Ann Inst Stat Math (2010) 62:1143–1173
DOI 10.1007/s10463-009-0224-6

Analysis of rounded data from dependent sequences

Baoxue Zhang · Tianqing Liu · Z. D. Bai

Received: 13 March 2008 / Revised: 30 October 2008 / Published online: 26 March 2009

Abstract Observations on continuous populations are often rounded when recorded due to the precision of the recording mechanism. However, classical statistical approaches have ignored the effect caused by the rounding errors. When the observations are independent and identically distributed, the exact maximum likelihood estimation (MLE) can be employed. However, if rounded data are from a dependent structure, the MLE of the parameters is difficult to calculate since the integral involved in the likelihood equation is intractable. This paper presents and examines a new approach to the parameter estimation, named as "short, overlapping series" (SOS), to deal with the α-mixing models in presence of rounding errors. We will establish the asymptotic properties of the SOS estimators when the innovations are normally distributed. Comparisons of this new approach with other existing techniques in the literature are also made by simulation with samples of moderate sizes.

Keywords ARMA(p, q) model · BRB corrections · Rounded data · Sheppard corrections · SOS estimation

B. Zhang · T. Liu · Z. D. Bai (✉)
Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, 130024 Changchun, China
e-mail: baizd@nenu.edu.cn

B. Zhang
e-mail: bxzhang@nenu.edu.cn

Z. D. Bai
Department of Statistics and Applied Probability, National University of Singapore, Singapore 117543, Singapore
e-mail: stabaizd@nus.edu.sg

1 Introduction

The rounding of data may be due to the precision of the measuring instruments in experiments or the nature of the recording/storage mechanism. In general, the rounding scheme is to record the measurements to the nearest multiples of the precision unit. Except for categorical data, all observations of continuous populations have to be rounded. It is obvious that the rounding errors affect the statistical inferences if the rounding error effect is ignored. However, it had not drawn serious attention in the old days because sample size was small and the inaccuracy caused by the rounding errors was tolerable relatively to the statistical problem. However, it is no longer the case nowadays, owing to the wide adoption of modern computers. We are able to collect, store and analyze data sets of huge sizes and/or large dimensions. For example, in a data mining problem, the sample size may be as large as several millions. Then, it is natural to ask whether the classical statistical inferences are still reliable when the rounding errors are ignored? It has been noted in the literature that the standard t test will reject the true hypothesis with a probability close to one if the data are rounded and the sample size is large. Therefore, the rounding errors have to be taken into account in all large sample problems.

Rounding error has received considerable attention in the literature (see, e.g., Dempster and Rubin 1983; Heitjan 1989). For rounded data from independent and identically distributed ($i.i.d$) samples, there are four approaches to deal with the rounding error. The first is the so-called Sheppard's correction for the sample covariance estimator whose derivation is based on the Taylor expansion of the likelihood function and hence it is also called the maximum-likelihood corrections. The earliest discussion on rounded data at least dates back to Sheppard (1898). Recent work can be found in Tricker (1984, 1990a,b), Tricker et al. (1998), and Vardeman and Lee (2005) who investigated the effect of the parameter estimation and control charts based on rounded data from normal and non-normal populations such as gamma and exponential. Dempster and Rubin (1983) compared the rounding effects in a rounded linear models using three well known statistical approaches: the Sheppard's correction, the BRB correction and the usual least squares estimation by considering the rounded data as unrounded. The third approach treats the rounding error as additive and uniformly distributed. In the above mentioned references, the formulae in BRB and Sheppard's corrections are equivalent to assuming the rounding errors as uniformly distributed over a symmetric interval with length equal to the rounding precision and are independent of the rounded values or the true value of the observed sample respectively. However, it is not difficult to examine the invalidity of these assumptions since the rounding errors and the rounded data are both functions of the original data. Heitjan (1989) gave a detailed review of the most popular approaches in dealing with rounded data. Heitjan and Rubin (1991) presented a general model for coarsened data, including rounded, heaped, censored and missing data. Bai et al. (2008) showed that the sample variance of the rounded data of a sample from a normal population is inconsistent. The sample mean is also inconsistent unless the true mean is a multiple of half a precision unit. This stimulates statisticians to reinstitute the mathematical model to "recover" the true sample values which may lead to better inferences based on the rounded data. All earlier works had failed to provide consistent estimation except the recent works by Lee and Vardeman (2001, 2002, 2003) who investigated confidence interval estimation of the parameters μ or σ^2 when a rounded sample came from the $N(\mu, \sigma^2)$ population. They also extended their results to the balanced one-way ANOVA with random effects. However, they did not obtain simultaneous interval estimation of the parameter μ and σ^2 since their method was based on an unjustified inversion of the rounded data likelihood ratio test. Hall (1982) and Coppejans (2003) examined the influence of rounding errors on nonparametric density estimation and proposed a weighted average of neighboring frequency estimators. Wright and Bray (2003) showed the danger of ignoring heaping before presenting a case-study of the ultrasound measurements. They analyzed a mixture model for rounded data by the Bayesian approach implemented by using the Gibbs sampler. The general issue of rounding continuous data has been indispensable for practitioners. Some recent references are made to Tricker et al. (1998), Vardeman and Lee (2005), Vardeman (2005) and Grimshaw et al. (2005).

The rounding errors in dependent sequences of random variables may cause an error of the statistical inferences in a more complicated way than that in iid cases. Jones (1980) theoretically discussed the problem of maximum likelihood estimation in ARMA models when data were rounded, Machak and Rose (1984, 1985) reported on studies of the problem of rounding errors in ARMA models, Stam and Cogger (1993) considered the effects of rounding errors in parameter estimation of autoregressive (AR) models, deriving appropriate adjustments for the estimates of the parameters when using rounded data, and Rose (1993) addressed the impact of rounding errors on model identification and parameter estimation for ARMA(1, 1). In practical hydrologic examples, such as the frequency distributions of rainfall or storm-water runoff and a flooding study of the American River, the Sheppard's correction techniques were employed, see Durrans et al. (2004). Durrans and Pitt (2004) extended the Sheppard's corrections to allow unequal rounding intervals at different instants and then applied their extension to mammalogical studies, where the date of parturition of animals were obtained from collections in intervals larger than one day. Many financial, economic, labor, public health variables are also continuous but *rounded*. In the stock market, the stock price P_t (or the return $R_t := \frac{P_t - P_{t-1}}{P_t - 1}$) of stock at date t is allowed to change only at discrete grids. Brock et al. (1987) proposed the so-called BDS test to examine the serial independence of a time series data. For this test, Krämer and Runde (1997) seriously pointed out that the BDS test would reject the true null hypothesis with a probability as high as 0.8 when the data was rounded to integers and 0.20 when rounded to one decimal digit. Such a phenomenon was also pointed out by Kozicki and Hoffman (2004). Bai et al. (2008) presented approximate maximum likelihood estimation and proved strong consistency and asymptotic normality of the new estimates.

The general problem will be formulated as follows. Consider a time series $\{X_t\}$ with a specified model, where X_t is the modelled random variable at time t. Suppose that $\{x_t, t = 1, \ldots, T\}$ is the realizations from time $t = 1$ to T and $\{\tilde{x}_t\}$ the recorded values of $\{x_t\}$. That is, $x_t = \widetilde{x}_t + \varepsilon_t$, where ε_t represents the rounding error. Suppose X_t is a random variable, let the rounded value be denoted by $\widetilde{X}_t$. It is easy to see that both $\widetilde{X}_t$ and ε_t are functions of X_t. Thus, ε_t cannot be independent of X_t or $\widetilde{X}_t$ unless in some specially constructed distributions. Also, ε_t is not uniformly distributed. Although the mathematical form of rounding error looks like that of the measurement error, the major difference lies in that the former cannot be independent of the recorded observations, or the unknown value to be measured. Thus, statistical techniques that have been proposed to deal with the measurement errors cannot be transplanted to the rounded data.

This article focuses on the estimation of the parameters in certain mixing sequences including the ARMA model under normality assumptions. As pointed out by Stam and Cogger (1993, p. 489), the maximum likelihood estimator based on the rounded data is intractable both numerically or by Monte Carlo, such as MCMC, because the integrals involved in the likelihood function have too many folds to compute even when the number of observations is moderately large. So they arrived at appropriate adjustments for the estimates using Taylor series expansion, following approaches of Lindley (1950) and Tallis (1967). In the literature, almost all works employ various corrections such as BRB and Sheppard under various model structures. We shall propose a new approach to calculate estimates of the parameters.

This article is organized as follows. In the next section, we propose a new method to estimate the parameters for an α-mixing sequence under assumption of a known parametric form of the distributions. This new method is named the “short, overlapping series” (SOS) approach because the small, manageable segments of data may have overlaps—this is part of how it maintains some efficiency. Then, we consider the small segments as independent random variables (or vectors) to obtain the pseudo maximum likelihood estimator of the parameters. We also illustrate our approach by some examples. In Sect. 3.1 we establish the asymptotic properties of that estimator under the regularity conditions. We will show the asymptotic properties when rounded data are from α-mixing sequences including some of the time series models. Furthermore, we give the estimators of the asymptotic covariance matrix for three cases in Sect. 3.2. In Sects. 3.3 and 3.4, as illustrations, we establish more detailed properties of the estimators for rounded data from IID and ARMA models. In Sect. 4, we present a modified version of the SOS method to further reduce the computation time of the parameter estimates. In Sects. 5 and 6, we compare our new method with the existing methods in the literature through simulation studies. Some discussion and comments are presented in Sect. 7. Some technical details are given in the Appendix.

2 SOS method

We first briefly show that the exact maximum likelihood approach could not be employed for rounded data from a dependent sequence with a continuously distribution.

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample, and suppose that its joint density belongs to a parametric family, $\{f_n(\mathbf{x}, \theta), \theta \in \Theta \subseteq R^p\}$. Suppose the sample data are observed in rounded form with rounding interval width h. Thus, the rounded sample data Y_i are defined by:

$$Y_i = y_i \qquad \text{if } y_i - h/2 \le X_i < y_i + h/2. \tag{1}$$

For the rounded data, the exact log-likelihood function is given by:

$$l_r(\theta;\mathbf{y}) = \log\left(\int_{y_n-h/2}^{y_n+h/2}\cdots\int_{y_1-h/2}^{y_1+h/2} f_n(\mathbf{x};\theta)\,\mathrm{d}x_1\cdots\mathrm{d}x_n\right) \tag{2}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$. The corresponding maximum likelihood estimator (MLE) $\hat{\theta}_r(\mathbf{y}) = (\hat{\theta}_{1r}(\mathbf{y}), \ldots, \hat{\theta}_{pr}(\mathbf{y}))$ should be found by maximizing (2) or solving the following systems of equations:

$$\frac{\partial l_r(\theta;\mathbf{y})}{\partial\theta_j}\,|_{\theta=\hat{\boldsymbol{\theta}}_r(\mathbf{y})} = 0, \quad j = 1, \ldots, p. \tag{3}$$

However, the MLE $\hat{\theta}_r(\mathbf{y})$ is actually intractable from the above equations because the value of the n fold integral in (2) might be too small to be stored as non-zero in any computer system. Even these values are not that small for small or moderately large n, the calculation to a high level of accuracy of the n fold integral is not an easy job. For example, if we require the 4th decimal of accuracy, then at least, the integration interval for each fold needs to be split into 10^4 subintervals. If the length of the series n is as small as 100, then the integral needs to calculate the density function values at 10^{400} grids. This is unaffordable for almost any existing computers. Therefore, various adjustments or corrections for the parameter estimation are proposed in the literature by using Taylor expansion of $l_r(\theta;\mathbf{y})$ in powers of h when it is small. The famous ones are the Sheppard's correction and the BRB correction. In Lindley (1950); Tallis (1967); Dempster and Rubin (1983), and Stam and Cogger (1993), such corrections are deeply discussed for multivariate distribution models, linear regression models, linear autoregressive models, etc. However, the major drawback of these corrections is that they do not provide consistent estimation of the parameters because h does not actually tend to 0 and hence the error caused by eliminating the higher order h terms in the Taylor expansion does not actually tend to 0. So far we have not found in the literature any consistent estimation of the parameters for rounded time series observations, until Bai et al. (2008).

