Regularized Estimation of Sinusoidal Frequencies

Arindam Panja*[1] and Debasis Kundu[2]

[1]Department of Mathematics, National Institute of Technology Durgapur, Durgapur, West Bengal–713209, India.

[2]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, Kanpur, Uttar Pradesh, India

Abstract

This article studies ridge-type regularized estimation for single and multiple sinusoidal regression models with unknown amplitudes and frequencies. Since frequency estimation is nonlinear and least squares estimators can be unstable under outlier contamination or finite-sample near-collinearity, ridge regularization is used as a shrinkage-based stabilization method for the amplitude parameters. The proposed estimator is computationally convenient and is obtained through a profile optimization procedure over the frequencies. We establish consistency and asymptotic normality for fixed-dimensional sinusoidal models under suitable assumptions on the error process and the tuning parameter. Simulation studies under random and block contamination show that ridge regularization often reduces mean squared error and stabilizes estimation in the considered settings. Comparisons with least squares, least absolute deviation, Huber, and Huber–ridge estimators are also provided. A real-data analysis using the AirPassengers series illustrates the practical usefulness of the proposed method for stable seasonal sinusoidal fitting under controlled contamination.

Keywords: Sinusoidal regression, ridge regularization, frequency estimation, shrinkage estimation, outlier contamination, consistency, asymptotic normality.

1 Introduction

In this paper, we consider the multiple sinusoidal model with additive noise,

$$z(t) = \sum_{k=1}^{p} A_k^0 \cos(\omega_k^0 t) + B_k^0 \sin(\omega_k^0 t) + e(t), \quad t = 1, 2, \cdots, n. \tag{1.1}$$

*Corresponding author e-mail: arindampnj@gmail.com
Email Address: kundu@iitk.ac.in (Debasis Kundu)

Each component k represents a sinusoidal signal with amplitude $A_k^{0^2} + B_k^{0^2}$ and frequency $0 < \omega_k^0 < \pi$. The error sequence $\{e(t)\}$ has mean zero and finite variance $\sigma^2 > 0$, with further assumptions specified later. The true signal is characterized by the parameters A_k^0, B_k^0, and ω_k^0, some of which may be unknown and need to be estimated. We denot A_k, B_k, and ω_k as arbitrary parameter values.

Parameter estimation in sinusoidal models is challenging due to the inherent nonlinearity of the model ([3, 22, 23, 25]). A classical approach is the ordinary least squares estimator (LSE). [14] established consistency and asymptotic normality of LSEs for model (1.1) under fairly general conditions on the error process. Although LSEs achieve optimal convergence rates, they are highly sensitive to outliers.

Outliers commonly occur in practical applications such as environmental studies (e.g., extreme weather events), financial time series (e.g., abrupt market shocks), biomedical signals (e.g., motion artifacts in EEG/ECG), and geophysical data (e.g., seismic events). Such extreme observations can distort parameter estimation and lead to misleading inference ([2, 5, 7, 8, 18, 21, 29, 30]).

To mitigate this issue, [13] proposed least absolute deviation estimators (LADEs), which possess desirable asymptotic properties under stringent assumptions on the error distribution. However, LADEs are computationally demanding, requiring optimization over $3p$ (or $2p$ when ω is known) dimensions, which makes them less practical for larger models.

In this work, we propose a regularized estimation approach based on ridge regression [10]. By augmenting the least squares criterion with a penalty on the amplitudes, ridge estimation reduces variance and improves mean squared error performance under contamination. In the sinusoidal setting, the penalty term corresponds directly to the amplitude $A_k^2 + B_k^2$, leading to a shrinkage effect that stabilizes estimation in the presence of outliers. This regularization improves numerical stability and preserves interpretability while maintaining computational feasibility even when frequencies are unknown. Related uses of ridge-type estimators in nonlinear settings can be found in ([17, 24, 28, 31]).

An important advantage of ridge estimation lies in the differentiability of the penalty term, which ensures smooth optimization, facilitates computation, and aids in deriving asymptotic properties.

The primary contributions of this paper are as follows. First, we introduce the concept of regularized estimation in sinusoidal models using ridge regression. Second, we develop estimation procedures for both single- and multiple-sinusoidal models. Third, we establish theoretical properties of the proposed estimators. The resulting estimators can be viewed as regularized estimators with improved stability under contamination. When no outliers are present, they closely approximate LSEs, whereas in contaminated settings they outperform both LSEs and LADEs.

The remainder of the paper is organized as follows. Section 2 presents the notation, preliminary definitions, and assumptions. Section 3.1 discusses the increasing-component setting with known frequencies, while Section 3.2 considers the corresponding unknown-frequency case as motivation for regularization. Section 4 develops the ridge-based estimation procedure for a fixed number of sinusoidal components. Section 4.3 establishes consistency, and Section 4.4 derives the asymptotic normality of the proposed estimators. Section 5 reports the simulation study, including comparisons with least squares, LAD, Huber, and Huber–ridge estimators. Section 6 summarizes the numerical performance, and the robustness diagnostics are discussed subsequently. The real-data application to the AirPassengers series is presented in Section 8. Finally, the concluding remarks are given in the last section.

2 Preliminaries

Definition 2.1 *Let* $\mathbf{a} = (a_1, a_2, ..., a_n)$ *an* n *dimensional vector then* $||.||_2$ *is* l_2 *norm for* $\mathbf{a}$ *is defined by[10]*

$$||\mathbf{a}||_2 = \left(\sum_{i=1}^{n} a_i^2\right)^{\frac{1}{2}}$$

Definition 2.2 *Consider the unconstrained optimization problem:*

$$\min_{x \in \mathbb{R}^n} f(x) \tag{2.1}$$

The Karush–Kuhn–Tucker(KKT) [1] conditions reduce to the following:

$$\nabla f(x^*) = 0 \tag{2.2}$$

To classify the critical point x^**, we analyze the Hessian:*

$$H(x) = \nabla^2 f(x)$$

We can express the second-order conditions as

$$\begin{aligned} H(x^*) \succ 0 &\quad \Rightarrow \quad x^* \textit{ is a local minimum,} \\ H(x^*) \prec 0 &\quad \Rightarrow \quad x^* \textit{ is a local maximum,} \\ H(x^*) \textit{ indefinite} &\quad \Rightarrow \quad x^* \textit{ is a saddle point.} \end{aligned}$$

Assumption 1:

$$e(t) = \sum_{j=-\infty}^{\infty} \alpha(j)\epsilon(t-j), \quad |\alpha(j)| < \infty \tag{2.3}$$

where $\epsilon(t)$'s are independent and identically distributed (i.i.d.) random variables with mean zero and finite variance. Here $=$ means $e(t)$ has the almost-sure representation.

lemma 2.1 *Let $e(t)$ be a stationary sequence that satisfies Assumption 1. then*

$$\lim_{N\to\infty} \sup_{\theta} \left| \frac{1}{N} \sum_{t=1}^{N} e(t)\cos(\theta t) \right| = 0 \quad a.s. \tag{2.4}$$

Proof: For the detailed proof, we refer to Lemma 1 of [14].

Before proceeding to the next section, we introduce the set of parameters listed below.

Notation	**Description**
A^0, B^0, ω^0	True parameters of the single–sinusoidal model
$\hat{A}_N, \hat{B}_N, \hat{\omega}_N$	Estimated parameters of the single–sinusoidal model
A_k^0, B_k^0, ω_k^0, $k = 1, \ldots, p$	True parameters of the kth component in the multiple–sinusoidal model
$\hat{A}_k, \hat{B}_k, \hat{\omega}_k$, $k = 1, \ldots, p$	Estimated parameters of the kth component in the multiple–sinusoidal model
$\mathbf{A}, \mathbf{B}, \boldsymbol{\omega}$	Arbitrary parameter vectors used in the estimation of the multiple–sinusoidal model

Table 1: Notation for parameters

3 Increasing Number of Sinusoidal Components

3.1 Known Frequencies

Consider the multiple sinusoidal model

$$z(t) = \sum_{k=1}^{P_n} \left[A_k^0 \cos(\omega_k^0 t) + B_k^0 \sin(\omega_k^0 t) \right] + e(t), \qquad t = 1, \ldots, n, \tag{3.1}$$

where

(i). the frequencies $\omega_1^0, \ldots, \omega_{P_n}^0 \in (0, \pi)$ are fixed and known;

(ii). A_k^0, B_k^0 are unknown real parameters;

(iii). P_n is allowed to increase with the sample size n;

(iv). $\{e(t)\}$ is a sequence of random errors satisfying

$$\mathbb{E}[e(t)] = 0, \qquad \sup_t \mathbb{E}[e(t)^2] = \sigma^2 < \infty.$$

Define the parameter vector. $\theta = (A_1^0, B_1^0, \ldots, A_{P_n}^0, B_{P_n}^0)^\top \in \mathbb{R}^{2P_n}$. Let $X \in \mathbb{R}^{n \times 2P_n}$ denote the design matrix with entries $X_{t,2k-1} = \cos(\omega_k t), X_{t,2k} = \sin(\omega_k t)$, for $t = 1, \ldots, n$ and $p = 1, \ldots, P_n$. Then the model reduces to a linear model that can be written in matrix form as $z = X\theta + e$.

Now, the normalized Gram matrix is given by $\frac{1}{n} X^\top X$. When P_n is fixed, and the frequencies are distinct, classical harmonic identities yield. $\frac{1}{n}\sum_{t=1}^{n} \cos(\omega_p t)\cos(\omega_q t) \longrightarrow \frac{1}{2}\mathbf{1}_{\{p=q\}}$, with analogous limits for sine–sine and sine–cosine terms. estimator This ensures consithe stency of ordinary least squares (OLS). When $P_n \to \infty$, however,

- near collinearity arises when $|\omega_p - \omega_q| = o(1/n)$;
- the minimum eigenvalue of $X^\top X$ approaches zero as P_n/n increases.

Consequently, $(X^\top X)^{-1}$ becomes unstable. The OLS estimator $\hat{\theta}_{\text{OLS}} = (X^\top X)^{-1} X^\top z$ exists only when $2P_n < n$. Even in this case, if $\frac{P_n}{n} \to \kappa \in (0,1)$, then $\mathbb{E}\big\|\hat{\theta}_{\text{OLS}} - \theta\big\|^2 \to \infty$, and OLS is not consistent, i.e., $\frac{1}{n}\mathbb{E}\big\|X(\hat{\theta}_{\text{OLS}} - \theta)\big\|^2 \not\to 0$. To stabilize estimation, we consider the ridge estimator

$$\hat{\theta}_{\lambda_n} = (X^\top X + \lambda_n I_{2P_n})^{-1} X^\top z, \tag{3.2}$$

where $\lambda_n > 0$ is a tuning parameter. Our next theorem establishes the consistency of the ridge estimator.

Theorem 3.1 *Consider the linear model $Y = X\theta + \varepsilon$, where $Y \in \mathbb{R}^n$, X is an $n \times P_n$ design matrix, $\theta \in \mathbb{R}^{P_n}$ is the true parameter vector, and $\varepsilon \in \mathbb{R}^n$ is a noise vector with $\mathbb{E}[\varepsilon] = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$. For $\lambda_n > 0$, define the ridge estimator $\hat{\theta}_{\lambda_n}$ as*

$$\hat{\theta}_{\lambda_n} = \arg\min_{\beta \in \mathbb{R}^{P_n}} \left\{ \frac{1}{n}\|Y - X\beta\|^2 + \lambda_n \|\beta\|^2 \right\} = \left(\frac{1}{n} X^\top X + \lambda_n I_{P_n} \right)^{-1} \frac{1}{n} X^\top Y.$$

Assume that:

(i). $P_n \to \infty$ and $P_n/n \to \kappa \in (0, \infty)$;
(ii). $\|\theta\|^2 = O(P_n)$;
(iii). $\lambda_n \asymp P_n/n$, i.e. there exist constants $0 < c_1 \le c_2 < \infty$ such that $c_1 P_n/n \le \lambda_n \le c_2 P_n/n$ for all large n;
(iv). the empirical Gram matrix $S_n := \frac{1}{n} X^\top X$ has eigenvalues uniformly bounded away from 0 and ∞, i.e. there exist constants $0 < m \le M < \infty$ such that $m \le \lambda_{\min}(S_n) \le \lambda_{\max}(S_n) \le M$ for all sufficiently large n.

Then the ridge estimator is consistent, that is

$$\frac{1}{n}\,\mathbb{E}\Big\| X\big(\hat{\theta}_{\lambda_n} - \theta\big) \Big\|^2 \longrightarrow 0 \qquad \text{as } n \to \infty.$$

Proof: See in Appendix A.

3.2 Unknown frequencies

Now consider the multiple sinusoidal model

$$z(t) = \sum_{k=1}^{P_n} \Big[A_k^0 \cos(\omega_k^0 t) + B_k^0 \sin(\omega_k^0 t) \Big] + e(t), \qquad t = 1, \dots, n, \tag{3.3}$$

where

- the frequencies $\omega_1^0, \ldots, \omega_{P_n}^0 \in (0, \pi)$ are Unknown;
- A_k^0, B_k^0 are unknown real parameters;
- P_n is allowed to increase with the sample size n;
- $\{e(t)\}$ is a sequence of random errors satisfying

$$\mathbb{E}[e(t)] = 0, \qquad \sup_t \mathbb{E}[e(t)^2] = \sigma^2 < \infty.$$

Scenario I: We conduct a Monte Carlo simulation study to examine the finite-sample performance of the proposed estimator in a high-dimensional nonlinear setting where the number of unknown parameters scales with the sample size. The experimental setup is as follows:

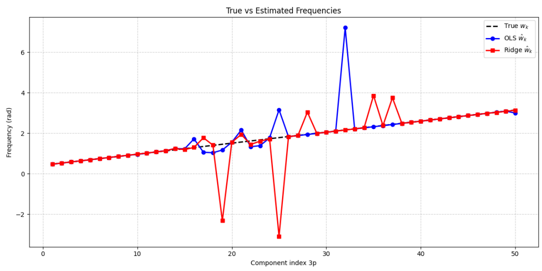
1. We consider $p = 150, 300, 600, 1500$ and choose the sample size such that $\frac{p}{N} = 0.6$ for each case.
2. For each p, we generate p sinusoidal components with decreasing magnitudes, and the corresponding frequencies are linearly spaced over $[0.15\pi, \pi]$.
3. The observations are contaminated with additive noise having standard deviation $\sigma = 30$, along with 30% outliers.
4. For each (p, N) configuration, we perform a single simulation run and estimate the parameters using a ridge-regularized ordinary least squares method.

