero-mean stationary process with spectral density $\boldsymbol{f}$, which is not observable. The vectors $\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k \in \mathbb{R}^d$ are the coefficient vectors of the periodic components, and $\theta_k \in (0, \pi)$ is the corresponding frequency. Under this model, the periodic structure is expressed as a sum of several periodic components. To appropriately capture the periodic structure, it is necessary to estimate the periodic components. However, the number of periodic components is typically unknown, and therefore the estimation of the number of periodic components becomes an important statistical problem. As a result, models are often fitted with an incorrect number of periodic components. This problem is often formulated as a model selection problem.

In model selection, it is important to evaluate the goodness-of-fit of competing models while accounting for model complexity. Information criteria are widely used for this purpose because they provide a trade-off between model fit and model complexity. Typical examples include the Akaike information criterion (Akaike (1974)), the Bayesian information criterion (Schwarz (1978)), and the Hannan-Quinn criterion (Hannan and Quinn (1979)). Various extensions of information criteria have also been proposed, such as the corrected AIC (Hurvich and Tsai (1989)) and the Generalised information criterion (Konishi and Kitagawa (1996)).

2010 *Mathematics Subject Classification.* MSC-62M10, MSC-62H12.
Key words and phrases. Multivariate autoregressive process, model misspecification, prediction error matrix.

However, in practical data analysis, the true model is typically unknown and the fitted model may therefore be misspecified. The properties of estimation and prediction under model misspecification have been extensively studied in the literature. White (1982) investigated the effects of model misspecification on estimation and inference based on maximum likelihood methods. Yamamoto (1976) derived a manageable expression for the asymptotic mean squared prediction error in multistep forecasting for finite-order autoregressive model with the estimated coefficients. Bhansali (1981) investigated the effects of order misspecification in autoregressive models fitted by least squares. Choi and Taniguchi (2001) studied prediction under model misspecification using a parametric spectral density. Yajima and Matsuda (2009) proposed a unified framework for hypothesis testing in multivariate linear time series that is robust to model misspecification. In such situations, it becomes important to understand how to evaluate model fit when the model is misspecified. In particular, Hosoya and Taniguchi (1982) studied the misspecified prediction error matrix at a theoretical level without regression components. In contrast, our study considers a harmonic regression framework and the prediction error matrix estimated via the least squares estimator (LSE) is theoretically characterized.

Furthermore, many studies have investigated periodicity estimation, most of them focus on univariate time series (e.g. Walker (1971), Quinn and Thomson (1991)). In contrast, in multivariate time series the presence of interactions among components makes the identification and estimation of periodic structures more challenging.

Therefore, in this study, we investigate the prediction error matrix under model misspecification in multivariate harmonic regression models. Specifically, we consider the prediction error $\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_t$, where $\hat{\boldsymbol{Y}}_t$ denotes the one-step-ahead linear predictor of $\boldsymbol{Y}_t$ constructed based on the model in (1.1). The covariance matrix of this prediction error is referred to as the prediction error matrix. In practical data analysis, the periodic components are typically unknown, and models are often fitted using periodic structures that differ from the true periodic components. Motivated by this issue, we focus on the one-step-ahead prediction error matrix constructed under such misspecified model and evaluate the properties of its prediction error matrix. In particular, we consider the case $r_0 = 0$, which corresponds to the model without periodic components. In this case, when $\boldsymbol{\mu} = \boldsymbol{0}$, it follows that $\boldsymbol{Y}_t = \boldsymbol{X}_t$, and thus the one-step-ahead linear prediction of $\boldsymbol{Y}_t$ is obtained from $\boldsymbol{X}_t$. In this study, we focus on the prediction error matrix under model misspecification.

The contribution of this paper is twofold. First, we derive a decomposition of the prediction error matrix for misspecified multivariate harmonic models into the prediction error matrix of vector autoregressive models and a component representing periodic structures. Second, the proposed decomposition provides a principled way to quantify the discrepancy between observed data and multivariate harmonic models. Furthermore, this framework provides a foundation for constructing information criteria for selecting the number of periodic components in multivariate time series.

Throughout this paper, due to the inherent complexity of the notation, we adopt the notation used in Lütkepohl (2005) to ensure clarity and consistency. Additionally, $\|\cdot\|$ denote the Frobenius norm, and for any matrix A, $A^\top$ denote the transpose of A.

This paper is organized as follows. In Section 2, we derive one-step-ahead prediction error matrices for autoregressive models and misspecified multivariate harmonic regression models with a single periodic component. In Section 3, we propose the prediction error matrix decomposition for misspecified multivariate harmonic regression models with a single periodic component. In Section 4, we apply our decomposition with the estimated parameters to the temperature data. All proofs are provided in Appendix.

2 Preliminary In this section, we consider two one-step-ahead prediction error matrices: one obtained from the correctly specified model and the other obtained from a misspecified model. We first introduce their theoretical definitions and then focus on their estimators obtained via least squares estimator (LSE), which are the primary focus of this paper. To clarify the setting, we consider the case where the observed data contain no periodic components. In this case, the observed series satisfies $\boldsymbol{Y}_t = \boldsymbol{X}_t$ under $\boldsymbol{\mu} = \boldsymbol{0}$ in (1.1). For this series, we consider two one-step-ahead linear prediction structures. The first corresponds to the correctly specified model, where $\boldsymbol{X}_t$ is predicted using a vector autoregressive (VAR) model. The second corresponds to the one-step-ahead linear prediction obtained from a misspecified harmonic regression model that includes periodic components.

Let $\{\boldsymbol{X}_t; t \in \mathbb{Z}\}$ be the residual dependent process as in (1.1). First, we focus on d-dimensional vector autoregressive model such as

$$\boldsymbol{X}_t = \sum_{j=1}^{p} A_j \boldsymbol{X}_{t-j} + \boldsymbol{U}_t, \tag{2.1}$$

where $\boldsymbol{X}_t = (X_{t1}, \dots, X_{td})^\top$, $\{A_j\}$ is a sequence of $d \times d$ matrices, $\boldsymbol{U}_t = (U_{t1}, \dots, U_{td})^\top$ is a sequence of i.i.d. zero-mean random vectors with covariance matrix Ξ.

The one-step-ahead linear predictor of $\boldsymbol{X}_t$ based on $\boldsymbol{X}_{t-1}, \dots, \boldsymbol{X}_{t-p}$ is defined as

$$\hat{\boldsymbol{X}}_t(p) = \sum_{j=1}^{p} A_j \boldsymbol{X}_{t-j}.$$

The corresponding prediction error is

$$\boldsymbol{e}_t^{\boldsymbol{X}}(p) = \boldsymbol{X}_t - \hat{\boldsymbol{X}}_t(p)$$

and the theoretical prediction error matrix is defined by

$$\Sigma_{\boldsymbol{X}}(p) = \mathbb{E}[\boldsymbol{e}_t^{\boldsymbol{X}}(p)\boldsymbol{e}_t^{\boldsymbol{X}}(p)^\top].$$

We now consider the estimator of prediction error matrix $\Sigma_{\boldsymbol{X}}(p)$ via the least squares estimator, denoted by $\hat{\Sigma}_{\boldsymbol{X}}(p)$. Let N be the number of observations of $\boldsymbol{X}_t$. The matrix form of (2.1) is

$$\boldsymbol{X} = A\boldsymbol{X}(p) + \boldsymbol{U}, \tag{2.2}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_N) \in \mathbb{R}^{d\times N}$, $A = (A_1, \dots, A_p) \in \mathbb{R}^{d\times dp}$, $\boldsymbol{X}(p) = (\boldsymbol{X}_1(p), \dots, \boldsymbol{X}_N(p)) \in \mathbb{R}^{dp\times N}$, $\boldsymbol{X}_t(p) = (\boldsymbol{X}_{t-1}^\top, \dots, \boldsymbol{X}_{t-p}^\top)^\top \in \mathbb{R}^{dp\times 1}$, $\boldsymbol{U} = (\boldsymbol{U}_1, \dots, \boldsymbol{U}_N) \in \mathbb{R}^{d\times N}$. Here, we impose the following assumption.

Assumption 2.1 (Lütkepohl (2005), p.16)**.** Let $A(z)$ be

$$A(z) = \det\left\{\sum_{j=0}^{p} A_j z^j\right\}, \qquad z \in \mathbb{C},$$

where $A_0 = \boldsymbol{I}_d$ and $\boldsymbol{I}_d$ is the $d \times d$ identity matrix. Then, we assume $A(z) = 0$ has no roots in $D = \{z \in \mathbb{C} : |z| \leq 1\}$.

This assumption ensures the VAR(p) model (2.2) is stationary. Let $\hat{\Sigma}_{\boldsymbol{X}}(p)$ denote the prediction error matrix by autoregressive model (2.2). The following proposition holds for $\hat{\Sigma}_{\boldsymbol{X}}(p)$.