Incorporating this problem, we propose the following procedure to estimate the parameters θ. From now on, we shall assume that the rounding scheme is to have the data rounded to integers without loss of generality. We split the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ into pieces of length $a+1$ as follows:

$$\mathbf{X} = \{(X_1, \ldots, X_{a+1}), (X_2, \ldots, X_{a+2}), \ldots, (X_{n-a}, \ldots, X_n)\}$$

and let $\mathbf{X}_t = (X_t, \ldots, X_{t+a})'$, where $a = 0, 1, 2, \ldots, n-1$, for which the realizations are denoted by $\mathbf{x}_t{=}(x_t, \ldots, x_{t+a})'$. Correspondingly, the rounded version of $\mathbf{X}$ vectors and observations are denoted by $\mathbf{Y}_t{=}(Y_t, \ldots, Y_{t+a})'$ and $\mathbf{y}_t = (y_t, \ldots, y_{t+a})'$, respectively. If the $\mathbf{X}$ sequence is mixing, then the sequence $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{n-a}\}$ is also mixing. When finding the parameter estimates, we just regard the $\mathbf{Y}$-sequence as iid $a+1$-dimensional random vectors. Because the small, manageable segments of data may have overlaps, we will name the method "short, overlapping series". Consequently, we name the sum of piece-wised log-likelihood functions as the SOS log-likelihood which is given by

$$l_{\text{sos}}(\theta) = \sum_{t=1}^{n-a} q(\mathbf{y}_t, \theta) = \sum_{t=1}^{n-a} \log p(\mathbf{y}_t, \theta), \tag{4}$$

where

$$p(\mathbf{y}_t, \theta) = \int_{\mathbf{y}_t - \mathbf{1}/2}^{\mathbf{y}_t + \mathbf{1}/2} f(\mathbf{x}_t; \theta)\, \mathrm{d}\mathbf{x}_t. \tag{5}$$

$\mathbf{1}$ is the $a+1$ dimensional vector of all elements 1 and $f(\mathbf{x}_t; \theta)$ is the density function of $a+1$ consecutive variables in the strong stationary process $\{x_t\}$. The corresponding SOS-score function is given by

$$\dot{l}_{\text{sos}}(\theta) = \sum_{t=1}^{n-a} \dot{q}(\mathbf{y}_t, \theta) = \sum_{t=1}^{n-a} \frac{\dot{p}(\mathbf{y}_t, \theta)}{p(\mathbf{y}_t, \theta)}, \tag{6}$$

where $\dot{l}_{\text{sos}}(\theta)$ denotes the derivative vector of $l_{\text{sos}}(\theta)$ with respect to components of θ. Note that when $a = 0$ and $a = n - 1$, the SOS-log-likelihood function become the exact log-likelihood for the rounded iid sequence and that for the whole dependent sequence, respectively. Under regularity conditions, the new estimator $\hat{\theta}_n$ of θ is a solution to the SOS-estimating equation $\dot{l}_{\text{sos}}(\theta) = 0$ and thus is consistent. When the SOS log-likelihood function $l_{\text{sos}}(\theta)$ is strictly concave, the solution is also unique. In some cases a solution with a close form can be found; but more often, a numerical solution can be obtained by iterative approaches such as Newton–Raphson, quasi-Newton Raphson and simulated annealing.

The Newton–Raphson algorithm iteratively updates the parameter estimates by the formula

$$\theta_k = \theta_{k-1} - \left\{ \frac{1}{n} \ddot{l}(\theta_{k-1}) \right\}^{-1} \left\{ \frac{1}{n} \dot{l}(\theta_{k-1}) \right\}, \quad k = 1, 2, \ldots, \tag{7}$$

where $\ddot{l}$ is the Hessian matrix of second derivatives of l_{sos} and θ_0 is an initial value. A variation on this is the Fisher scoring algorithm in which the factor $-\frac{1}{n}\ddot{l}_{\text{sos}}$ is replaced by its expectation. Start from good initial values, both methods converge rapidly to the SOS estimator $\hat{\theta}_n$. Meanwhile, we shall also discuss the estimate of the asymptotic covariance matrix of $\hat{\theta}_n$. In the following sections, the asymptotic properties of the SOS-estimators are analyzed.

Remark 1 Bai et al. (2008) considered the rounded data from AR(p) or MA(p) sequences and proposed that for each $\ell = 1, \ldots, k$, choose $a+1$ consecutive variables $(X_\ell, \ldots, X_{a+\ell})$. And after each $k - a - 1$ consecutive variables, we choose another $a+1$ consecutive variables, and so on. In such a way, we obtain $m = [n/k]$ sequences of $(a+1)$-vectors with spacings $k - a - 1$ variables between two consecutive vectors. When k (or $k - a - 1$) is large, each subsequence can be considered as one of iid vectors and hence we may get an approximate MLE θ_ℓ. Finally, we use the average of the k estimates as the final estimate of the parameter θ. This method seems more intuitive than the method described above. However, it has been proved that the asymptotic variance is the same as the estimator proposed in this paper.

Remark 2 If a searching approach is employed to find the maximum of the SOS-log-likelihood (4), for each value of θ and t, we need to compute the $a + 1$ fold integral for $p(\mathbf{y}_t, \theta)$. If the Newton–Raphson approach (7) is used, at each recursion step, for each t, we need to compute $a+1$ fold integrals for $p(\mathbf{y}_t, \theta)$, $\dot{p}(\mathbf{y}_t, \theta)$ and $\ddot{p}(\mathbf{y}_t, \theta)$. In practice, only $a = 1$ or 2 can be used for a numerical integral approach to compute the estimates. If a is relatively large, one may need to use a Monte-Carlo approach (say, the MCMC approach, or the E-M algorithm) to approximate these integrals. We alert readers that the Monte–Carlo method will have a very slow convergence rate and large computational errors. It is even more time consuming than the numerical integration technique if we require the same computation accuracy. Thus, if a has to be large (see next remark), we need to further find alternate methods to reduce the number of integral folds for faster computation. For details, see Sect. 4.

Remark 3 Obviously, if a larger a is chosen, the SOS-log-likelihood is closer to the exact log-likelihood and hence the efficiency of the SOS-estimates should be theoretically higher. However, a larger a will exponentially increase the computation errors and computation time. Thus, one may wish to choose a to be as small as possible to reduce computation error and save time. However, a small a may cause an identifiability problem. This depends on how many parameters can be identified by the function $p(\mathbf{y}_t; \theta)$. For the normal distribution, the mean value and the $(a+1)\times(a+1)$ covariance matrix can be solved from $p(\mathbf{y}_t; \theta)$ and its derivatives. For an ARMA(p, q) model, the model coefficients and the innovation variance can only be identified by a covariance matrix of order $(p + q + 1) \times (p + q + 1)$. Thus, for ARMA$(p, q)$ models, the smallest choice of a is $p + q$. To reduce the computation time when $p + q + 1$ is large, see the modified SOS method in Sect. 4.

3 Asymptotics

In this section, we first consider the asymptotics of the SOS-estimators, including the consistency and asymptotic normality as $n \to \infty$. Then, we give consistent estimators of the asymptotic covariance matrix of the SOS estimators.

3.1 Consistency and asymptotic normality

The following two results describe the asymptotic behavior of $\hat{\theta}_n$. First, we list a set of regularity conditions which are parallel to those popularly used in the literature for MLE and then show the consistency result. Finally, we state an asymptotic normality result with an explicit formula of the asymptotic variance. Without loss of generality, suppose that the original sample is $x_1, \ldots, x_n$. Under this assumption, the SOS-log-likelihood function of the $n - a$ rounded observations $y_1, \ldots, y_{n-a}$ is

$$l_{\text{sos}}(\mathbf{y}, \theta) = \sum_{t=1}^{n-a} \log p(\mathbf{y}_t, \theta) := \sum_{t=1}^{n-a} q(\mathbf{y}_t, \theta)$$

To establish the asymptotic properties, we need the following regularity conditions related to the SOS-log-likelihood function:

1. The parameter space Θ is an open set of $\mathbb{R}^k$ and the true value θ_0 is an inner point of Θ.
2. The sample space, i.e., $\mathscr{X} = \{\mathbf{x};\ f_{a+1}(\mathbf{x};\theta) > 0\}$, is independent of the parameters.
3. $\lambda\left(\{\mathbf{x} : p(\mathbf{y},\theta) \neq p(\mathbf{y},\theta')\}\right) > 0\ \ \forall\theta \neq \theta'$, where λ is the Lebesgue measure in $\mathbb{R}^{a+1}$.
4. $\{X_t\}$ is strictly stationary and α-mixing.
5. At the true value θ_0,

$$\psi(\theta_0;\theta) = \int f_{a+1}(\mathbf{x};\theta_0)\log p(\mathbf{y},\theta)d\mathbf{x}$$

   exists for all $\theta \in \Theta$ and is continuous in θ, where $p(\mathbf{y},\theta)$ is the probability defined in (5).
6. For each compact subset $\overline{\Theta}$ of Θ, there is an integrable function $h(\mathbf{y}) = h_{\overline{\Theta}}(\mathbf{y})$ such that

$$\frac{\|\dot{f}_{a+1}(\mathbf{x};\theta)\|_\infty}{f(\mathbf{x};\theta)} \le h(\mathbf{y})$$

   for all $\theta \in \overline{\Theta}$, where $\mathbf{y}$ is the vector of rounded $\mathbf{x}$ and the norm $\|\cdot\|_\infty$ denotes the maximum absolute value of entries of the indicated vector or matrix. Here, that the function $h(\mathbf{y})$ is integrable means $Eh(\mathbf{Y}) = \sum_{\mathbf{j}} h(\mathbf{j})p(\mathbf{j},\theta_0) < \infty$.
7. Except for a null set of $\mathscr{X}$, for each fixed $\mathbf{x}$,

$$\sup_{\theta\in\overline{\Theta}_m^c} f(\mathbf{x};\theta) \to 0,\ \ m \to \infty,$$

   where $\{\overline{\Theta}_m\}$ is an increasing sequence of compact subspaces $\overline{\Theta}_m$ of Θ such that $\cup\overline{\Theta}_m = \Theta$.
8. For some constant $\delta > 0$,

$$Eh(\mathbf{Y})^{2+\delta} < \infty,$$

   where the function $h(\mathbf{y})$ is as defined in condition 6.
9. The α-mixing coefficients satisfy

$$\sum_{n=1}^{\infty}\alpha_n^{\delta/(2+\delta)} < \infty.$$

10. The integration in $\mathbf{x}$ and the differentiation in θ about $f(\mathbf{x},\theta)$ are interchangeable.

Comments on the regularity conditions

- Conditions 1–5 are commonly assumed for almost all parametric statistical problems. Especially, condition 2 is necessary for the identifiability of parameters.
- In usual MLE problems, condition 6 is assumed to be

$$\|\dot{f}_{a+1}(\mathbf{x};\theta)\|_\infty \le g(\mathbf{x})$$

  where $g(\mathbf{x})$ is integrable with respect to the Lebesgue measure. Correspondingly, our h is similar to $g(\mathbf{x})/f(\mathbf{x},\theta)$. Thus, it is equivalent to $h(\mathbf{y})$ being integrable with respect to $f(\mathbf{x},\theta)d\mathbf{x}$. The difference lies in that $\mathbf{x}$ is changed to $\mathbf{y}$, the rounded value. This is not a strong restriction because $h(\mathbf{x}) \le Kh(\mathbf{y})$ for some constant K unless $h(\mathbf{x})$ has a rate higher than $e^{b\|\mathbf{x}\|}$ for any $b > 0$.
- Condition 7 assumes the density value tends to 0 as the parameter tends to the boundary of the parameter space. In some references, this condition is assumed to be $f(\mathbf{x},\theta) \to 0$ as $\|\theta\| \to \infty$, e.g., Chen (1997).
- Conditions 8–10 are commonly assumed for the central limit theorems of MLE.

Theorem 1 *Under regularity conditions* 1–6*, the local SOS–MLE is consistent, that is, for some compact subspace* $\overline{\Theta}$*, if* θ_0 *is an inner point of* $\overline{\Theta}$*, then*

$$\widehat{\theta_n} \to \theta_0 \quad a.s.$$

where

$$\hat{\theta}_n = \arg\left\{\sup_{\theta\in\overline{\Theta}} l_{\text{sos}}(\theta)\right\}.$$

Theorem 2 *Under regularity conditions* 1–7*, the global SOS–MLE is strong consistent.*

In hypothesis testing or confidence interval estimation, it is important to establish the asymptotic distribution. To this end, we shall prove the following theorem.

