Censored Time Series Analysis with Autoregressive Moving Average Models
Author(s): Jung Wook Park, Marc G. Genton and Sujit K. Ghosh
Source: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 35, No. 1 (Mar., 2007), pp. 151-168
Published by: Statistical Society of Canada
Stable URL: https://www.jstor.org/stable/20445244

The Canadian Journal of Statistics
Vol. 35, No. 1, 2007, Pages 151–168
La revue canadienne de statistique

Censored time series analysis with autoregressive moving average models

Jung Wook PARK, Marc G. GENTON and Sujit K. GHOSH

Key words and phrases: Censored time series; Fisher information; Gibbs sampler; imputation; truncated multivariate normal distribution.

MSC 2000: Primary 62M10; secondary 62C10.

Abstract: The authors consider time series observations with data irregularities such as censoring due to a detection limit. Practitioners commonly disregard censored data cases which often result in biased estimates. The authors present an attractive remedy for handling autocorrelated censored data based on a class of autoregressive and moving average (ARMA) models. In particular, they introduce an imputation method well suited for fitting ARMA models in the presence of censored data. They demonstrate the effectiveness of their technique in terms of bias, efficiency, and information loss. They also describe its adaptation to a specific context of meteorological time series data on cloud ceiling height, which are measured subject to the detection limit of the recording device.

Analyse de séries chronologiques censurées au moyen de modèles autorégressifs à moyenne mobile

Résumé : Les auteurs s'intéressent à des séries chronologiques dont certaines observations sont censurées lorsqu'elles dépassent un certain seuil. Règle générale, les praticiens ne tiennent pas compte des cas de censure, ce qui induit souvent un biais d'estimation. Les auteurs présentent une façon élégante de traiter des données autocorrélées censurées au moyen d'une classe de modèles autorégressifs à moyenne mobile (ARMA). En particulier, ils proposent une méthode d'imputation bien adaptée à l'ajustement de modèles ARMA en présence de censure. Ils démontrent la bonne performance de leur technique aux plans du biais, de l'efficacité et de la perte d'information. Ils montrent aussi comment elle peut être adaptée au contexte spécifique de données météorologiques temporelles sur la hauteur du plafond nuageux, dont l'évaluation est limitée par la capacité de l'instrument de mesure.

1. INTRODUCTION

Observations collected over time or space are often autocorrelated rather than independent. Time series data analysis deals with temporally collected observations by modelling their autocorrelations. Autoregressive moving average (ARMA) models for time series data developed by Box & Jenkins (1970) have been widely used as a basic approach. The ARMA(p, q) model for a time series, $\{Y_t,\ t = 0, \pm 1, \pm 2, \ldots\}$ is defined as

$$\widetilde{Y}_t = \rho_1 \widetilde{Y}_{t-1} + \cdots + \rho_p \widetilde{Y}_{t-p} + \varepsilon_t - \psi_1 \varepsilon_{t-1} - \cdots - \psi_q \varepsilon_{t-q}, \tag{1}$$

where $\widetilde{Y}_t = Y_t - \mu$, μ is the mean parameter, $\rho_1, \ldots, \rho_p$ are the autoregressive parameters, and $\psi_1, \ldots, \psi_q$ are the moving average parameters. The error process ε_t is assumed to be white noise with mean 0 and variance σ^2. This model reduces to an AR(p) model when $q = 0$, and to an MA(q) model when $p = 0$.

Time series measurements are often observed with data irregularities, e.g., observations with a detection limit. For instance, a monitoring device usually has a detection limit and it records the limit value when the true value exceeds/precedes the detection limit. This is often called censoring. Censoring occurs in various situations in physical sciences, business, and economics. The measurement of rainfall is often limited due to the size of the gauge and signals may intentionally be limited for convenience in storage or processing (for further examples, see Robinson 1980). Our motivation is the analysis of cloud ceiling height defined as the distance from the ground to the bottom of a cloud measured in hundreds of feet. Cloud ceiling height is one of the major factors contributing to weather-related accidents and one of the major causes of flight delays. Moreover, adverse ceiling conditions could seriously affect the general aviation pilot since they could lead to the risk of disorientation, loss of control, and flight into obstructed terrain. Thus an accurate determination of the cloud ceiling height is important. Our data on ceiling height, collected by the National Center for Atmospheric Research (NCAR), was observed hourly in San Francisco and recorded during the month of March 1989, consisting of $n = 716$ observations, see Figure 1 for a plot of the log-transformed data whose original scale is hundred (100) feet. The observations have a detection limit at about 4.79 (12,000 feet in original scale) and hence we consider the data as a right-censored time series. Due to this upper detection limit of the recording device, many observations are censored: the censoring rate is 41.62%. This will potentially lead to biased estimates when we implement classical analysis tools that ignore censoring.