The results show that the ridge estimator successfully captures the overall linear trend of the true frequencies even under substantial noise and contamination For smaller dimensions ($p = 150, 300$), the estimates exhibit higher variability with noticeable deviations As both p and N increase, the estimates become more stable and concentrate closer to the true frequencies, indicating improved finite-sample performance However, occasional large deviations persist even for larger p, reflecting the continuing impact of heavy noise and outliers.

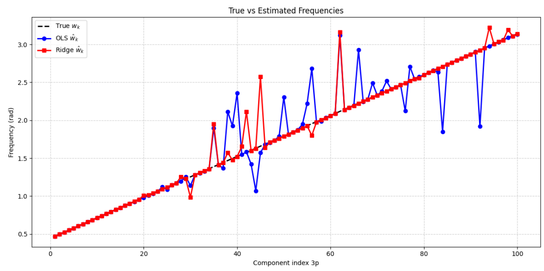
Scenario II: We conduct a simulation study to investigate the finite-sample performance of the proposed estimator in a high-dimensional nonlinear setting where the number of unknown parameters varies while the sample size is fixed. The experimental setup is as follows:

1. We consider $p \in \{600, 900, 1200, 1500\}$ and fix the sample size at $N = 1500$.
2. For each p, p sinusoidal components with decreasing magnitudes are generated, and the corresponding frequencies are linearly spaced over $[0.75\pi, \pi]$.
3. Observations are contaminated with additive noise having standard deviation $\sigma = 30$, along with 30% outliers.
4. For each value of p, a single simulation run is performed using a ridge-regularized ordinary least squares estimator.

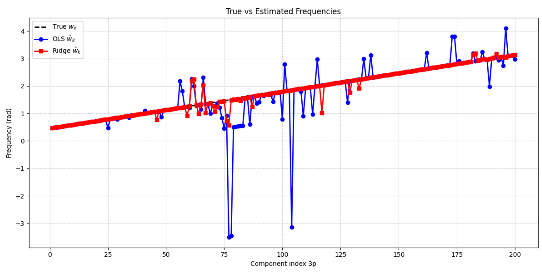
The results indicate that, for fixed N, increasing p leads to a clear deterioration in estimation accuracy For $p = 600$, the estimator still captures the overall trend of the true frequencies but



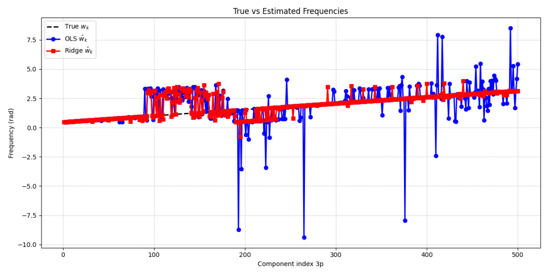
(a) $p = 150, N = 250$



(b) $p = 300, N = 500$



(c) $p = 600, N = 1000$



(d) $p = 1500, N = 2500$

Figure 1: Components vs Sample Size.

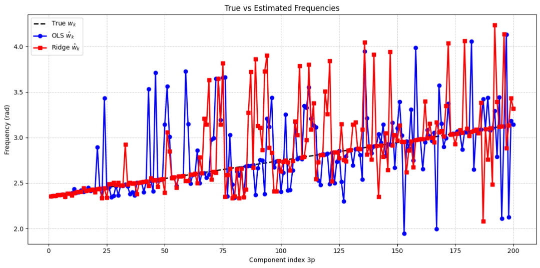
exhibits noticeable variability and spikes due to noise and outliers As p increases to 900 and 1200, the ordinary least squares estimator becomes increasingly unstable. In contrast, ridge regularization effectively shrinks extreme deviations and produces smoother estimates closer to the true frequencies For $p = 1500$, OLS suffers from severe variance inflation. In contrast, ridge continues to provide comparatively stable estimates. However, some large errors persist This suggests that regularization alleviates—but does not eliminate—the challenges posed by high dimensionality, noise, and contamination.

These findings highlight the difficulties inherent in high-dimensional settings and further motivate the use of regularization Accordingly, the main focus of this work is to develop a regularized estimator (with improved stability under contamination) for the unknown frequencies and amplitudes when p is fixed In the next section, we develop ridge-based estimators for the sinusoidal model under the assumption that the number of components p is fixed.

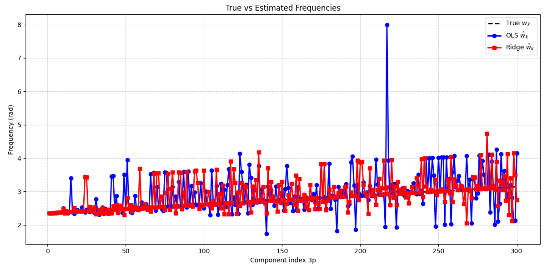
4 Fixed numbers of Sinusoidal components

4.1 Known frequencies

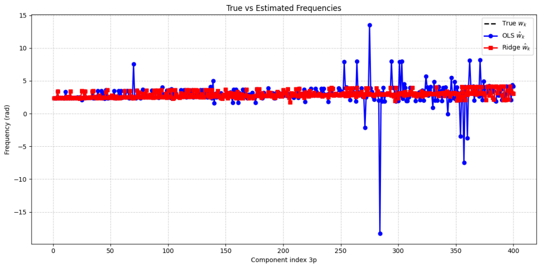
We first consider the case when the frequencies are known. In this situation, the sinusoidal model becomes linear in the unknown amplitude parameters, and hence the estimation problem reduces to a standard least squares problem. Since the theoretical properties of least squares estimators are well established in the literature, we only briefly discuss this case and focus mainly on the more challenging problem of unknown frequencies in the next subsection.



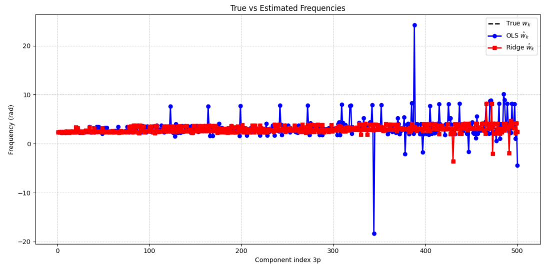
(a) $p = 600, N = 1500$



(b) $p = 900, N = 1500$



(c) $p = 1200, N = 1500$



(d) $p = 1500, N = 1500$

Figure 2: Components.

For a fixed number of sinusoidal components p, suppose that the frequency vector $\boldsymbol{\omega}^0 = (\omega_1^0, \omega_2^0, \ldots, \omega_p^0)^T$ is known. Then model (1.1) can be written in matrix form as

$$\mathbf{Z} = \mathbf{X}_p(\boldsymbol{\omega}^0)\boldsymbol{\psi}^0 + \mathbf{E}, \tag{4.1}$$

where $\boldsymbol{\psi}^0 = (\eta_1^{0T}, \eta_2^{0T}, \ldots, \eta_p^{0T})^T, \qquad \eta_j^0 = (A_j^0, B_j^0)^T$, and $\mathbf{X}_p(\boldsymbol{\omega}^0) = \left[X(\omega_1^0), X(\omega_2^0), \ldots, X(\omega_p^0)\right]$.

Therefore, for known frequencies, estimation of the unknown coefficients reduces to the minimization of $Q(\boldsymbol{\psi}) = \left(\mathbf{Z} - \mathbf{X}_p(\boldsymbol{\omega}^0)\boldsymbol{\psi}\right)^T \left(\mathbf{Z} - \mathbf{X}_p(\boldsymbol{\omega}^0)\boldsymbol{\psi}\right)$. The corresponding least squares estimator is

$$\hat{\boldsymbol{\psi}}_{LS} = \left[\mathbf{X}_p(\boldsymbol{\omega}^0)^T \mathbf{X}_p(\boldsymbol{\omega}^0)\right]^{-1} \mathbf{X}_p(\boldsymbol{\omega}^0)^T \mathbf{Z}, \tag{4.2}$$

provided that $\mathbf{X}_p(\boldsymbol{\omega}^0)^T \mathbf{X}_p(\boldsymbol{\omega}^0)$ is nonsingular.

If a ridge penalty is imposed, then the estimator becomes

$$\hat{\boldsymbol{\psi}}_R = \left[\mathbf{X}_p(\boldsymbol{\omega}^0)^T \mathbf{X}_p(\boldsymbol{\omega}^0) + \Lambda\right]^{-1} \mathbf{X}_p(\boldsymbol{\omega}^0)^T \mathbf{Z}, \tag{4.3}$$

where Λ is the penalty matrix. In the special case of a common penalty parameter, $\Lambda = \lambda I$.

Since the theoretical properties of least squares estimators, as well as their ridge-regularized counterparts, are well established for the known-frequency case, we do not pursue them further here. Instead, we focus on the estimation problem when the frequencies are unknown, which leads to a nonlinear optimization problem.

4.2 Unknown frequencies

First, we consider a single sinusoidal model, i.e., when $p = 1$. For $p = 1$ matrix form of (1.1) can be written as

$$Z = X(\omega^0)\eta^0 + E \tag{4.4}$$

Where $Z = [z(1), \cdots, z(n)]^T, E = [e(1), e(2), \cdots e(n)]^T, \eta^0 = [A^0, B^0]^T$ and

$$X(\omega^0) = \begin{bmatrix} \cos(\omega^0) & \sin(\omega^0) \\ \cos(2\omega^0) & \sin(2\omega^0) \\ \cdots & \cdots \\ \cos(n\omega^0) & \sin(n\omega^0) \end{bmatrix} \tag{4.5}$$

The Ridge estimates of A^0, B^0, ω^0 can be obtained by minimizing

$$Q_N^R(\theta) \quad = \quad \sum_{i=1}^{N} (Z(t) - A\cos(\omega t) - B\sin(\omega t))^2 + \lambda\left(A^2 + B^2\right) \tag{4.6}$$

with respect to $\theta = (A, B, \omega)$.

In vector notation, which can be written as

$$||Z - X(\omega)\eta||_2^2 + \lambda\eta^T\eta \tag{4.7}$$

$\lambda \geq 0$ is the regularization parameter that controls the amount of shrinkage. It is to be mentioned that $\lambda = 0$ corresponds to ordinary least squares (OLS) estimates.

Next we are going to propose a sequential approach to solve the nonlinear optimization problem (4.7) First for any fixed ω we are going to minimize (4.7).

$$\min_{\text{arg}} \left\{||Z - X(\omega)\eta||_2^2 + \lambda\eta^T\eta\right\} \tag{4.8}$$

Suppose $\hat{\eta} = \left(\hat{A}, \hat{B}\right)$ is a solution of the above optimization problem, then from the KKT stationarity condition 2.2, the sub-differential must be zero, i.e.

$$\begin{aligned} &-2X(\omega)^T Z + 2\hat{\eta}^T X(\omega)^T X(\omega) + 2\hat{\eta}\lambda = 0 \\ &\implies 2\hat{\eta}^T X(\omega)^T X(\omega) + 2\hat{\eta}\lambda = 2X(\omega)^T Z \\ &\implies \hat{\eta}(\omega) = \left(X(\omega)^T X(\omega) + \lambda\mathbf{I}\right)^{-1} X(\omega)^T Z \end{aligned} \tag{4.9}$$

Substituting $\hat{\eta}(\omega)$ in 4.8 we have

$$
\begin{aligned}
R(\omega) &= \left(Z - X(\omega)\left(X(\omega)^T X(\omega) + \lambda\mathbf{I}\right)^{-1} X(\omega)^T Z\right)^T \\
&\quad \times \left(Z - X(\omega)\left(X(\omega)^T X(\omega) + \lambda\mathbf{I}\right)^{-1} X(\omega)^T Z\right) + \lambda\eta^T\eta \qquad (4.10)
\end{aligned}
$$

Therefore Ridge estimate of $\omega, \hat{\omega}$ is obtained by minimizing (4.10) and then plugging in $\hat{\omega}$ into equation (4.9), η is estimated as $\hat{\eta}(\hat{\omega})$. In each case, expanding window cross-validation is often used in choosing λ[12] More details detailed on expanding window cross-validation are discussed in section 5.4.

Now we continue our approach for $p > 1$. For $p > 1$ the model equation 4.2 can be written as

$$
\mathbf{Z} = \mathbf{X_p}(\omega)\psi + \mathbf{E} \qquad (4.11)
$$

where $\omega = \left(\omega_1^0, \omega_2^0, \cdots, \omega_p^0\right), \psi = \left(\eta_1^0, \eta_2^0, \cdots, \eta_p^0\right)^T, \eta_j^0 = (A_j^0, B_j^0)^T$ and $\mathbf{X_p}(\omega)$ takes the form $\mathbf{X_p}(\omega) = \left[X(\omega_1^0), X(\omega_2^0), \cdots, X(\omega_p^0)\right]$. The matrix $X(\omega_j^0)$ is the same as that $X(\omega)$ defined in (4.5), replacing ω by ω_j^0. It is to be note that Equation (4.32) can also be written as

$$
\mathbf{Z} = \sum_{\mathbf{j=1}}^{\mathbf{p}} \mathbf{X}(\omega_{\mathbf{j}}^{\mathbf{0}})\eta_{\mathbf{j}}^{\mathbf{0}} + \mathbf{E} \qquad (4.12)
$$

Now the ridge estimates of ω and η are obtained by minimizing:

$$
\begin{aligned}
U(\eta, \omega) &= \left(\mathbf{Z} - \sum_{\mathbf{j=1}}^{\mathbf{p}} \mathbf{X}(\omega_{\mathbf{j}})\eta_{\mathbf{j}}\right)^T \left(\mathbf{Z} - \sum_{\mathbf{j=1}}^{\mathbf{p}} \mathbf{X}(\omega_{\mathbf{j}})\eta_{\mathbf{j}}\right) + \lambda\psi^{\mathbf{T}}\psi \\
&= \left(\mathbf{Z} - \sum_{\mathbf{j=1}}^{\mathbf{p}} \mathbf{X}(\omega_{\mathbf{j}})\eta_{\mathbf{j}}\right)^T \left(\mathbf{Z} - \sum_{\mathbf{j=1}}^{\mathbf{p}} \mathbf{X}(\omega_{\mathbf{j}})\eta_{\mathbf{j}}\right) + \sum_{\mathbf{i=1}}^{\mathbf{p}} \lambda_{\mathbf{j}}\eta_{\mathbf{j}}^{\mathbf{T}}\eta_{\mathbf{j}} \qquad (4.13)
\end{aligned}
$$