Theorem 3 *Under regularity conditions* 1–10*,*

$$\sqrt{n-a}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_k\left(0, I^{-1}(\theta_0)V(\theta_0)I^{-1}(\theta_0)\right)$$

where $N_k(\mu,\Sigma)$ *denotes a multivariate normal variable of dimension* k *with mean vector* μ *and covariance matrix* Σ*, and*

$$I(\theta_0) = E\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\bigg|_{\theta=\theta_0},$$

$$V(\theta_0) = I(\theta_0) + \sum_{t=1}^{\infty} E\left[\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_{t+1},\theta)}{\partial\theta'} + \frac{\partial q(\mathbf{Y}_{t+1},\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\right]\bigg|_{\theta=\theta_0}.$$

The asymptotic covariance of $\hat{\theta}_n$ is needed to estimate when the asymptotic normality of $\hat{\theta}_n$ is used in hypothesis testing or confidence interval estimation. We will discuss it for three cases in the next subsection.

3.2 Estimating the asymptotic covariance matrix

As discussed in Sect. 3.1, in the application of the asymptotic normality of $\hat{\theta}_n$, there is a need of a consistent estimate of the asymptotic covariance matrix of $\hat{\theta}_n$. In this subsection, we will discuss estimates of the asymptotic covariance matrix in three cases in accordance with the correlation of score functions $\frac{\partial q(\mathbf{Y}_t,\theta_0)}{\partial\theta}$ as follows:

Case 1 $\{\frac{\partial q(\mathbf{Y}_t,\theta_0)}{\partial\theta}\}$ are uncorrelated for different t.

In this case, $V(\theta_0) = I(\theta_0) = E\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\big|_{\theta=\theta_0}$. The moment method suggests estimating $V(\theta_0)$ by $\hat{V}_n = \frac{1}{n-a}\sum_{t=1}^{n-a}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta'}\big|_{\theta=\hat{\theta}_n}$. It is easy to prove that $\hat{V}_n - V_n(\theta_0) \xrightarrow{p} 0$, where $V_n(\theta_0)$ is $\hat{V}_n$ with $\hat{\theta}_n$ replaced by θ_0. Then, the convergence of $V_n(\theta_0)$ is an easy consequence of the ergodicity of α-mixing sequences. For example, in Sect. 3 (Rounding an IID Sample), we can estimate $I(\theta_0)$ and $V(\theta_0)$ by

$$\hat{I}_n = \frac{1}{n-a}\sum_{t=1}^{n-a}\frac{1}{p^2(\mathbf{y}_t;\theta)}\frac{\partial p(\mathbf{y}_t;\theta)}{\partial\theta}\frac{\partial p(\mathbf{y}_t;\theta)}{\partial\theta'}\bigg|_{\theta=\hat{\theta}_n}.$$

Case 2 $\{\frac{\partial q(\mathbf{Y}_t,\theta_0)}{\partial\theta}\}$ is m-dependent.

In this case,

$$\begin{aligned}V(\theta_0) = {} & E\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\bigg|_{\theta=\theta_0}\\ & +\sum_{t=1}^{m}E\left[\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_{1+t},\theta)}{\partial\theta'}+\frac{\partial q(\mathbf{Y}_{1+t},\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\right]\bigg|_{\theta=\theta_0}.\end{aligned}$$

Therefore, by the moment method, we may estimate $V(\theta_0)$ by

$$\begin{aligned}\hat{V}_n = {} & \frac{1}{n-a}\sum_{t=1}^{n-a}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta'}\bigg|_{\theta=\hat{\theta}_n}\\ & +\frac{1}{n-a}\sum_{\tau=1}^{m}\sum_{t=1}^{n-a-\tau}\left[\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta}\frac{\partial q(\mathbf{y}_{t+\tau},\theta)}{\partial\theta'}+\frac{\partial q(\mathbf{y}_{t+\tau},\theta)}{\partial\theta}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta'}\right]\bigg|_{\theta=\hat{\theta}_n}.\end{aligned}$$

Then by the Ergodic theorem and $\hat{\theta}_n - \theta_0 \xrightarrow{a.s.} 0$, it is easy to show that

$$\hat{V}_n - V_n(\theta_0) \xrightarrow{a.s.} 0.$$

For example, in Sect. 3.4, for the rounded MA(q) model, we can estimate $V(\theta_0)$ by

$$\hat{V}_n = \frac{1}{n-a}\sum_{t=1}^{n-a}\frac{1}{p^2(\mathbf{y}_t;\theta)}\frac{\partial p(\mathbf{y}_t;\theta)}{\partial\theta}\frac{\partial p(\mathbf{y}_t;\theta)}{\partial\theta'}\Bigg|_{\theta=\hat{\theta}_n} + \frac{1}{n-a}\sum_{\tau=1}^{q}\sum_{t=1}^{n-a-\tau}\frac{1}{p(\mathbf{y}_t;\theta)}\frac{1}{p(\mathbf{y}_{t+\tau};\theta)} \times\left[\frac{\partial p(\mathbf{y}_t;\theta)}{\partial\theta}\frac{\partial p(\mathbf{y}_{t+\tau};\theta)}{\partial\theta'}+\frac{\partial p(\mathbf{y}_t;\theta)}{\partial\theta}\frac{\partial p(\mathbf{y}_{t+\tau};\theta)}{\partial\theta'}\right]\Bigg|_{\theta=\hat{\theta}_n}.$$

Case 3 $\{\frac{\partial q(\mathbf{Y}_t,\theta_0)}{\partial\theta}\}$ are weakly dependent.

In this case, we have

$$V(\theta_0) = E\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\Bigg|_{\theta=\theta_0} + \sum_{t=1}^{\infty}E\left[\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_{t+1},\theta)}{\partial\theta'}+\frac{\partial q(\mathbf{Y}_{t+1},\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\right]\Bigg|_{\theta=\theta_0}$$

By the moment approach, we can estimate $V(\theta)$ by

$$\hat{V}_n = \frac{1}{n-a}\sum_{t=1}^{n-a}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta'}\Bigg|_{\theta=\hat{\theta}_n} + \frac{1}{n-a}\sum_{\tau=1}^{m_n}\omega_{n\tau}\sum_{t=1}^{n-a-\tau}\left[\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta}\frac{\partial q(\mathbf{y}_{t+\tau},\theta)}{\partial\theta'}+\frac{\partial q(\mathbf{y}_{t+\tau},\theta)}{\partial\theta}\frac{\partial q(\mathbf{y}_t,\theta)}{\partial\theta'}\right]\Bigg|_{\theta=\hat{\theta}_n},$$

where $\omega_{n\tau}$ is a reducing factor which tends to 1 for all fixed τ and tends to 0 as $\tau\to\infty$ with certain rate. For example, by Theorem 6.20 and Lemma 6.22 of White (2001) and under the regularity conditions guaranteeing $\hat{\theta}_n-\theta_0\overset{a.s.}{\to}0$, we may choose $m_n\to\infty$ with a rate $m_n=o(n^{1/4})$ and $\omega_{n\tau}=1-\tau/(m_n+1)$ for $\tau=1,2,\ldots,m_n$, and we have

$$\hat{V}_n - V_n(\theta_0)\overset{a.s.}{\to}0.$$

3.3 Rounding IID data

As mentioned in the Introduction, without loss of generality, from now on we will assume that the observations are rounded to the nearest integers. Suppose that the original sample $\mathbf{X}_1,\ldots,\mathbf{X}_n$ are iid d-dimension random vectors. If $a=0$, then the SOS MLE simply reduces to the exact MLE. Therefore, the general theory for MLE holds. In more detailed, under this assumption, the log-likelihood of the rounded observations $\mathbf{y}_1, \ldots, \mathbf{y}_n$ is

$$l_n(\mathbf{y}; \theta) = \sum_{t=1}^{n} \log p(\mathbf{y}_t, \theta) = \sum_{\mathbf{j}=-\infty}^{\infty} f_{\mathbf{j}} q_{\mathbf{j}}(\theta)$$

where θ is a k-dimensional parameter vector, $q_{\mathbf{j}}(\theta) = \log p(\mathbf{j}; \theta)$ with

$$p(\mathbf{j}; \theta) = P\left(\mathbf{X}_t \in \prod_{h=1}^{d} (j_h - 0.5, j_h + 0.5]\right), \quad \mathbf{j} = (j_1, j_2, \ldots j_d)'$$

and $f_{\mathbf{j}}$ is the frequency of $\mathbf{j}$ among $y_1, \ldots, y_n$.

By Theorems 1 and 2, we have the following corollaries:

Corollary 1 *Under regularity conditions* 1–3, 5 *and* 6, *we have*

$$\widehat{\theta} \to \theta_0, \quad \textit{almost surely}.$$

Corollary 2 *Under regularity conditions* 1–3 *and* 5–7,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, I^{-1}(\theta_0))$$

where the Fisher information matrix is given by

$$I(\theta_0) = \sum_{\mathbf{j}=-\infty}^{\infty} \frac{1}{p_{\mathbf{j}}(\theta)} \frac{\partial p_{\mathbf{j}}(\theta)}{\partial \theta} \frac{\partial p_{\mathbf{j}}(\theta)}{\partial \theta'}\bigg|_{\theta_0}$$

Corollaries 1 and 2 show that the maximum likelihood estimates of the parameters based on rounded data are consistent and asymptotically normal. An estimate of $I(\theta_0)$ can be obtained by replacing θ_0 with $\hat{\theta}_n$.

3.4 Rounding the ARMA(p, q) and AR(p) models

In this subsection, we first spell out the ARMA model, paying particular interest to its distributional properties with an unrounded (true) sample. Following that, we derive the SOS estimator and its properties.

Suppose that $\{X_t\}$ is a causal ARMA(p, q) process

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = c + \epsilon_t - \vartheta_1 \epsilon_{t-1} - \cdots - \vartheta_q \epsilon_{t-q}, \tag{8}$$

where $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$ for $t = 1, \ldots, n$. It is well known that the causality condition is that $1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| \leq 1$.

Lemma 1 *Let* $\{X_t\}$ *satisfy the causal ARMA*(p, q) *process* (8)*, or equivalently,*

$$\phi(B)X_t = c + \vartheta(B)\epsilon_t$$

where

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

and

$$\vartheta(B) = 1 - \vartheta_1 B - \cdots - \vartheta_q B^q,$$

and B *is the backward-shift operator, that is,*

$$B^j X_t = X_{t-j}, \quad j = 0, \pm 1, \pm 2, \ldots$$

Then, the causal expression of X_t *can be given as follows:*

$$X_t = \mu + \psi(B)\epsilon_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

where

$$\psi(B) = \frac{\vartheta(B)}{\phi(B)} = \sum_{j=0}^{\infty} \psi_j B^j, \qquad \mu = \frac{c}{1 - \phi_1 - \cdots - \phi_p},$$

and $\sum_{j=0}^{\infty} | \psi_j | < \infty$.

This lemma can be found in Brockwell and Davis (1991), Chapter 3.

Proposition 1 *Under the assumptions of Lemma* 1 *and the assumptions that* $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$, *the following assertions hold:*

- *For any fixed m,* $(X_t, X_{t-1}, \ldots, X_{t-m})'$ *is of multivariate normal distribution* $N(\mu \cdot \mathbf{1}_{m+1}, \mathbf{\Sigma}_{(m+1)\times(m+1)})$ *with* $\mu = c/(1 - \phi_1 - \cdots - \phi_m)$, $\gamma_i = E(X_{t+i} - \mu)(X_t - \mu)$, *and*

$$\mathbf{\Sigma}_{(m+1)\times(m+1)} = \sigma^2 \cdot V_{p+1} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_m \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_m & \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{bmatrix}.$$

- $\{X_t\}$ *is a strictly stationary Gaussian time series.*

Proof That $\mathbf{X} = (X_0, X_1, \ldots, X_m)'$ is an $m+1$ dimensional normal distribution can be seen from the causal expression $X_t = \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$. The expression of μ can be obtained by taking expectation on both sides of (8). The expression of Σ can be obtained by definition of γ_i. Further, by the causal expression of X_t, we have

$$\gamma_i = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+i}.$$

□

For a general ARMA(p, q) model, we need to take $a = p + q$ and apply the theory discussed in Sect. 3.1. Now, as an illustration, consider the case of ARMA(1, 1). Based on the general procedure of SOS, split the rounded data as follows:

$$(Y_1, Y_2, Y_3), (Y_2, Y_3, Y_4), \ldots, (Y_{n-2}, Y_{n-1}, Y_n).$$

For any integer j_1, j_2, and j_3, let $\mathbf{j} = (j_1, j_2, j_3)'$, $A_{\mathbf{j}} = \prod_{h=1}^{3}(j_h - 0.5, j_h + 0.5]$ and $\mathbf{X}_t = (X_t, X_{t+1}, X_{t+2})'$. Then the probability of the event $\{Y_1 = j_1, Y_2 = j_2, Y_3 = j_3\}$ is found as follows:

$$\begin{aligned} P(\mathbf{X}_t \in A_{\mathbf{j}}) &\overset{\triangle}{=} P\left(Y_t = j_1, Y_{t+1} = j_2, Y_{t+2} = j_3\right) \\ &= \int_{(X_1, X_2, X_3) \in A_j} \frac{1}{2\pi |\mathbf{\Sigma}_{3\times 3}|^{\frac{1}{2}}} \cdot \exp\left[-\frac{1}{2}\begin{pmatrix} x_1 - \mu \\ x_2 - \mu \\ x_3 - \mu \end{pmatrix}' \mathbf{\Sigma}_{3\times 3}^{-1} \begin{pmatrix} x_1 - \mu \\ x_2 - \mu \\ x_3 - \mu \end{pmatrix}\right] \mathrm{d}x. \end{aligned}$$