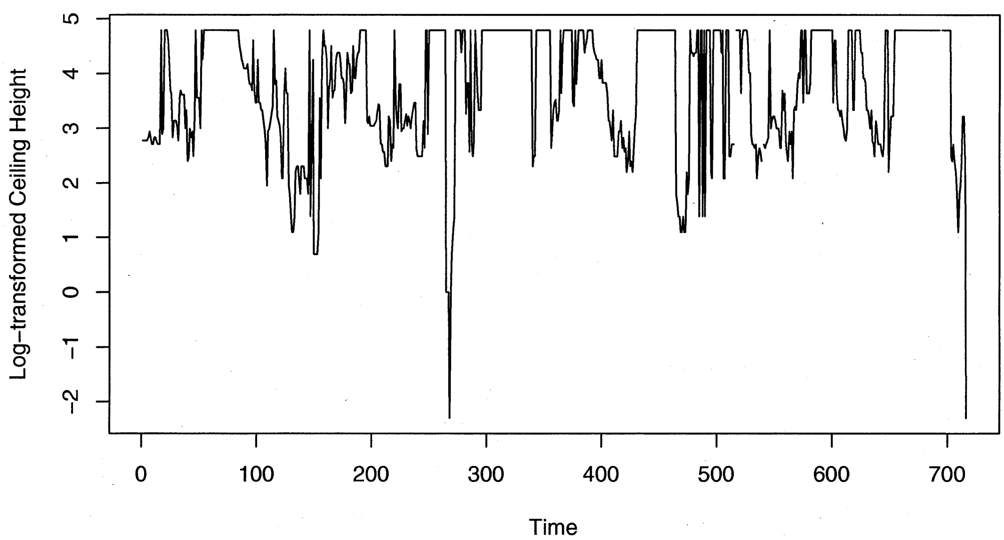

FIGURE 1: Censored time series of log-transformed hourly cloud ceiling height in San Francisco during March 1989.

There may be two naive approaches to handling censoring. One is to discard the censored observations and the other is to treat the censored values as observed. However, both approaches produce biased and inefficient estimates and lead to unreliable results; see the results based on simulated data in Section 5. Thus it is necessary to use an alternative approach to improve the performance of the parameter estimation for time series containing censored observations. One of the simplest methods to account for censoring is to substitute the censored values with a constant exceeding/preceding the detection limit. This has been done under the belief that the censoring rate might not be large and the effect on the inference might be insignificant (Helsel 1990). However, such beliefs are not true in general, and the results will depend highly on the strength of such model assumptions. Thus the key issue in estimating the parameters of time series models based on censored data is to obtain estimates that are at least (asymptotically) unbiased and more efficient than some of the ad hoc methods described above.

Robinson (1980) suggested imputing the censored part with its conditional expectation, given the completely observed part. Since the conditional expectation has the form of multiple incomplete integrals, he would group the data vector so that each subgroup includes one censored observation, and thus requires a single integral. However, the method may not be feasible for many consecutive censored observations. Zeger & Brookmeyer (1986) suggested a full likelihood estimation and approximate method for an autoregressive time series model. However, the authors have pointed out that the method may not be feasible when the censoring rate is very high (Zeger & Brookmeyer 1986, p. 728). To overcome this limitation the authors have suggested the use of pseudo-likelihood estimation. Hopke, Liu & Rubin (2001) used multiple imputation based on a Bayesian approach. However, little explanation was provided about the theoretical properties of the estimators, such as unbiasedness and efficiency.

We present an attractive remedy for handling censored data based on a class of Gaussian ARMA models. In particular, we introduce an imputation method well suited for fitting ARMA models. In this method, the observed time series data are regarded as a realization from a multivariate normal distribution. The censored values are then imputed using the conditional multivariate normal distribution given the observed part. Note that our method does not require the specification of prior distributions and is flexible in obtaining estimates using standard techniques once the censored observations have been imputed.