Proposition 2.2. Under Assumption 2.1,

$$\hat{\Sigma}_{\boldsymbol{X}}(p) = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top - \frac{1}{N^2}\boldsymbol{X}\boldsymbol{X}(p)^\top \hat{\Gamma}_{\boldsymbol{X}(p)}^{-1}\boldsymbol{X}(p)\boldsymbol{X}^T,$$

where

$$\hat{\Gamma}_{\boldsymbol{X}(p)}^{-1} = \left(\frac{1}{N}\boldsymbol{X}(p)\boldsymbol{X}(p)^T\right)^{-1},$$

assuming $\hat{\Gamma}_{\boldsymbol{X}(p)}$ is nonsingular.

The proof is provided in Appendix A.

Now, we move to the main discussion of one-step-ahead prediction error matrix of misspecifying multivariate harmonic regression model. For simplicity, let us recall d-dimensional harmonic regression model (1.1) with a single periodic component at frequency $\theta \in (0, \pi)$:

$$\boldsymbol{Y}_t = \boldsymbol{\alpha} \cos(t\theta) + \boldsymbol{\beta} \sin(t\theta) + \boldsymbol{X}_t, \qquad t \in \mathbb{Z}, \tag{2.3}$$

where $\{\boldsymbol{X}_t : t \in \mathbb{Z}\}$ is a zero-mean stationary process with the spectral density matrix $\boldsymbol{f}$. In practice, the periodic components are usually unknown and must be estimated from the data. As a result, models may include periodic components that are not present in the true data generating process, leading to model misspecification. This situation corresponds to fitting a harmonic regression model with unnecessary periodic components, that is, overfitting the periodic structure. In this case, prediction is carried out using a model that includes periodic components even though the true data generating process contains none.

Here, we consider a misspecified multivariate harmonic regression model in which the stationary process $\boldsymbol{X}_t$ is modeled using the observed process $\boldsymbol{Y}_t$ that contains periodic components:

$$\boldsymbol{X}_t = \sum_{j=1}^{p} B_j \boldsymbol{Y}_{t-j} + \boldsymbol{V}_t, \tag{2.4}$$

where $\boldsymbol{X}_t = (X_{t1}, \dots, X_{td})^\top$, $\boldsymbol{Y}_t = (Y_{t1}, \dots, Y_{td})^\top$, $\{B_j\}$ is a sequence of $d \times d$ matrices, $\boldsymbol{V}_t = (\boldsymbol{V}_{t1}, \dots, \boldsymbol{V}_{td})$ is a sequence of i.i.d. zero-mean random vectors with covariance matrix Ξ.

The one-step-ahead linear predictor of $\boldsymbol{X}_t$ based on the misspecified harmonic regression model is

$$\hat{\boldsymbol{X}}_t^{\boldsymbol{Y}}(p) = \sum_{j=1}^{p} B_j \boldsymbol{Y}_{t-j}.$$

The corresponding prediction error is defined as

$$\boldsymbol{e}_t^Y(p) = \boldsymbol{X}_t - \hat{\boldsymbol{X}}_t^{\boldsymbol{Y}}(p)$$

The theoretical prediction error matrix is defined as

$$\Sigma_{\boldsymbol{Y}}(p) = \mathbb{E}[\boldsymbol{e}_t^Y(p) \boldsymbol{e}_t^Y(p)^\top].$$

Now, let us consider the estimator of the prediction error matrix $\Sigma_{\boldsymbol{Y}}(p)$ via the least squares estimator, denoted by $\hat{\Sigma}_{\boldsymbol{Y}}(p)$.

The matrix form of (2.4) is

$$\boldsymbol{X} = B\boldsymbol{Y}(p) + \boldsymbol{V}, \tag{2.5}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_N) \in \mathbb{R}^{d\times N}$, $B = (B_1, \dots, B_p) \in \mathbb{R}^{d\times dp}$, $\boldsymbol{X}(p) = (\boldsymbol{X}_1(p), \dots, \boldsymbol{X}_N(p)) \in \mathbb{R}^{dp\times N}$, $\boldsymbol{Y}(p) = (\boldsymbol{Y}_1(p), \dots, \boldsymbol{Y}_N(p)) \in \mathbb{R}^{dp\times N}$, $\boldsymbol{Y}_t(p) = (\boldsymbol{Y}_{t-1}^\top, \dots, \boldsymbol{Y}_{t-p}^\top)^\top \in \mathbb{R}^{dp\times 1}$, and $\boldsymbol{V} = (\boldsymbol{V}_1, \dots, \boldsymbol{V}_N) \in \mathbb{R}^{d\times N}$.

Let $\hat{\Sigma}_{\boldsymbol{Y}}(p)$ be the prediction error matrix by misspecifying the true model (2.2) with multivariate harmonic regression model (2.5). The following proposition holds for $\hat{\Sigma}_{\boldsymbol{Y}}(p)$.

Proposition 2.3. Under Assumption 2.1,

$$\hat{\Sigma}_{\boldsymbol{Y}}(p) = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top - \frac{1}{N^2}\boldsymbol{X}\boldsymbol{Y}(p)^\top \hat{\Gamma}_{\boldsymbol{Y}(p)}^{-1} \boldsymbol{Y}(p)\boldsymbol{X}^T,$$

where

$$\hat{\Gamma}_{\boldsymbol{Y}(p)}^{-1} = \left(\frac{1}{N}\boldsymbol{Y}(p)\boldsymbol{Y}(p)^T\right)^{-1}.$$

The proof is provided in Appendix B.

3 Prediction error matrix decomposition In this section, we demonstrate the relationship between one-step-ahead prediction error matrix by autoregressive model and that by misspecifying the true model (2.2) with the multivariate harmonic regression model (2.5) with a single periodic component.

We introduce the following assumptions to establish probability 1 bounds for a range of relevant statistics.

Assumption 3.1 (Brillinger (2001), Assumption 2.6.3)**.** Let C_k be

$$C_k = \sup_{a_1,\ldots,a_k} \sum_{\nu_1,\ldots,\nu_k} |c_{a_1,\ldots,a_k}(\nu_1,\ldots,\nu_{k-1})|,$$

where $c_{a_1,\ldots,a_k}(\nu_1,\ldots,\nu_{k-1})$ is the kth order cumulants of $\boldsymbol{X}_t$. Assume that

$$\sum_k C_k z^k/k! < \infty,$$

for z in a neighborhood of 0.

For $1 \leq i \leq d$, let α_i and β_i be the ith element of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (2.3), respectively. We define $\rho_i = (\alpha_i^2 + \beta_i^2)^{1/2}$ and ω_i satisfies

$$\cos(\omega_i) = \frac{\alpha_i}{\rho_i}, \qquad \sin(\omega_i) = \frac{\beta_i}{\rho_i}, \qquad i = 1,\ldots,d.$$

Let us recall that

$$\hat{\Gamma}_{\boldsymbol{X}(p)} = \left(\frac{1}{N}\boldsymbol{X}(p)\boldsymbol{X}(p)^\top\right), \qquad \hat{\Gamma}_{\boldsymbol{Y}(p)} = \left(\frac{1}{N}\boldsymbol{Y}(p)\boldsymbol{Y}(p)^\top\right).$$

Before stating the main decomposition result, we introduce the following technical lemma concerning $\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}$ and $\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}$.

Lemma 3.2. Under Assumption 3.1,

$$\left\|\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)} - \left\{\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)} - \hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}(\boldsymbol{I}_2 + \boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\right\}\right\| = O_p\left(p^3\Big(\frac{\log N}{N}\Big)^{1/2}\right),$$

where $\boldsymbol{I}_2$ is the 2×2 identity matrix and for $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_d)^\top$,

$$\boldsymbol{J} = \big(\boldsymbol{\xi}(\theta)\otimes\boldsymbol{\rho}(\boldsymbol{\omega}), \boldsymbol{\xi}(-\theta)\otimes\boldsymbol{\rho}(-\boldsymbol{\omega})\big) \in \mathbb{C}^{dp\times 2},$$

$$\boldsymbol{\xi}(\theta) = (e^{\mathrm{i}\theta},\ldots,e^{\mathrm{i}\theta p})^\top \in \mathbb{C}^{p\times 1}, \qquad \boldsymbol{\rho}(\boldsymbol{\omega}) = \left(\frac{\rho_1}{2}e^{-\mathrm{i}\omega_1},\ldots,\frac{\rho_d}{2}e^{-\mathrm{i}\omega_d}\right)^\top \in \mathbb{C}^{d\times 1}.$$

The proof of Lemma 3.2 is provided in Appendix C. To ensure the decomposition of one-step-ahead prediction error matrix under misspecification in multivariate harmonic regression models, we impose the following assumption.

Assumption 3.3. Let $p := p_N$ be the order of vector autoregressive model (2.1) or the misspecifying multivariate harmonic regression model (2.4). We assume that $p \to \infty$ and $p^4/N \to 0$, as $N \to \infty$.

Based on Lemmas 2.2, 2.3, and 3.2, we arrive at the following theorem.

