Improvement in Modified Least Squares Estimation for Fitting a Sinusoidal Regression Model with AR (1) Error

Wannapa Pukdee* and Atchara Namburi

Faculty of Sciences and Engineering, Kasetsart University Chalermphrakiat Sakon Nakhon Province Campus, Sakon Nakhon, 47000, Thailand

* Corresponding author. E-mail address: wannapa.p@ku.th

Received: 24 March 2021; Revised: 7 May 2021; Accepted: 13 May 2021

Abstract

Sinusoidal functions are widely used in many areas, such as physics, engineering, and gene expression to describe correlated data along with time. A sinusoidal model with correlated error is fitted using a modified two-stage least squares method by modifying the weight matrix of the correlation coefficient based on residuals from the one-way ANOVA model proposed by Pukdee, Polsen, and Baksh (2020). By using that modification, a conditional least squares model with the AR (1) error is modified and proposed as an alternative method. A Monte Caro simulation study is made of an effect of error mis-specifications and this finding might be beneficial for some applications.

Keywords: autoregressive process, conditional least squares, one-way ANOVA model

Introduction

To analyze data collected over time, a nonlinear function $\mathbf{f}(\mathbf{t}_i;\boldsymbol{\theta})$ with correlated errors $\boldsymbol{\varepsilon}_i$ is widely used,

$$\mathbf{y}_i = \mathbf{f}(\mathbf{t}_i;\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_i;\ i = 1,\ldots,r, \tag{1}$$

where $\mathbf{y}_i = (y_{i,1},\ldots,y_{i,n})'$ is a vector of response observations at a time vector $\mathbf{t}_i = (t_{i,1},\ldots,t_{i,n})'$ set as an independent variable vector, $\boldsymbol{\theta}$ is an unknown parameter vector for the i^{th} dataset and each observed data for subject i is measured for r replicates. By assuming repeated measures at each n time point, the correlated error vector $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1},\ldots,\varepsilon_{i,n})'$ is based on a stationary autoregressive process of order p, AR(p) (Asikgil and Erar, 2013),

$$\varepsilon_{ij} = \delta_{ij} + \rho_1\varepsilon_{i,j-1} + \rho_2\varepsilon_{i,j-2} + \ldots + \rho_p\varepsilon_{i,j-p};\ \ j = 1,\ldots,n, \tag{2}$$

where let $(\rho_1,\ldots,\rho_p)' \in [-1,1]$ be the correlation coefficient vector and δ_{ij} be independent and identically distributed (IID) errors with mean 0 and variance σ^2. In the case of $p = 1$ in (2), the AR (1) process of the error vector has a vector of mean $\mathbf{0}$ and the variance-covariance matrix $\sigma^2 \boldsymbol{V}_i$ which

$$\boldsymbol{V}_i = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix},$$

is the $n \times n$ known matrix and the inverse matrix $\boldsymbol{V}_i^{-1} = \boldsymbol{R}_i' \boldsymbol{R}_i$ by using a least squares transformation (Seber and Wild, 2003). Here, this is a weight matrix,

$$\boldsymbol{R}_i = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}, \tag{3}$$

and it can yield the IID normal error vector $\boldsymbol{\delta}_i = \boldsymbol{R}_i \boldsymbol{\varepsilon}_i \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{I}_i\right)$.

The transformation methods applied to fit nonlinear regression models, with correlated errors assumed to be AR (1) errors, are conditional least squares (CLS) estimation (Bates and Watts, 1988) and a two-stage least squares (TSLS) method (Seber and Wild, 2003). Pukdee, Polsen, and Baksh (2018) presented the TSLS and CLS methods for estimating the period parameters of sinusoidal models whether fitted models are correctly specified or not correctly specified with the weight AR (1) matrix. They found that the TSLS method produces more underestimated standard errors than the standard deviation for the period parameter estimates. To overcome that problem, Pukdee, Polsen, and Baksh (2020) modify the two-stage least squares (MTSLS) method by using residuals from the one-way ANOVA model and calculating the correlation coefficient into the weight matrix. They reported that the standard error for the period parameter of sinusoidal models was more accurate.