The paper is organized as follows. In Section 2, we define various forms of censored time series, in particular censored ARMA models and their multivariate normal representation when the white noise is Gaussian. Then we introduce the imputation method for estimating the parameters of censored Gaussian ARMA models. In Section 3, we describe an algorithm for generating random samples from a truncated multivariate distribution and its adaptation to implement the imputation method. In Section 4, we discuss how censoring affects the efficiency of parameter estimation in terms of the observed Fisher information matrix. A simulation study was conducted based on an AR(1) process to study the performance of the imputation method and the results are presented in Section 5. Our imputation method is of course not limited to AR processes and in Section 6 we illustrate its use on the meteorological time series of cloud ceiling height, measured with a detection limit, using an ARMA(1,1) model. Finally, in Section 7, we discuss some general issues related to the imputation method.

2. MODELS FOR CENSORED TIME SERIES

2.1. Censored ARMA models.

Consider a time series $\{Y_t,\ t = 0, \pm 1, \pm 2, \ldots\}$. In many practical situations we may not be able to observe Y_t directly. Instead, we may observe Y_t only when Y_t precedes/exceeds a constant value c. Let X_t be the value we observe instead of Y_t due to censoring. Then there are mainly three types of censoring, given by:

$$X_t = \begin{cases} \min(Y_t, c), & \text{in case of left censoring,} \\ \max(Y_t, c), & \text{in case of right censoring,} \\ \operatorname{median}(Y_t, c_1, c_2), & \text{in case of interval censoring,} \end{cases} \tag{2}$$

where the constants $c, c_1, c_2 \in \mathbf{R}$ are the cutoff values, that is, the detection limits. If the detection limit is a random variable, say C_t, rather than a constant, then it is random censoring. Although our proposed method can be applied to randomly censored data, we restrict our discussions to fixed censoring.

For an ARMA(p,q) model defined in (1), we call the process $\{X_t\}$ defined by (2) a censored autoregressive and moving average model, and denote it by CENARMA(p,q). The models corresponding to AR(p) and MA(q) are denoted by CENAR(p) and CENMA(q), respectively. It is straightforward to see that for the types of censoring given by (2), the process $\{Y_t\}$ and $\{X_t\}$ have different distributions. Thus we cannot use X_t directly to make inference about the parameters of the model described by Y_t.

2.2. Conditional distribution of the censored part of an ARMA.

Notice that if $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ is a realization from a stationary stochastic process described by an ARMA(*p*,*q*) model with Gaussian white noise, we can write

$$\mathbf{Y} \sim \mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3}$$

where N_n represents an n-dimensional multivariate normal distribution with mean $\boldsymbol{\mu} = \mu \mathbf{1}_n$ and stationary $n \times n$ covariance matrix $\boldsymbol{\Sigma}$ whose elements are given by $\{\boldsymbol{\Sigma}\}_{ij} = \gamma(|i-j|) = \gamma(h)$, where $\gamma(h)$ is the autocovariance function at lag h. By using a permutation matrix $\mathbf{P}$, we can rearrange the order of the data so that the data vector can be partitioned into an observed part $\mathbf{Y}_O$ and a censored part $\mathbf{Y}_C$ given by

$$\mathbf{PY} = \begin{pmatrix} \mathbf{P}_O \\ \mathbf{P}_C \end{pmatrix} \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_O \\ \mathbf{Y}_C \end{pmatrix}. \tag{4}$$

Then $\mathbf{PY}$ also follows a multivariate normal distribution given by

$$\mathbf{PY} \sim \mathsf{N}_n \left(\mu \mathbf{P} \mathbf{1}_n, \begin{bmatrix} \mathbf{P}_O \boldsymbol{\Sigma} \mathbf{P}_O^T & \mathbf{P}_O \boldsymbol{\Sigma} \mathbf{P}_C^T \\ \mathbf{P}_C \boldsymbol{\Sigma} \mathbf{P}_O^T & \mathbf{P}_C \boldsymbol{\Sigma} \mathbf{P}_C^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{OO} & \boldsymbol{\Sigma}_{OC} \\ \boldsymbol{\Sigma}_{CO} & \boldsymbol{\Sigma}_{CC} \end{bmatrix} \right). \tag{5}$$