Theorem 3.4. Suppose that Assumptions 3.1 and 3.3 hold. We obtain

$$\left\| \hat{\Sigma}_{\boldsymbol{Y}}(p) - \left\{ \hat{\Sigma}_{\boldsymbol{X}}(p) + \boldsymbol{C}\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\boldsymbol{C}^{*} \right\} \right\| = O_P\left(p^2 \frac{\log N}{N} \right), \tag{3.1}$$

where $\boldsymbol{C}$ is the $d \times 2$ matrix such as

$$\left(\Big(\sum_{j=1}^{p} \hat{A}_j e^{\mathrm{i}\theta j} \Big) \boldsymbol{\rho}(\boldsymbol{\omega}), \Big(\sum_{j=1}^{p} \hat{A}_j e^{-\mathrm{i}\theta j} \Big) \boldsymbol{\rho}(-\boldsymbol{\omega}) \right),$$

and $\hat{A}_1, \ldots, \hat{A}_p$ is the estimated coefficient matrix of (2.1), and $\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}$ is 2×2 diagonal matrix such as

$$\begin{pmatrix} \left(p\boldsymbol{\rho}^*(\boldsymbol{\omega})\boldsymbol{f}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}) \right)^{-1} & 0 \\ 0 & \left(p\boldsymbol{\rho}^*(-\boldsymbol{\omega})\boldsymbol{f}(\theta)\boldsymbol{\rho}(-\boldsymbol{\omega}) \right)^{-1} \end{pmatrix},$$

where $\boldsymbol{f}(\theta)$ is the spectral density matrix of $\{\boldsymbol{X}_t\}$ at frequency θ corresponding to the periodic component.

The proof of Theorem 3.4 is provided in Appendix D. When p is larger than N, Theorem 3.4 does not hold. Thus, there is a trade-off between p and N. In particular, if p becomes too large relative to the sample size N, the estimation error of the autoregressive coefficients increases and the approximation in Theorem 3.4 may no longer hold. On the other hand, if p is chosen too small, the autoregressive model may fail to adequately capture the dependence structure of the stationary process $\boldsymbol{X}_t$, resulting in a large approximation error. This type of trade-off between approximation accuracy and estimation variability is well known in autoregressive modeling (see e.g. Lütkepohl (2005)). In practical applications, the order p is typically selected using standard model selection criteria for autoregressive models such as AIC or BIC.

In summary, when p is smaller than N, the one-step-ahead prediction error matrix for multivariate harmonic model $\hat{\Sigma}_{\boldsymbol{Y}}(p)$ can be decomposed into the estimated prediction error matrix for the stationary process $\hat{\Sigma}_{\boldsymbol{X}}(p)$ and the periodic component $\boldsymbol{C}\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\boldsymbol{C}^{*}$.

4 Data Analysis In this section, we apply the proposed prediction error matrix decomposition to real data. Time series analysis of temperature data has been studied in the literature; for instance Akdi and Ünlü (2021). Specifically, we utilize the Japanese daily average temperature of four Japanese cities; Tokyo, Osaka, Nagoya, and Sendai. The dataset spans two years, from January 1, 2019, to December 30, 2020. In other words, the number of observations N is 730.

Let us consider the multivariate harmonic regression model with multiple periodic components;

$$\boldsymbol{Y}_t = \sum_{k=1}^{r} \{ \boldsymbol{\alpha}_k \cos(t\theta_k) + \boldsymbol{\beta}_k \sin(t\theta_k) \} + \boldsymbol{X}_t, \quad t \in \mathbb{Z}, \tag{4.1}$$

where $\theta_k \in (0, \pi)$ $(k = 1, \ldots, r)$ and $\boldsymbol{X}_t$ is a zero-mean stationary process.

In Section 3, the one-step-ahead linear prediction error matrices $\hat{\Sigma}_{\boldsymbol{X}}(p)$ and $\hat{\Sigma}_{\boldsymbol{Y}}(p)$ are defined based on the stationary component $\boldsymbol{X}_t$. However, since the stationary component $\boldsymbol{X}_t$ is unobservable in practice, it must be estimated from the observation. For a fixed number of periodicities r, we first estimate the periodic parameters $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$, θ_k from the observed data using the estimation procedure proposed in Sagawa et al. (2026). Let us denote the estimated periodic parameter by $\hat{\boldsymbol{\alpha}}_k$, $\hat{\boldsymbol{\beta}}_k$, $\hat{\theta}_k$, respectively. The estimated $\boldsymbol{X}_t$, denoted by $\hat{\boldsymbol{X}}_t$, is represented as

$$\hat{\boldsymbol{X}}_t(r) = \boldsymbol{Y}_t - \sum_{k=1}^{r} \{ \hat{\boldsymbol{\alpha}}_k \cos(t\hat{\theta}_k) + \hat{\boldsymbol{\beta}}_k \sin(t\hat{\theta}_k) \}.$$

Using the estimated stationary component $\hat{\boldsymbol{X}}_t(r)$, we compute the one-step-ahead linear prediction error matrices. Let us define $\Psi_{\hat{\boldsymbol{X}}}(p)$ and $\Psi_{\boldsymbol{Y}}(p)$ as

$$\Psi_{\hat{\boldsymbol{X}}}(r,p) = \mathbb{E}\left[\left(\hat{\boldsymbol{X}}_t(r) - \sum_{j=1}^{p} \tilde{A}_j \hat{\boldsymbol{X}}_{t-j}(r)\right)\left(\hat{\boldsymbol{X}}_t(r) - \sum_{j=1}^{p} \tilde{A}_j \hat{\boldsymbol{X}}_{t-j}(r)\right)^{\top}\right],$$

$$\Psi_{\boldsymbol{Y}}(r,p) = \mathbb{E}\left[\left(\hat{\boldsymbol{X}}_t(r) - \sum_{j=1}^{p} \tilde{B}_j \boldsymbol{Y}_{t-j}\right)\left(\hat{\boldsymbol{X}}_t(r) - \sum_{j=1}^{p} \tilde{B}_j \boldsymbol{Y}_{t-j}\right)^{\top}\right],$$

respectively. Let $\hat{\Psi}_{\hat{\boldsymbol{X}}}(r,p)$ and $\hat{\Psi}_{\boldsymbol{Y}}(r,p)$ be the corresponding estimated prediction error matrices.

$$\Psi_{\hat{\boldsymbol{X}}}(r,p) = \frac{1}{N}\hat{\boldsymbol{X}}(r)\hat{\boldsymbol{X}}(r)^{\top} - \frac{1}{N^2}\hat{\boldsymbol{X}}(r)\hat{\boldsymbol{X}}(r,p)^{\top}\hat{\Gamma}^{-1}_{\hat{\boldsymbol{X}}(r,p)}\hat{\boldsymbol{X}}(r,p)\hat{\boldsymbol{X}}(r)^{T},$$

$$\Psi_{\boldsymbol{Y}}(r,p) = \frac{1}{N}\hat{\boldsymbol{X}}(r)\hat{\boldsymbol{X}}(r)^{\top} - \frac{1}{N^2}\hat{\boldsymbol{X}}(r)\boldsymbol{Y}(p)^{\top}\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\boldsymbol{Y}(p)\hat{\boldsymbol{X}}(r)^{T},$$

where $\hat{\boldsymbol{X}}(r) = (\hat{\boldsymbol{X}}_1(r), \ldots, \hat{\boldsymbol{X}}_N(r)) \in \mathbb{R}^{d\times N}$, $\hat{\boldsymbol{X}}(r,p) = (\hat{\boldsymbol{X}}_1(r,p), \ldots, \hat{\boldsymbol{X}}_N(r,p)) \in \mathbb{R}^{dp\times N}$, $\hat{\boldsymbol{X}}_t(r,p) = (\hat{\boldsymbol{X}}_{t-1}^{\top}(r), \ldots, \hat{\boldsymbol{X}}_{t-p}(r)^{\top})^{\top} \in \mathbb{R}^{dp\times 1}$, $\hat{\Gamma}^{-1}_{\hat{\boldsymbol{X}}(r,p)} = \left(N^{-1}\hat{\boldsymbol{X}}(r,p)\hat{\boldsymbol{X}}(r,p)^{T}\right)^{-1}$.

Additionally, in Theorem 3.4, the periodic component $\boldsymbol{C}(r)\mathcal{F}_{\boldsymbol{X}}(r)\boldsymbol{C}^*(r)$ includes the true parameters, $\boldsymbol{f}$, θ_k, $\boldsymbol{\rho}(\boldsymbol{\omega}_k)$, $(k = 1, \ldots, r)$. $\hat{A}_1, \ldots, \hat{A}_p$. Thus, we replace them with their estimator. The quantities $\boldsymbol{\rho}(\boldsymbol{\omega}_k)$ and $\boldsymbol{\omega}_k$ are obtained from the relations $\rho_k \cos(\omega_k) = \alpha_k$ and $\rho_k \sin(\omega_k) = \beta_k$. These parameters are determined from the estimated periodic coefficient parameters $\hat{\boldsymbol{\alpha}}_k$ and $\hat{\boldsymbol{\beta}}_k$. The estimators $\hat{\boldsymbol{\alpha}}_k$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\theta}_k$ are the same as those used in construction of the estimated stationary component $\hat{\boldsymbol{X}}_t(r)$. The VAR coefficient matrices $\hat{\tilde{A}}_1, \ldots, \hat{\tilde{A}}_p$ are estimated by fitting a VAR(p) model to $\hat{\boldsymbol{X}}_t(r)$ such as

$$\hat{\tilde{A}} = \hat{\boldsymbol{X}}(r)\hat{\boldsymbol{X}}(r,p)^{\top}\left(\hat{\boldsymbol{X}}(r,p)\hat{\boldsymbol{X}}(r,p)^{\top}\right)^{-1},$$

where $\hat{\tilde{A}} = (\tilde{A}_1, \ldots, \tilde{A}_p) \in \mathbb{R}^{d\times dp}$. The spectral density matrix $\boldsymbol{f}$ is estimated by

$$\hat{\boldsymbol{f}}(r,\theta) = \frac{1}{2\pi N}\sum_{h=-N+1}^{N-1}\sum_{t=1}^{N-h}(\hat{\boldsymbol{X}}_{t+h}(r) - \bar{\boldsymbol{X}}(r))(\hat{\boldsymbol{X}}_t(r) - \bar{\boldsymbol{X}}(r))^{\top}e^{\mathrm{i}h\theta},$$

where $\bar{\boldsymbol{X}}(r) = N^{-1}\sum_{t=1}^{N}\hat{\boldsymbol{X}}_t(r)$. In this data analysis, we set $1 \le p \le 5$ and $0 \le r \le 9$.