Consequently, the objective of this article is to apply the correlation coefficient based on the one-way ANOVA model for modifying the CLS model. The modified CLS (MCLS) method is an alternative approach and is used to estimate all parameters of a sinusoidal model where correlated errors are both correctly and incorrectly specified. In addition, this study presents when the true value of the correlation coefficient is known and is used in the TSLS method called TrueTSLS. Sinusoidal regression models which display cyclical patterns are widely used to analyze data in many areas; for example, an engineering study presented by Liang, Ren, Sun, and Zhu (2018) transformed the three parameters of a sinusoidal curve model into a one-dimension search for a frequency parameter. In gene expression research, Izumo, Johnson, and Yamazaki (2003) studied the relationship between circadian rhythms and the temperature fitted by a sine wave. In the literature, sinusoidal regression models play the main rule in the frequency or period parameter of correlated data over time. This can be used to predict the time interval for the process to repeat itself for experimental units after treatment, such as body's response to the effects of drugs. Sinusoidal models are fitted by the methods and then are compared based on simulations.

Methodology

Sinusoidal Regression Functions

Sinusoidal regression functions displaying cyclic patterns are used to analysis data. To describe that pattern, the components of simple oscillations consist of amplitude, period, and phase. The amplitude is the expected maximum value from the response. The period explains the time needed for the response process to repeat itself. The phase is the time point at the beginning of the cycle. For example, responses of circadian gene expression are measurements of light intensity over time. Therefore, to analyze the responses, Pukdee, Polsen, and Baksh (2018) used the sinusoidal function modified from Kyriacou and Hall (1980) by adding a linear trend as,

$$f(t_{ij};\boldsymbol{\theta}) = \alpha + \beta t_{ij} + (a_s + a\exp(-dt_{ij}))\sin(\frac{2\pi t_{ij}}{\tau} + \Phi), \quad (4)$$

where the vector of unknown parameters is $\boldsymbol{\theta} = (\alpha, \beta, a, a_s, d, \tau, \Phi)'$, α and β are an intercept and a slope of the regression line, respectively, a_s is the amplitude adjustment that can be increased or decreased by using $a(\exp - dt)$, in which a is the amplitude, d is a damping parameter, τ is the period, and Φ is the phase of the sine curve.

Two-stage Least Squares Method with True ρ^*

If the model (1) with the AR (1) error is transformed under the IID error process, $\boldsymbol{\delta}_i \sim (\mathbf{0}, \sigma^2 \boldsymbol{I}_i)$, this can build a new model, presented by Seber and Wild (2003) as,

$$\mathbf{z}_i = \mathbf{g}(\mathbf{t}_i;\boldsymbol{\theta}) + \boldsymbol{\delta}_i;\ i = 1,\ldots,r, \quad (5)$$

where $\mathbf{z}_i = \boldsymbol{R}_i\mathbf{y}_i$, $\mathbf{g}(\mathbf{t}_i;\boldsymbol{\theta}) = \boldsymbol{R}_i\mathbf{f}(\mathbf{t}_i;\boldsymbol{\theta})$, $\boldsymbol{\delta}_i = \boldsymbol{R}_i\boldsymbol{\varepsilon}_i$. Let $\mathbf{z}_i = (z_{i,1},\ldots,z_{i,n})'$, $\mathbf{g}(\mathbf{t}_i;\boldsymbol{\theta}) = (g(t_{i,1};\boldsymbol{\theta}),\ldots,g(t_{i,n};\boldsymbol{\theta}))'$ and $\boldsymbol{\delta}_i = (\delta_{i,1},\ldots,\delta_{i,n})'$, so the transformed model (5) can be rewritten as,

$$z_{ij} = g(t_{ij};\boldsymbol{\theta}) + \delta_{ij};\ i = 1,\ldots,r, \text{ and } j = 1,\ldots,n,$$

where

$$z_{ij} = \begin{cases} (1-\rho^2)^{\frac{1}{2}} y_{ij} \\ y_{ij} - \rho y_{i,j-1} \end{cases} \text{ and } g(t_{ij};\boldsymbol{\theta}) = \begin{cases} (1-\rho^2)^{\frac{1}{2}} f(t_{ij};\boldsymbol{\theta}) & ;\ j = 1 \\ f(t_{ij};\boldsymbol{\theta}) - \rho f(t_{i,j-1};\boldsymbol{\theta}) & ;\ j = 2,\ldots,n, \end{cases} \quad (6)$$