It follows that the conditional distribution of $\mathbf{Y}_C$ given $\mathbf{Y}_O$ is also a multivariate normal distribution (see, for example, Anderson 1984) whose mean and covariance matrix are functions of $\mathbf{Y}_O$ and the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Applying the permutation matrix to the observed data $\mathbf{X} = (X_1, \ldots, X_n)^T$, we have

$$\mathbf{PX} = \begin{pmatrix} \mathbf{P}_O \\ \mathbf{P}_C \end{pmatrix} \mathbf{X} = \begin{pmatrix} \mathbf{X}_O \\ \mathbf{X}_C \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \mathbf{Y}_O \\ \mathbf{X}_C \end{pmatrix},$$

where $\stackrel{d}{=}$ represents equality in distribution. The purpose of deriving the above conditional multivariate normal distribution is to find an appropriate substitute for $\mathbf{X}_C$. The basic idea is to replace $\mathbf{X}_C$ by sampling values from the conditional distribution of $\mathbf{Y}_C$ given $\mathbf{X}_O$ and $\mathbf{X}_C$. This is a truncated multivariate normal distribution:

$$(\mathbf{Y}_C \,|\, \mathbf{X}_O, \mathbf{X}_C \in R_C) \sim \mathsf{TN}_{n_C}(\boldsymbol{\nu}, \boldsymbol{\Delta}, R_C),$$

where n_C is the number of censored observations, TN_{n_C} is the truncated multivariate normal distribution of dimension n_C, and R_C is the censoring region. The parameters $\boldsymbol{\nu}$ and $\boldsymbol{\Delta}$ are the conditional mean and covariance of a nontruncated version of a conditional multivariate normal distribution.

We illustrate our method with an AR(1) process in order to keep the notations simple, but the extension to ARMA(*p*,*q*) models is straightforward. Suppose we have an AR(1) time series process with mean μ, variance σ^2, and autocorrelation ρ. We can consider the data as a random vector from a multivariate normal distribution as in (3) where $\boldsymbol{\mu} = \mu \mathbf{1}_n$ and $\{\boldsymbol{\Sigma}\}_{ij} = \frac{\sigma^2}{1-\rho^2} \rho^{|i-j|}$, $i, j = 1, \ldots, n$. For example, the conditional distribution of X_k given other observations is a univariate normal distribution whose mean is a function of X_{k-1} and X_{k+1} since the inverse of the covariance matrix $\boldsymbol{\Sigma}$ is tridiagonal. If X_k is censored, then the conditional distribution is a truncated univariate normal distribution.

3. IMPUTATION METHOD

3.1. Generation from a truncated multivariate distribution.

Let $g(\mathbf{x})$ be the density of a random vector defined on the n-dimensional Euclidean space $\mathbf{R}^n$, and let

$$f(\mathbf{x}) = \frac{g(\mathbf{x})I_A(\mathbf{x})}{\int_A g(\mathbf{z})\,d\mathbf{z}}, \quad A \subseteq \mathbf{R}^n$$

denote the density of a truncated distribution, where $I_A(\mathbf{x}) = 1$ if $\mathbf{x} \in A$ and $I_A(\mathbf{x}) = 0$ otherwise. Assume that $A = \bigotimes_{j=1}^n A_j$, where $\bigotimes$ denotes the Cartesian product and $A_j \subseteq \mathbf{R}$, $j = 1, \ldots, n$, and it is easy to sample from full conditionals $g(x_i \mid \mathbf{x}_{-i})I_{A_i}(x_i)$, where the vector $\mathbf{x}_{-i} \doteq (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)^T$. Note that $f(x_i \mid \mathbf{x}_{-i}) \propto g(x_i \mid \mathbf{x}_{-i})I_{A_i}(x_i)$. We describe a Gibbs sampling (Gelfand & Smith 1990) to generate random samples from $f(\mathbf{x})$:

(0) Set $\mathbf{x}^{(0)} = (x_1^{(0)}, \ldots, x_n^{(0)})^T$.

(1) For, $k = 1, 2, \ldots$, generate

$$\begin{aligned}
&x_1^{(k)} \sim g(x_1 \mid \mathbf{x}_{-1}^{(k-1)})I_{A_1}(x_1)\\
&x_2^{(k)} \sim g(x_2 \mid x_1^{(k)}, x_3^{(k-1)}, \ldots, x_n^{(k-1)})I_{A_2}(x_2)\\
&\vdots\\
&x_n^{(k)} \sim g(x_n \mid \mathbf{x}_{-n}^{(k)})I_{A_n}(x_n)
\end{aligned}$$

(2) Repeat until the iterates converge to the stationary distribution with density $f(\mathbf{x})$.