With the estimated parameters, the difference between the prediction error matrix in general setting replaced by the estimated parameters is

$$\tilde{\Delta}(r,p) := \|\hat{\Psi}_{\boldsymbol{Y}}(p) - \{\hat{\Psi}_{\hat{\boldsymbol{X}}}(p) + \hat{\boldsymbol{C}}(r)\hat{\mathcal{F}}_{\hat{\boldsymbol{X}}}(r)\hat{\boldsymbol{C}}^*(r)\}\|, \tag{4.2}$$

where $\hat{\boldsymbol{C}}(r)$ is $d \times 2r$ matrix with main column vectors

$$\left(\sum_{j=1}^{p}\hat{\tilde{A}}_j e^{\mathrm{i}\hat{\theta}_k j}\right)\hat{\boldsymbol{\rho}}_k(\hat{\boldsymbol{\omega}}_k), \quad \left(\sum_{j=1}^{p}\hat{\tilde{A}}_j e^{-\mathrm{i}\hat{\theta}_k j}\right)\hat{\boldsymbol{\rho}}_k(-\hat{\boldsymbol{\omega}}_k), \quad k = 1, \ldots, r,$$

and $\hat{\mathcal{F}}_{\hat{\boldsymbol{X}}}(r)$ is also redefined as $2r \times 2r$ diagonal matrix with main elements

$$\left(p\hat{\boldsymbol{\rho}}_k^*(\hat{\boldsymbol{\omega}}_k)\hat{\boldsymbol{f}}(\hat{\theta}_k)\hat{\boldsymbol{\rho}}_k(\hat{\boldsymbol{\omega}}_k)\right)^{-1}, \quad \left(p\hat{\boldsymbol{\rho}}_k^*(-\hat{\boldsymbol{\omega}}_k)\hat{\boldsymbol{f}}(\hat{\theta}_k)\hat{\boldsymbol{\rho}}_k(-\hat{\boldsymbol{\omega}}_k)\right)^{-1}, \quad k = 1, \ldots, r.$$

In this data analysis, let us consider the value of the difference $\tilde{\Delta}(r, p)$. For $1 \leq p \leq 5$, Figure 1 illustrates the plot of (4.2) using the daily average temperature data from four Japanese cities with $0 \leq r \leq 9$. To further assess the impact of the number of periodicities, we additionally reported the values of the difference $\tilde{\Delta}(r, p)$ for $0 \leq r \leq 9$ and $1 \leq p \leq 5$.

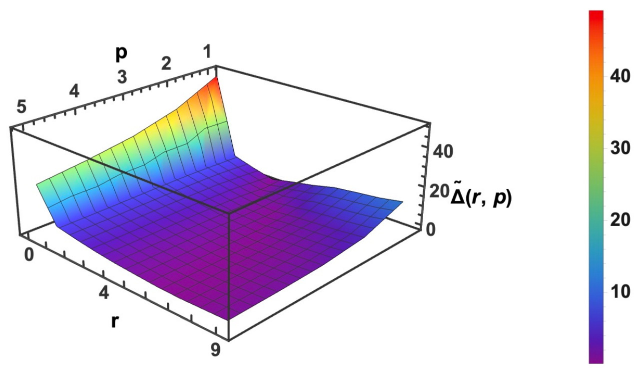

Figure 1: The plot of the difference $\tilde{\Delta}(r, p)$ in (4.2) when using temperature data of Japanese four cities and $1 \leq p \leq 5$ with $0 \leq r \leq 9$.

r \ p	1	2	3	4	5
0	49.2	37.9	31.0	26.8	23.5
1	6.45	7.08	6.93	6.45	6.26
2	3.09	4.56	4.77	4.48	4.40
3	**0.367**	2.50	3.03	2.90	2.86
4	2.78	0.960	1.67	1.70	1.69
5	4.90	**0.483**	0.808	0.940	0.971
6	7.66	1.64	**0.221**	**0.194**	0.268
7	9.05	2.44	0.688	0.220	**0.0764**
8	10.3	3.20	1.18	0.587	0.354
9	12.4	4.45	1.89	1.06	0.655

Table 1: The value of difference $\tilde{\Delta}(r, p)$ for $0 \leq r \leq 9$ and $1 \leq p \leq 5$. The bold numbers indicate the value of difference $\tilde{\Delta}(r, p)$ takes the smallest value for fixed p and $0 \leq r \leq 9$.

From Figure 1 and Table 1, it is observed that the value of the difference $\tilde{\Delta}(r, p)$ tends to decrease as both r and p increase. Furthermore, Table 1 shows that the values highlighted in bold indicate the minimum values of $\tilde{\Delta}(r, p)$ over $0 \leq r \leq 9$ for each value of p. We observe that the minimum value of $\tilde{\Delta}(r, p)$ tends to decrease as p and the number of periodicities increase. This suggests that larger values of r and p tend to produce smaller values of $\tilde{\Delta}(r, p)$. However, this does not necessarily indicate better model performance, since overly complex models may lead to overfitting. Therefore, the selection of r and p should balance goodness-of-fit and model parsimony.

The numbers of periodicities corresponding to the bold values are $r = 3$ for $p = 1$, $r = 5$ for $p = 2$, $r = 6$ for $p = 3$ and $p = 4$, and $r = 7$ for $p = 5$. For each value of r, the corresponding estimated periodicities are summarized in Table 2.

r	Detected periodicities [months]
3	12, 6.0, 3.5 (107 days)
5	12, 6.0, 3.5 (107 days), 4.0, 8.0
6	12, 6.0, 3.5 (107 days), 4.0, 8.0, 2.4 (73 days)
7	12, 6.0, 3.5 (107 days), 4.0, 8.0, 2.4 (73 days), 1.1 (33 days)

Table 2: Estimated periodicities (in months) for $r = 3, 5, 6, 7$ (1 month $\approx$ 30.44 days).

Table 2 shows that the periodic components that can be detected under each fixed number of periodicities r. This illustrates how model selection proceeds when the number of periodic components is fixed. For $r = 3$, the estimated periodicities consist of an annual cycle (12 months), a semiannual cycle (6 months), and a cycle of approximately 3.5 months. When $r = 5$, the estimated periodicities include the annual and semiannual cycle, together with cycles of approximately 3.5, 4, and 8 months. For $r = 6$, an additional periodicity of approximately 2.4 months is detected, and for $r = 7$, a further cycle of approximately 1.1 months is identified.

It is worth noting that the periodicities detected for smaller values of r are nested within those obtained for larger values of r. In particular, the periodicities detected when $r = 3$ are all included in the sets obtained for $r = 5, 6, 7$; similarly, the periodicities detected for $r = 5$ are contained in those for $r = 6, 7$, and the periodicities for $r = 6$ are contained in those for $r = 7$.

Moreover, for each value of p, the minimum value of $\tilde{\Delta}(r, p)$ decreases as p increases and as the number of periodicities becomes larger. This indicates that deeper model fitting, in terms of p, enables the detection of smaller-scale periodic components. In this data analysis, allowing larger values of r, such as $r = 6, 7$, leads to the identification of short-period components that contribute to a further reduction in prediction error.

At the same time, detecting all minor periodicities may increase the risk of overfitting. Therefore, these results suggest that period detection should be terminated before excessively small periodic components are included. From a broader perspective, our proposal may serve as a useful building block for developing an information criterion to estimate the number of periodic components in multivariate time series.

Remark 4.1. In Sagawa et al. (2026), the number of periodicities in the daily average temperature data of Kyoto, Japan, was estimated to be 3 by using the proposed information criterion for functional time series. This is consistent with the present analysis, where the minimum of $\tilde{\Delta}(r, p)$ is obtained for $r = 3$ when $p = 1$, indicating three dominant periodic components.

5 Conclusion We investigated the prediction error matrix of multivariate harmonic models under the misspecified situation. Additionally, we derived an approximate formula that decomposes the prediction error matrix into the sum of the prediction error matrix of vector autoregressive models and a matrix consisting of periodic components. In data analysis, we demonstrated the practicality of our proposal, applying it to temperature data. Moreover, the proposed framework may serve as a useful building block for developing information criteria to estimate the number of periodic components. Future work includes constructing such criteria and investigating their theoretical and empirical properties. The proofs regarding the prediction error matrix replaced by the estimated parameters and in general setting are available upon request to the author.