Let ψ_0 and ω_0 be the true values of ψ_0 and ω, respectively, and $\hat{\psi}_0$ and $\hat{\omega}_0$ denote the LSEs of ψ_0 and ω_0, respectively. Then for any given $\omega, \hat{\psi}$ can be written as

$$
\hat{\psi}(\omega) = \left[\mathbf{X_p}(\omega)^{\mathbf{T}}\mathbf{X_p}(\omega) + \lambda\mathbf{I}\right]^{-1} \mathbf{X_p}(\omega)^{\mathbf{T}}\mathbf{Z} \qquad (4.14)
$$

where

$$\hat{\psi}(\omega) = \begin{bmatrix} \hat{\eta}(\omega_1) \\ \hat{\eta}(\omega_2) \\ \cdots \\ \hat{\eta}(\omega_p) \end{bmatrix} = \begin{bmatrix} \left[\mathbf{X}(\omega_1)^\mathbf{T}\mathbf{X}(\omega_1) + \lambda_1\mathbf{I}\right]^{-1}\mathbf{X}(\omega_1)^\mathbf{T}\mathbf{Z} \\ \left[\mathbf{X}(\omega_2)^\mathbf{T}\mathbf{X}(\omega_2) + \lambda_2\mathbf{I}\right]^{-1}\mathbf{X}(\omega_2)^\mathbf{T}\mathbf{Z} \\ \cdots \\ \left[\mathbf{X}(\omega_\mathbf{p})^\mathbf{T}\mathbf{X}(\omega_\mathbf{p}) + \lambda_\mathbf{p}\mathbf{I}\right]^{-1}\mathbf{X}(\omega_\mathbf{p})^\mathbf{T}\mathbf{Z} \end{bmatrix} \tag{4.15}$$

Now plugging $\hat{\psi}(\omega)$ in (4.13) we have

$$\begin{aligned} S(\omega) = U(\hat{\psi}(\omega), \omega) &= \left[\mathbf{Z} - \sum_{j=1}^{p}\left(\mathbf{X}(\omega_\mathbf{j})^\mathbf{T}\mathbf{X}(\omega_\mathbf{j}) + \lambda_\mathbf{j}\mathbf{I}\right)^{-1}\mathbf{X}(\omega_\mathbf{j})^\mathbf{T}\mathbf{Z}\right]^T \\ &\times \left[\mathbf{Z} - \sum_{j=1}^{p}\left(\mathbf{X}(\omega_\mathbf{j})^\mathbf{T}\mathbf{X}(\omega_\mathbf{j}) + \lambda_\mathbf{j}\mathbf{I}\right)^{-1}\mathbf{X}(\omega_\mathbf{j})^\mathbf{T}\mathbf{Z}\right] \\ &= \left[\mathbf{Z} - \sum_{j=1}^{p}\mathbf{P}_\mathbf{X}(\omega_\mathbf{j})\mathbf{Z}\right]^T \left[\mathbf{Z} - \sum_{j=1}^{p}\mathbf{P}_\mathbf{X}(\omega_\mathbf{j})\mathbf{Z}\right] \end{aligned} \tag{4.16}$$

Where $\mathbf{P}_\mathbf{X}(\omega_\mathbf{j}) = \left(\mathbf{X}(\omega_\mathbf{j})^\mathbf{T}\mathbf{X}(\omega_\mathbf{j}) + \lambda_\mathbf{j}\mathbf{I}\right)^{-1}\mathbf{X}(\omega_\mathbf{j})^\mathbf{T}$.

4.3 Asymptotic Properties

In this section, we investigate the asymptotic behavior of the proposed estimators The theoretical results are derived under the following set of assumptions:

- The number of sinusoidal components p is fixed;
- The true frequencies ω_k^0 are distinct and lie in the interval $(0, \pi)$;
- The error sequence $\{e(t)\}$ is a zero-mean process with finite variance;
- The regularization parameter λ_N satisfies suitable growth conditions as $N \to \infty$.

Under these assumptions, we establish the consistency and asymptotic normality properties of the Ridge estimators for the single sinusoidal model.

Theorem 4.1 *Let $\hat{\theta} = (\hat{A}_N, \hat{B}_N, \hat{\omega}_N)$ denote the Ridge estimator of the nonlinear regression model given in (4.7). Then, $\hat{\theta}$ is a strongly consistent estimator of the true parameter vector $\theta^0 = (A^0, B^0, \omega^0)$.*

Proof: See in Appendix B.

4.4 Asymptotic Normality

In this section, we prove the asymptotic normality of $\hat{\theta}$. First, we calculates all first and second order partial derivatives of $\hat{\theta}_N^R$.

$$\frac{\partial Q_N^R(\theta^0)}{\partial A} = -2\sum_{t=1}^{N} e(t)\cos(\omega^0 t) + 2\hat{\lambda}_N A^0 \tag{4.17}$$

$$\frac{\partial Q_N^R(\theta^0)}{\partial B} = -2\sum_{t=1}^{N} e(t)\sin(\omega^0 t) + 2\hat{\lambda}_N B^0 \tag{4.18}$$

$$\frac{\partial Q_N^R(\theta^0)}{\partial \omega} = 2\sum_{t=1}^{N} te(t)\left(A^0\sin(\omega^0 t) - B^0\cos(\omega^0 t)\right) \tag{4.19}$$

$$\frac{\partial Q_N^R(\bar{\theta})}{\partial A^2} = 2\sum_{t=1}^{N} \cos^2(\bar{\omega}t) + 2\hat{\lambda}_N \tag{4.20}$$

$$\frac{\partial Q_N^R(\bar{\theta})}{\partial B^2} = 2\sum_{t=1}^{N} \sin^2(\bar{\omega}t) + 2\hat{\lambda}_N \tag{4.21}$$

$$\begin{aligned}\frac{\partial Q_N^R(\bar{\theta})}{\partial \omega^2} &= 2\sum_{t=1}^{N} t^2\left[\left(A^0\cos(\omega^0 t) + B^0\sin(\omega^0 t) - \bar{A}\cos(\bar{\omega}t) - \bar{B}\sin(\bar{\omega}t) + e(t)\right)\right.\\ &\times \left.\left(\bar{A}\cos(\bar{\omega}t + \bar{B}\sin(\bar{\omega}t))\right) + \left(\bar{A}\sin(\bar{\omega}t + \bar{B}\cos(\bar{\omega}t))\right)^2\right] \qquad (4.22)\\ \frac{\delta^2 Q_N^R(\bar{\theta})}{\delta\omega\delta A} &= 2\sum_{t=1}^{N}\left[t\sin(\omega t)\left(A^0\cos(\omega^0 t) + B^0\sin(\omega^0 t)\right) - \bar{A}\cos(\bar{\omega}t)\right.\\ &\left.-\bar{B}\sin(\bar{\omega}t) + X(t) - \cos(\bar{\omega}t)\left(\bar{A}\sin(\bar{\omega}t) - \bar{B}\cos(\bar{\omega}t)\right)\right] \qquad (4.23)\end{aligned}$$

$$\begin{aligned}\frac{\delta^2 Q_N^R(\bar{\theta})}{\delta\omega\delta B} &= -2\sum_{t=1}^{N}\left[\cos(\omega t)\left(A^0\cos(\omega^0 t) + B^0\sin(\omega^0 t)\right) - \bar{A}\cos(\bar{\omega}t)\right.\\ &\left.-\bar{B}\sin(\bar{\omega}t) + X(t) - \sin(\bar{\omega}t)\left(\bar{A}\sin(\bar{\omega}t) - \bar{B}\cos(\bar{\omega}t)\right)\right] \qquad (4.24)\\ \frac{\delta^2 Q_N^R(\bar{\theta})}{\delta A\delta B} &= 2\sum_{t=1}^{N}\sin(\bar{\omega}t)\cos(\bar{\omega}t) + 2 \qquad (4.25)\end{aligned}$$

Let's define

$$
\begin{aligned}
\sigma_{11} &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\cos^2(\omega^0 t) = \frac{1}{2} \\
\sigma_{22} &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sin^2(\omega^0 t) = \frac{1}{2} \\
\sigma_{33} &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}t^2\left(A_0\sin(\omega^0 t) - B^0\cos(\omega^0 t)\right)^2 = \frac{1}{6}\left(A^{0^2} + B^{0^2}\right) \\
\sigma_{13} &= \sigma_{31} = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}B^0 t\cos^2(\omega^0 t) = \frac{1}{4}B^0
\end{aligned}
$$

Now, if $\bar{\omega}\to\omega^0$, $\bar{A}\to A^0$ and $\bar{B}\to B^0$ a.s. We have the following results.

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\cos^2(\bar{\omega}t) = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\cos^2(\omega^0 t) = \frac{1}{2} \tag{4.26}
$$

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sin^2(\bar{\omega}t) = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sin^2(\omega^0 t) = \frac{1}{2} \tag{4.27}
$$

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{2N^3}\frac{\delta^2 Q_N^R(\bar{\theta})}{\delta\omega^2} &= \lim_{N\to\infty}\frac{1}{N^3}\sum_{t=1}^{N}t^2(\bar{A}\sin(\bar{\omega}t) + \bar{B}\cos(\bar{\omega}t))^2 \\
&= \lim_{N\to\infty}\frac{1}{N^3}\sum_{t=1}^{N}t^2(A^0\sin(\omega^0 t) - B^0\cos(\omega^0 t))^2 \\
&= \frac{1}{6}(A^{0^2} + B^{0^2})
\end{aligned} \tag{4.28}
$$

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{2N^2}\frac{\delta^2 Q_N^R(\bar{\theta})}{\delta\omega\delta A} &= \lim_{N\to\infty}\frac{1}{N^2}\sum_{t=1}^{N}t\cos(\omega t)(\bar{A}\sin(\bar{\omega}t) - \bar{B}\cos(\bar{\omega}t)) \\
&= \lim_{N\to\infty}\frac{1}{N^2}\sum_{t=1}^{N}tB^0\cos^2(\omega^0 t) = \frac{1}{4}B^0
\end{aligned} \tag{4.29}
$$

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{2N^2}\frac{\delta^2 Q_N^R(\bar{\theta})}{\delta\omega\delta B} &= \lim_{N\to\infty}\frac{1}{N^2}\sum_{t=1}^{N}t\sin(\omega t)(\bar{A}\sin(\bar{\omega}t) - \bar{B}\cos(\bar{\omega}t)) \\
&= \lim_{N\to\infty}\frac{1}{N^2}\sum_{t=1}^{N}tA^0\sin^2(\omega^0 t) = \frac{1}{4}A^0
\end{aligned} \tag{4.30}
$$

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{2N^2}\frac{\delta^2 Q_N^R(\bar{\theta})}{\delta A\delta B} &= \frac{2}{N}\sum_{t=1}^{N}\sin(\bar{\omega}t)\cos(\bar{\omega}t) \\
&= \frac{2}{N}\sum_{t=1}^{N}\sin(\omega^0 t)\cos(\omega^0 t) = 0.
\end{aligned} \tag{4.31}
$$

Now, we proposed our next theorem

Theorem 4.2 *Under the assumptions of Theorem 4.1, suppose that the tuning parameter* λ_N *satisfies* $\lambda_N \to \lambda \in [0, \infty)$ *as* $N \to \infty$. *Then,* $\left\{ \frac{(\hat{A}_N - A^0)}{\sqrt{N}}, \frac{(\hat{B}_N - B^0)}{\sqrt{N}}, \frac{(\hat{\omega}_N - \omega^0)}{N\sqrt{N}} \right\}$ *converges in distribution to a trivariate normal distribution with mean vector zero and dispersion matrix* $\sigma^2 \gamma \Sigma^{-1}(\lambda)$, *where* $\gamma = \left| \sum_{j=-\infty}^{\infty} \alpha(j) \cos(\omega^0 j) \right|^2 + \left| \sum_{j=-\infty}^{\infty} \alpha(j) \sin(\omega^0 j) \right|^2$, *and* $\Sigma^{-1}(\lambda)$ *is given by*

$$\Sigma^{-1}(\lambda) = \frac{1}{(A^{0^2} + B^{0^2})(16\lambda^2 + 10\lambda + 1)}$$
$$\times \begin{bmatrix} 2(8\lambda+1)(A^{0^2} + 4B^{0^2}) & -6A^0B^0 & -12B^0 \\ -6A^0B^0 & 2(8\lambda+1)(4A^{0^2} + B^{0^2}) & 12A^0 \\ -12B^0 & 12A^0 & 24(2\lambda+1) \end{bmatrix}.$$

Proof: See in Appendix C.

We now extend the results from the single sinusoidal model to the multiple sinusoidal case. Consider the model

$$Y(t) = \sum_{k=1}^{M} \left(A_k^0 \cos(\omega_k^0 t) + B_k^0 \sin(\omega_k^0 t) \right) + X(t), \quad t = 1, \ldots, N, \tag{4.32}$$

where A_k^0, B_k^0 are the true amplitudes and $\omega_k^0 \in (0, \pi)$ are distinct frequencies for $k = 1, \ldots, M$. The noise process $\{X(t)\}$ satisfies Assumption 1. Let $A = (A_1, \ldots, A_M)$, $B = (B_1, \ldots, B_M)$, and $\omega = (\omega_1, \ldots, \omega_M)$, and define $\Phi = (A, B, \omega)$. We consider the ridge-type estimator obtained by minimizing the penalized least squares criterion

$$R_N(\Phi) = \sum_{t=1}^{N} \left(Y(t) - \sum_{k=1}^{M} (A_k \cos(\omega_k t) + B_k \sin(\omega_k t)) \right)^2 + \sum_{k=1}^{M} \lambda_{k,N} \left(A_k^2 + B_k^2 \right),$$

where the tuning parameters $\lambda_{k,N} \geq 0$ satisfy $\lambda_{k,N} \to \lambda_k \in [0, \infty)$, as $N \to \infty$, $k = 1, \ldots, M$. We now establish the consistency and asymptotic normality of the ridge estimators.

Theorem 4.3 *Let* $\hat{\Phi}_k = (\hat{A}_k, \hat{B}_k, \hat{\omega}_k)$ *denote the ridge estimator of* $\Phi_k = (A_k^0, B_k^0, \omega_k^0)$ *for* $k = 1, \ldots, M$. *Then, under the above assumptions,* $\hat{\Phi}_k$ *is a strongly consistent estimator of* Φ_k.

Proof. See Appendix D.

Theorem 4.4 *Under the assumptions of Theorem 4.3, and assuming that* $\lambda_{k,N} \to \lambda_k \in [0, \infty)$ *for each* $k = 1, \ldots, M$, *we have* $\left(N^{1/2}(\hat{A}_k - A_k^0),\ N^{1/2}(\hat{B}_k - B_k^0),\ N^{3/2}(\hat{\omega}_k - \omega_k^0) \right)_{k=1}^{M}$ *converges in distribution to a* $3M$*-variate normal distribution with mean vector zero and dispersion matrix* $\sigma^2 \Gamma^{-1}(\lambda)\, \mathbf{G}\, \Gamma^{-1}(\lambda)$, *where* $\Gamma(\lambda)$ *is the limiting block information matrix depending on* $\{\lambda_k\}_{k=1}^{M}$, *and* $\mathbf{G}$ *is the asymptotic covariance matrix of the scaled score vector.*

Proof: Proof is immediately follows from the proof of Theorem 4.3.