can be replaced by firstly assuming to know ρ in (6) with the true ρ^* and fitted by secondly minimizing the error sum of squares,

$$S_{\text{TrueTSLS}}(\boldsymbol{\theta}) = (1-\rho^{*2})\left(y_{i1} - f(t_{i1};\boldsymbol{\theta})\right)^2 + \sum_{i=1}^{r}\sum_{j=2}^{n}\left(y_{ij} - \rho^* y_{i,j-1} - f(t_{ij};\boldsymbol{\theta}) + \rho^* f(t_{i,j-1};\boldsymbol{\theta})\right)^2, \quad (7)$$

with respect to $\boldsymbol{\theta}$.

Modified Two-stage Least Squares Method

To modify the previous TSLS method for a practical approach, when ρ is unknown, it can be estimated by using residuals from the one-way ANOVA model (Montgomery, 2001),

$$y_{ij} = \mu_j + \varepsilon_{ij},$$

where μ_j is the mean response from the j^{th} time group and ε_{ij} is the IID error and the residual is $\hat{\varepsilon}_{ij} = y_{ij} - \hat{\mu}_j$, in which $\hat{\mu}_j$ is the sample mean. Next, to estimate ρ for the i^{th} experimental unit, the lag-1 autoregressive estimation is simply given by Abraham and Ledolter (1983) as,

$$\hat{\rho}_i = \frac{\sum_{j=2}^{n} \hat{\varepsilon}_{ij}\hat{\varepsilon}_{i,j-1}}{\sum_{j=1}^{n} \hat{\varepsilon}_{ij}^2}.$$

To modify the model (6), $\hat{\rho} = mean(\hat{\rho}_1, \ldots, \hat{\rho}_r)$ is the mean of $\hat{\rho}_i$ for r replicates; this modification can be used to build the modified two-stage least squares (MTSLS) model as,

$$z_{i,j} = \begin{cases} (1-\hat{\rho}^2)^{\frac{1}{2}} y_{i,j} \\ y_{i,j} - \hat{\rho} y_{i,j-1} \end{cases} \text{ and } g(t_{i,j};\boldsymbol{\theta}) = \begin{cases} (1-\hat{\rho}^2)^{\frac{1}{2}} f(t_{i,j};\boldsymbol{\theta}) & ;\ j=1 \\ f(t_{i,j};\boldsymbol{\theta}) - \hat{\rho} f(t_{i,j-1};\boldsymbol{\theta}) & ;\ j=2,\ldots,n, \end{cases} \tag{8}$$

and its error sum of squares is,

$$S_{\text{MTSLS}}(\boldsymbol{\theta}) = (1-\hat{\rho}^2)\left(y_{i1} - f(t_{i1};\boldsymbol{\theta})\right)^2 + \sum_{i=1}^{r}\sum_{j=2}^{n}\left(y_{ij} - \hat{\rho} y_{i,j-1} - f(t_{ij};\boldsymbol{\theta}) + \hat{\rho} f(t_{i,j-1};\boldsymbol{\theta})\right)^2, \tag{9}$$

and ordinary least squares (OLS) estimation can be used for estimating $\boldsymbol{\theta}$.

Modified Conditional Least Squares Method

Similarly, if the model (8) can be reduced by omitting the first pair $\left(z_{i,1}, g(t_{i,1};\boldsymbol{\theta})\right)$ based on the first order autoregressive process, the modified conditional least squares (MCLS) model can be shown by,

$$z_{ij} = y_{ij} - \hat{\rho} y_{i,j-1} \text{ and } g(t_{ij};\boldsymbol{\theta}) = f(t_{ij};\boldsymbol{\theta}) - \hat{\rho} f(t_{i,j-1};\boldsymbol{\theta}) \quad ;\ j=2,\ldots,n,$$

and the error sum of squares function of the MCLS model,

$$S_{\text{MCLS}}(\boldsymbol{\theta}) = \sum_{i=1}^{r}\sum_{j=2}^{n}\left(y_{ij} - \hat{\rho} y_{i,j-1} - f(t_{ij};\boldsymbol{\theta}) + \hat{\rho} f(t_{i,j-1};\boldsymbol{\theta})\right)^2. \tag{10}$$

To be the same as the previous method, the function (10) is minimized with respect to $\boldsymbol{\theta}$. The estimators $\hat{\boldsymbol{\theta}}$ are asymptotic properties to OLS estimators (Gallant and Goebel, 1976; Bender and Heinemann, 1995).