Considering $\mathbf{Y}_1, \ldots, \mathbf{Y}_{n-2}$ as a sample of iid 3-dimensional random vectors, the SOS log-likelihood is

$$l_{\text{sos}}(\mathbf{y}; \theta) = \sum_{t=1}^{n-2} \log p(\mathbf{y}_t; \theta) = \sum_{\mathbf{j}=-\infty}^{\infty} f_{\mathbf{j}} q(\mathbf{j}, \theta) \tag{9}$$

where θ is a 4-dimensional (μ, σ^2, ϕ_1 and ϑ_1) parameter vector, $\mathbf{j} = (j_1, j_2, j_3)'$ and

$$f_{\mathbf{j}} = \sum_{t=1}^{n-2} I_{\{\mathbf{Y}_t = \mathbf{j}\}}, \quad q(\mathbf{j}; \theta) = \log p(\mathbf{j}; \theta),$$

$$p(\mathbf{j}; \theta) = P\left(\mathbf{X}_t \in \prod_{t=1}^{3} (j_t - 0.5, j_t + 0.5]\right).$$

By Theorems 1 and 2, we have the following corollaries.

Corollary 3 *Suppose conditions of Proposition 1 hold, then we have*

$$\widehat{\theta}_n \to \theta_0 \quad a.s.$$

Corollary 4 *Suppose conditions of Proposition 1 hold, then we have,*

$$\sqrt{n-a}\left(\hat{\theta}_n-\theta_0\right)\xrightarrow{d}N\left(0,I^{-1}(\theta_0)V(\theta_0)I^{-1}(\theta_0)\right)$$

where

$$I(\theta_0)=\sum_{\mathbf{j}=-\infty}^{\infty}\frac{1}{p(\mathbf{j};\theta)}\frac{\partial p(\mathbf{j};\theta)}{\partial\theta}\frac{\partial p(\mathbf{j};\theta)}{\partial\theta'}\Bigg|_{\theta_0}$$

$$V(\theta_0)=I(\theta_0)+\sum_{t=1}^{\infty}\sum_{\mathbf{j},\mathbf{k}=-\infty}^{\infty}\frac{p_{t,\mathbf{j},\mathbf{k}}(\theta)}{p_{\mathbf{j}}(\theta)p_{\mathbf{k}}(\theta)}\left[\frac{\partial p_{\mathbf{j}}(\theta)}{\partial\theta}\frac{\partial p_{\mathbf{k}}(\theta)}{\partial\theta'}+\frac{\partial p_{\mathbf{k}}(\theta)}{\partial\theta}\frac{\partial p_{\mathbf{j}}(\theta)}{\partial\theta'}\right]\Bigg|_{\theta_0},$$

where

$$p_{t,\mathbf{j},\mathbf{k}}(\theta)=P\left(\mathbf{X}_1\in A_{\mathbf{j}},\mathbf{X}_{1+t}\in A_{\mathbf{k}}\right).$$

4 Ramification of the SOS approach

As argued in Remark 3 in Sect. 2, the parameters may not be identifiable in the SOS method if the a chosen too small or the computation too time consuming if a is large. Now, we propose the following modified procedure to obtain the SOS Estimator (MSOSE) of parameters θ.

1. Split the data as the following:

$$\begin{array}{cccccccc}
(X_1,X_2) & (X_2,X_3) & (X_3,X_4) & \cdots & (X_{n-3},X_{n-2}) & (X_{n-2},X_{n-1}) & (X_{n-1},X_n)\\
(X_1,X_3) & (X_2,X_4) & (X_3,X_5) & \cdots & (X_{n-3},X_{n-1}) & (X_{n-2},X_n) & \\
(X_1,X_4) & (X_2,X_5) & (X_3,X_6) & \cdots & (X_{n-3},X_n) & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \\
(X_1,X_{n-1}) & (X_2,X_n) & & & & & \\
(X_1,X_n) & & & & & &
\end{array}$$

2. For brevity, let $\mathbf{X}_{it}=(X_t,X_{t+i})'$, where $i=1,2,\ldots,n-1$; $t=1,2,\ldots,n-i$. Consider $\{\mathbf{X}_{it}\}$ as a sample of independent 2-dimensional random vectors with $1\le i\le a$, $1\le t\le n-i$, where a is a fixed integer in $\{1,2,\ldots,n-1\}$. Then maximizing the SOS log-likelihood $\sum_i\sum_t q_i(\mathbf{y}_{it},\theta)$ where $q_i(\mathbf{y}_{it},\theta)=\log p_i(\mathbf{y}_{it},\theta)$, $p_i(\mathbf{y}_{it},\theta)$ is the probability that $\mathbf{X}_{it}$ falls in the square with center $\mathbf{y}_{it}$ and edge length of rounding unit. We then obtain an MSOSE of parameters θ, denoted by $\widehat{\theta}_n$.

As mentioned in Sect. 2, without loss of generality, suppose that the original sample is $X_1,\ldots,X_n$. Under this assumption, the modified SOS log-likelihood of the observations $\{\mathbf{y}_{it}\}$ is

$$l_{m\text{sos}}(\mathbf{y},\theta) = \sum_{t=1}^{n-1} \log p_1(\mathbf{y}_{1t},\theta) + \sum_{t=1}^{n-2} \log p_2(\mathbf{y}_{2t},\theta) + \cdots + \sum_{t=1}^{n-a} \log p_a(\mathbf{y}_{at},\theta)$$
$$= \sum_{t=1}^{n-1} q_1(\mathbf{y}_{1t},\theta) + \sum_{t=1}^{n-2} q_2(\mathbf{y}_{2t},\theta) + \cdots + \sum_{t=1}^{n-a} q_a(\mathbf{y}_{at},\theta)$$

where a is fixed.

Similar to the discussion given in Sect. 3, we have the following theorems.

Theorem 4 *Under regularity conditions* 1–7*,*

$$\widehat{\theta}_n \to \theta_0 \quad a.s.$$

Theorem 5 *Under regularity conditions* 1–10*,*

$$\sqrt{n-a}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, G)$$

where the matrix G is given in the proof of these theorems which is postponed to the Appendix.

5 Simulation results

5.1 AR(1) model

Consider the following two AR(1) models:

$$(a)\ X_t = c + \phi X_{t-1} + \epsilon_t; \quad (b)\ \ X_t = \phi X_{t-1} + \epsilon_t \tag{10}$$

where $\epsilon_t \overset{i.i.d.}{\sim} N(0,\sigma^2)$. Data $X_1,\ldots,X_n$ are generated from the above models and denote the corresponding rounded data by $Y_1,\ldots,Y_n$. The AR(1) models of the forms in (10) are fitted by various methods and the simulation results are summarized in Tables 1 and 2 for models (a) and (b) of (10) respectively. The first two columns show the parameter notation and the true values of the parameters. In the simulation, the sample size is 10,000 for model (a) and 100 for model (b) and repeated 100 times for each model. A traditional procedure for the unrounded data is the so-called conditional maximum likelihood estimator (CMLE). The third column, labeled as "unrounded sample", shows the average of the estimates of parameters by the CMLE based on the unrounded data $\mathbf{X}$ and the corresponding square roots of the estimated MSE (mean squared error) in parentheses. The fourth columns, labeled as "uncorrected, rounded", presents the average of the estimates of parameters by the CMLE based on the rounded data $\mathbf{Y}$ and the corresponding square roots of the MSE. The fifth columns, labeled "A–K corrected", presents the average of estimates and their square roots of MSE, adjusted by the formulae (3.1.1)–(3.1.3) for the rounded data for autoregressive processes, suggested by Stam and Cogger (1993). The last column, labeled as "SOS method", presents the simulation results by the new method.

Table 1 Simulations for the model (a) of (10)

	True model	Unrounded sample	Uncorrected rounded	*A–K* corrected	SOS method
μ	1.0	1.0000 (0.0031)	1.0000 (0.0026)	1.0004 (0.0027)	1.0000 (0.0050)
ϕ	0.3	0.2988 (0.0107)	0.0976 (0.2028)	0.1015 (0.1988)	0.3035 (0.0314)
σ^2	0.0625	0.0624 (0.0009)	0.0556 (0.0073)	−0.0280 (0.0905)	0.0625 (0.0018)
μ	1.0	1.0006 (0.0048)	1.0009 (0.0040)	1.0000 (0.0037)	1.0005 (0.0058)
ϕ	0.5	0.5003 (0.0086)	0.2261 (0.2743)	0.2269 (0.2736)	0.5093 (0.0249)
σ^2	0.0625	0.0626 (0.0010)	0.0792 (0.0169)	−0.0086 (0.0711)	0.0619 (0.0022)
μ	1.0	0.9985 (0.0123)	0.9990 (0.0128)	1.0000 (0.0121)	1.0000 (0.0122)
ϕ	0.8	0.8007 (0.0058)	0.5799 (0.2204)	0.5840 (0.2163)	0.8079 (0.0122)
σ^2	0.0625	0.0626 (0.0008)	0.1545 (0.0920)	0.0422 (0.0205)	0.0600 (0.0033)
μ	2.25	2.2501 (0.0035)	2.1861 (0.0821)	2.1679 (0.0822)	2.2557 (0.0092)
ϕ	0.3	0.2991 (0.009)	0.1559 (0.1445)	0.1592 (0.1411)	0.3076 (0.0208)
σ^2	0.0625	0.0625 (0.0009)	0.1403 (0.0778)	0.0549 (0.0081)	0.0595 (0.0043)
μ	2.25	2.2504 (0.0047)	2.1890 (0.0612)	2.1890 (0.0613)	2.2526 (0.0071)
ϕ	0.5	0.4988 (0.0089)	0.2930 (0.2073)	0.2995 (0.2008)	0.5130 (0.0216)
σ^2	0.0625	0.0626 (0.0009)	0.1489 (0.0864)	0.0577 (0.0060)	0.0602 (0.0034)
μ	2.25	2.2506 (0.0113)	2.2402 (0.0149)	2.2387 (0.0196)	2.2491 (0.0126)
ϕ	0.8	0.7997 (0.0055)	0.5914 (0.2089)	0.5972 (0.2031)	0.8099 (0.0129)
σ^2	0.0625	0.0626 (0.0009)	0.1673 (0.1049)	0.0549 (0.0086)	0.0596 (0.0039)
μ	1	1.0014 (0.0154)	1.0014 (0.0156)	0.9979 (0.0137)	1.0023 (0.0142)
ϕ	0.3	0.2993 (0.0094)	0.2781 (0.0238)	0.3022 (0.0118)	0.3011 (0.0098)
σ^2	1	0.9998 (0.0146)	1.0908 (0.0921)	0.9991 (0.0167)	0.9971 (0.0158)
μ	1	0.9982 (0.0197)	0.9982 (0.0196)	1.0004 (0.0178)	1.0034 (0.0189)
ϕ	0.5	0.5004 (0.0091)	0.4712 (0.0303)	0.5050 (0.0103)	0.4991 (0.0094)
σ^2	1	1.0011 (0.0147)	1.1036 (0.1048)	1.0017 (0.0159)	0.9995 (0.0141)
μ	1	1.0231 (0.1056)	1.0233 (0.1058)	1.0109 (0.1016)	0.9881 (0.0982)
ϕ	0.9	0.9000 (0.0051)	0.8859 (0.0151)	0.9041 (0.0060)	0.9006 (0.0046)
σ^2	1	1.0008 (0.0150)	1.1506 (0.1516)	1.0020 (0.0184)	1.0003 (0.0156)
μ	1	0.9996 (0.0070)	0.1000 (0.0075)	1.0007 (0.0073)	1.0000 (0.0087)
ϕ	0.3	0.3000 (0.0111)	0.2292 (0.0716)	0.2347 (0.0660)	0.3019 (0.0105)
σ^2	0.25	0.2500 (0.0037)	0.3341 (0.0842)	0.2466 (0.0061)	0.2481 (0.0052)
μ	1	1.0009 (0.0099)	1.0005 (0.0102)	1.0011 (0.0107)	1.0002 (0.0105)
ϕ	0.5	0.4999 (0.0084)	0.3999 (0.3482)	0.4091 (0.0915)	0.5025 (0.0121)
σ^2	0.25	0.2498 (0.0036)	0.3482 (0.0983)	0.2510 (0.0051)	0.2474 (0.0055)
μ	1	1.0005 (0.0264)	0.9998 (0.0266)	0.9981 (0.0245)	1.0022 (0.0261)
ϕ	0.8	0.7998 (0.0052)	0.7140 (0.0863)	0.7252 (0.0752)	0.8005 (0.0067)
σ^2	0.25	0.2497 (0.0034)	0.3806 (0.1307)	0.2558 (0.0079)	0.2472 (0.0064)
μ	2.25	2.250 (0.0077)	0.2480 (0.0088)	2.2480 (0.0077)	2.2506 (0.0070)
ϕ	0.3	0.3005 (0.0102)	0.2304 (0.0703)	0.2352 (0.0658)	0.3025 (0.0136)
σ^2	0.25	0.2497 (0.0037)	0.3390 (0.0892)	0.2515 (0.0050)	0.2488 (0.0053)