Remark. The assumption $\lambda_{k,N} \to \lambda_k \in [0, \infty)$ ensures a non-degenerate limiting distribution. If $\lambda_{k,N} \to \infty$ for any k, the corresponding components of the asymptotic variance collapse to zero, leading to degeneracy of the limiting distribution.

5 Simulation study

This section describes the simulation design used to evaluate the proposed penalized estimator and to compare it with standard and robust alternatives for model 4.32. To strengthen the methodological comparison, we include not only classical benchmarks such as OLS and LAD, but also a robust ridge-type competitor based on a Huber objective This comparison is motivated by the robust ridge literature, particularly the ridge-type M-estimator of Silvapulle [27] and the penalized robust ridge/MM framework of Maronna [19]. Our implementation is not an exact reproduction of those procedures; rather, the Huber–ridge estimator serves as a direct robust-shrinkage benchmark in the same spirit Accordingly, any performance gain of the proposed estimator should be interpreted relative to both robustification and shrinkage, rather than attributed to one mechanism alone.

5.1 Data-generating model

We consider the multiple Sinusoidal model

$$y_t = \mu_t + e_t, \qquad \mu_t = \sum_{j=1}^{p} \{A_j \cos(\omega_j t) + B_j \sin(\omega_j t)\}, \qquad t = 1, 2, \ldots, n, \tag{5.1}$$

where A_j and B_j are amplitude parameters, $\omega_j \in (0, \pi)$ are unknown angular frequencies, and e_t is the error term. For a parameter vector $\boldsymbol{\theta} = (A_1, B_1, \omega_1, \ldots, A_p, B_p, \omega_p)^\top$, the fitted value at time t is

$$\hat{y}_t(\boldsymbol{\theta}) = \sum_{j=1}^{p} \{A_j \cos(\omega_j t) + B_j \sin(\omega_j t)\}, \tag{5.2}$$

and the residual is

$$r_t(\boldsymbol{\theta}) = y_t - \hat{y}_t(\boldsymbol{\theta}). \tag{5.3}$$

The simulation focuses on the two-component case ($p = 2$) with

$$(A_1, A_2) = (5.0, 1.5), \qquad (B_1, B_2) = (5.0, 1.5), \qquad (\omega_1, \omega_2) = (0.85\pi,\, 0.50\pi).$$

Sample sizes are taken as $n \in \{500, 600, 700, 800, 900, 1000\}$.

Intrinsic multicollinearity in the harmonic design. The present study does not impose a separate artificial collinearity parameter. Instead, multicollinearity enters *structurally* through the harmonic design matrix

$$\mathbf{X}(\boldsymbol{\omega}) = \begin{bmatrix} \cos(\omega_1 \cdot 1) & \sin(\omega_1 \cdot 1) & \cdots & \cos(\omega_p \cdot 1) & \sin(\omega_p \cdot 1) \\ \cos(\omega_1 \cdot 2) & \sin(\omega_1 \cdot 2) & \cdots & \cos(\omega_p \cdot 2) & \sin(\omega_p \cdot 2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cos(\omega_1 n) & \sin(\omega_1 n) & \cdots & \cos(\omega_p n) & \sin(\omega_p n) \end{bmatrix}, \tag{5.4}$$

whose columns may be strongly dependent in finite samples, particularly when multiple oscillatory components generate similar patterns over the observed time span Thus, the simulation studies the instability caused by the intrinsic near-collinearity of sinusoidal regressors, which provides the main motivation for ridge-type penalization.

5.2 Error distributions and contamination schemes

For each replication, the clean signal μ_t in (5.1) is perturbed by an additive error term e_t To evaluate robustness under different departures from normality, we consider two contamination settings: (i) heavy-tailed (t-distributed) contamination and (ii) normally distributed contamination with inflated variance. The error process is generated from the mixture model $e_t = (1 - I_t)\,\varepsilon_t + I_t\,\eta_t$, where $I_t \sim \text{Bernoulli}(\alpha)$, and α denotes the contamination proportion Thus, approximately $m = \lfloor \alpha n \rfloor$ observations are contaminated. For the uncontaminated observations, the baseline noise component is generated as $\varepsilon_t \sim N(0, \sigma^2)$, $\sigma^2 = 1$. Two different distributions are then used for the contaminated component η_t.

1. **Heavy-tailed (t-distributed) contamination:** In this setting, contaminated observations are generated from a scaled Student's t distribution: $\eta_t \sim \sigma_{\text{out}}\, t_\nu$, where t_ν denotes the Student's t distribution with ν degrees of freedom. This produces heavy-tailed shocks and occasional extreme observations, thereby creating a challenging setting for least-squares-based estimation procedures.
2. **Normally distributed contamination:** In this case, contaminated observations are generated from a Gaussian distribution with inflated variance: $\eta_t \sim N(0, \sigma^2_{\text{out}})$. This contamination mechanism preserves Gaussian symmetry while introducing unusually large observations through variance inflation.

For both contamination settings, we consider $\alpha \in \{0.1, 0.3\}, \quad \sigma_{\text{out}} \in \{5, 8\}$. In addition to the contamination distribution, two contamination layouts are studied.

1. **Random contamination:** the m contaminated indices are selected uniformly at random from $\{1, \ldots, n\}$. This produces scattered outliers throughout the time series.
2. **Middle-block contamination:** the m contaminated observations form a contiguous block positioned near the center of the series Specifically, if $m = \lfloor \alpha n \rfloor$, the contaminated set is
$$\left\{ \left\lfloor \frac{n-m}{2} \right\rfloor + 1,\ \left\lfloor \frac{n-m}{2} \right\rfloor + 2, \ldots, \left\lfloor \frac{n-m}{2} \right\rfloor + m \right\}.$$

Hence, the phrase "centrally positioned outliers" refers to a single consecutive contaminated segment in the middle of the observed time series rather than randomly selected central observations.

The distinction between the two contamination layouts is important because they affect estimation differently Random contamination disperses atypical observations across the entire sample, whereas middle-block contamination creates a localized region of distortion that may interfere more strongly with finite-sample frequency estimation and signal recovery.

5.3 Estimators compared

Let $\boldsymbol{\theta}$ denote the unknown parameter vector We compare five estimators.

1. **Ordinary least squares (OLS)** The OLS estimator minimizes the residual sum of squares:
$$\hat{\boldsymbol{\theta}}_{\text{OLS}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{n} r_t(\boldsymbol{\theta})^2. \tag{5.5}$$

2. **Least absolute deviations (LAD)** The LAD estimator minimizes the ℓ_1 criterion
$$\hat{\boldsymbol{\theta}}_{\text{LAD}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{n} |r_t(\boldsymbol{\theta})|. \tag{5.6}$$
For numerical optimization, a smooth approximation is used in implementation:
$$\hat{\boldsymbol{\theta}}_{\text{LAD}} \approx \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{n} \sqrt{r_t(\boldsymbol{\theta})^2 + \varepsilon}, \qquad \varepsilon > 0 \text{ small}. \tag{5.7}$$

3. **Huber M-estimator** To reduce sensitivity to outliers, we consider Huber's loss
$$\rho_c(u) = \begin{cases} \frac{1}{2}u^2, & |u| \le c, \\ c|u| - \frac{1}{2}c^2, & |u| > c, \end{cases} \tag{5.8}$$
with tuning constant $c > 0$. Let s_0 be a robust residual scale estimated from the median absolute deviation (MAD):
$$s_0 = 1.4826 \times \text{MAD}(r_t), \qquad \text{MAD}(r_t) = \text{median}\left(|r_t - \text{median}(r_t)|\right). \tag{5.9}$$
The Huber estimator is then
$$\hat{\boldsymbol{\theta}}_{\text{Huber}} = \arg\min_{\boldsymbol{\theta}} s_0^2 \sum_{t=1}^{n} \rho_c\left(\frac{r_t(\boldsymbol{\theta})}{s_0}\right). \tag{5.10}$$

4. **Huber–ridge estimator** As a robust-shrinkage benchmark, we combine Huber's loss

with a ridge penalty on the amplitudes:

$$\hat{\boldsymbol{\theta}}_{\mathrm{HR}} = \arg\min_{\boldsymbol{\theta}} \left[s_0^2 \sum_{t=1}^{n} \rho_c\left(\frac{r_t(\boldsymbol{\theta})}{s_0}\right) + \lambda \sum_{j=1}^{p} (A_j^2 + B_j^2) \right], \tag{5.11}$$

where $\lambda \geq 0$ is the regularization parameter. This estimator is included to represent the robust ridge literature in the spirit of Silvapulle [27] and Maronna [19], while remaining directly comparable to the proposed estimator in the present nonlinear harmonic setting.

5. **Proposed penalized estimator** The proposed estimator applies ridge regularization to the squared-error objective:

$$\hat{\boldsymbol{\theta}}_{\mathrm{PR}} = \arg\min_{\boldsymbol{\theta}} \left[\sum_{t=1}^{n} r_t(\boldsymbol{\theta})^2 + \lambda \sum_{j=1}^{p} (A_j^2 + B_j^2) \right]. \tag{5.12}$$

The penalty shrinks the amplitude coefficients toward zero and is intended to reduce variance inflation caused by instability in the harmonic design.

5.4 Tuning parameter selection

For the penalized estimators, the regularization parameter λ is selected from the grid $\Lambda = \{0.001,\ 0.005,\ 0.01,\ 0.05,\ 0.1,\ 0.25,\ 0.5,\ 1.0,\ 2.0\}$. We use an expanding-window cross-validation procedure that preserves time ordering.

1. Define the candidate set $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_L\}$.
2. Split the series $\{y_1, \ldots, y_n\}$ into an initial training window and K ordered validation blocks.
3. For each $\lambda \in \Lambda$:
   (a) For each split $k = 1, 2, \ldots, K$:
      i. fit the penalized model using observations up to the end of the kth training window;
      ii. predict the observations in the next validation block $\mathcal{V}_k$;
      iii. compute
      $$\mathrm{MSE}_k(\lambda) = \frac{1}{|\mathcal{V}_k|} \sum_{t \in \mathcal{V}_k} (y_t - \hat{y}_t)^2 .$$
   (b) Average the validation errors: $\mathrm{CV}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \mathrm{MSE}_k(\lambda)$.
4. Select $\hat{\lambda} = \arg\min_{\lambda \in \Lambda} \mathrm{CV}(\lambda)$.

Table 2: MSE comparison of estimators for randomly t-distributed errors ($\sigma^2 = 5$) under 10% and 30% contamination.

		10% Contamination					30% Contamination				
Parameter	n	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**
A_1	500	2.15E-03	5.43E-05	3.59E-05	**4.55E-06**	**4.55E-06**	3.72E-03	6.84E-05	1.82E-03	**6.71E-06**	**3.71E-06**
	1000	1.74E-03	1.15E-05	5.19E-06	2.45E-06	**1.82E-06**	2.77E-03	1.30E-05	2.21E-04	3.83E-06	**2.51E-06**
B_1	500	3.99E-04	1.70E-04	1.06E-03	3.95E-04	**2.81E-05**	1.98E-02	6.52E-03	1.17E-03	5.65E-04	**5.65E-05**
	1000	2.93E-04	1.06E-04	6.77E-05	3.39E-05	**2.39E-05**	2.04E-04	5.61E-05	1.54E-04	4.97E-05	**1.23E-05**
ω_1	500	3.64E-10	2.00E-10	1.70E-10	**1.66E-10**	**1.66E-10**	1.74E-03	1.31E-03	1.36E-03	1.45E-03	**1.26E-03**
	1000	3.02E-10	1.08E-10	3.11E-12	**2.99E-12**	**2.99E-12**	3.02E-05	2.91E-05	**2.53E-05**	3.02E-05	**2.53E-05**
A_2	500	1.12E-04	2.56E-05	**2.44E-07**	2.49E-03	**2.44E-07**	7.80E-03	2.45E-04	**5.42E-04**	1.83E-02	**5.42E-04**
	1000	1.05E-04	1.47E-05	**8.82E-08**	1.99E-03	**8.82E-08**	2.23E-04	1.90E-04	**4.38E-05**	3.71E-03	**4.38E-05**
B_2	500	5.43E-03	2.15E-03	3.59E-05	1.83E-03	**4.55E-06**	5.90E-02	1.96E-02	**1.69E-03**	1.76E-02	**1.69E-03**
	1000	1.74E-03	1.15E-05	**1.82E-05**	1.45E-03	**1.82E-06**	2.81E-03	6.10E-04	**1.94E-05**	8.53E-03	**1.64E-05**
ω_2	500	5.43E-03	2.15E-03	3.59E-05	1.83E-03	**4.55E-06**	1.86E-06	1.58E-07	5.75E-09	**5.09E-11**	**5.09E-11**
	1000	1.74E-02	1.15E-05	**1.82E-06**	2.45E-03	**1.82E-06**	9.47E-09	1.54E-09	**7.54E-12**	2.03E-07	**7.54E-12**

Table 3: MSE comparison of estimators for t-distributed errors ($\sigma^2 = 8$) under 10% and 30% contamination in middle part.