Monte Carlo Simulation

In the Monte Carlo simulation study to compare the performance of the above methods, each simulated dataset y_{ij} for four replicates $i = 1, \ldots, 4$ is generated under the sinusoidal function (4) with known parameters as given by,

$$y_{ij} = 330 - 3t_{ij} + (0.5 + 180\exp(-0.07t_{ij}))\sin\left(\frac{2\pi t_{ij}}{24} + 0.31\right) + \varepsilon_{ij}\,, \qquad (11)$$

where let $t_{ij} = 0, 1.5, \ldots, 78$, $n = 53$ with the following error structures,

IID, $\varepsilon_{ij} = \delta_{ij}$; $j = 1, \ldots, n$,

AR (1), $\varepsilon_{ij} = \begin{cases} \delta_{ij} & ;\ j = 1 \\ \delta_{ij} + 0.25\varepsilon_{i,j-1} & ;\ j = 2, \ldots, n, \end{cases}$

AR (2), $\varepsilon_{ij} = \begin{cases} \delta_{ij} & ;\ j = 1, 2 \\ \delta_{ij} + 0.25\varepsilon_{i,j-1} + 0.1\varepsilon_{i,j-2} & ;\ j = 3, \ldots, n, \end{cases}$

where independent errors are normally distributed $\delta_{ij} \sim N(0,5)$. Each sample size with a total of Monte Carlo runs $M = 20000$ and parameter estimates $\hat{\boldsymbol{\theta}}$ of those methods are based on the Gauss-Newton iterative algorithm using the R software (Ritz and Streibig, 2008; Crawley, 2013) with setting the known parameters $\boldsymbol{\theta}^*$ in (11) as initial values. They are assessed and compared first based on the percentage bias of the estimator of the parameter as follows,

$$\%\text{Bias} = 100\left(\frac{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*}{\boldsymbol{\theta}^*}\right),$$

where $\hat{\boldsymbol{\theta}} = \frac{1}{20000}\sum_{m=1}^{20000}\hat{\boldsymbol{\theta}}_m$ is the mean of the parameter estimate, $\hat{\boldsymbol{\theta}}_m$, which is obtained from the m^{th} simulation run $(m = 1, 2 \ldots, 20000)$. Secondly, the efficiency of the method is measured using the root mean square error as estimated by,

$$\text{RMSE} = \sqrt{\left(\text{SD}(\hat{\boldsymbol{\theta}})\right)^2 + \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)^2},$$

where the standard deviation is $\text{SD}(\hat{\boldsymbol{\theta}}) = \sqrt{\frac{1}{20000-1}\sum_{m=1}^{20000}(\hat{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}})^2}$. Thirdly, the standard error for parameter estimates is,

$$\text{SE}(\hat{\boldsymbol{\theta}}) = \frac{1}{20000}\sum_{m=1}^{20000}\text{SE}(\hat{\boldsymbol{\theta}}_m),$$

where $\text{SE}(\hat{\boldsymbol{\theta}}_m)$ is estimated from the m^{th} simulated times. Finally, to assess the statistical inference validity, the empirical coverage probability is the proportion of times that the nominal 95% confidence interval covers the parameters $\boldsymbol{\theta}^*$, as given by,

$$P\left(\hat{\boldsymbol{\theta}}_m - t_{0.025,v}\text{SE}(\hat{\boldsymbol{\theta}}_m) \le \boldsymbol{\theta}^* \le \hat{\boldsymbol{\theta}}_m + t_{0.025,v}\text{SE}(\hat{\boldsymbol{\theta}}_m)\right) = 0.95$$

where let $t_{0.025,v}$ be the upper 0.025 quantile of t distribution with degrees of freedom $v = nr - p$ which is dependent on each method.

Results

The simulation results are obtained from the two-stage least squares estimation with true ρ^* (TrueTSLS), as a theoretical part, the modified conditional least squares (MCLS) method and modified two-stage least squares (MTSLS) method, as practical parts in fitting the sinusoidal regression model with correlated error based on the autoregressive process of order one, AR(1), and evaluated using the percentage bias (%Bias), root mean squares error (RMSE), standard error (SE), and coverage probability for the parameter estimates.