Table 1 continued

	True model	Unrounded sample	Uncorrected rounded	*A–K* corrected	SOS method
μ	2.25	2.2510 (0.0099)	2.2508 (0.0104)	2.2527 (0.0105)	2.2494 (0.0113)
ϕ	0.5	0.4980 (0.0083)	0.3984 (0.1020)	0.4097 (0.0908)	0.5028 (0.0109)
σ^2	0.25	0.2501 (0.0039)	0.3500 (0.1001)	0.2524 (0.0060)	0.2473 (0.0054)
μ	2.25	2.2524 (0.0254)	2.2524 (0.0252)	2.2491 (0.0229)	2.2574 (0.0260)
ϕ	0.8	0.7998 (0.0056)	0.7144 (0.0860)	0.7252 (0.0753)	0.8004 (0.0076)
σ^2	0.25	0.2499 (0.0035)	0.3808 (0.1309)	0.2541 (0.0066)	0.2471 (0.0061)

Table 2 Simulations for the model (b) of (10)

	True model	Unrounded sample	Uncorrected rounded	*A–K* corrected	SOS method
ϕ	−0.8	−0.81 (0.058)	−0.58 (0.131)	−0.78 (0.145)	−0.76 (0.113)
σ^2	0.0625	0.06 (0.009)	0.16 (0.041)	0.05 (0.042)	0.07 (0.027)
ϕ	−0.5	−0.51 (0.096)	−0.23 (0.161)	−0.47 (0.371)	−0.39 (0.291)
σ^2	0.0625	0.06 (0.009)	0.07 (0.029)	−0.01 (0.029)	0.07 (0.019)
ϕ	−0.2	−0.20 (0.100)	−0.06 (0.093)	−0.15 (0.241)	−0.15 (0.266)
σ^2	0.0625	0.06 (0.009)	0.05 (0.023)	−0.03 (0.022)	0.058 (0.016)
ϕ	0	−0.03 (0.091)	−0.01 (0.083)	−0.04 (0.244)	0.02 (0.262)
σ^2	0.0625	0.06 (0.009)	0.05 (0.019)	−0.04 (0.019)	0.06 (0.015)
ϕ	0.2	0.19 (0.095)	0.05 (0.112)	0.12 (0.370)	0.10 (0.234)
σ^2	0.0625	0.06 (0.010)	0.05 (0.023)	−0.03 (0.023)	0.06 (0.014)
ϕ	0.5	0.46 (0.086)	0.16 (0.139)	0.33 (0.314)	0.46 (0.275)
σ^2	0.0625	0.06 (0.009)	0.07 (0.029)	−0.01 (0.028)	0.06 (0.021)
ϕ	0.8	0.77 (0.072)	0.52 (0.146)	0.74 (0.174)	0.77 (0.108)
σ^2	0.0625	0.06 (0.010)	0.15 (0.035)	0.04 (0.035)	0.06 (0.026)
ϕ	−0.8	−0.80 (0.059)	−0.70 (0.039)	−0.78 (0.083)	−0.78 (0.088)
σ^2	0.25	0.24 (0.059)	0.38 (0.077)	0.25 (0.061)	0.25 (0.058)
ϕ	−0.5	−0.51 (0.089)	−0.42 (0.089)	−0.50 (0.099)	−0.52 (0.122)
σ^2	0.25	0.25 (0.037)	0.35 (0.052)	0.25 (0.052)	0.25 (0.058)
ϕ	−0.2	−0.19 (0.101)	−0.15 (0.105)	−0.19 (0.130)	−0.20 (0.127)
σ^2	0.25	0.25 (0.036)	0.33 (0.050)	0.25 (0.050)	0.25 (0.045)
ϕ	0	−0.01 (0.099)	−0.01 (0.109)	−0.01 (0.138)	0.01 (0.129)
σ^2	0.25	0.25 (0.040)	0.33 (0.052)	0.25 (0.052)	0.24 (0.045)
ϕ	0.2	0.18 (0.105)	0.15 (0.105)	0.19 (0.129)	0.20 (0.147)
σ^2	0.25	0.25 (0.036)	0.33 (0.051)	0.24 (0.051)	0.24 (0.043)
ϕ	0.5	0.49 (0.073)	0.39 (0.092)	0.47 (0.104)	0.48 (0.102)
σ^2	0.25	0.25 (0.035)	0.35 (0.048)	0.26 (0.048)	0.25 (0.048)
ϕ	0.8	0.77 (0.070)	0.68 (0.087)	0.76 (0.081)	0.79 (0.076)

Table 2 continued

	True model	Unrounded sample	Uncorrected rounded	A–K corrected	SOS method
σ^2	0.25	0.25 (0.038)	0.38 (0.057)	0.26 (0.058)	0.24 (0.048)
ϕ	−0.8	−0.80 (0.060)	−0.77 (0.067)	−0.79 (0.064)	−0.79 (0.059)
σ^2	1	1.00 (0.130)	1.13 (0.148)	1.00 (0.149)	0.98 (0.153)
ϕ	−0.5	−0.50 (0.073)	−0.47 (0.077)	−0.49 (0.079)	−0.49 (0.092)
σ^2	1	1.02 (0.152)	1.12 (0.173)	1.02 (0.173)	0.99 (0.155)
ϕ	−0.2	−0.21 (0.103)	−0.20 (0.105)	−0.21 (0.113)	−0.20 (0.105)
σ^2	1	0.99 (0.139)	1.07 (0.157)	0.98 (0.158)	1.00 (0.152)
ϕ	0	−0.02 (0.104)	−0.01 (0.101)	−0.02 (0.109)	−0.01 (0.113)
σ^2	1	1.00 (0.160)	1.09 (0.179)	1.01 (0.179)	1.00 (0.163)
ϕ	0.2	0.19 (0.081)	0.18 (0.082)	0.19 (0.088)	0.19 (0.104)
σ^2	1	0.98 (0.146)	1.06 (0.161)	0.98 (0.161)	1.01 (0.156)
ϕ	0.5	0.48 (0.092)	0.46 (0.094)	0.48 (0.097)	0.50 (0.082)
σ^2	1	0.98 (0.134)	1.08 (0.138)	0.98 (0.139)	1.00 (0.131)
ϕ	0.8	0.78 (0.066)	0.75 (0.073)	0.78 (0.069)	0.80 (0.060)
σ^2	1	0.99 (0.145)	1.13 (0.159)	0.99 (0.157)	0.98 (0.149)
ϕ	−0.8	−0.80 (0.058)	−0.79 (0.060)	−0.80 (0.059)	−0.79 (0.061)
σ^2	4	3.94 (0.596)	4.09 (0.612)	3.95 (0.611)	3.94 (0.532)
ϕ	−0.5	−0.52 (0.076)	−0.51 (0.079)	−0.52 (0.079)	−0.47 (0.094)
σ^2	4	3.98 (0.633)	4.08 (0.654)	3.98 (0.654)	3.92 (0.447)
ϕ	−0.2	−0.21 (0.096)	−0.20 (0.099)	−0.21 (0.100)	−0.19 (0.109)
σ^2	4	4.10 (0.589)	4.20 (0.600)	4.11 (0.600)	3.95 (0.550)
ϕ	0	−0.01 (0.105)	−0.02 (0.100)	−0.02 (0.102)	0.00 (0.110)
σ^2	4	3.96 (0.563)	4.04 (0.580)	3.96 (0.580)	3.89 (0.541)
ϕ	0.2	0.18 (0.096)	0.18 (0.096)	0.18 (0.097)	0.20 (0.110)
σ^2	4	4.13 (0.562)	4.20 (0.570)	4.11 (0.571)	3.98 (0.546)
ϕ	0.5	0.48 (0.096)	0.47 (0.096)	0.48 (0.097)	0.48 (0.084)
σ^2	4	3.99 (0.564)	4.10 (0.572)	4.00 (0.573)	3.98 (0.539)
ϕ	0.8	0.78 (0.066)	0.77 (0.068)	0.78 (0.067)	0.80 (0.064)
σ^2	4	3.95 (0.558)	4.08 (0.586)	3.95 (0.586)	3.94 (0.548)

Table 1 presents results of a simulation study on the model (a) of (10) with parameter combinations of $\mu = \frac{c}{1-\phi} = 1, 2.25$, $\phi = 0.3, 0.5, 0.8, 0.9$ and $\sigma^2 = 0.0625, 0.25, 1.0$.

One can see that the SOSE significantly improves the CMLE. Comparing with the A–K correction, one sees that the SOSE is also clearly better than the A–K correction when σ^2 is small and slightly better than the latter when σ^2 is large. A serious draw-back of the A–K correction is that the estimate of σ^2 may take negative values.

Table 2 shows the simulation results for the model (b) of (10) with parameter combinations of $\phi = -0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8$ and $\sigma^2 = 0.0625, 0.25, 1, 4$.

The simulation results show that both A–K and SOS are significantly improves the CMLE. The simulation results show that the SOS method produces smaller MSE in 25 cases and slightly larger MSE in 2 cases among total 28 cases. We believe that the SOS is slightly better than A–K correction and exceptions are due to the randomness since the length size is only 100.

5.2 AR(2) model and ARMA(1,1) model

In order to save computation time, we present the modified SOSE (MSOSE) as proposed in Sect. 4. We shall adopt this method to the following models AR(2) and ARMA(1,1):

$$\text{(a)}\ \ X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t; \quad \text{(b)}\ \ X_t - \phi X_{t-1} = c + \epsilon_t - \vartheta \epsilon_{t-1}. \tag{11}$$

where $\epsilon_t \sim \ iid\ N(0, \sigma^2)$. We generate $X_1, \ldots, X_n$ first from the models and then round them to $Y_1, \ldots, Y_n$.