Acknowledgments I would like to express my deepest gratitude to Professor Yan Liu for his insightful advice and support throughout this research. This work was supported by JST SPRING, Grant Number JPMJSP2128, Waseda Research Institute for Science and Engineering, Grant-in-Aid for Young Scientists (Early Bird), and JEES-Mitsubishi Corporation Science and Technology Scholarship for Students, Scholarship Number MITSUSCI2508.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Akdi, Y., Ünlü, K.D., 2021. Periodicity in precipitation and temperature for monthly data of turkey. Theoretical and Applied Climatology 143, 957–968.

Bhansali, R., 1981. Effects of not knowing the order of an autoregressive process on the mean squared error of prediction—I. Journal of the American Statistical Association 76, 588–597.

Brillinger, D.R., 2001. Time Series: Data Analysis and Theory. SIAM.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer Science & Business Media.

Choi, I.B., Taniguchi, M., 2001. Misspecified prediction for time series. Journal of Forecasting 20, 543–564.

Hannan, E., Mackisack, M., 1986. A law of the iterated logarithm for an estimate of frequency. Stochastic Processes and their Applications 22, 103–109.

Hannan, E., Wahlberg, B., 1989. Convergence rates for inverse toeplitz matrix forms. Journal of Multivariate Analysis 31, 127–135.

Hannan, E.J., 1970. Multiple Time Series. John Wiley & Sons.

Hannan, E.J., 1973. The estimation of frequency. Journal of Applied Probability 10, 510–519.

Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological) 41, 190–195.

Hosoya, Y., Taniguchi, M., 1982. A central limit theorem for stationary processes and the parameter estimation of linear processes. The Annals of Statistics , 132–153.

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

Konishi, S., Kitagawa, G., 1996. Generalised information criteria in model selection. Biometrika 83, 875–890.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer Science & Business Media.

Priestley, M.B., 1988. The spectral analysis of time series.

Quinn, B., Thomson, P., 1991. Estimating the frequency of a periodic function. Biometrika 78, 65–74.

Sagawa, R., Liu, Y., Patilea, V., 2026. An information criterion for detecting periodicities in functional time series. arXiv preprint arXiv:2603.21161 .

Schorfheide, F., 2005. VAR forecasting under misspecification. Journal of Econometrics 128, 99–136.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics , 461–464.

Shumway, R.H., Stoffer, D.S., 2006. Time series analysis and its applications: with R examples. Springer.

Walker, A., 1971. On the estimation of a harmonic component in a time series with stationary independent residuals. Biometrika 58, 21–36.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica: Journal of the econometric society , 1–25.

Yajima, Y., Matsuda, Y., 2009. On nonparametric and semiparametric testing for multivariate linear time series. The Annals of Statistics 37, 3529–3554.

Yamamoto, T., 1976. Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. Journal of the Royal Statistical Society Series C: Applied Statistics 25, 123–127.

Appendix A. Proof of Proposition 2.2

Proof of Proposition 2.2. First, we estimate the coefficient matrix A of (2.2). By utilizing vec operator, (2.2) is written as

$$\boldsymbol{x} = (\boldsymbol{X}(p)^\top \otimes \boldsymbol{I}_d)\boldsymbol{a} + \boldsymbol{u},$$

where $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$, $\boldsymbol{a} = \mathrm{vec}(A)$, $\boldsymbol{u} = \mathrm{vec}(\boldsymbol{U})$, and $\boldsymbol{I}_d$ is $d \times d$ identity matrix. By least squares method, the estimated coefficient vector $\boldsymbol{a}$, denoted by $\hat{\boldsymbol{a}}$, is

$$\begin{aligned}\hat{\boldsymbol{a}} &= \big((\boldsymbol{X}(p)\boldsymbol{X}(p)^\top)^{-1}\boldsymbol{X}(p) \otimes \boldsymbol{I}_d\big)\boldsymbol{x} \\ &= \mathrm{vec}(\boldsymbol{X}\boldsymbol{X}(p)^\top\big(\boldsymbol{X}(p)\boldsymbol{X}(p)^\top\big)^{-1})\end{aligned}$$

The estimation of A, denoted by $\hat{A}$, is

$$\hat{A} = \boldsymbol{X}\boldsymbol{X}(p)^\top\big(\boldsymbol{X}(p)\boldsymbol{X}(p)^\top\big)^{-1}, \tag{A.1}$$

(e.g. Lütkepohl, 2005, p.72). Therefore, we obtain

$$\begin{aligned}\hat{\Sigma}_{\boldsymbol{X}}(p) &= \frac{1}{N}(\boldsymbol{X} - \hat{A}\boldsymbol{X}(p))(\boldsymbol{X} - \hat{A}\boldsymbol{X}(p))^\top \\ &= \frac{1}{N}\boldsymbol{X}\big(\boldsymbol{I}_N - \boldsymbol{X}(p)^\top(\boldsymbol{X}(p)\boldsymbol{X}(p)^\top)^{-1}\boldsymbol{X}(p)\big)\boldsymbol{X}^\top \\ &= \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top - \frac{1}{N^2}\boldsymbol{X}\boldsymbol{X}(p)^\top\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top,\end{aligned}$$

which completes the proof of Proposition 2.2. □

Appendix B. Proof of Proposition 2.3

Proof of Proposition 2.3. Let us consider the estimation of the coefficient matrix B of (2.5). By utilizing vec operator, (2.5) is written as

$$\boldsymbol{x} = (\boldsymbol{Y}(p)^\top \otimes \boldsymbol{I}_d)\boldsymbol{b} + \boldsymbol{v},$$

where $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$, $\boldsymbol{b} = \mathrm{vec}(B)$, $\boldsymbol{v} = \mathrm{vec}(\boldsymbol{V})$, and $\boldsymbol{I}_d$ is $d \times d$ identity matrix. By least squares method, the estimated coefficient vector $\boldsymbol{b}$, denoted by $\hat{\boldsymbol{b}}$, is

$$\begin{aligned}\hat{\boldsymbol{b}} &= \big((\boldsymbol{Y}(p)\boldsymbol{Y}(p)^\top)^{-1}\boldsymbol{Y}(p) \otimes \boldsymbol{I}_d\big)\boldsymbol{x} \\ &= \mathrm{vec}(\boldsymbol{X}\boldsymbol{Y}(p)^\top\big(\boldsymbol{Y}(p)\boldsymbol{Y}(p)^\top\big)^{-1}).\end{aligned}$$

The estimation of B in (2.5), denoted by $\hat{B}$, is

$$\hat{B} = \boldsymbol{X}\boldsymbol{Y}(p)^\top \left(\boldsymbol{Y}(p)\boldsymbol{Y}(p)^\top\right)^{-1}.$$

(e.g. Lütkepohl, 2005, p.72). Thus,

$$\begin{aligned}\hat{\Sigma}_{\boldsymbol{Y}}(p) &= \frac{1}{N}(\boldsymbol{X} - \hat{B}\boldsymbol{Y}(p))(\boldsymbol{X} - \hat{B}\boldsymbol{Y}(p))^\top \\ &= \frac{1}{N}\boldsymbol{X}\left(\boldsymbol{I}_N - \boldsymbol{Y}(p)^\top(\boldsymbol{Y}(p)\boldsymbol{Y}(p)^\top)^{-1}\boldsymbol{Y}(p)\right)\boldsymbol{X}^\top \\ &= \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top - \frac{1}{N^2}\boldsymbol{X}\boldsymbol{Y}(p)^\top\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\boldsymbol{Y}(p)\boldsymbol{X}^\top,\end{aligned}$$

which completes the proof of Proposition 2.3. □

Appendix C. Proof of Lemma 3.2 Before providing the proof of Lemma 3.2, we introduce the following lemmas.