Table 1 Percentage bias for parameter estimates

$\hat{\theta}$	Errors	%Bias		
		TrueTSLS	MTSLS	MCLS
$\hat{\alpha}$	IID	-0.0018	-0.0017	-0.0007
	AR(1)	**-0.0066**	**-0.0065**	**-0.0052**
	AR(2)	-0.0056	-0.0053	-0.0065
$\hat{\beta}$	IID	0.0004	0.0011	-0.0007
	AR(1)	**-0.0032**	**-0.0038**	**-0.0019**
	AR(2)	-0.0025	-0.0032	0.0018
$\hat{a}_s$	IID	-1.1281	-0.8728	-0.4687
	AR(1)	**-2.6312**	**-2.6377**	**-1.1718**
	AR(2)	-3.9272	-3.8800	-1.5986
$\hat{a}$	IID	0.0254	0.0266	0.0339
	AR(1)	**0.0589**	**0.0553**	**0.0781**
	AR(2)	0.0685	0.0699	0.0934
$\hat{d}$	IID	0.0465	0.0493	0.0530
	AR(1)	**0.0592**	**0.0579**	**0.0868**
	AR(2)	0.0652	0.0703	0.0959
$\hat{\tau}$	IID	0.0028	0.0027	0.0050
	AR(1)	**0.0070**	**0.0078**	**0.0061**
	AR(2)	-0.0036	-0.0037	-0.0032
$\hat{\Phi}$	IID	0.0423	0.0414	0.0005
	AR(1)	**0.0351**	**0.0425**	**-0.0035**
	AR(2)	0.0785	0.0760	-0.0306

Table 1 shows the simulation results when the sinusoidal model with the correlated error AR (1) assumed to be correct is fitted by these three methods. Although the %Bias values for the first three parameter estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{a}_s$ are negative, for the last four estimates $\hat{a}$, $\hat{d}$, $\hat{\tau}$ and $\hat{\Phi}$ are positive, and their performances are quite good because they are close to zero. However, $\hat{a}_s$ are biased by approximately -1.17% by MCLS and -2.63% by TrueTSLS and MTSLS.

Although the correlated datasets generated by the models with IID and AR (2) errors are fitted for mis-specifying error cases, almost all values of %Bias of all parameter estimates produced by these three methods are close to zero. However, the %Bias values of $\hat{a}_s$ are biased and around -1.59% obtained from MCLS, -3.88% from MTSLS, and -3.92% from TrueTSLS

Table 2 Root mean squares error and standard error for parameter estimates

$\hat{\theta}$	Errors	RMSE			SE		
		TrueTSLS	MTSLS	MCLS	TrueTSLS	MTSLS	MCLS
	IID	0.8489	0.8495	0.8889	0.8409	0.8444	0.8830
$\hat{\alpha}$	AR(1)	**1.0966**	**1.0981**	**1.1759**	**1.0974**	**1.0860**	**1.1673**
	AR(2)	1.2541	1.2548	1.3458	1.1009	1.1171	1.2046
	IID	0.0175	0.0175	0.0181	0.0176	0.0176	0.0182
$\hat{\beta}$	AR(1)	**0.0231**	**0.0231**	**0.0243**	**0.0229**	**0.0227**	**0.0239**
	AR(2)	0.0260	0.0261	0.0275	0.0230	0.0233	0.0246
	IID	0.9037	0.9047	0.9094	0.8911	0.8942	0.8994
$\hat{a}_s$	AR(1)	**1.1426**	**1.1436**	**1.1591**	**1.1359**	**1.1234**	**1.1333**
	AR(2)	1.2316	1.2325	1.2520	1.1390	1.1504	1.1601
	IID	3.0080	3.0133	3.1477	3.0192	3.0238	3.1623
$\hat{a}$	AR(1)	**3.7842**	**3.7871**	**3.9697**	**3.7893**	**3.7479**	**3.9396**
	AR(2)	4.1036	4.1065	4.3238	3.8001	3.8317	4.0214
	IID	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022
$\hat{d}$	AR(1)	**0.0027**	**0.0027**	**0.0028**	**0.0027**	**0.0027**	**0.0028**
	AR(2)	0.0029	0.0030	0.0959	0.0027	0.0028	0.0029
	IID	0.1099	0.1099	0.1360	0.1090	0.1092	0.1348
$\hat{\tau}$	AR(1)	**0.1348**	**0.1348**	**0.1862**	**0.1339**	**0.1325**	**0.1827**
	AR(2)	0.1425	0.1426	0.2021	0.1342	0.1353	0.1888
	IID	0.0127	0.0128	0.0190	0.0127	0.0127	0.0191
$\hat{\Phi}$	AR(1)	**0.0145**	**0.0145**	**0.0278**	**0.0147**	**0.0146**	**0.0274**
	AR(2)	0.0145	0.0146	0.0298	0.0148	0.0148	0.0284