As an illustration, consider model (b) of (11). We have

$$(X_t, X_{t-1})' \sim N(\mu \mathbf{1}, \mathbf{\Sigma}_1), \quad (X_t, X_{t-2})' \sim N(\mu \mathbf{1}, \mathbf{\Sigma}_2)$$

where $\mu = c/(1-\phi)$,

$$\mathbf{\Sigma}_1 = \begin{bmatrix} R(0) & R(1) \\ R(1) & R(0) \end{bmatrix}, \quad \mathbf{\Sigma}_2 = \begin{bmatrix} R(0) & R(2) \\ R(2) & R(0) \end{bmatrix},$$

and

$$\begin{cases} R(0) = \dfrac{1+\vartheta^2 - 2\vartheta\phi}{1-\phi^2}\sigma^2, \\ R(1) = \dfrac{\phi(1+\vartheta^2) - \vartheta(1+\phi^2)}{1-\phi^2}\sigma^2, \\ R(2) = \dfrac{\phi^2(1+\vartheta^2) - \phi\vartheta(1+\phi^2)}{1-\phi^2}\sigma^2. \end{cases}$$

Based on the above argument, split the data $Y_1, Y_2, \ldots, Y_n$ as follows:

$$\begin{array}{l} (Y_1, Y_2)\ (Y_2, Y_3)\ (Y_3, Y_4)\ \cdots\ (Y_{n-3}, Y_{n-2})\ (Y_{n-2}, Y_{n-1})\ (Y_{n-1}, Y_n) \\ (Y_1, Y_3)\ (Y_2, Y_4)\ (Y_3, Y_5)\ \cdots\ (Y_{n-3}, Y_{n-1})\ \ \ (Y_{n-2}, Y_n) \end{array}$$

Then the MSOS log-likelihood of the rounded observations is

$$l_{\mathrm{msos}}(y; \theta) = \sum_{k=1}^{2} \sum_{t=1}^{n-k} \log p_k(y_t, y_{t+k}, \theta) = \sum_{k=1}^{2} \sum_{t=1}^{n-k} q_k(y_t, y_{t+k}, \theta) \tag{12}$$

where $\theta = (\mu, R(0), R(1), R(2))'$ and

$$p_k(y_t, y_{t+k}, \theta) = P_\theta\left(X_t \in (y_t - 0.5, y_t + 0.5]; X_{t+k} \in (y_{t+k} - 0.5, y_{t+k} + 0.5]\right),$$
$$q_k(y_t, y_{t+k}, \theta) = \log p_k(y_t, y_{t+k}, \theta).$$

If we write $p_k(i, j, \theta)$ as $p_{ijk}(\theta)$, then

$$p_{ijk}(\theta) = \int_{i-0.5}^{i+0.5}\int_{j-0.5}^{j+0.5} \frac{1}{2\pi|\Sigma_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\begin{pmatrix} x_1-\mu \\ x_2-\mu \end{pmatrix}' \Sigma_k^{-1} \begin{pmatrix} x_1-\mu \\ x_2-\mu \end{pmatrix}\right] \mathrm{d}x_2 \mathrm{d}x_1$$

where

$$\Sigma_k = \begin{pmatrix} R(0) & R(k) \\ R(k) & R(0) \end{pmatrix}, \quad k = 1, 2.$$

The MSOSE will processed through the following steps.

Step 1 Obtain the MSOSE of parameters μ, $R(0)$, $R(1)$, and $R(2)$ by maximizing (12), and denoted them by $\hat{\mu}$, $\hat{R}(0)$, $\hat{R}(1)$ and $\hat{R}(2)$, respectively.

Step 2 Solve the equations

$$\begin{aligned} \hat{R}(0) &= \frac{1+\vartheta^2-2\vartheta\phi}{1-\phi^2}\sigma^2, \\ \hat{R}(1) &= \frac{\phi(1+\vartheta^2)-\vartheta(1+\phi^2)}{1-\phi^2}\sigma^2, \\ \hat{R}(2) &= \frac{\phi^2(1+\vartheta^2)-\phi\vartheta(1+\phi^2)}{1-\phi^2}\sigma^2. \end{aligned} \tag{13}$$

Then we obtain the estimators of ϕ, ϑ and σ^2, denoted by $\hat{\phi}$, $\hat{\vartheta}$ and $\hat{\sigma}^2$, respectively.

Remark 4 Generally, the system of Eqs. (13) does have a solution provided $\begin{pmatrix} \hat{R}(0) & \hat{R}(1) \\ \hat{R}(1) & \hat{R}(2) \end{pmatrix}$ is positive definite. Because of the strong consistency of the MSOSE and of the limiting matrix being positive definite, we know that for almost all large n, the equations (13) does have a solution. However, sometimes, due to random effect, the matrix $\begin{pmatrix} \hat{R}(0) & \hat{R}(1) \\ \hat{R}(1) & \hat{R}(2) \end{pmatrix}$ may not be positive definite and hence the equation system (13) may not have any solutions. In this case, we may minimize the mean square errors, that is,

$$\min_{\mu,\phi,\vartheta,\sigma^2}\left(\left(\hat{R}(0)-\frac{1+\vartheta^2-2\vartheta\phi}{1-\phi^2}\sigma^2\right)^2+\left(\hat{R}(1)-\frac{\phi(1+\vartheta^2)-\vartheta(1+\phi^2)}{1-\phi^2}\sigma^2\right)^2 + \left(\hat{R}(2)-\frac{\phi^2(1+\vartheta^2)-\phi\vartheta(1+\phi^2)}{1-\phi^2}\sigma^2\right)^2\right)$$

For this set of models, the simulation comparison is made only between CMLE and MSOSE since the A–K correction does not apply to these models. The simulations are based on 100 datasets of sample lengths 1,000 and 10,000 as shown in Tables 3 and 4. The simulation result show that the MSOSE is much better than the CMLE in the sense of smaller biases and square root of MSE (Tables 3, 4).

6 Applications

We shall revisit the data analysis given in Table Series C of the book *Time Series Analysis* (*Forecasting and Control*) (Box et al. 1994, pp. 544) to illustrate our estimation procedures. The 226 data are records of chemical process temperature readings per minute. From Table Series C, we can see that the 226 chemical temperature data take values at most to one digit after decimal point. But we know that temperature is continuous. Thus, the 226 values in the data set are in fact having rounded. If conventional methods are directly used to analyze the rounded data, this will cause some serious errors, as explored in the Introduction. In order to avoid such problems, we use our estimation procedures to deal with the rounded data. As pointed out in the book (pp. 189) that data in Table Series C are suggested to be an AR(1) process about $\phi = 0.8$ after taking the first difference. That is, the data satisfy the following model

$$\nabla x_t = c + \phi \nabla x_{t-1} + \epsilon_t,$$

Table 3 Simulations for the model (a) of (11)

Method	$\hat{c}(\sqrt{\text{MSE}})$	$\hat{\phi}_1(\sqrt{\text{MSE}})$	$\hat{\phi}_2(\sqrt{\text{MSE}})$	$\hat{\sigma}^2(\sqrt{\text{MSE}})$
$(c = 1, \phi_1 = 0.8, \phi_2 = -0.15, \sigma^2 = 1.0, n = 1,000)$				
CMLE	1.0549 (0.1015)	0.7303 (0.0759)	−0.0973 (0.0627)	1.1306 (0.1396)
MSOSE	1.0137 (0.0768)	0.7984 (0.0355)	−0.1518 (0.0373)	0.9959 (0.0493)
$(c = 1, \phi_1 = 0.8, \phi_2 = -0.15, \sigma^2 = 1.0, n = 10,000)$				
CMLE	1.0463 (0.0539)	0.7339 (0.0667)	−0.0997 (0.0514)	1.1346 (0.1356)
MSOSE	1.0001 (0.0270)	0.8010 (0.0107)	−0.1506 (0.0120)	0.9988 (0.01659)
$(c = 1, \phi_1 = 0.5, \phi_2 = 0.24, \sigma^2 = 1.0, n = 1,000)$				
CMLE	1.0800 (0.1289)	0.4716 (0.0418)	0.2472 (0.0326)	1.1070 (0.1177)
MSOSE	1.0093 (0.0955)	0.4987 (0.0333)	0.2389 (0.0356)	1.0102 (0.0510)
$(c = 1, \phi_1 = 0.5, \phi_2 = 0.24, \sigma^2 = 1.0, n = 10,000)$				
CMLE	1.0756 (0.0830)	0.4759 (0.0260)	0.2446 (0.0098)	1.1089 (0.1102)
MSOSE	0.9953 (0.0309)	0.5025 (0.0118)	0.2384 (0.0116)	0.9967 (0.0154)

Table 4 Simulations for the model (b) of (11)

Method	$\hat{c}(\sqrt{\text{MSE}})$	$\hat{\phi}(\sqrt{\text{MSE}})$	$\hat{\vartheta}(\sqrt{\text{MSE}})$	$\hat{\sigma}^2(\sqrt{\text{MSE}})$
$(c = 0.5, \phi = 0.7, \vartheta = -0.5, \sigma^2 = 1.0, n = 1,000)$				
CMLE	0.4984 (0.0623)	0.7019 (0.0269)	−0.3619 (0.1422)	1.1976 (0.2038)
MSOSE	0.5078 (0.0700)	0.6945 (0.0265)	−0.5143 (0.0695)	0.9868 (0.0605)
$(c = 0.5, \phi = 0.7, \vartheta = -0.5, \sigma^2 = 1.0, n = 10,000)$				
CMLE	0.5057 (0.0230)	0.6975 (0.0082)	−0.3663 (0.1340)	1.2141 (0.2147)
MSOSE	0.4980 (0.0183)	0.7002 (0.0081)	−0.5038 (0.0206)	0.9947 (0.0196)
$(c = 0.5, \phi = 0.9, \vartheta = -0.5, \sigma^2 = 1.0, n = 1,000)$				
CMLE	0.5248 (0.0932)	0.8958 (0.0155)	−0.3389 (0.1639)	1.2620 (0.2696)
MSOSE	0.5281 (0.0925)	0.8952 (0.0153)	−0.5101 (0.0660)	0.9859 (0.0696)
$(c = 0.5, \phi = 0.9, \vartheta = -0.5, \sigma^2 = 1.0, n = 10,000)$				
CMLE	0.5041 (0.0307)	0.8996 (0.0051)	−0.0083 (0.1620)	1.2570 (0.2576)
MSOSE	0.5004 (0.0254)	0.8999 (0.0044)	−0.5058 (0.0206)	0.9902 (0.0232)

Table 5 Comparisons of CMLE, *A–K* and SOS methods for a real data set given in Box et al.

	Uncorrected rounded	*A–K* corrected	SOS method
c	−0.0092	−0.0092	−0.0061
ϕ	0.8074	0.8074	0.8274
σ^2	0.0178	0.0165	0.0164

where $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$ and ∇ denotes the first order difference operator, that is, $\nabla x_t = x_t - x_{t-1}$. Table 5 provides parameter estimates of $\theta = (c, \phi, \sigma^2)$ using the CMLE, *A–K* and SOS methods, respectively. In this example, a is chosen as 1 for the SOS method, the computation time is about 6 seconds. The computation time for the CMLE and *A–K* methods is about 1 second each by using S-Plus. From Table 5, we see that the SOSE of ϕ is 0.8274 and the CMLE of ϕ is 0.8074, giving the difference of 0.02. Note that $\sqrt{n/\sigma^2} \times 0.02 = 2.3426 >> 1.96$, showing that the difference between the two estimation approaches is significant. The real data set suggests that the SOS method is superior to the conventional methods when dealing with the rounded data.

7 Comments and conclusions

As mentioned in the Introduction, rounded data are so popularly encountered in almost every discipline and huge data sets are often met in many situations. Therefore, the rounding error must be taken into account when statistical inferences are made to such problems. From the analysis of the real data given in book by Box et al. the error due to the rounding effect is significant even when the sample size is as small as 225. Our simulation also shows that the rounding errors can have more important effect on dependent sequences such as ARMA(p) models.

On the other hand, although our discussion is limited to rounding errors, the idea of our SOS method applies to the other coarsened errors such as heaped, ceilinged and trimmed error which are frequently met in applications and having more serious effects on statistical inferences involving large data sets.

Although this paper has dealt with some important rounded data problems, many rounded data problems have not yet been dealt with, for example, the handling of rounded data in all kinds of non-parametric statistical models. Further investigation is needed.

Acknowledgments The research of the first two authors was partially supported by the NSFC grant 10871037 and that of the third author was partially supported by the NFSC grant 10871036 and the NUS grant R-155-000-079-112. The authors thank the anonymous referees and the associate editor for their careful reading of the paper and their invaluable comments and suggestions which greatly improve the readability of it.

Appendix

Definition 1 For a sequence of random vectors $\{Z_t\}$, let $\mathscr{F}_\tau^t \equiv \sigma(Z_\tau, \ldots, Z_t)$, $\mathscr{F}_{-\infty}^t \equiv \sigma(\ldots, Z_t)$, $\mathscr{F}_\tau^\infty \equiv \sigma(Z_\tau, \ldots)$ be the σ-fields generated by the random variables indicated in the brackets, and define the mixing coefficients

$$\phi_m \equiv \sup_\tau \sup_{\{F \in \mathscr{F}_{-\infty}^\tau, G \in \mathscr{F}_{\tau+m}^\infty : P(F>0)\}} |P(G|F) - P(G)|$$

and

$$\alpha_m \equiv \sup_\tau \sup_{\{F \in \mathscr{F}_{-\infty}^\tau, G \in \mathscr{F}_{\tau+m}^\infty\}} |P(G \cap F) - P(G)P(F)|.$$

If, for the sequence $\{Z_t\}$, $\phi(m) \to 0$ as $m \to \infty$, $\{Z_t\}$ is called ϕ-mixing. If, for the sequence $\{Z_t\}$, $\alpha(m) \to 0$ as $m \to \infty$, $\{Z_t\}$ is called α-mixing.