		10% Contamination					30% Contamination				
Parameter	n	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**
A_1	500	6.11E-03	3.01E-03	8.39E-04	**9.10E-05**	**9.10E-05**	2.75E-02	1.76E-02	**6.30E-03**	9.82E-03	**6.30E-03**
	1000	2.81E-04	6.39E-05	3.59E-04	**8.45E-05**	**8.45E-05**	2.66E-03	2.18E-03	1.63E-03	2.79E-04	**4.64E-04**
B_1	500	2.34E-03	5.73E-04	5.98E-03	**3.56E-03**	**3.56E-05**	2.07E-02	1.24E-02	8.36E-03	**5.04E-03**	**5.04E-03**
	1000	1.69E-04	5.43E-06	**2.88E-04**	5.14E-04	**2.88E-08**	8.99E-04	7.27E-05	2.61E-03	5.84E-04	**6.84E-05**
ω_1	500	2.71E-04	2.55E-04	2.58E-04	2.71E-04	**2.58E-04**	2.38E-02	2.26E-02	2.38E-02	2.51E-02	**2.35E-02**
	1000	2.92E-05	2.69E-05	2.93E-05	3.02E-05	**2.92E-05**	2.70E-04	2.42E-04	2.72E-04	**2.70E-04**	**2.70E-04**
A_2	500	4.50E-03	3.88E-03	2.81E-03	2.25E-02	**1.04E-03**	1.10E-01	1.43E-02	**8.83E-02**	1.90E-01	**8.83E-02**
	1000	2.09E-04	3.22E-03	1.44E-06	7.16E-04	**4.04E-07**	1.96E-03	1.13E-03	5.54E-03	2.44E-03	**7.21E-04**
B_2	500	4.59E-03	4.35E-03	2.02E-03	1.96E-02	**3.73E-03**	1.14E-01	1.07E-01	**8.40E-02**	1.62E-01	**8.40E-02**
	1000	3.86E-03	7.00E-05	**8.63E-07**	6.89E-04	**8.63E-07**	4.76E-02	8.21E-04	4.17E-03	1.77E-03	**3.19E-04**
ω_2	500	8.41E-07	4.11E-08	1.29E-07	**1.06E-08**	**1.06E-08**	9.67E-06	3.25E-06	2.02E-06	**1.15E-07**	**1.15E-07**
	1000	1.89E-08	3.09E-09	1.99E-14	1.90E-09	**1.37E-14**	8.52E-08	5.72E-09	2.40E-07	1.89E-08	**8.07E-09**

6 Performance analysis

6.1 T-tailed

6.2 Normal Outliers

Table 4: MSE comparison of estimators for randomly G-distributed errors ($\sigma^2 = 5$) under 10% and 30% contamination. The minimum MSE in each row is highlighted in bold.

		10% Contamination					30% Contamination				
Parameter	n	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**
A_1	500	2.87E-04	2.32E-04	2.58E-04	6.17E-04	**1.15E-04**	1.01E-03	8.01E-04	1.68E-03	7.66E-04	**6.34E-04**
	1000	2.31E-04	2.16E-04	1.13E-04	**1.26E-04**	**1.06E-04**	3.86E-04	3.05E-04	8.92E-04	**3.59E-04**	**3.59E-04**
B_1	500	1.04E-04	8.79E-05	2.47E-04	1.89E-05	**5.50E-06**	7.62E-04	1.92E-04	6.58E-04	7.45E-05	**6.32E-06**
	1000	2.07E-05	1.17E-05	1.17E-05	1.15E-05	**3.23E-06**	5.03E-07	2.57E-05	1.52E-05	5.19E-05	**1.66E-07**
ω_1	500	3.01E-05	2.53E-05	3.01E-05	2.56E-05	**2.43E-05**	3.84E-05	2.77E-05	3.95E-05	3.76E-05	**2.94E-05**
	1000	2.16E-11	4.21E-12	6.78E-13	**7.68E-14**	**7.68E-14**	2.93E-11	6.06E-12	**1.11E-12**	2.56E-11	**1.12E-12**
A_2	500	1.33E-03	6.82E-05	1.09E-03	**3.37E-05**	**3.37E-05**	5.56E-03	1.36E-04	4.21E-03	6.80E-05	**2.94E-06**
	1000	1.91E-05	5.47E-06	4.71E-06	7.13E-06	**1.48E-06**	1.92E-03	1.23E-04	1.74E-04	1.29E-05	**4.57E-05**
B_2	500	3.25E-04	2.25E-04	1.64E-04	1.14E-04	**1.10E-05**	4.84E-04	2.99E-04	4.85E-04	1.83E-04	**9.86E-04**
	1000	2.03E-04	1.46E-04	7.79E-06	8.62E-06	**1.36E-06**	3.55E-04	2.35E-04	**1.63E-06**	1.45E-04	**1.63E-06**
ω_2	500	7.41E-11	8.89E-11	1.01E-08	**1.51E-11**	**1.51E-11**	8.08E-11	9.33E-11	5.99E-08	2.82E-11	**1.96E-11**
	1000	5.75E-13	2.13E-11	2.50E-13	3.59E-13	**1.73E-13**	2.84E-11	2.07E-11	1.35E-11	**1.24E-11**	1.93E-12

Table 5: MSE comparison of estimators for middle G-distributed errors ($\sigma^2 = 8$) under 10% and 30% contamination. The minimum MSE in each row is highlighted in bold.

		10% Contamination					30% Contamination				
Parameter	n	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**	**OLS**	**LAD**	**Huber_M**	**Huber_Ridge**	**Seq_Ridge**
A_1	500	6.81E-04	3.30E-05	2.17E-06	**3.72E-07**	**3.72E-07**	9.59E-03	8.12E-03	**2.83E-03**	2.21E-02	**2.83E-03**
	1000	2.50E-05	1.65E-05	1.42E-06	1.25E-07	**7.28E-08**	2.74E-03	**1.98E-04**	2.36E-03	8.67E-04	**1.98E-04**
B_1	500	1.18E-03	1.71E-07	**4.89E-08**	1.90E-04	**4.89E-08**	1.43E-02	8.67E-03	**4.90E-04**	7.79E-03	**4.90E-04**
	1000	1.03E-04	2.61E-05	**1.15E-09**	5.70E-06	**1.15E-09**	8.18E-04	1.60E-04	4.16E-04	1.40E-03	**4.78E-05**
ω_1	500	2.68E-04	**1.11E-04**	2.58E-04	2.68E-04	**1.11E-04**	**2.92E-02**	2.04E-02	2.03E-02	2.19E-02	**1.92E-02**
	1000	1.59E-10	3.58E-12	1.78E-12	3.09E-12	**1.09E-12**	3.89E-03	3.20E-03	3.44E-03	3.53E-03	**3.31E-03**
A_2	500	2.61E-03	2.01E-03	8.28E-04	3.62E-03	**3.95E-06**	6.83E-02	2.06E-01	**5.71E-02**	1.59E-01	**5.71E-02**
	1000	1.90E-03	1.57E-05	6.60E-04	1.37E-06	**3.30E-08**	1.05E-02	**2.14E-03**	2.37E-02	1.64E-02	**2.14E-03**
B_2	500	9.67E-03	3.45E-04	**3.99E-04**	1.37E-03	**3.99E-04**	6.97E-02	8.75E-02	**4.75E-02**	1.70E-01	**4.75E-02**
	1000	1.68E-04	4.05E-06	1.33E-07	**3.81E-10**	**3.81E-10**	4.69E-02	1.70E-02	2.10E-02	**1.40E-02**	**1.40E-02**
ω_2	500	2.27E-07	1.53E-08	1.61E-08	**5.97E-11**	**5.97E-11**	2.12E-06	1.21E-06	1.08E-06	**2.47E-07**	**2.47E-07**
	1000	3.03E-09	6.57E-13	1.28E-14	**6.74E-16**	**6.74E-16**	1.94E-07	1.73E-08	1.33E-07	9.52E-08	**7.73E-08**

The detailed numerical results are reported in the tables that follow; to avoid repetition, we summarize only the main points here First, the comparison should be read as a contrast between different *robustness mechanisms*: LAD targets absolute-deviation robustness, Huber-type methods target bounded-influence robustness, and ridge-type penalties target variance reduction under instability of the fitted harmonic basis For this reason, the study does not claim that ridge regularization universally outperforms LAD Rather, the appropriate interpretation is:

6.3 Discussion of simulation results

The simulation results in Tables 2–5 reveal clear patterns across contamination level, sample size, contamination variance, and estimator performance.

Contamination proportion: The MSE increases as contamination rises from 10% to 30% under both heavy-tailed and Gaussian settings The deterioration is most pronounced for OLS, reflecting its high sensitivity to outliers Robust methods (LAD and Huber) exhibit greater stability, while ridge-type estimators typically achieve lower MSE due to shrinkage-induced variance control, although the extent varies with parameters and contamination structure.

Sample size: Increasing the sample size from $n = 500$ to $n = 1000$ reduces MSE for all estimators, consistent with asymptotic theory The improvement is particularly evident for the frequency parameters ω_1 and ω_2 Robust and penalized methods show better finite-sample stability under contamination.

Contamination variance: An increase in contamination variance ($\sigma_{\text{out}} = 5$ to 8) leads to higher MSE across all methods The effect is stronger under middle-block contamination than under random contamination due to localized distortion OLS is most affected, whereas Huber and ridge-type methods demonstrate greater robustness, with penalized estimators often maintaining relatively low MSE.

Joint effects: Contamination proportion and sample size interact: larger samples mitigate, but do not eliminate, the adverse effects of contamination Under severe and high-variance contamination ($30\%, \sigma_{\text{out}} = 8$), OLS performs poorly, while robust and penalized estimators remain comparatively stable.

Method comparison: Overall performance under contamination can be summarized as

$$\text{Seq\_Ridge} \approx \text{Huber\_Ridge} \succ \text{Huber\_M} \succ \text{LAD} \succ \text{OLS}.$$

Seq_Ridge and Huber_Ridge frequently attain the smallest MSE, highlighting the role of shrinkage in stabilizing estimation under multicollinearity and contamination Huber_M performs well under heavy-tailed errors but is sometimes less stable than penalized variants LAD improves upon OLS but is not consistently optimal.

These findings do not imply universal dominance of ridge-based methods; rather, they indicate that, under the contamination designs and tuning choices considered, ridge regularization enhances stability and reduces MSE when least-squares estimation becomes unstable.

Interpretation of the Huber-ridge benchmark: The Huber-ridge estimator helps separate the effects of robustness and shrinkage Improvements over OLS indicate gains from regularization, while improvements over Huber-ridge suggest advantages beyond standard robust-shrinkage formulations.

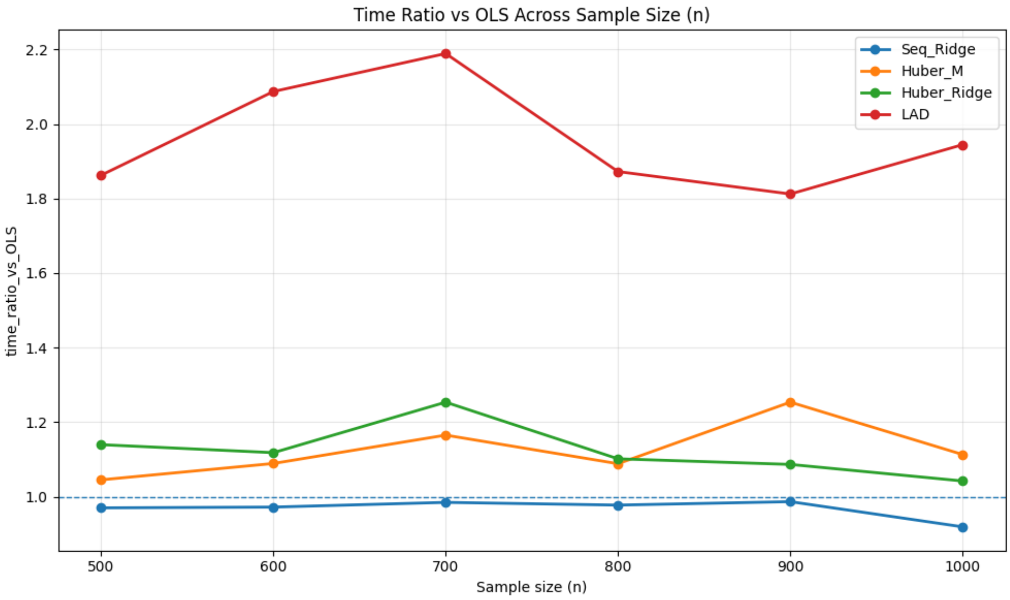


Figure 3: Computational time ratio of the competing estimators relative to OLS across different sample sizes

Computational efficiency. Figure 3 shows that LAD is the most computationally expensive due to non-smooth optimization Huber-based methods incur a moderate additional cost due to iterative weighting In contrast, Seq_Ridge remains computationally close to OLS while providing improved robustness and variance stabilization, indicating a favorable accuracy-efficiency trade-off.

7 Robustness Analysis

In this section, we study the robustness properties of different estimators used in sinusoidal regression models In particular, we focus on two fundamental robustness measures:

1. The **Influence Function (IF)**, which quantifies sensitivity to infinitesimal contamination.
2. The **Breakdown Point (BP)**, which measures resistance to large contamination.

7.1 Influence Function

The concept of the influence function was introduced by Hampel [9] as a tool to assess the local robustness of an estimator. It measures the effect of an infinitesimal contamination at point z on the estimator. Further developments and applications can be found in [11, 20]. Let F denote the true data distribution and let $T(F)$ be a statistical functional corresponding to an estimator Consider a contaminated distribution $F_{\varepsilon,z} = (1-\varepsilon)F + \varepsilon\Delta_z$, where Δ_z is a point mass at z and ε is a small contamination fraction. The influence function (IF) is defined as

$$\mathrm{IF}(z; T, F) = \lim_{\varepsilon \to 0} \frac{T(F_{\varepsilon,z}) - T(F)}{\varepsilon}. \tag{7.1}$$

Since the exact analytical IF is difficult to derive for nonlinear sinusoidal models, we adopt an **empirical influence function (EIF)** Let $\hat{\theta}$ denote the fitted parameters based on the original sample y, and $\hat{\theta}^{(z,i)}$ denote the fitted parameters when the i-th observation is replaced by z The empirical influence is measured through the perturbed fitted signal: $\Delta\hat{y}(z) = \hat{y}^{(z,i)} - \hat{y}$. We summarize the effect using the relative root mean squared change:

$$\mathrm{EIF}(z) = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{y}_t^{(z,i)} - \hat{y}_t\right)^2}}{\mathrm{sd}(\hat{y})}. \tag{7.2}$$

A smaller value of $\mathrm{EIF}(z)$ indicates greater robustness. Figure 4 shows the variation of EIF across contamination levels z for different estimators. From the figure, we observe:

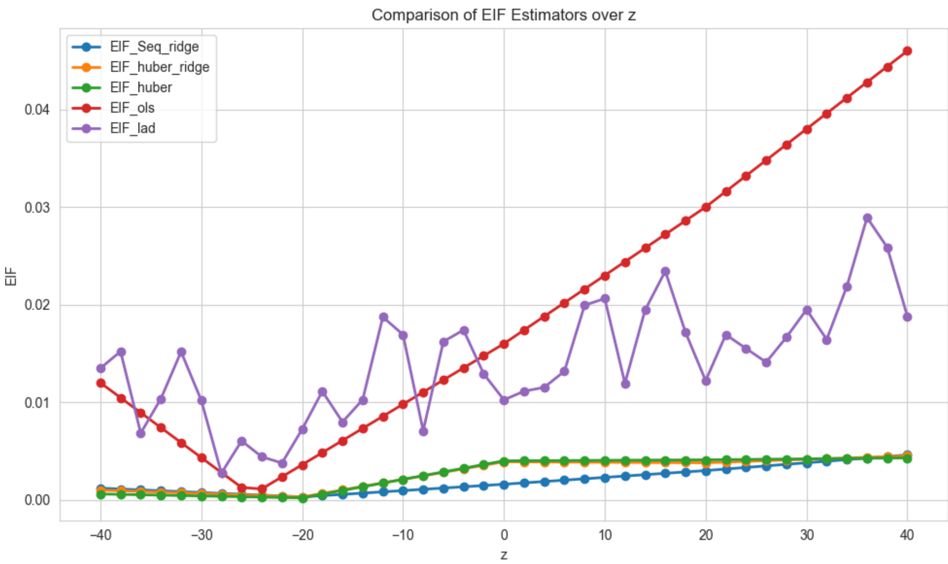


Figure 4: Empirical Influence Function comparison across estimators.