It can be seen in Table 2 that the values of RMSE produced using the TrueTSLS and MTSLS methods are similar and more efficient than those of MCLS. Since the values of SE from the TureTSLS, MTSLS, and MCLS methods are slightly underestimated compared to RMSE, they produce good coverage probabilities which are slightly under 0.95, and of course the best one is for the TrueTSLS method, as depicted in Figure 1. However, the SE values of $\hat{\Phi}$ estimated using the TrueTSLS and MTSLS methods are overestimated and their coverage probabilities are also more than 95%, as shown in Figure 1 (g).

For mis-specifying error cases, IID and AR (2) errors, the RMSE under TrueTSLS and MTSLS are still similar and smaller than the one for MCLS. Because the SE values obtained from all methods are lower than the SD, the coverage probabilities for all parameters are significantly under the nominal probability 0.95 for the AR (2) error structure, as seen in Figure 1 (a)-(f). On the other hand, the coverage probabilities for Φ obtained from TrueTSLS and MTSLS are larger than 0.95, as shown in Figure 1 (g).

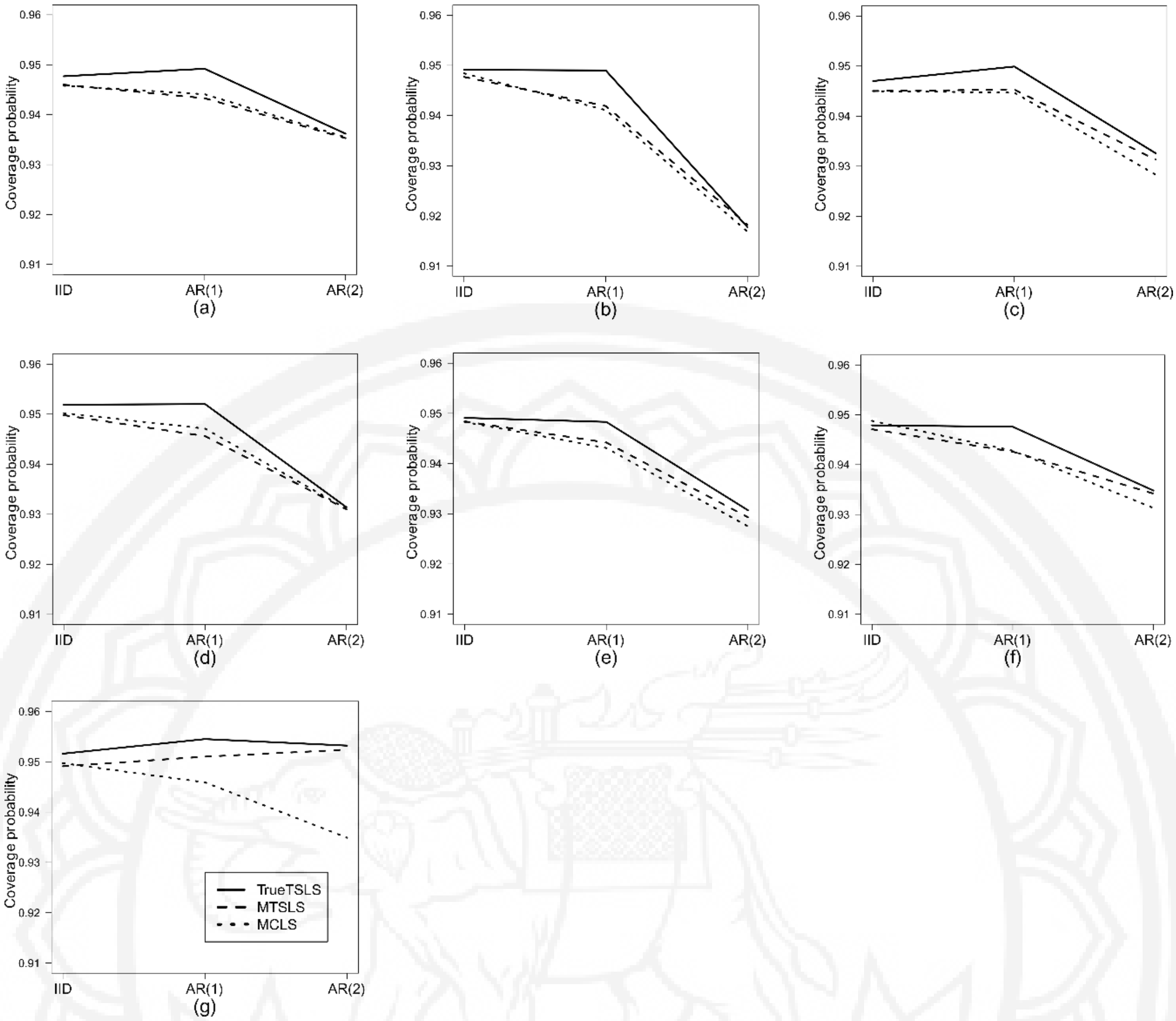

Figure 1 Coverage probability for parameters: (a) intercept α , (b) slope β , (c) amplitude adjustment a_s , (d) amplitude a , (e) damping d , (f) period τ and (g) phase Φ

Example