Lemma 2 *Let $(\Omega, \mathscr{F}, P)$ be a probability space and let g be an $\mathscr{F}$-measurable function into $\mathbb{R}^k$. Define $Y_t \equiv g(\ldots, Z_{t-1}, Z_t, Z_{t+1}, \ldots)$, where Z_t is a $q \times 1$-dimensional random vector. If $\{Z_t\}$ is strictly stationary, then $\{Y_t\}$ is also strictly stationary.*

Proof See Stout (1974, p. 170, p. 182). □

Lemma 3 *Let $(\Omega, \mathscr{F}, P)$ be a probability space and let g be an $\mathscr{F}$-measurable function into $\mathbb{R}^k$. Define $Y_t \equiv g(Z_t, Z_{t+1}, \ldots, Z_{t+\tau})$, where τ is finite. If the sequence of $q \times 1$ vectors $\{Z_t\}$ is ϕ-mixing (α-mixing) of size $-a$, $a > 0$, then $\{Y_t\}$ is also ϕ-mixing (or α-mixing) with the same size $-a$.*

Proof See White and Domowitz (1984, Lemma 2.1). □

Lemma 4 *Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\{Z_t\}$ be a strictly stationary sequence. If $\alpha(m) \to 0$ as $m \to \infty$. Then $\{Z_t\}$ is ergodic.*

Proof See Rosenblatt (1956). □

Lemma 5 *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space and* $\{Z_t\}$ *be a strictly stationary ergodic scalar sequence with* $E|Z_t| < \infty$*, then* $\bar{Z}_n \equiv \frac{1}{n}\sum_{t=1}^{n} Z_t \overset{a.s.}{\rightarrow} \mu \equiv EZ_t$.

Proof See Stout (1974, p. 181). □

Proof of Theorem 1 Under regularity condition 4, for each fixed $\theta \in \Theta$, by Lemmas 2–4, it is easy to see that $q(\mathbf{Y}_t, \theta)$ is a strictly stationary ergodic sequence. Under regularity condition 5, from Lemma 5, as $n \to \infty$, we have

$$\begin{aligned}\psi_n(\mathbf{Y}, \theta) &\equiv \frac{1}{n-a}\sum_{t=1}^{n-a} q(\mathbf{Y}_t, \theta) \overset{a.s.}{\rightarrow} \psi(\theta_0, \theta) = Eq(\mathbf{Y}_t, \theta)\\ &= \sum_{\mathbf{j}} p(\mathbf{j}, \theta_0)\log p(\mathbf{j}, \theta).\end{aligned} \tag{14}$$

The regularity condition 3 guarantees that $\theta = \theta_0$ is the only maximizer of $\psi(\theta_0, \theta)$.

Next, we show that the convergence (14) is uniform for $\theta \in \overline{\Theta}$, where $\overline{\Theta}$ is an arbitrarily given compact subset of Θ. Choose a finite subset $\{\theta_i, i = 1, \ldots, m\}$ from $\overline{\Theta}$. Then, we have

$$\begin{aligned}\sup_{\theta\in\overline{\Theta}} |\psi_n(\mathbf{Y}, \theta) - \psi(\theta_0, \theta)| &\le \max_{i\le m} |\psi_n(\mathbf{Y}, \theta_i) - \psi(\theta_0, \theta_i)|\\ &+ \sup_{\theta\in\overline{\Theta}} \min_{i\le m} [|\psi_n(\mathbf{Y}, \theta) - \psi_n(\mathbf{Y}, \theta_i)| + |\psi(\theta_0, \theta) - \psi(\theta_0, \theta_i)|].\end{aligned} \tag{15}$$

The first term tends to 0 by (14). By regularity condition 6, the third term can be made arbitrarily small by suitably choosing the finite set $\{\theta_i\}$. To complete our assertion, we only need to show that the second term in (15) can be made arbitrarily small by suitably choosing the finite set $\{\theta_i\}$. By regularity condition 6 and the mean value theorem, we have

$$\begin{aligned}&\sup_{\theta\in\overline{\Theta}} \min_{i\le m} |\psi_n(\mathbf{Y}, \theta) - \psi_n(\mathbf{Y}, \theta_i)|\\ &\le \sup_{\theta\in\overline{\Theta}} \min_{i\le m} \frac{1}{n-a}\sum_{t=1}^{n-a}(h(\mathbf{Y}_t))\|\theta - \theta_i\| \overset{a.s.}{\longrightarrow} (Eh(\mathbf{Y}_1)) \sup_{\theta\in\overline{\Theta}} \min_{i\le m} \|\theta - \theta_i\|\end{aligned}$$

which proves our assertion.

For any $\varepsilon > 0$, by the continuity of $\psi(\theta_0, \theta)$, we have

$$\inf_{\substack{\|\theta-\theta_0\| > \varepsilon\\ \theta\in\overline{\Theta}}} (\psi(\theta_0, \theta_0) - \psi(\theta_0, \theta)) > 0.$$

From this and the uniform convergence of (14), we conclude that with probability 1, when n is large

$$\sup_{\substack{\|\theta-\theta_0\| > \varepsilon\\ \theta\in\overline{\Theta}}} \psi_n(\mathbf{Y}, \theta) < \psi_n(\mathbf{Y}, \theta_0).$$

This proves that any maximum points $\hat{\theta}_n$ of $\psi_n(\mathbf{Y}, \theta)$ for $\theta \in \overline{\Theta}$ must be in the ball $\{\|\theta - \theta_0\| \leq \varepsilon\}$. Consequently, Theorem 1 is proved. □

Proof of Theorem 2 For a constant $b < \psi(\theta_0, \theta_0)$, define

$$A_m = \left\{ \mathbf{x} \in \mathscr{X};\ \max_{\theta \in \overline{\Theta}_m^c} q(\mathbf{y}; \theta) > b \right\}.$$

By regularity condition 7, for each $\mathbf{x} \in \mathscr{X}$, there is an integer m such that $\sup_{\theta \in \overline{\Theta}_m^c} q(\mathbf{y}; \theta) < b$. Therefore,

$$P(A_m) \to 0, \quad \text{as } m \to \infty. \tag{16}$$

Therefore,

$$\begin{aligned} &\sup_{\theta \in \overline{\Theta}_m^c} \frac{1}{n-a} \sum_{t=1}^{n-a} q(\mathbf{y}_t; \theta) \\ &\quad \leq \frac{b}{n-a} \sum_{t=1}^{n-a} 1(\mathbf{X}_t \in A_m^c) \xrightarrow{a.s.} b(1 - P(A_m)). \end{aligned}$$

Choose m so large that $b(1 - P(A_m)) < \psi(\theta_0, \theta_0)$. Then, with probability 1 for all large n,

$$\sup_{\theta \in \overline{\Theta}_m^c} l_{\text{sos}}(\mathbf{Y}, \theta) < l_{\text{sos}}(\mathbf{Y}, \theta_0).$$

This shows that $\hat{\theta}_n \in \overline{\Theta}_m$. By Theorems 1 and 2 follows. □

Lemma 6 *Let $0 < \delta < \infty$ be fixed. Suppose that the stationary sequence $\{Z_t\}$ satisfies the α-mixing condition with $EZ_t = 0$, $E|Z_t|^{2+\delta} < \infty$ $(0 < \delta < \infty)$, and $\sum_{n=1}^{\infty} \alpha_n^{\delta/(2+\delta)} < \infty$. Put $S_n = \sum_{t=1}^{n} Z_t$. Then $\lim_{n\to\infty} n^{-1} E S_n^2 = \sigma^2$, $0 \leq \sigma^2 < \infty$. If $\sigma^2 > 0$, $S_n / \sigma\sqrt{n}$ converges in distribution to the standard normal.*

Proof This is the Corollary 5.1 of Hall and Heyde (1980). □

Proof of Theorem 3 Let $l_{\text{sos}}(\mathbf{Y}, \theta)$ be the SOS log-likelihood based on the sample $(Y_1, \ldots, Y_{a+1}), (Y_2, \ldots, Y_{a+2}), \ldots, (Y_{n-a}, \ldots, Y_n)$ and let θ_0 be the true value of θ. By the mean-value-theorem, we have

$$0 = \left.\frac{\partial l_{\text{sos}}(\mathbf{Y}, \theta)}{\partial \theta}\right|_{\theta = \hat{\theta}_n} = \left.\frac{\partial l_{\text{sos}}(\mathbf{Y}, \theta)}{\partial \theta}\right|_{\theta = \theta_0} + \left[\left.\frac{\partial^2 l_{\text{sos}}(\mathbf{Y}, \theta)}{\partial \theta \partial \theta'}\right|_{\theta = \hat{\theta}_n^*} \right] \cdot (\hat{\theta}_n - \theta_0)$$

where $\hat{\theta}_n^*$ is a point on the segment connecting θ_0 and $\hat{\theta}_n$. Then we obtain

$$\left(\hat{\theta}_n - \theta_0\right) = \left[- \left.\frac{\partial^2 l_{\text{sos}}(\mathbf{Y}, \theta)}{\partial \theta \partial \theta'}\right|_{\theta = \hat{\theta}_n^*} \right]^{-1} \left.\frac{\partial l_{\text{sos}}(\mathbf{Y}, \theta)}{\partial \theta}\right|_{\theta = \theta_0}.$$

Consider the convergence of $-\frac{1}{n-a}\frac{\partial^2 l_{\mathrm{sos}}(\mathbf{Y},\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}_n^*}$. Let

$$W_t = -\frac{\partial^2 q(\mathbf{Y}_t,\theta))}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}$$

for $t = 1,\ldots,n-a$. Then we get

$$-\frac{1}{n-a}\frac{\partial^2 l_{\mathrm{sos}}(\mathbf{Y},\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0} = -\frac{1}{n-a}\sum_{t=1}^{n-a} W_t.$$

First, we notice that

$$\begin{aligned}&-\frac{1}{n-a}E\frac{\partial^2 l_{\mathrm{sos}}(\mathbf{Y},\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}\\ &\quad = -EW_1 = -E\frac{\partial^2 q(\mathbf{Y}_t,\theta)}{\partial\theta\partial\theta'} = E\frac{\partial q(\mathbf{Y}_t,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_t,\theta)}{\partial\theta'}\Big|_{\theta=\theta_0} = I(\theta_0)\end{aligned}$$

where $I(\theta_0)$ is the Fisher information matrix of θ_0 based on the joint pdf of $(Y_1,\ldots,$ $Y_{a+1})$. By the fact that each $\widehat{\theta}_n$ is consistent with θ_0, adopting to the proof of Theorem 2, we can show that

$$-\frac{1}{n-a}\frac{\partial^2 l_{\mathrm{sos}}(\mathbf{Y},\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}_n^*} \xrightarrow{a.s.} I(\theta_0).$$

Now, we discuss the limiting distribution of $\frac{1}{\sqrt{n-a}}\frac{\partial l_{\mathrm{sos}}(\mathbf{Y},\theta)}{\partial\theta}\Big|_{\theta=\theta_0}$ which can be expressed as

$$\frac{1}{\sqrt{n-a}}\sum_{t=1}^{n-a}\frac{\partial q(\mathbf{Y}_t,\theta)}{\partial\theta}\Big|_{\theta=\theta_0}.$$

By regularity conditions, the assumptions of Lemma 6 are satisfied. And by this lemma, we conclude that

$$\frac{1}{\sqrt{n-a}}\sum_{t=1}^{n-a}\frac{\partial q(\mathbf{Y}_t,\theta)}{\partial\theta}\Big|_{\theta=\theta_0} \xrightarrow{L} N(0, V(\theta_0)).$$

Here, the matrix $V(\theta_0)$ can be derived by using

$$\begin{aligned}&Var\left(\frac{1}{\sqrt{n-a}}\sum_{t=1}^{n-a} Z_t\right) = I(\theta_0) + \sum_{t=1}^{n-a-1}\left(1-\frac{t}{n-a}\right)E\left[Z_1 Z_{t+1}' + Z_{t+1}Z_1'\right]\\ &\to V(\theta_0) = I(\theta_0) + \sum_{t=1}^{\infty} E\left[\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_{t+1},\theta)}{\partial\theta'} + \frac{\partial q(\mathbf{Y}_{t+1},\theta)}{\partial\theta}\frac{\partial q(\mathbf{Y}_1,\theta)}{\partial\theta'}\right]\Bigg|_{\theta=\theta_0}\end{aligned}$$

Finally, we obtain

$$\sqrt{n-a}\left(\hat{\theta}-\theta_0\right) \xrightarrow{L} N\left(0, I^{-1}(\theta_0)V(\theta_0)I^{-1}(\theta_0)\right). \qquad \square$$