- The OLS estimator exhibits rapidly increasing influence, indicating high sensitivity to contamination.

- The LAD estimator shows moderate stability but remains relatively noisy.
- The Huber and Huber-Ridge estimators provide strong robustness with bounded influence.
- The proposed Ridge estimator (Seq_ridge) shows the lowest and smoothest influence, demonstrating superior stability.

7.2 Breakdown Behaviour

The breakdown point is a global robustness measure introduced by [6] It quantifies the smallest fraction of contamination that can cause an estimator to take arbitrarily large values A comprehensive discussion of breakdown properties in regression can be found in [20, 26]. The breakdown point of an estimator is defined as the largest fraction of contamination that the estimator can tolerate before it yields arbitrarily large errors. Formally, let $\hat{\theta}(y)$ denote an estimator and let $y^{(m)}$ denote a contaminated sample where $m = \lfloor \alpha n \rfloor$ observations are replaced by arbitrary values The finite-sample breakdown point is defined as

$$\varepsilon^* = \min \left\{ \alpha : \sup_{y^{(m)}} \|\hat{\theta}(y^{(m)})\| = \infty \right\}. \tag{7.3}$$

In practice, we operationalize breakdown by checking whether:

- parameters become excessively large,
- frequency estimates hit boundary values,
- or the fitted error explodes relative to a baseline fit.

The breakdown rate is then estimated as the probability that the estimator fails under repeated contamination at level α. Figure 5 shows the breakdown rate as a function of contamination fraction α. The results clearly indicate:

- The OLS estimator breaks down rapidly even at low contamination levels.
- The LAD estimator, although more robust than OLS, eventually breaks down at moderate contamination.
- The Huber estimator provides improved resistance but still deteriorates as contamination increases.
- The Huber-Ridge estimator enhances stability due to regularization.
- The proposed Ridge estimator (Seq_ridge) shows the slowest increase in breakdown rate, indicating the highest robustness.

Overall, the empirical analysis demonstrates that incorporating both regularization and robustness-aware parameter selection significantly improves resistance to contamination The proposed Ridge estimator consistently achieves lower influence and higher breakdown stability compared to classical methods.

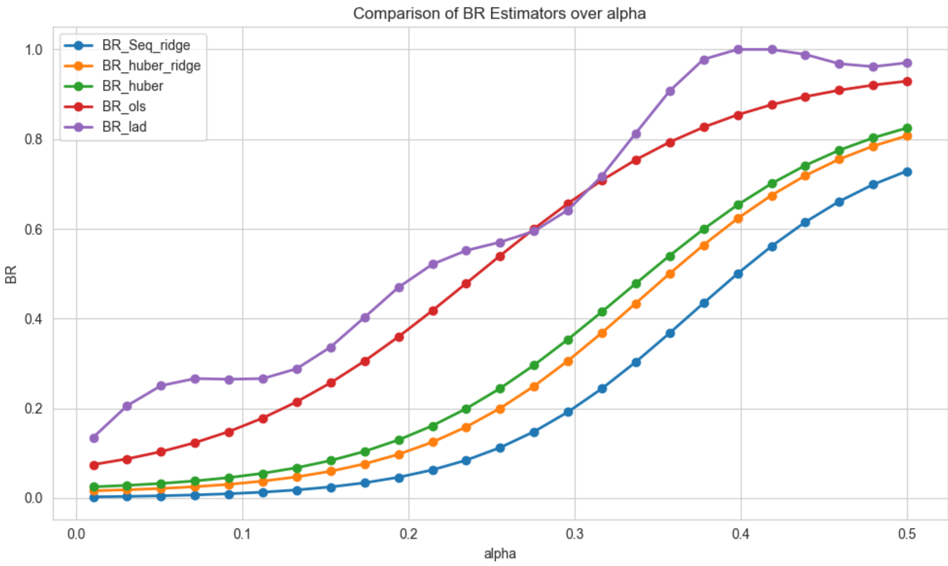


Figure 5: Finite-sample breakdown behaviour across estimators.

8 Real-data analysis: AirPassengers data

In this section, we apply the proposed robust sinusoidal modeling procedure to the classical Air Passengers dataset. The data consist of monthly totals of international airline passengers from January 1949 to December 1960, giving $n = 144$ observations The series is widely used as a benchmark for seasonal time-series modeling because it exhibits a clear upward trend and strong annual seasonality.

The purpose of this analysis is twofold First, we examine whether the sinusoidal representation provides an interpretable seasonal description of a real monthly time series Second, we evaluate the robustness of the competing estimators under a controlled contamination mechanism Importantly, the contamination considered in this section is artificially introduced by us Therefore, the contaminated data should not be interpreted as evidence that the original `AirPassengers` series contains genuine outliers Rather, the contamination experiment is designed as a reproducible stress test for assessing the ability of the proposed method to recover the underlying seasonal structure when anomalous observations are present.

Figure 6 displays the original passenger counts and the corresponding log-transformed series The raw data show an increasing trend together with increasing seasonal amplitude The logarithmic transformation reduces the changing amplitude and makes an additive trend-plus-seasonality representation more appropriate.

8.1 Log transformation and trend removal

Let P_t denote the original passenger count at time t, where $t = 1, \ldots, n$ Since the original series exhibits increasing amplitude over time, we first apply the logarithmic transformation $Z_t =$

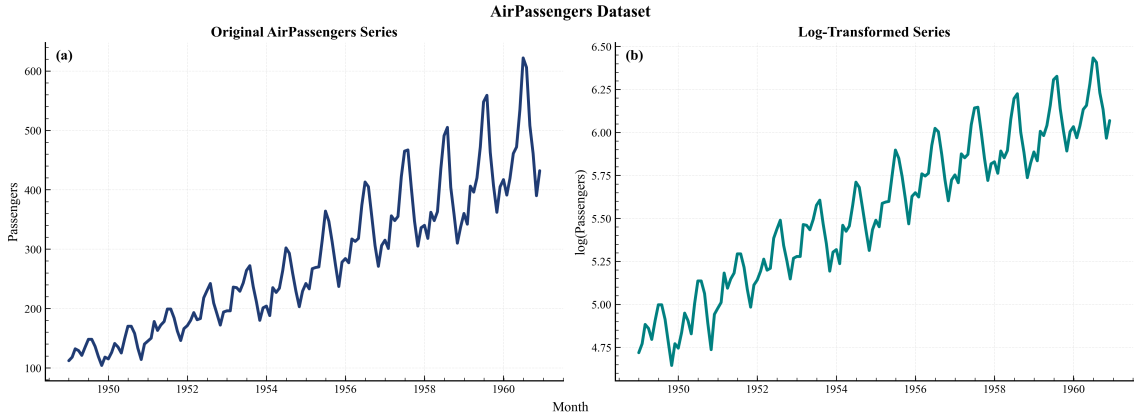


Figure 6: AirPassengers dataset. Panel (a) shows the original monthly passenger counts from January 1949 to December 1960. Panel (b) shows the corresponding log-transformed series.

$\log(P_t)$. The log transformation reduces the multiplicative effect between trend and seasonality and allows the subsequent sinusoidal model to be fitted on an approximately additive scale.

To remove the long-term growth component, we fit a linear trend to the log-transformed data:

$$Z_t = \alpha_0 + \alpha_1 t + e_t, \qquad t = 1, \dots, n, \tag{8.1}$$

where α_0 and α_1 are estimated by ordinary least squares. The estimated trend is

$$\widehat{\tau}_t = \widehat{\alpha}_0 + \widehat{\alpha}_1 t. \tag{8.2}$$

The detrended log-passenger series is then defined as

$$Y_t = Z_t - \widehat{\tau}_t. \tag{8.3}$$

Finally, the detrended series is centered by subtracting its sample mean:

$$Y_t^{(c)} = Y_t - \bar{Y}, \qquad \bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t. \tag{8.4}$$

The centered detrended series $Y_t^{(c)}$ is treated as the clean baseline seasonal signal extracted from the original data.

This preprocessing step is important because the sinusoidal model is intended to capture the periodic seasonal component rather than the long-term growth trend Thus, all subsequent sinusoidal fits are performed after log transformation, linear detrending, and centering.

8.2 Construction of the controlled contamination experiment

The original AirPassengers series is not assumed to be naturally contaminated To evaluate robustness in a controlled and reproducible manner, we artificially contaminate the clean detrended series $Y_t^{(c)}$.

Let $\mathcal{O}$ denote the set of contaminated time indices We randomly select 20% of the observations without replacement Since $n = 144$, this gives $n_{\text{out}} = \lfloor 0.20 \times 144 \rfloor = 28$ contaminated observations A fixed random seed is used to ensure reproducibility.

The contaminated signal is generated as

$$\widetilde{Y}_t = Y_t^{(c)} + \xi_t I(t \in \mathcal{O}), \tag{8.5}$$

where $I(\cdot)$ is the indicator function and ξ_t is an additive outlier shock. The magnitude of the shock is determined relative to the variability of the clean detrended signal:

$$\xi_t = S_t U_t, \qquad S_t \in \{-1, +1\}, \qquad U_t \sim \text{Unif}(3\sigma_Y, 6\sigma_Y), \tag{8.6}$$

where σ_Y is the sample standard deviation of $\{Y_t^{(c)}\}_{t=1}^n$. Thus, both positive and negative outliers are introduced, with magnitudes between three and six standard deviations of the clean detrended signal Figure 7 shows the resulting contaminated series The blue curve represents the clean detrended signal, the orange curve represents the contaminated signal, and the red points indicate the injected outlier locations This figure makes clear that the outliers are imposed by design and are not claimed to be naturally occurring anomalies in the original airline-passenger data.

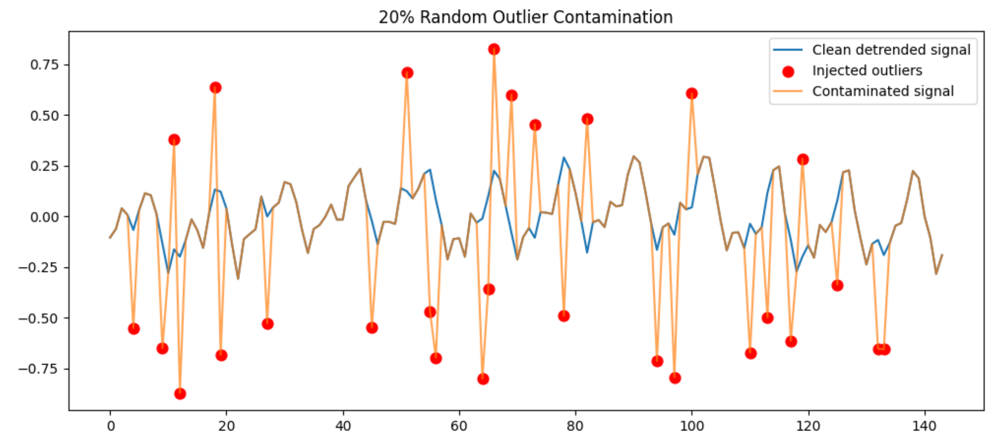


Figure 7: Controlled contamination of the detrended log AirPassengers series.

8.3 Seasonal sinusoidal model

After detrending and centering, the seasonal component is modeled by a finite sinusoidal expansion For a given number of sinusoidal components p, the fitted model is

$$Y_t^{(c)} = \beta_0 + \sum_{j=1}^{p} [a_j \cos(\omega_j t) + b_j \sin(\omega_j t)] + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{8.7}$$

where ω_j denotes the jth angular frequency, a_j and b_j are the corresponding cosine and sine coefficients, and ε_t is the error term.

For monthly data, the dominant seasonal frequency corresponds to an annual cycle Therefore, the first few harmonics have a direct interpretation in terms of seasonal periodicities In particular, for the selected value $p = 3$, the retained components correspond to annual and low-order seasonal harmonics, approximately representing periods of 12 months, 6 months, 4 months. Hence, the selected frequencies are interpretable as the annual seasonal cycle and its harmonics The additional harmonics allow the model to represent asymmetry and sharper seasonal peaks in the airline-passenger pattern.

8.4 Selection of the number of sinusoidal components

The number of sinusoidal components p is selected using information criteria We consider candidate values $p = 1, \ldots, 6$ and compute the residual sum of squares (RSS), Akaike information criterion (AIC), and Bayesian information criterion (BIC) The results are reported in Table 6.

Table 6: AIC Table

p	RSS	AIC
1	11.5288	49.0593
2	11.2746	49.8486
3	10.8574	**48.4194**
4	10.7026	50.3517
5	10.5149	51.8037
6	10.5149	55.8034

From Table 6, AIC attains its minimum value at $p = 3$, while BIC attains its minimum at $p = 1$ The BIC-selected model captures only the dominant annual seasonal component and therefore provides a more parsimonious representation. However, visual inspection of the AirPassengers series in Figure 6 indicates that the seasonal pattern is not purely sinusoidal; the yearly peaks are relatively sharp and asymmetric. For this reason, the additional harmonics selected by AIC are useful for representing the seasonal shape more accurately. We therefore take $p = 3$ for all subsequent comparisons. This choice gives a model that is still simple and interpretable, while allowing sufficient flexibility to capture annual seasonality and its low-order harmonic structure.

8.5 Selection of the ridge tuning parameter

For ridge-based robust procedures, the tuning parameter λ controls the amount of shrinkage imposed on the sinusoidal coefficients. Larger values of λ produce smoother fitted curves by shrinking the estimated coefficients toward zero, whereas smaller values of λ allow more flexible fits. The tuning parameter is selected using the procedure described in Section 5.4. Applying this procedure to the contaminated AirPassengers experiment with $p = 3$ gives $\lambda_{\text{Huber-Ridge}} = 1.09$, and $\lambda_{\text{Seq-Ridge}} = 1.57$. These values are then used in the Huber-Ridge and Sequential Ridge fits, respectively. The selected penalties provide a balance between preserving the dominant seasonal signal and reducing spurious oscillations caused by the injected outliers.