In this section, the proposed methods are applied to synthetic circadian rhythms over time courses presented by Yang and Su (2010) to estimate the period parameter in circadian models. The data were generated by the sinusoidal model as shown in Figure 2. The synthesis time-series data were fitted by the cosine function as given by,

$$f(t_{ij};\boldsymbol{\theta}) = 500\exp(-0.01t_{ij}) + 100\text{SNR}\exp(-0.01t_{ij})\cos\left(\frac{2\pi t_{ij}}{\tau} - \Phi\right),$$

where SNR is a signal-to-noise ratio parameter, Φ is the phase and τ is the period of the cosine wave. The initial values are taken from sampling intervals of SNR $\in \left[\frac{1}{5}, 1\right]$ and $\Phi \in (0, 2\pi]$. The period interval is between 20 and 28 hours. MTSLS and MCLS procedures produce the nominal 95% confidence interval for the circadian period τ , 24.0342 $\pm$ 0.4562 and 24.0354 $\pm$ 0.4665. The residual standard errors $\hat{\sigma}$ are 41.26 and 41.73, respectively.

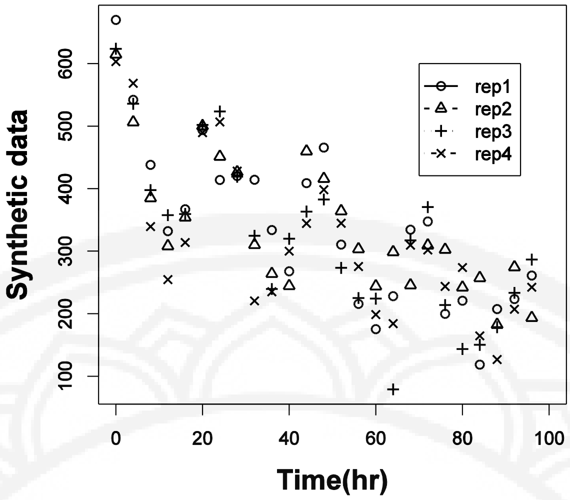

Figure 2 The example of synthetic datasets for four replicated generated under the following sinusoidal model $y_{ij} = 500\exp(-0.01t_{ij}) + 140\exp(-0.01t_{ij})\cos\left(\frac{2\pi t_{ij}}{24}\right) + \delta_{ij}$, where let $t_{ij} = 0, 4, \ldots, 96$ with 4 hour intervals and δ_{ij} is IID with mean $\mu = 0$ and high standard deviation $\sigma = 40$