Lemma C.1 (Brillinger (2001), Theorem 4.5.1)**.** Under Assumption 3.1, it holds that

$$\sup_{\lambda}\left\|\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{X}_t\exp(\mathrm{i}t\lambda)\right\| = O_p\left(\left(\frac{\log N}{N}\right)^{1/2}\right).$$

Lemma C.2. Under Assupmtion 3.1, for any ℓ $(0 \le \ell \le p-1)$, the $d \times d$ matrix $N^{-1}\sum_{t=1}^{N}\boldsymbol{Y}_t\boldsymbol{Y}_{t-\ell}^\top$ is formulated by

$$\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{Y}_t\boldsymbol{Y}_{t-\ell}^\top = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{X}_t\boldsymbol{X}_{t-\ell}^\top + \boldsymbol{D}(\ell) + O\left(\left(\frac{\log N}{N}\right)^{1/2}\right),$$

where

$$\boldsymbol{D}(\ell) = \left(\frac{\rho_i\rho_j}{2}\sum_{t=1}^{N}\cos\left(\ell\theta + (\omega_i - \omega_j)\right)\right)_{i,j=1,\dots,d},$$

$\rho_i = (\alpha_i^2 + \beta_i^2)^{1/2}$ and ω_i satisfies

$$\cos(\omega_i) = \frac{\alpha_i}{\rho_i}, \qquad \sin(\omega_i) = \frac{\beta_i}{\rho_i}.$$

Proof of Lemma C.2. For any ℓ $(0 \le \ell \le p-1)$, $N^{-1}\sum_{t=1}^{N}\boldsymbol{Y}_t\boldsymbol{Y}_{t-\ell}^\top$ is expressed as

$$\begin{aligned}&\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{Y}_t\boldsymbol{Y}_{t-\ell}^\top \\ &= \frac{1}{N}\sum_{t=1}^{N}\left\{\left(\boldsymbol{\alpha}\cos(t\theta) + \boldsymbol{\beta}\sin(t\theta)\right) + \boldsymbol{X}_t\right\}\left\{\left(\boldsymbol{\alpha}\cos((t-\ell)\theta) + \boldsymbol{\beta}\sin((t-\ell)\theta)\right) + \boldsymbol{X}_{t-\ell}\right\}^\top \\ &= \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{X}_t\boldsymbol{X}_{t-\ell}^\top + \text{(a)} + \text{(b)} + \text{(c)},\end{aligned}$$

where

$$(\mathrm{a}) := \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{X}_t \left\{\boldsymbol{\alpha}\cos((t-\ell)\theta) + \boldsymbol{\beta}\sin((t-\ell)\theta)\right\}^\top,$$

$$(\mathrm{b}) := \frac{1}{N}\sum_{t=1}^{N} \left\{\boldsymbol{\alpha}\cos(t\theta) + \boldsymbol{\beta}\sin(t\theta)\right\} \boldsymbol{X}_{t-\ell}^\top,$$

$$(\mathrm{c}) := \frac{1}{N}\sum_{t=1}^{N} \left\{\boldsymbol{\alpha}\cos(t\theta) + \boldsymbol{\beta}\sin(t\theta)\right\}\left\{\boldsymbol{\alpha}\cos((t-\ell)\theta) + \boldsymbol{\beta}\sin((t-\ell)\theta)\right\}^\top.$$

Now, we evaluate the terms (a)–(c) as $N \to \infty$.

(I) The evaluation of (a)
(a) is $d \times d$ matrix and , for $1 \le i, j \le d$, its arbitrary (i, j)element is

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N} X_{ti}\left\{\alpha_j\cos((t-\ell)\theta) + \beta_j\sin((t-\ell)\theta)\right\} \\
&\quad = \frac{1}{N}\left\{\sum_{t=1}^{N} X_{ti}\frac{\alpha_j - \beta_j\mathrm{i}}{2}e^{\mathrm{i}(t-\ell)\theta} + \sum_{t=1}^{N} X_{ti}\frac{\alpha_j + \beta_j\mathrm{i}}{2}e^{-\mathrm{i}(t-\ell)\theta}\right\}.
\end{aligned}$$

By Lemma C.1,

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{t=1}^{N} X_{ti}\left\{\alpha_j\cos((t-\ell)\theta) + \beta_j\sin((t-\ell)\theta)\right\}\right| \\
&\le \left|\frac{1}{N}\sum_{t=1}^{N} X_{ti}\frac{\alpha_j - \beta_j\mathrm{i}}{2}e^{\mathrm{i}(t-\ell)\theta}\right| + \left|\frac{1}{N}\sum_{t=1}^{N} X_{ti}\frac{\alpha_j + \beta_j\mathrm{i}}{2}e^{-\mathrm{i}(t-\ell)\theta}\right| \\
&= \frac{\alpha_j^2+\beta_j^2}{2}\left|\frac{1}{N}\sum_{t=1}^{N} X_{ti}e^{\mathrm{i}(t-\ell)\theta}\right| + \frac{\alpha_j^2+\beta_j^2}{2}\left|\frac{1}{N}\sum_{t=1}^{N} X_{ti}e^{-\mathrm{i}(t-\ell)\theta}\right| \\
&= O_p\left(\left(\frac{\log N}{N}\right)^{1/2}\right).
\end{aligned}$$

Therefore,

$$(\mathrm{a}) = O_p\left(\left(\frac{\log N}{N}\right)^{1/2}\right). \tag{C.1}$$

(II) The evaluation of (b)
Similarly to (a), (b) is evaluated by

$$(\mathrm{b}) = O_p\left(\left(\frac{\log N}{N}\right)^{1/2}\right). \tag{C.2}$$

(III) The evaluation of (c)
For $1 \le i, j \le d$, (i, j)- element of (c) is expressed as

$$\frac{1}{N}\sum_{t=1}^{N} \left\{\alpha_i\cos(t\theta) + \beta_i\sin(t\theta)\right\}\left\{\alpha_j\cos((t-\ell)\theta) + \beta_j\sin((t-\ell)\theta)\right\}.$$

Let us recall

$$\rho_j = (\alpha_j^2 + \beta_j^2)^{1/2},$$

and ω_j satisfies

$$\cos(\omega_j) = \frac{\alpha_j}{\rho_j}, \qquad \sin(\omega_j) = \frac{\beta_j}{\rho_j},$$

and $\omega_i = \omega_j$ for $i = j$; $\omega_i \neq \omega_j$ for $i \neq j$. By utilizing the above notations, (i, j)- element of (c) is

$$\begin{aligned}&\frac{1}{N}\sum_{t=1}^{N}\{\rho_i \cos(t\theta + \omega_i)\}\left\{\rho_j \cos((t-\ell)\theta + \omega_j)\right\}\\&= \rho_i\rho_j \frac{1}{N}\sum_{t=1}^{N} \cos(t\theta + \omega_i)\cos((t-\ell)\theta + \omega_j).\end{aligned}$$

Here, we focus on

$$\frac{1}{N}\sum_{t=1}^{N}\cos(t\theta + \omega_j)\cos((t-\ell)\theta + \omega_j) = (A) + (B) + (C) + (D),$$

where

$$\begin{aligned}(A) &:= \frac{1}{N}\sum_{t=1}^{N}\cos(t\theta)\cos(\omega_i)\cos((t-\ell)\theta)\cos(\omega_j),\\(B) &:= -\frac{1}{N}\sum_{t=1}^{N}\cos(t\theta)\cos(\omega_i)\sin((t-\ell)\theta)\sin(\omega_j),\\(C) &:= -\frac{1}{N}\sum_{t=1}^{N}\sin(t\theta)\sin(\omega_i)\cos((t-\ell)\theta)\cos(\omega_j),\\(D) &:= \frac{1}{N}\sum_{t=1}^{N}\sin(t\theta)\sin(\omega_i)\sin((t-\ell)\theta)\sin(\omega_j).\end{aligned}$$

To evaluate each term (A)–(D), we introduce the following three formulas:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\cos(t\theta)\cos((t-\ell)\theta) = \frac{1}{2}\cos(\ell\theta), \qquad 0 < \theta < \pi, \tag{C.3}$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\cos(t\theta)\sin(t\theta) = 0, \qquad 0 < \theta < \pi, \tag{C.4}$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sin(t\theta)\sin(t\theta) = \frac{1}{2}, \qquad 0 < \theta < \pi. \tag{C.5}$$

(i) The evaluation of (A)
By (C.3), as $N \to \infty$,

$$\lim_{N\to\infty} (A) = \frac{1}{2} \cos(\omega_i) \cos(\omega_j) \cos(\ell\theta), \qquad 0 < \theta < \pi.$$

(ii) The evaluation of (B)
(B) is expressed as

$$-\frac{1}{N} \cos(\omega_i) \sin(\omega_j) \sum_{t=1}^{N} \cos(t\theta) \sin((t-\ell)\theta)$$
$$= -\cos(\omega_i) \sin(\omega_j) \frac{1}{N} \sum_{t=1}^{N} \cos(t\theta) \big(\sin(t\theta) \cos(\ell\theta) - \cos(t\theta) \sin(\ell\theta) \big).$$

By (C.3) and (C.4), as $N \to \infty$, we obtain

$$\lim_{N\to\infty} (B) = \frac{1}{2} \cos(\omega_i) \sin(\omega_j) \sin(\ell\theta), \qquad 0 < \theta < \pi.$$

(iii) The evaluation of (C)
(C) is expressed as

$$-\frac{1}{N} \sin(\omega_i) \cos(\omega_j) \sum_{t=1}^{N} \sin(t\theta) \cos((t-\ell)\theta)$$
$$= -\sin(\omega_i) \cos(\omega_j) \frac{1}{N} \sum_{t=1}^{N} \sin(t\theta) \big(\cos(t\theta) \cos(\ell\theta) + \sin(t\theta) \sin(\ell\theta) \big).$$

By (C.4) and (C.5), as $N \to \infty$, we obtain

$$\lim_{N\to\infty} (C) = -\frac{1}{2} \sin(\omega_i) \cos(\omega_j) \sin(\ell\theta), \qquad 0 < \theta < \pi.$$

(iv) The evaluation of (D)
(D) is expressed by

$$\frac{1}{N} \sin(\omega_i) \sin(\omega_j) \sum_{t=1}^{N} \sin(t\theta) \sin((t-\ell)\theta)$$
$$= \sin(\omega_i) \sin(\omega_j) \frac{1}{N} \sum_{t=1}^{N} \sin(t\theta) \big(\sin(t\theta) \cos(\ell\theta) - \cos(t\theta) \sin(\ell\theta) \big).$$