8.6 Comparison of fitted seasonal signals

We compare five estimation methods on the contaminated detrended series $\widetilde{Y}_t$: least squares estimation (LSE), least absolute deviation estimation (LAD), Huber M-estimation, Huber-Ridge estimation, and Sequential Ridge estimation. The fitted curves are shown in Figure 8. In each panel, the blue curve denotes the contaminated detrended series and the red dashed curve denotes the fitted seasonal signal.

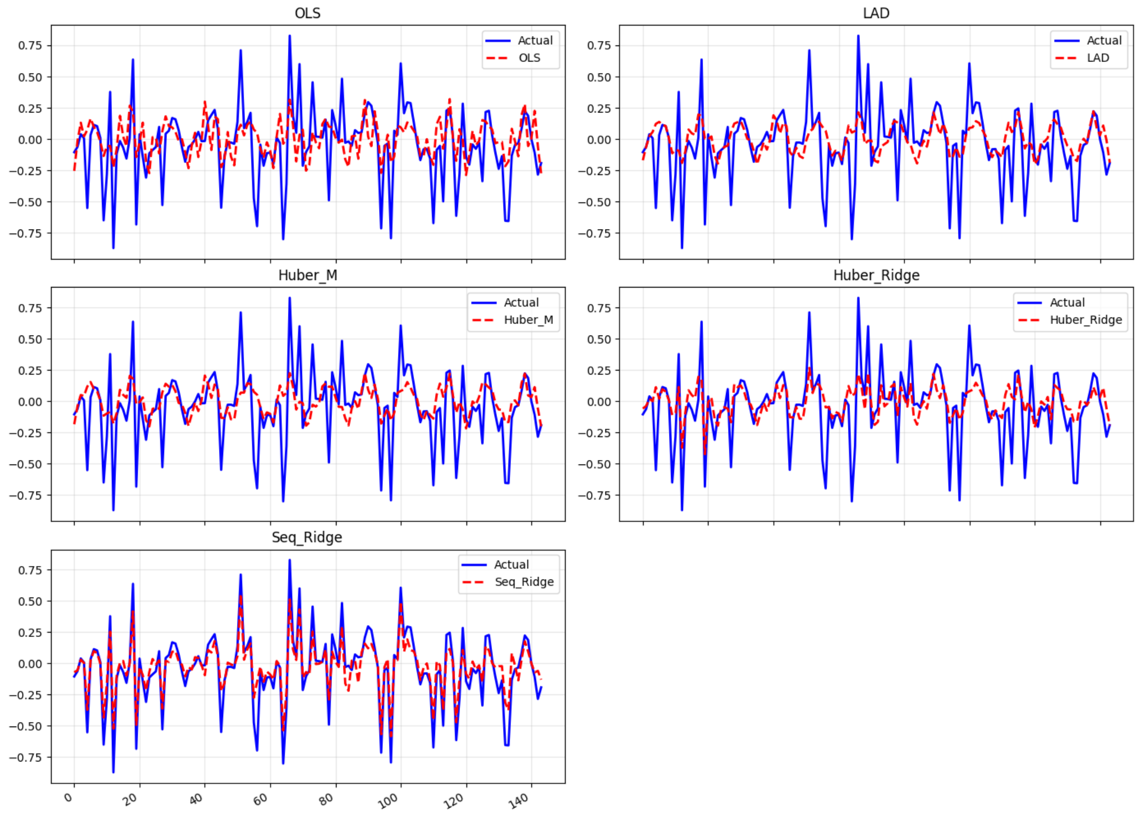


Figure 8: Fitted seasonal signals for the contaminated detrended AirPassengers series.

The main observations from Figure 8 are as follows.

1. **LSE:** The LSE fit captures the broad oscillatory structure, but it is visibly affected by the injected outliers. Since squared-error loss gives large residuals high influence, the fitted curve is pulled toward several contaminated observations. This makes the LSE fit less stable as an estimate of the underlying clean seasonal signal.
2. **LAD:** The LAD fit is less sensitive to extreme observations than LSE because it minimizes absolute deviations. The fitted curve is therefore less strongly distorted by large positive or negative shocks. However, the LAD fit can still show local irregularity when several contaminated observations occur close together.
3. **Huber M-estimation:** The Huber fit provides an intermediate behavior between LSE and LAD. It follows the main seasonal pattern while limiting the influence of large residuals. Compared with LSE, the Huber fitted curve is less attracted to extreme outliers and gives a more stable representation of the central seasonal movement.
4. **Huber-Ridge:** The Huber-Ridge fit combines residual downweighting with coefficient shrinkage. As a result, the fitted signal is smoother than the unpenalized robust fits. The ridge penalty reduces excessive oscillatory behavior caused by contaminated observations and stabilizes the estimated seasonal component.
5. **Sequential Ridge:** The Sequential Ridge fit preserves the main cyclical structure while providing stronger regularization against outlier-driven fluctuations. Compared with LSE, LAD, and Huber estimation, the fitted curve is more controlled and less reactive to individual anomalous points. This indicates that ridge regularization improves recovery of the underlying seasonal signal under contamination.

Overall, Figure 8 shows that robust methods reduce the influence of artificial outliers, while ridge-based methods further smooth and stabilize the fitted seasonal signal. Thus, ridge regularization changes the fitted curve in a practically meaningful way: it discourages the model from following contaminated observations too closely and produces a more regular estimate of the seasonal component.

9 Conclusion

This paper proposed a ridge-type regularized estimator for sinusoidal regression models, mainly as a shrinkage-based stabilization method rather than a classical robust estimator. For fixed-dimensional models with unknown frequencies, consistency and asymptotic normality were established under suitable assumptions. Simulation studies and the AirPassengers example show that ridge regularization can improve stability and reduce MSE under the considered contamination settings. Future work may study other penalties and formal robustness properties such as influence functions and breakdown behavior.

10 Declarations

Competing interests : The authors declare that they have no competing interests.

References

[1] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2006). *Nonlinear programming: theory and algorithms*. John wiley & sons.

[2] Caton, R. C., Pavlis, G. L., Thomson, D. J., and Vernon III, F. L. (2020). Methods for the robust computation of the long-period seismic spectrum of broad-band arrays. *Geophysical Journal International*, 222(3):1480–1501.

[3] Chan, Y., Lavoie, J., and Plant, J. (1981). A parameter estimation approach to estimation of frequencies of sinusoids. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(2):214–219.

[4] Chatfield, C. (1977). Introduction to statistical time series.

[5] De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

[6] Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. *A Festschrift for Erich L. Lehmann*, pages 157–184.

[7] Fiszeder, P., Małecka, M., and Molnár, P. (2024). Robust estimation of the range-based garch model: Forecasting volatility, value at risk and expected shortfall of cryptocurrencies. *Economic Modelling*, 141:106887.

[8] Grané, A. and Veiga, H. (2014). Outliers, garch-type models and risk measures: A comparison of several approaches. *Journal of Empirical Finance*, 26:26–40.

[9] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons.

[10] Hastie, T., Tibshirani, R., Friedman, J., et al. (2009). The elements of statistical learning.

[11] Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*. John Wiley & Sons, 2nd edition.

[12] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

[13] Kim, T.-S., Kim, H.-K., and Park, S.-H. (2000). Asymptotic properties of lad esimators of a nonlinear time series regression model. *Journal of the Korean Statistical Society*, 29(2):187–199.

[14] Kundu, D. (1997). Asymptotic theory of the least squares estimators of sinusoidal signal. *Statistics: A Journal of Theoretical and Applied Statistics:*, 30(3):221–238.

[15] Kundu, D. and Mitra, A. (1995). Estimating the parameters of exponentially damped/undamped sinusoids in noise: A non-iterative approach. *Signal processing*, 46(3):363–368.

[16] Kundu, D. and Mitra, A. (1996). Asymptotic theory of least squares estimator of a nonlinear time series regression model. *Communications in Statistics-Theory and Methods*, 25(1):133–141.

[17] Lim, C. (2015). Robust ridge regression estimators for nonlinear models with applications to high throughput screening assay data. *Statistics in medicine*, 34(7):1185–1198.

[18] Liu, W., Chen, R., Cai, H., and Luo, W. (2016). Robust statistical methods for impulse noise suppressing of spread spectrum induced polarization data, with application to a mine site, gansu province, china. *Journal of Applied Geophysics*, 135:397–407.

[19] Maronna, R. A. (2011). Robust ridge regression for high-dimensional data. *Technometrics*, 53(1):44–53.

[20] Maronna, R. A., Martin, D. R., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)*. John Wiley & Sons.

[21] Melman, T. and Victor, J. D. (2016). Robust power spectral estimation for eeg data. *Journal of neuroscience methods*, 268:14–22.

[22] Nandi, S. (2012). Estimation of parameters of two-dimensional sinusoidal signal in heavy-tailed errors. *Journal of Statistical Planning and Inference*, 142(10):2799–2808.

[23] Nandi, S., Prasad, A., and Kundu, D. (2010). An efficient and fast algorithm for estimating the parameters of two-dimensional sinusoidal signals. *Journal of statistical planning and inference*, 140(1):153–168.

[24] Olaluwoye, O. T., Lukman, A. F., Alrasheedi, M. A., Nzomo, W. N., and Farghali, R. A. (2025). Robust estimation methods for addressing multicollinearity and outliers in beta regression models. *Scientific Reports*, 15(1):11649.

[25] Prasad, A., Kundu, D., and Mitra, A. (2008). Sequential estimation of the sum of sinusoidal model parameters. *Journal of Statistical Planning and Inference*, 138(5):1297–1313.

[26] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. John Wiley & Sons.

[27] Silvapulle, M. J. (1991). Robust ridge regression based on an m-estimator. *Australian Journal of Statistics*, 33(3):319–333.

[28] Varin, S. and Panagiotakos, D. B. (2020). A review of robust regression in biomedical science research. *Archives of Medical Science*, 16(5):1267–1269.

[29] von Storch, H. and Zwiers, F. W. (1999). *Statistical Methods in Climate Research.* Cambridge University Press, Cambridge.

[30] Von Storch, H. and Zwiers, F. W. (2002). *Statistical analysis in climate research.* Cambridge university press.

[31] Yeap, B. Y. and Davidian, M. (2001). Robust two-stage estimation in hierarchical nonlinear models. *Biometrics*, 57(1):266–272.

Appendix A

Proof 10.1 *We work conditionally on the design matrix X and take expectations with respect to the noise ε. By definition,*

$$\hat{\theta}_{\lambda_n} = (S_n + \lambda_n I_{P_n})^{-1}\Big(S_n\theta + \frac{1}{n}X^\top\varepsilon\Big), \quad \text{where } S_n = \frac{1}{n}X^\top X.$$

Hence

$$\begin{aligned}\hat{\theta}_{\lambda_n} - \theta &= (S_n + \lambda_n I)^{-1}\Big(S_n\theta + \tfrac{1}{n}X^\top\varepsilon\Big) - \theta \\ &= \big[(S_n + \lambda_n I)^{-1}S_n - I\big]\theta + (S_n + \lambda_n I)^{-1}\frac{1}{n}X^\top\varepsilon.\end{aligned}$$

Using

$$(S_n + \lambda_n I)^{-1}S_n - I = -\lambda_n(S_n + \lambda_n I)^{-1},$$

we obtain the standard bias + noise *decomposition:*

$$\hat{\theta}_{\lambda_n} - \theta = -\lambda_n(S_n + \lambda_n I)^{-1}\theta + (S_n + \lambda_n I)^{-1}\frac{1}{n}X^\top\varepsilon.$$

Define $b_n := -\lambda_n(S_n + \lambda_n I)^{-1}\theta, \qquad \eta_n := (S_n + \lambda_n I)^{-1}\frac{1}{n}X^\top\varepsilon$, so that $\hat{\theta}_{\lambda_n} - \theta = b_n + \eta_n$. The (normalized) prediction error can be written as

$$\frac{1}{n}\big\|X(\hat{\theta}_{\lambda_n} - \theta)\big\|^2 = (\hat{\theta}_{\lambda_n} - \theta)^\top S_n(\hat{\theta}_{\lambda_n} - \theta) = (b_n + \eta_n)^\top S_n(b_n + \eta_n).$$

Taking expectation with respect to ε and using $\mathbb{E}[\eta_n] = 0$ and $\mathbb{E}[\eta_n b_n^\top] = 0$, we obtain

$$\begin{aligned}\frac{1}{n}\mathbb{E}\big\|X(\hat{\theta}_{\lambda_n} - \theta)\big\|^2 &= b_n^\top S_n b_n + \mathbb{E}\big[\eta_n^\top S_n \eta_n\big] \\ &=: B_n + V_n.\end{aligned}$$

Thus, we need to control the bias term B_n and the variance term V_n. Recall $b_n = -\lambda_n(S_n + \lambda_n I)^{-1}\theta$. Hence

$$B_n = b_n^\top S_n b_n = \lambda_n^2\,\theta^\top(S_n + \lambda_n I)^{-1}S_n(S_n + \lambda_n I)^{-1}\theta.$$

Using the spectral bounds on S_n and on $(S_n+\lambda_n I)^{-1}$, there exists a constant $C_1 > 0$ independent of n such that

$$B_n \le C_1\lambda_n^2\,\|\theta\|^2.$$

By assumption, $\|\theta\|^2 = O(P_n)$ and $\lambda_n \asymp P_n/n$, so

$$B_n \le C_2\frac{P_n^2}{n^2}\,P_n = C_2\,\frac{P_n^3}{n^2},$$

for some constant $C_2 > 0$. Under the stated growth regime and the eigenvalue conditions, this term can be shown to vanish, i.e. $B_n \to 0$ as $n \to \infty$. We have

$$\eta_n = (S_n + \lambda_n I)^{-1} \tfrac{1}{n} X^\top \varepsilon,$$

and, using $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$*,*

$$\begin{aligned}\mathrm{Var}(\eta_n) &= (S_n + \lambda_n I)^{-1} \mathrm{Var}\left(\tfrac{1}{n} X^\top \varepsilon\right)(S_n + \lambda_n I)^{-1} \\ &= (S_n + \lambda_n I)^{-1} \Big(\tfrac{\sigma^2}{n} S_n\Big)(S_n + \lambda_n I)^{-1}.\end{aligned}$$

Hence

$$\begin{aligned}V_n &= \mathbb{E}[\eta_n^\top S_n \eta_n] = \mathrm{tr}\left(S_n \mathrm{Var}(\eta_n)\right) \\ &= \frac{\sigma^2}{n} \mathrm{tr}\left[S_n (S_n + \lambda_n I)^{-1} S_n (S_n + \lambda_n I)^{-1}\right].\end{aligned}$$

Using again the spectral bounds on S_n and $(S_n + \lambda_n I)^{-1}$, there exists a constant $C_3 > 0$ such that

$$V_n \leq C_3 \frac{1}{n} \mathrm{tr}(S_n) = C_3 \frac{1}{n} \sum_{j=1}^{P_n} \lambda_j(S_n) \leq C_4 \frac{P_n}{n}$$

for some $C_4 > 0$. Since $P_n/n \to \kappa \in (0, \infty)$, this term is of order P_n/n and can be controlled suitably in the high-dimensional limit. Combining the bounds for B_n and V_n, we have

$$\frac{1}{n} \mathbb{E}\big\| X(\hat{\theta}_{\lambda_n} - \theta)\big\|^2 = B_n + V_n \longrightarrow 0,$$

under the stated growth conditions on P_n, λ_n and $\|\theta\|^2$, together with the spectral assumptions on S_n.