Discussion

Almost all the bias results, obtained from the above three methods, TrueTSLS, MTSLS, and MCLS assuming that the correlated error is a stationary AR(1) process when the fitted sinusoidal model with correlated errors whether correctly specified or not, are asymptotically unbiased estimates. This has the advantage of least squares estimation. The root mean squares errors, standard errors and coverage probabilities indicate that if we know the true values of the correlation coefficients, the TrueTSLS method is the best choice. However, if we do not know the correlation coefficients and the correlated errors are mis-specified, the MTSLS and MCLS methods are comparable. In addition, the MCLS method is considerably less efficient because its degree of freedom is reduced by the first order of the autoregressive process. It should be noted that they are much less efficient when the autoregressive error process is high-order, which corresponded with Gallant and Goebel (1976). In the example, the proposed methods can be applied to the synthetic dataset shown by Yang and Su (2010), and here they are used to produce period estimates and the confidence interval.

Conclusion and Suggestions

In this paper, proposed methods for practical use to fit the sinusoidal model where the error is correlated, based on the AR (1) process are the modified two-stage least squares (MTSLS) method and modified conditional least squares (MCLS) estimation using pure errors to compute the correlation coefficient in the weight matrix. This paper presents the two-stage least squares when it is theoretically assumed to know the correlation coefficient of errors (TrueTSLS). The simulation results suggest that these methods tend to produce asymptotically unbiased estimators for parameters. The TrueTSLS method produces better confidence intervals than the MTSLS and MCLS methods, but for mis-specification cases, the MTSLS and MCLS methods are comparable with the TrueTSLS method. Therefore, this work suggests that the MTSLS and MCLS methods can produce reliable estimates and confidence intervals and can be useful for analysis for any situations. Further work will extend modified least squares methods to cover nonlinear regression models with AR(p) errors.

Acknowledgments

The Kasetsart University, Chalermphrakiat Sakon Nakhon Province Campus (KUCSC) provided financial support for this paper.

References

Abraham, B., & Ledolter, L. (1983). *Statistical methods for forecasting*. New York: John Wiley & Sons.

Asikgil, B., & Erar, A. (2013). Polynomial tapered two-stage least squares method in nonlinear regression. *Applied Mathematics and Computation*, *219*, 9743-9754.

Bates, D. M., & Watts, D.G. (1988). *Nonlinear regression analysis and its applications*. New York: John Wiley & Sons.

Bender, R., & Heinemann, L. (1995). Fitting nonlinear regression models with correlated errors to individual pharmacodynamics data using SAS software. *Journal of Pharmacokinetics and Biopharmaceutics*, *23*(1), 87-100.

Crawley, M. J. (2013). *The R book*. West Sussex: John Wiley & Sons.

Gallant, A. R., & Goebel, J. J. (1976). Nonlinear regression with autoregressive errors. *Journal of the American Statistical Association*, *71*(356), 961-967.

Izumo, M., Johnson, C. H., & Yamazaki, S. (2003). Circadian gene expression in mammalian fibroblasts revealed by real-time luminescence reporting: Temperature compensation and damping. *The National Academy of Sciences of the USA*, *100*(26), 16089-16094.

Kyriacou, C. P, & Hall, J.C. (1980). Circadian rhythm mutations in Drosophila melanogaster affect short-term fluctuations in the male's courtship song. *The National Academy of Sciences of the USA*, *77*(11), 6729-6733.

Liang, Z., Ren, D., Sun, J., & Zhu, Z. (2018). Fitting algorithm of sine wave with partial period waveforms and non-uniform sampling based on least-square method. *Journal of Physics Conference Series*, *1149*(1), 1-8.

Montgomery, D. C. (2001). *Design and analysis of experiments*. New York: John Wiley & Sons.

Pukdee, W., Polsen, O., & Baksh, M. F. (2018) Improved methods for the analysis of circadian rhythms in correlated gene expression data. *Songklanakarin Journal of Science and Technology*, *40*(3), 692-700.

Pukdee, W., Polsen, O., & Baksh, M. F. (2020). A Modified Two-Stage Method for Parameter Estimation in Sinusoidal Models of Correlated Gene Expression Profiles. *Thailand Statistician*, *18*(1), 77-89.

Ritz, C., & Streibig, J. C. (2008). *Nonlinear regression with R*. New York: Springer.

Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear regression*. New York: Wiley.

Yang, R., & Su, Z. (2010). Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioresource Technology*, *26*(1), 168-174.