By (C.4) and (C.5), as $N \to \infty$, we obtain

$$\lim_{N\to\infty} (D) = \frac{1}{2} \sin(\omega_i) \sin(\omega_j) \cos(\ell\theta), \qquad 0 < \theta < \pi.$$

Accordingly, as $N \to \infty$, we obtain

$$(A) + (D) = \frac{1}{2} \cos(\ell\theta) \cos(\omega_i - \omega_j),$$

$$(B) + (C) = -\frac{1}{2}\sin(\ell\theta)\sin(\omega_i - \omega_j),$$

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N}\cos(t\theta + \omega_j)\cos((t-\ell)\theta + \omega_j) \\
&= \rho_i\rho_j\{(A) + (B) + (C) + (D)\} \\
&= \frac{\rho_i\rho_j}{2}\left\{\cos(\ell\theta)\cos(\omega_i - \omega_j) - \sin(\ell\theta)\sin(\omega_i - \omega_j)\}\right) \\
&= \frac{\rho_i\rho_j}{2}\cos\bigl(\ell\theta + (\omega_i - \omega_j)\bigr).
\end{aligned}$$

Therefore,

$$\text{(c)} = \left(\frac{\rho_i\rho_j}{2}\cos\bigl(\ell\theta + (\omega_i - \omega_j)\bigr)\right)_{i,j=1,\dots,d} := \boldsymbol{D}(\ell), \tag{C.6}$$

where $\omega_i = \omega_j$ for $i = j$; $\omega_i \neq \omega_j$ for $i \neq j$.

In summary, from (C.1), (C.2) and (C.6), we obtain

$$\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{Y}_t\boldsymbol{Y}_{t-\ell}^\top = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{X}_t\boldsymbol{X}_{t-\ell}^\top + \boldsymbol{D}(\ell) + O_p\left(\left(\frac{\log N}{N}\right)^{1/2}\right),$$

which completes the proof of Lemma C.2. □

Now, we provide the proof of Lemma 3.2. We state the outline of the proof. The proof is divided into 2 steps. First, we evaluate $\hat{\Gamma}_{\boldsymbol{Y}(p)}$ by utilizing $\hat{\Gamma}_{\boldsymbol{X}(p)}$. Secondly, we consider $\hat{\Gamma}_{\boldsymbol{Y}(p)}^{-1}$.

Proof of Lemma 3.2. First, let us consider the evaluation of $\hat{\Gamma}_{\boldsymbol{Y}(p)}$ with $\hat{\Gamma}_{\boldsymbol{X}(p)}$. $\hat{\Gamma}_{\boldsymbol{Y}(p)}$ is $p \times p$ block matrix such as

$$\begin{aligned}
\hat{\Gamma}_{\boldsymbol{Y}(p)} &= \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{Y}_t(p)\boldsymbol{Y}(p)^\top \\
&= \frac{1}{N}\sum_{t=1}^{N}\begin{pmatrix} \boldsymbol{Y}_t\boldsymbol{Y}_t^\top & \boldsymbol{Y}_t\boldsymbol{Y}_{t-1}^\top & \cdots & \boldsymbol{Y}_t\boldsymbol{Y}_{t-p+1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Y}_{t-p+1}\boldsymbol{Y}_t^\top & \boldsymbol{Y}_{t-p+1}\boldsymbol{Y}_{t-1}^\top & \cdots & \boldsymbol{Y}_{t-p+1}\boldsymbol{Y}_{t-p+1}^\top \end{pmatrix}.
\end{aligned}$$

Let us recall that

$$\boldsymbol{J} = \bigl(\boldsymbol{\xi}(\theta)\otimes\boldsymbol{\rho}(\boldsymbol{\omega}), \boldsymbol{\xi}(-\theta)\otimes\boldsymbol{\rho}(-\boldsymbol{\omega})\bigr) \in \mathbb{C}^{dp\times 2},$$

$$\boldsymbol{\xi}(\theta) = (e^{\mathrm{i}\theta},\dots,e^{\mathrm{i}\theta p})^\top \in \mathbb{C}^{p\times 1}, \qquad \boldsymbol{\rho}(\boldsymbol{\omega}) = \left(\frac{\rho_1}{2}e^{-\mathrm{i}\omega_1},\dots,\frac{\rho_d}{2}e^{-\mathrm{i}\omega_d}\right)^\top \in \mathbb{C}^{d\times 1}.$$

By using $\boldsymbol{J}$ and Lemma C.2, we obtain

$$\hat{\Gamma}_{\boldsymbol{Y}(p)} = \hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^* + O_p\left(p\left(\frac{\log N}{N}\right)^{1/2}\right). \tag{C.7}$$

Remark C.3. $\boldsymbol{J}\boldsymbol{J}^*$ is $dp \times dp$ matrix such as

$$\boldsymbol{J}\boldsymbol{J}^* = \begin{pmatrix} \boldsymbol{D}(0) & \boldsymbol{D}(1) & \cdots & \boldsymbol{D}(p-1) \\ \boldsymbol{D}(-1) & \boldsymbol{D}(0) & \cdots & \boldsymbol{D}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{D}(-p+1) & \boldsymbol{D}(-p+2) & \cdots & \boldsymbol{D}(0) \end{pmatrix} \in \mathbb{C}^{dp\times dp}.$$

Now, we condider $\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}$. From (C.7), we obtain

$$\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)} - (\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)^{-1} = \hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\{\hat{\Gamma}_{\boldsymbol{Y}(p)} - (\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)\}(\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)^{-1}.$$

Accordingly,

$$\begin{aligned}\|\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)} - (\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)^{-1}\| &\le \|\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\|\|\hat{\Gamma}_{\boldsymbol{Y}(p)} - (\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)\|\|(\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)^{-1}\| \\ &= \|\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\|\left\|O_p\left(p\Big(\frac{\log N}{N}\Big)^{1/2}\right)\right\|\|(\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)^{-1}\| \\ &= O_p\left(p^3\Big(\frac{\log N}{N}\Big)^{1/2}\right).\end{aligned}$$

Thus, we obtain

$$\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)} = (\hat{\Gamma}_{\boldsymbol{X}(p)} + \boldsymbol{J}\boldsymbol{J}^*)^{-1} + O_p\left(p^3\Big(\frac{\log N}{N}\Big)^{1/2}\right).$$

By Matrix Inverse Lemma, we obtain

$$\hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)} = \hat{\Gamma}^{-1}_{\boldsymbol{X}(p)} - \hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}(\boldsymbol{I}_2 + \boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)} + O_p\left(p^3\Big(\frac{\log N}{N}\Big)^{1/2}\right),$$

which completes the proof of Lemma 3.2. □

Appendix D. Proof of Theorem 3.4 Before providing the proof of Theorem 3.4, we introduce the following lemma.

Lemma D.1. Under Assuption 3.1, we obtain

$$\hat{\Sigma}_{\boldsymbol{Y}}(p) = \hat{\Sigma}_{\boldsymbol{X}}(p) + T(p) + O_p\left(p^2\frac{\log N}{N}\right),$$

where

$$T(p) = \frac{1}{N^2}\boldsymbol{X}\boldsymbol{X}(p)^\top\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}(\boldsymbol{I}_2 + \boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top.$$

Proof of Lemma D.1. First, we formulate $\hat{\Sigma}_{\boldsymbol{Y}}(p)$ in Proposition 2.3 with $\hat{\Sigma}_{\boldsymbol{X}}(p)$. Let us recall that the periodic component in (2.3) is expressed as

$$\boldsymbol{\alpha}\cos(t\theta) + \boldsymbol{\beta}\sin(t\theta) = \boldsymbol{Q}_t(\theta)\boldsymbol{\psi},$$

where $\boldsymbol{I}_d$ is $d \times d$ identity matrix, and

$$\begin{aligned}\boldsymbol{\psi} &= (\boldsymbol{\alpha}, \boldsymbol{\beta})^\top \in \mathbb{R}^{2d\times 1}, \\ \boldsymbol{Q}_t(\theta) &= (\boldsymbol{q}_t(\theta)^\top \otimes \boldsymbol{I}_d) \in \mathbb{R}^{d\times 2d}, \\ \boldsymbol{q}_t(\theta) &= (\cos(t\theta), \sin(t\theta))^\top \in \mathbb{R}^{2\times 1}.\end{aligned}$$