Hence the ridge estimator $\hat{\theta}_{\lambda_n}$ is prediction-consistent:

$$\frac{1}{n} \mathbb{E}\big\| X(\hat{\theta}_{\lambda_n} - \theta)\big\|^2 \to 0.$$

Appendix B

Proof of Theorem (4.1).
Let $\hat{\theta}_N = (\hat{A}_N, \hat{B}_N, \hat{\omega}_N)$ to be the ridge estimates of $\theta^0 = (A^0, B^0, \omega^0)$ obtained by minimizing

$$Q_N^R(\theta) = \sum_{i=1}^{N} (Z(t) - A\cos(\omega t) - B\sin(\omega t))^2 + \hat{\lambda}_N \left(A^2 + B^2\right) \tag{10.1}$$

with respect to $\theta = (A, B, \omega)$. Here $\hat{\lambda}_N$ is the regularization parameter under ridge regression chosen optimally by cross-validation. Let us now consider

$$\begin{aligned}
\frac{1}{N}\left[Q_N^R(\theta) - Q_N^R(\theta^0)\right] &= \frac{1}{N}\sum_{i=1}^{N}(Z(t) - A\cos(\omega t) - B\sin(\omega t))^2 + \frac{\hat{\lambda}_N}{N}\left(A^2 + B^2\right) \\
&- \frac{1}{N}\sum_{t=1}^{N}(e(t))^2 - \frac{\hat{\lambda}_N}{N}\left(A^{0^2} + B^{0^2}\right) \\
&= \frac{1}{N}\sum_{t=1}^{N}\left(A^0\cos(\omega^0 t) + B^0\sin(\omega^0 t) - A\cos(\omega t) - B\sin(\omega t)\right)^2 \\
&\quad + \frac{2}{N}\sum_{t=1}^{N}e(t)\left(A^0\cos(\omega^0 t) + B^0\sin(\omega^0 t) - A\cos(\omega t) - B\sin(\omega t)\right) \\
&\quad + \frac{\hat{\lambda}_N}{N}\left[(A^2 - A^{0^2}) + (B^2 - B^{0^2})\right] \\
&= f_N^R(A, B, \omega) + g_N^R(A, B, \omega) + h_N^R(A, B, \omega) \qquad (10.2)
\end{aligned}$$

Now with the help of lemma 2.1 and calculation done in [14], we can easily conclude that

$$\lim_{N\to\infty}\sup_{\theta\in S_{\delta,M}} g_N^R(A, B, \omega) = 0 \ \text{ a.s.} \qquad (10.3)$$

where $\delta > 0$ the set $S_{\delta,M}$ is defined as

$$\begin{aligned}
S_{\delta,M} = &\Big\{(A, B, \omega); |A - A^0| \geq \delta, |A| \leq M, |B| \leq M \\
&\text{or } |B - B^0| \geq \delta, |A| \leq M, |B| \leq M \\
&\text{or } |\omega - \omega^0| \geq \delta, |A| \leq M, |B| \leq M\Big\}. \qquad (10.4)
\end{aligned}$$

Now from [16] it follows that for all $\delta > 0$

$$\lim_{N\to\infty}\sup_{\theta\in S_{\delta,M}} f_N(A, B, \omega) > 0. \qquad (10.5)$$

Also all $\delta > 0$ from 10.4 it follows that

$$\lim_{N\to\infty}\sup_{\theta\in S_{\delta,M}} h_N(A, B, \omega) > 0. \qquad (10.6)$$

Therefore from (10.3) - (10.6) we can conclude that

$$\lim_{N\to\infty}\inf_{S_{\delta,M}} \frac{1}{N}[Q_N^R(\theta) - Q_N^R(\theta^0)] > 0 \qquad (10.7)$$

Now suppose $(\hat{A}_0, \hat{B}_0, \hat{\omega}_0)$ be the Ridge estimates of (A^0, B^0, ω^0) and they are not consistent. Therefore either

Case I: For all sub-sequences $\{N_k\}$ of $\{N\}$, $|\hat{A}_{N_k}| + |\hat{B}_{N_k}|$ tends to infinity or

Case II: There exists a $\delta > 0$ and a $M < \infty$ and a sub-sequence $\{N_k\}$ such that $(\hat{A}_{N_k}, \hat{B}_{N_k}, \hat{\omega}_{N_k}) \in S_{\delta,M}$ for all $k = 1, 2, \ldots$. Now

$$Q^R_{N_k}(\hat{A}_{N_k}, \hat{B}_{N_k}, \hat{\omega}_{N_k}) - Q^R_{N_k}(A^0, B^0, \omega^0) \leq 0 \tag{10.8}$$

as $(\hat{A}_{N_k}, \hat{B}_{N_k}, \hat{\omega}_{N_k})$ is the ridge regularized estimates of (A_0, B_0, ω_0) when $N = N_k$. Observe that as $K \to \infty$, for both cases, the left-hand side of (10.8) converges to a strictly positive number, that is a contradiction. Therefore, the ridge regularized estimates of the model (4.7) have to be strongly consistent.

Appendix C: Proof of Theorem 4.2

Let

$$Q'_N(\theta) = \left(\frac{\partial Q^R_N(\theta)}{\partial A}, \frac{\partial Q^R_N(\theta)}{\partial B}, \frac{\partial Q^R_N(\theta)}{\partial \omega} \right),$$

and denote by $Q''_N(\theta)$ the corresponding 3×3 Hessian matrix.

Using a Taylor series expansion around θ^0, we obtain

$$Q'_N(\hat{\theta}) - Q'_N(\theta^0) = (\hat{\theta} - \theta^0)\, Q''_N(\bar{\theta}), \tag{10.9}$$

where $\bar{\theta}$ lies on the line segment joining $\hat{\theta}$ and θ^0. Since $Q'_N(\hat{\theta}) = 0$, equation (10.9) yields

$$(\hat{\theta} - \theta^0) = -Q'_N(\theta^0)\, [Q''_N(\bar{\theta})]^{-1}. \tag{10.10}$$

Define the scaling matrix

$$\mathcal{D} = \begin{bmatrix} N^{-1/2} & 0 & 0 \\ 0 & N^{-1/2} & 0 \\ 0 & 0 & N^{-3/2} \end{bmatrix}.$$

Multiplying both sides of (10.10) by $\mathcal{D}^{-1}$ gives

$$(\hat{\theta} - \theta^0)\mathcal{D}^{-1} = -Q'_N(\theta^0)\mathcal{D} \left[\mathcal{D} Q''_N(\bar{\theta})\mathcal{D}\right]^{-1}. \tag{10.11}$$

From (4.26)–(4.31) and the assumption $\lambda_N \to \lambda \in [0, \infty)$, we have

$$\mathcal{D} Q''_N(\bar{\theta})\mathcal{D} \xrightarrow{P} 2\Sigma(\lambda),$$

where

$$\Sigma(\lambda) = \begin{bmatrix} \frac{1}{2}+\lambda & 0 & \frac{1}{4}B^0 \\ 0 & \frac{1}{2}+\lambda & -\frac{1}{4}A^0 \\ \frac{1}{4}B^0 & -\frac{1}{4}A^0 & \frac{1}{6}(A^{0^2}+B^{0^2}) \end{bmatrix}.$$

Since $16\lambda^2+10\lambda+1 > 0$ for all $\lambda \geq 0$, the matrix $\Sigma(\lambda)$ is nonsingular whenever $A^{0^2}+B^{0^2} > 0$, and its inverse is given by

$$\Sigma^{-1}(\lambda) = \frac{1}{(A^{0^2}+B^{0^2})(16\lambda^2+10\lambda+1)}$$

$$\times \begin{bmatrix} 2(8\lambda+1)(A^{0^2}+4B^{0^2}) & -6A^0B^0 & -12B^0 \\ -6A^0B^0 & 2(8\lambda+1)(4A^{0^2}+B^{0^2}) & 12A^0 \\ -12B^0 & 12A^0 & 24(2\lambda+1) \end{bmatrix}.$$

Next, by the central limit theorem for stationary stochastic processes (see [4]), we have

$$Q_N'(\theta^0)\mathcal{D} \xrightarrow{d} \mathcal{N}_3\Big(0,\, 4\sigma^2\gamma\Sigma(\lambda)\Big),$$

where

$$\gamma = \left| \sum_{j=-\infty}^{\infty} \alpha(j)\cos(\omega^0 j) \right|^2 + \left| \sum_{j=-\infty}^{\infty} \alpha(j)\sin(\omega^0 j) \right|^2.$$

Combining this with (10.11) and applying Slutsky's theorem, we obtain

$$(\hat{\theta}-\theta^0)\mathcal{D}^{-1} \xrightarrow{d} \mathcal{N}_3\big(0,\, \sigma^2\gamma\Sigma^{-1}(\lambda)\big).$$

This completes the proof. $\square$

Remark. If λ_N diverges to infinity, the matrix $\Sigma^{-1}(\lambda_N)$ collapses to zero, leading to a degenerate limiting distribution. Hence, the assumption $\lambda_N \to \lambda \in [0,\infty)$ is essential for obtaining a non-degenerate asymptotic normal distribution.

Appendix D

Proof of the Theorem 4.3.

With the help of Lemma 2.1 and using similar techniques as that of ([15]), the results can be established.

Let's denote the $1 \times 3M$ vector $\mathcal{R}'_N(\Phi)$ as follows:

$$\mathcal{R}'_N(\Phi) = \left(\frac{\partial \mathcal{R}_N(\Phi)}{\partial \mathbf{A}}, \frac{\partial \mathcal{R}_N(\Phi)}{\partial \mathbf{B}}, \frac{\partial \mathcal{R}_N(\Phi)}{\partial \omega} \right)$$

and $R''_N(\theta)$ denotes the $3M \times 3M$ matrix which contains the double derivative of $R_N(\Phi)$. Now we have

$$R'_N(\hat{\Phi}_k) - R'_N(\Phi_k) = (\hat{\Phi}_k - \Phi_k) R''_N(\bar{\Phi}) \tag{10.12}$$

where $\bar{\Phi} = (\bar{A}, \bar{B}, \bar{\omega})$ is a point in the line joining $\hat{\Phi}_k$ and Φ_0. Since $R'_N(\hat{\Phi}_N) = 0$, we have

$$(\hat{\Phi}_k - \Phi_k) = -R'_N(\Phi_k)[R''_N(\bar{\Phi})]^{-1}. \tag{10.13}$$

Let's define the $3M \times 3M$ diagonal matrix V whose first $2M$ diagonal elements are $N^{-1/2}$ and the last M diagonal elements are $N^{-3/2}$. Therefore, we can write (10.13) as

$$(\hat{\Phi}_k - \Phi_k)V^{-1} = -R'_N(\Phi_k)V^{-1}\left[V R''_N(\bar{\Phi})V^{-1}\right]^{-1}. \tag{10.14}$$

Now, using similar arguments as in Section 3, we can say that.

$$R'_N(\Phi_k)V \to N_{3M}(0, 4\sigma^2 \mathcal{G})$$

where $\mathcal{G}$ is a $3M \times 3M$ matrix and it has the following structure:

$$\mathcal{G} = \begin{bmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} & \mathcal{G}_{13} \\ \mathcal{G}_{21} & \mathcal{G}_{22} & \mathcal{G}_{23} \\ \mathcal{G}_{31} & \mathcal{G}_{32} & \mathcal{G}_{33} \end{bmatrix} \tag{10.15}$$

where each of the $\mathcal{G}_{ij}$ is an $M \times M$ matrix and

$$
\begin{aligned}
\mathcal{G}_{11} &= \mathcal{G}_{22} = \operatorname{diag}\left\{\left(\frac{1}{2}+\hat{\lambda}_1\right)c_1,\cdots,\left(\frac{1}{2}+\hat{\lambda}_M\right)c_M\right\}\\
\mathcal{G}_{13} &= \mathcal{G}_{31} = \operatorname{diag}\left\{\frac{1}{4}B_1^0 c_1,\cdots,\frac{1}{4}B_M^0 c_M\right\}\\
\mathcal{G}_{23} &= \mathcal{G}_{32} = -\operatorname{diag}\left\{\frac{1}{4}A_1^0 c_1,\cdots,\frac{1}{2}A_M^0 c_M\right\}\\
\mathcal{G}_{33} &= \frac{1}{6}\operatorname{diag}\left\{d_1,\cdots,d_M\right\}\\
\mathcal{G}_{12} &= \mathcal{G}_{21} = 0
\end{aligned}
$$

where

$$
c_k = \left|\sum_{j=-\infty}^{\infty}\alpha(j)\cos(\omega_k^0 j)\right|^2 + \left|\sum_{j=-\infty}^{\infty}\alpha(j)\sin(\omega_k^0 j)\right|^2
$$

are and $d_K = c_K\left[(A_K^0)^2 + (B_K^0)^2\right]$ for $K = 1,\ldots,M$. Observe that

$$
\lim_{N\to\infty} VR_N''(\bar{\Phi})V = \lim_{N\to\infty} VR_N''(\Phi_k)V = 2\Gamma
$$

Here the $3M \times 3M$ matrix Γ is

$$
\Gamma = \begin{bmatrix} (\frac{1}{2}+\hat{\lambda}_M)I_M & 0 & S_1 \\ 0 & (\frac{1}{2}+\hat{\lambda}_M)\frac{1}{2}I_M & S_2 \\ S_1 & S_2 & S_3 \end{bmatrix}
$$

where S_1, S_2, S_3 are $M \times M$ diagonal matrices as follows:

$$
S_1 = \frac{1}{4}\operatorname{diag}\{B_1^0,\ldots,B_M^0\}
$$

$$
S_2 = -\frac{1}{4}\operatorname{diag}\{A_1^0,\ldots,A_M^0\}
$$

$$
S_3 = \frac{1}{6}\operatorname{diag}\{d_1,\ldots,d_M\}
$$

I_M is the identity matrix of order M.