Proof of Corollaries 3 and 4 We only need to show that conditions 1 − 10 of Sect. 3 are satisfied. Ibragimov and Linnik (1971) show that a Gaussian autoregressive average ARMA(p, q) process $(p, q \in N)$ has $\alpha_m \to 0$, and that, as $m \to \infty, \alpha_m$ approaches 0 exponentially. Thus, for the Gaussian ARMA(p, q) process, the conditions for Lemma 6 is satisfied. That is, the conditions 5 and 9 are satisfied. From Proposition 1, we know that the Gaussian ARMA(p, q) process is strictly stationary. So, condition 4 is satisfied. As other conditions are easy to verify, the details of the proofs are omitted. □

Proof of Theorem 4 is similar to the proof of Theorem 1. □

Proof of Theorem 5 Let $l_{\text{msos}}(\mathbf{Y}, \theta)$ be the modified SOS log-likelihood based on the sample $\mathbf{Y}_{it}$ for $1 \le i \le a$, $1 \le t \le n-i$, where a is a finite constant that takes values in $\{1, 2, \ldots, n-1\}$. Let θ_0 be the true value of θ. The Mean value theorem yields

$$0 = \left.\frac{\partial l_{\text{msos}}(\mathbf{Y}, \theta)}{\partial \theta}\right|_{\hat{\theta}_n} = \left.\frac{\partial l_{\text{msos}}(\mathbf{Y}, \theta)}{\partial \theta}\right|_{\theta_0} + \left.\left[\frac{\partial^2 l_{\text{msos}}(\mathbf{Y}, \theta)}{\partial \theta \partial \theta'}\right]\right|_{\hat{\theta}_n^*} \cdot (\hat{\theta}_n - \theta_0)$$

where $\hat{\theta}_n^*$ is a point on the segment connecting θ_0 and $\hat{\theta}_n$. Then we obtain

$$(\hat{\theta}_n - \theta_0) = \left.\left[-\frac{\partial^2 l_{\text{msos}}(\mathbf{Y}, \theta)}{\partial \theta \partial \theta'}\right]^{-1}\right|_{\hat{\theta}_n^*} \left.\frac{\partial l_{\text{msos}}(\mathbf{Y}, \theta)}{\partial \theta}\right|_{\theta_0}.$$

We first consider the convergence of $-\frac{1}{n}\left.\frac{\partial^2 l_{\text{msos}}(\mathbf{Y}, \theta)}{\partial \theta \partial \theta'}\right|_{\hat{\theta}_n^*}$. Let

$$W_{it} = -\left.\frac{\partial^2 q_i(\mathbf{Y}_{it}, \theta)}{\partial \theta \partial \theta'}\right|_{\theta_0}$$

for $1 \le i \le a$, $1 \le t \le n-i$. Then we obtain

$$-\frac{1}{n}\left.\frac{\partial^2 l_{\text{msos}}(\mathbf{Y}, \theta)}{\partial \theta \partial \theta'}\right|_{\theta_0} = -\frac{1}{n}\sum_{i=1}^{a}\sum_{t=1}^{n-i} W_{it}.$$

First, we notice that

$$-\frac{1}{n}E\frac{\partial^2 l_{\mathrm{msos}}(\mathbf{Y},\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta_0} = -\sum_{i=1}^{a}\left(1-\frac{i}{n}\right)EW_{i1}$$
$$\to -\sum_{i=1}^{a}E\frac{\partial^2 q_i(\mathbf{Y}_{it},\theta)}{\partial\theta\partial\theta'} = \sum_{i=1}^{a}E\frac{\partial q_i(\mathbf{Y}_{it},\theta)}{\partial\theta}\frac{\partial q_i(\mathbf{Y}_{it},\theta)}{\partial\theta'}\bigg|_{\theta_0}$$

Let $I(\theta_0) = \sum_{i=1}^{a} I_i(\theta_0) = \sum_{i=1}^{a} E\frac{\partial q_i(\mathbf{Y}_{it},\theta))}{\partial\theta}\frac{\partial q_i(\mathbf{Y}_{it},\theta)}{\partial\theta'}\Big|_{\theta_0}$, where $I_i(\theta_0)$ is the Fisher information matrix of θ_0 based on the joint pdf of (X_1, X_{1+i}). Using the fact that each $\widehat{\theta}_n$ is consistent with θ_0. In a similar manner to the proof of Theorem 4, we can show that

$$-\frac{1}{n}\frac{\partial^2 l_{\mathrm{msos}}(\mathbf{Y},\theta)}{\partial\theta\partial\theta'}\bigg|_{\hat{\theta}_n^*} \xrightarrow{a.s.} I(\theta_0).$$

Next, we discuss the asymptotic properties of $\frac{1}{\sqrt{n}}\frac{\partial l_{\mathrm{msos}}(\mathbf{Y},\theta)}{\partial\theta}\Big|_{\theta_0}$ which can be expressed as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{a}\sum_{t=1}^{n-i}\frac{\partial q_i(\mathbf{Y}_{it},\theta)}{\partial\theta}\bigg|_{\theta_0}.$$

Now, let $U_{i,t} = \frac{\partial q_i(\mathbf{Y}_{it},\theta)}{\partial\theta}|_{\theta_0}$ for $1 \le i \le a$, $1 \le t \le n-i$. Then we have

$$\frac{1}{\sqrt{n}}\frac{\partial l_{\mathrm{msos}}(\mathbf{Y},\theta)}{\partial\theta}\bigg|_{\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{a}\sum_{t=1}^{n-i}U_{i,t}$$

where $EU_{i,t} = 0$ for $1 \le i \le a$, $1 \le t \le n-i$. Let

$$\{Z_t\}_{t=1}^{s} = \{U_{1,1}, U_{1,2}, \ldots, U_{1,a}, U_{1,a+1}, U_{2,3}, U_{2,4}, \ldots, U_{2,a+1}, U_{2,a+2}, \ldots, U_{n-1,n}\}$$

where $s = an - \frac{a(a+1)}{2}$. From Lemma 3, $\{Z_t\}$ is α-mixing with mixing coefficients of the same order.

Furthermore, we see that

$$Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{s}Z_t\right) = \sum_{i=1}^{a}\left(1-\frac{i}{n}\right)I_i(\theta_0) + \sum_{j=1}^{s}\sum_{k=j+1}^{s}E\left[Z_jZ_k' + Z_kZ_j'\right]$$
$$\to V(\theta_0) = I(\theta_0) + \sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}E\left[Z_jZ_k' + Z_kZ_j'\right]$$

Applying Lemma 6, it follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{s} Z_t \xrightarrow{L} N\left(0, V(\theta_0)\right).$$

Hence, we obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} N\left(0, G\right).$$

where $G = I^{-1}(\theta_0)V(\theta_0)I^{-1}(\theta_0)$. □

References

Bai, Z. D., Zheng, S. R., Zhang, B. X., Hu, G. R. (2008). *Statistical analysis for rounded data*. Department of Statistics, Northeast Normal University. *Journal of Statistical Planning and Inference* (to appear).

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1994). *Time series analysis: forecasting and control* (3rd ed.). New York: Prentice-Hall.

Brock, W. A., Dechert, W., Scheinkman, J. (1987). *A test for independence based on the correlation dimension*. Working paper, University of Winconsin at Madison, University of Houston, and University of Chicago.

Brockwell, P. J., Davis, R. A. (1991). *Time series: theory and methods* (2nd ed.). New York: Springer.

Chen, X. (1997). *Introduction to mathematical statistics* (1st ed., 2nd Print, in Chinese). Beijing: Science Press.

Coppejans, M. (2003). Effective nonparametric estimation in the case of severely discretized data. *Journal of Econometrics*, *117*, 331–367.

Dempster, A. P., Rubin, D. B. (1983). Rounding error in regression: The appropriateness of sheppard's corrections. *Journal of the Royal Statistical Society, Series B*, *45*, 51–59.

Durrans, S. R., Pitt, R. (2004). Maximum likelihood estimators for coarsely resolved precipitation data. *Journal of Hydrologic Engineering*, *9*, 13–27.

Durrans, S. R., Burian, S. J., Pitt, R. (2004). Enhancement of precipitation data for small storm hydrologic prediction. *Journal of Hydrology*, *299*, 180–185.

Grimshaw, S. D., McDonald, J., McQueen, G. R., Thorley, S. (2005). Estimating hazard functions for discrete lifetimes. *Communications in Statistics, Part B-Simulation and Computation*, *34*, 451–463.

Hall, P. (1982). The influence of rounding errors on some nonparametric estimators of a density and its derivatives. *SIAM Journal on Applied Mathematics*, *42*, 390–399.

Hall, P., Heyde, C. C. (1980). *Martingale limit theory and its application.* New York: Academic Press.

Heitjan, D. F. (1989). Inference from grouped continuous data: a review. *Statistical Science*, *4*, 164–183.

Heitjan, D. F., Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics*, *19*, 2244–2253.

Ibragimov, I. A., Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Groningen: Wolters-Noordhoff.

Johnson, D. S., Barry, R. P., Bowyer, R. T. (2004). Estimating timing of life-history events with coarse data. *Journal of Mammalogy*, *85*, 932–939.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, *22*, 389–395.

Kozicki, S., Hoffman, B. (2004). Rounding error: a distorting influence on index data. *Journal of Money, Credit, and Banking*, *36*(3), 319–338.

Krämer, W., Runde, R. (1997). Chaos and the compass rose. *Economics Letters*, *54*, 113–118.

Lee, C. S., Vardeman, S. B. (2001). Interval estimation of a normal process mean from rounded data. *Journal of Quality Technology*, *33*, 335–348.

Lee, C. S., Vardeman, S. B. (2002). Interval estimation of a normal process standard deviation from rounded data. *Communications in Statistics, Part B-Simulation and Computation*, *31*, 13–34.

Lee, C. S., Vardeman, S. B. (2003). Confidence interval based on rounded data from the balanced one-way normal random effects model. *Communications in Statistics, Part B-Simulation and Computation*, *32*, 835–856.

Lindley, D. V. (1950). Grouping corrections and maximum likelihood equations. *Proceedings of the Cambridge Philosophical Society*, *46*, 106–110.
Machak, J. A., Rose, E. L. (1984). Uniform additive outliers in ARMA models. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section* (pp. 577–580).
Machak, J. A., Rose, E. L. (1985). Robust identification and estimation of time series. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section*.
Mcleish, D. L. (1975). A maximal inequality and dependent strong laws. *Annals of Probability*, *3*, 829–839.
Rose, E. L. (1993). Some effects of rounding errors on ARMA (1,1) models. *Communications in Statistics, Part B-Simulation and Computation*, *22*, 155–174.
Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, *42*, 43–47.
Sheppard, W. F. (1898). On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a ccale. *Proceedings of the London Mathematical Society*, *29*, 353–380.
Stam, A., Cogger, K. O. (1993). Rounding errors in autoregressive processes. *International Journal of Forecasting*, *9*, 487–508.
Stout, W. F. (1974). *Almost sure convergence*. New York: Academic Press.
Tallis, G. M. (1967). Approximate maximum likelihood estimates from grouped data. *Technometrics*, *9*, 599–606.
Tricker, T. (1984). Effects of rounding data sampled from the exponential distribution. *Journal of Applied Statistics*, *11*, 54–87.
Tricker, A. R. (1990a). The effect of rounding on the power level of certain normal test statistics. *Journal of Applied Statistics*, *17*, 219–228.
Tricker, A. R. (1990b). The effect of rounding on the significance level of certain normal test statistics. *Journal of Applied Statistics*, *17*, 31–38.
Tricker, A. R. (1992). Estimation of parameters for rounded data from non-normal distributions. *Journal of Applied Statistics*, *19*, 465–471.
Tricker, A., Coates, E., Okell, E. (1998). The effect on the R chart of precision of measurement. *Journal of Quality Technology*, *30*, 232–239.
Vardeman, S. B. (2005). Sheppard's correction for variances and the quantization noise model. *IEEE Transactions on Instrumentation and Measurement*, *54*, 2117–2119.
Vardeman, S. B., Lee, C. S. (2005). Likelihood-based statistical estimation from quantization data. *IEEE Transactions on Instrumentation and Measurement*, *54*, 409–414.
White, H. (2001). *Asymptotic theory for econometricians* (2nd ed.). New York: Academic Press.
White, H., Domowitz, I. (1984). Nonlinear regression with dependent observations. *Econometrica*, *52*, 143–162.
Wright, D. E., Bray, I. (2003). A mixture model for rounded data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *52*, 3–13.