Thus, $\boldsymbol{Y}(p)$ is expressed as

$$\begin{aligned}\boldsymbol{Y}(p) &= \begin{pmatrix}\boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \cdots & \boldsymbol{Y}_N \\ \boldsymbol{Y}_0 & \boldsymbol{Y}_1 & \cdots & \boldsymbol{Y}_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Y}_{2-p} & \boldsymbol{Y}_{3-p} & \cdots & \boldsymbol{Y}_{N-p+1}\end{pmatrix} \\ &= \begin{pmatrix}\boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_N \\ \boldsymbol{X}_0 & \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_{2-p} & \boldsymbol{X}_{3-p} & \cdots & \boldsymbol{X}_{N-p+1}\end{pmatrix} + \begin{pmatrix}\boldsymbol{Q}_1(\theta)\boldsymbol{\psi} & \boldsymbol{Q}_2(\theta)\boldsymbol{\psi} & \cdots & \boldsymbol{Q}_N(\theta)\boldsymbol{\psi} \\ \boldsymbol{Q}_0(\theta)\boldsymbol{\psi} & \boldsymbol{Q}_1(\theta)\boldsymbol{\psi} & \cdots & \boldsymbol{Q}_{N-1}(\theta)\boldsymbol{\psi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Q}_{2-p}(\theta)\boldsymbol{\psi} & \boldsymbol{Q}_{3-p}(\theta)\boldsymbol{\psi} & \cdots & \boldsymbol{Q}_{N-p+1}(\theta)\boldsymbol{\psi}\end{pmatrix} \\ &:= \boldsymbol{X}(p) + \boldsymbol{Q}(p).\end{aligned}$$

Accordingly,

$$\begin{aligned}\hat{\Sigma}_{\boldsymbol{Y}}(p) &= \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top - \frac{1}{N^2}\boldsymbol{X}(\boldsymbol{X}(p)+\boldsymbol{Q}(p))^\top \hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}(\boldsymbol{X}(p)+\boldsymbol{Q}(p))\boldsymbol{X}^\top \\ &= \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top - \frac{1}{N^2}\boldsymbol{X}\boldsymbol{X}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top + (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}),\end{aligned}$$

where

$$\begin{aligned}(\mathrm{I}) &:= -\frac{1}{N^2}\boldsymbol{X}\boldsymbol{X}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\boldsymbol{Q}(p)\boldsymbol{X}^\top, \\ (\mathrm{II}) &:= -\frac{1}{N^2}\boldsymbol{X}\boldsymbol{Q}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top, \\ (\mathrm{III}) &:= \frac{1}{N^2}\boldsymbol{X}\boldsymbol{Q}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\boldsymbol{Q}(p)\boldsymbol{X}^\top.\end{aligned}$$

By Lemma C.1, we obtain

$$(\mathrm{I}) = O_p\left(p^2\frac{(\log N)^{1/2}}{N}\right), \quad (\mathrm{II}) = O_p\left(p^2\frac{(\log N)^{1/2}}{N}\right), \quad (\mathrm{III}) = O_p\left(p^2\frac{\log N}{N}\right).$$

Therefore,

$$\hat{\Sigma}_{\boldsymbol{Y}}(p) = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top - \frac{1}{N^2}\boldsymbol{X}\boldsymbol{X}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{Y}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top + O_p\left(p^2\frac{\log N}{N}\right).$$

From Lemma 3.2, $\hat{\Sigma}_{\boldsymbol{Y}}(p)$ is

$$\hat{\Sigma}_{\boldsymbol{Y}}(p) = \hat{\Sigma}_{\boldsymbol{X}}(p) + T(p) + O_p\left(p^2\frac{\log N}{N}\right),$$

where

$$T(p) = \frac{1}{N^2}\boldsymbol{X}\boldsymbol{X}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}(\boldsymbol{I}_2 + \boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top,$$

which completes the proof of Lemma D.1. □

Now, let us provide the proof of Theorem 3.4.

Proof of Theorem 3.4. Let us divide $T(p)$ into 3 terms such as

$$\frac{1}{N}\boldsymbol{X}\boldsymbol{X}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}, \quad (\boldsymbol{I}_2 + \boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}, \quad \frac{1}{N}\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top.$$

Now, we evaluate 3 terms, respectively.

(i) The evaluation of $N^{-1}\boldsymbol{X}\boldsymbol{X}(p)^\top \hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}$
By (A.1), we obtain

$$\begin{aligned}\frac{1}{N}\boldsymbol{X}\boldsymbol{X}(p)\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)} &= \boldsymbol{X}\boldsymbol{X}(p)^\top(\boldsymbol{X}(p)\boldsymbol{X}(p)^\top)^{-1} \\ &= \hat{A} \in \mathbb{R}^{d\times dp}.\end{aligned}$$

Accordingly,

$$\begin{aligned}\frac{1}{N}\boldsymbol{X}\boldsymbol{X}(p)\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J} &= \hat{A}\boldsymbol{J}\\ &= \hat{A}\big(\boldsymbol{\xi}(\theta)\otimes\boldsymbol{\rho}(\boldsymbol{\omega}),\boldsymbol{\xi}(-\theta)\otimes\boldsymbol{\rho}(-\boldsymbol{\omega})\big)\\ &= \left(\Big(\sum_{j=1}^{p}\hat{A}_j e^{\mathrm{i}\theta j}\Big)\boldsymbol{\rho}(\boldsymbol{\omega}),\Big(\sum_{j=1}^{p}\hat{A}_j e^{-\mathrm{i}\theta j}\Big)\boldsymbol{\rho}(-\boldsymbol{\omega})\right)\\ &:= \boldsymbol{C}.\end{aligned}$$

(ii) The evaluation of $(\boldsymbol{I}_2+\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}$

We consider $\boldsymbol{I}_2+\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}$. Let $W(\theta)$ be

$$W(\theta)=(e^{-\mathrm{i}\theta}\boldsymbol{I}_d,\ldots,e^{-\mathrm{i}p\theta}\boldsymbol{I}_d)^*\in\mathbb{C}^{dp\times d},$$

where $\boldsymbol{I}_d$ is $d\times d$ identity matrix. Any $dp\times 1$ block elements of $\boldsymbol{J}$ is

$$\boldsymbol{\xi}(\theta)\otimes\boldsymbol{\rho}(\boldsymbol{\omega})=W(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}).$$

From Theorem 1.2 in Hannan and Wahlberg (1989), we obtain

$$\frac{1}{p}(W(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}(W(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))=\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega})+o(1),$$

$$(W(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}(W(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))=p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))+o(p),$$

where $\boldsymbol{f}$ is the spectral density matrix of $\boldsymbol{X}_t$ at the frequency θ. $\boldsymbol{I}_2+\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J}$ is 2×2 diagonal matrix such as

$$\begin{pmatrix}1+p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega})) & 0\\ 0 & 1+p(\boldsymbol{\rho}(-\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(-\theta)\boldsymbol{\rho}(-\boldsymbol{\omega}))\end{pmatrix}.$$

Thus, $(\boldsymbol{I}_2+\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}$ is

$$\begin{pmatrix}\big(1+p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))\big)^{-1} & 0\\ 0 & \big(1+p(\boldsymbol{\rho}(-\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(-\theta)\boldsymbol{\rho}(-\boldsymbol{\omega}))\big)^{-1}\end{pmatrix}.$$

The (1,1)-element of $(\boldsymbol{I}_2+\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}$ is

$$\begin{aligned}&\Big(1+p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))\Big)^{-1}\\ &=\left\{\Big(p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))\Big)\left(1+\frac{1}{p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))}\right)\right\}^{-1}\\ &=\Big(p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))\Big)^{-1}\left(1+\frac{1}{p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))}\right)^{-1}\\ &=\Big(p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))\Big)^{-1}+o(p^{-2}).\end{aligned}$$

Therefore, we obtain

$$(\boldsymbol{I}_2+\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{J})^{-1}=\mathcal{F}_{\boldsymbol{X}}+o(p^{-2}),$$

where $\mathcal{F}_{\boldsymbol{X}}$ is 2×2 diagonal matrix such as

$$\begin{pmatrix}\big(p(\boldsymbol{\rho}(\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(\boldsymbol{\omega}))\big)^{-1} & 0\\ 0 & \big(p(\boldsymbol{\rho}(-\boldsymbol{\omega})^*\boldsymbol{f}^{-1}(\theta)\boldsymbol{\rho}(-\boldsymbol{\omega}))\big)^{-1}\end{pmatrix}.$$

(iii) The evaluation of $N^{-1}\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{X}(p)\boldsymbol{X}^\top$
Similarly to (i),

$$\frac{1}{N}\boldsymbol{J}^*\hat{\Gamma}^{-1}_{\boldsymbol{X}(p)}\boldsymbol{X}(p)\boldsymbol{X} = \left(\Big(\sum_{j=1}^{p}\hat{A}_j e^{-\mathrm{i}\theta j}\Big)\boldsymbol{\rho}(-\boldsymbol{\omega}), \Big(\sum_{j=1}^{p}\hat{A}_j e^{\mathrm{i}\theta j}\Big)\boldsymbol{\rho}(\boldsymbol{\omega})\right)^\top.$$
$$= \boldsymbol{C}^*.$$

In summary, we obtain

$$\hat{\Sigma}_{\boldsymbol{Y}}(p) = \hat{\Sigma}_{\boldsymbol{X}}(p) + \boldsymbol{C}\mathcal{F}_{\boldsymbol{X}}\boldsymbol{C}^* + O_p\left(p^2\frac{\log N}{N}\right),$$

which completes the proof of Theorem 3.4. □

Communicated by *Masanobu Taniguchi*

Rinka Sagawa
Department of Applied Mathematics,Waseda University
3-4-1, Okubo, Sshinjuku-ku, 169-8555, Tokyo, Japan
cherry.kyubin3@akane.waseda.jp