In particular, when $g(\mathbf{x})$ is the density of $\mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $A_j = [c_j, \infty]$, we can sample from $g(x_j \mid \mathbf{x}_{-j})$ by generating $U_j \sim \mathsf{U}(0,1)$ and setting

$$x_j = \nu_j + \tau_j \Phi^{-1}\left[U_j\left\{1 - \Phi\left(\frac{c_j - \nu_j}{\tau_j}\right)\right\} + \Phi\left(\frac{c_j - \nu_j}{\tau_j}\right)\right],$$

where $\nu_j = \mu_j - \boldsymbol{\sigma}_j^T \boldsymbol{\Sigma}_{-j}^{-1}(\mathbf{x}_{-j} - \boldsymbol{\mu}_{-j})$, $\tau_j^2 = \sigma_{jj} - \boldsymbol{\sigma}_j^T \boldsymbol{\Sigma}_{-j}^{-1}\boldsymbol{\sigma}_j$, $\boldsymbol{\mu}_{-j} = (\mu_1, \ldots, \mu_{j-1}, \mu_{j+1}, \ldots, \mu_n)^T$, $\boldsymbol{\sigma}_j = (\sigma_{1j}, \sigma_{2j}, \ldots, \sigma_{j-1,j}, \sigma_{j+1,j}, \ldots, \sigma_{nj})^T$, and $\boldsymbol{\Sigma}_{-j}$ is the matrix $\boldsymbol{\Sigma}$ with jth row and column deleted (see Robert 1995).

3.2. Imputation algorithm to fit a CENAR(1) model.

The main idea of the algorithm is to update the parameter estimates by imputing the censored values with the conditional sample. The method is mainly divided into two parts:

(i) Data augmentation; and

(ii) Parameter estimation.

For data augmentation, we need to use a random sample generated from a truncated multivariate normal distribution. For parameter estimation we can use any conventional method.

We describe the imputation algorithm to fit the CENAR(1) model as an illustration. The extension to CENARMA(p, q) models is straightforward.

Step 1. Compute the permutation matrix $\mathbf{P}_C$ and $\mathbf{P}_O$ so that we can construct $\mathbf{X}_C$ and $\mathbf{X}_O$ similar to (4).

Step 2. Obtain the initial estimates $\hat{\mu}^{(0)}$, $\hat{\rho}^{(0)}$, and $\{\hat{\sigma}^{(0)}\}^2$ (see Step 6, for example). Then construct the following mean vector and covariance matrix $\widehat{\boldsymbol{\Sigma}}^{(0)}$,

$$\begin{aligned}
\hat{\boldsymbol{\mu}}^{(0)} &= \hat{\mu}^{(0)}\mathbf{1}_n,\\
\{\widehat{\boldsymbol{\Sigma}}^{(0)}\}_{ij} &= \frac{\{\hat{\sigma}^{(0)}\}^2}{1 - \{\hat{\rho}^{(0)}\}^2}\{\hat{\rho}^{(0)}\}^{|i-j|}, \quad i, j = 1, \ldots, n.
\end{aligned}$$

Step 3. Calculate the conditional mean $\hat{\boldsymbol{\nu}}^{(0)}$ and variance $\hat{\boldsymbol{\Delta}}^{(0)}$ of the censored part using the following relationships:

$$\begin{aligned}\hat{\boldsymbol{\nu}}^{(0)} &= \hat{\boldsymbol{\mu}}_C^{(0)} + \widehat{\boldsymbol{\Sigma}}_{CO}^{(0)}(\widehat{\boldsymbol{\Sigma}}_{OO}^{(0)})^{-1}(\mathbf{x}_O - \hat{\boldsymbol{\mu}}_O^{(0)}),\\ \hat{\boldsymbol{\Delta}}^{(0)} &= \widehat{\boldsymbol{\Sigma}}_{CC}^{(0)} - \widehat{\boldsymbol{\Sigma}}_{CO}^{(0)}(\widehat{\boldsymbol{\Sigma}}_{OO}^{(0)})^{-1}\widehat{\boldsymbol{\Sigma}}_{OC}^{(0)},\end{aligned}$$

where the covariances are defined in (5).

Step 4. Generate a random sample $\mathbf{x}_C^{(1)}$ from $\mathsf{TN}_{n_C}(\boldsymbol{\nu}^{(0)}, \boldsymbol{\Delta}^{(0)}, R_C)$, where $R_C = (c,\infty)^{n_C} = (c,\infty) \times (c,\infty) \times \cdots \times (c,\infty)$, see Section 3.1.

Step 5. Construct the augmented data from the observed part and imputed sample for the censored part

$$\mathbf{X}^{(1)} = \mathbf{P}^{-1}\begin{pmatrix}\mathbf{X}_O\\ \mathbf{X}_C^{(1)}\end{pmatrix},$$

where $\mathbf{P}$ is defined in (4).

Step 6. Re-estimate the parameters μ, ρ, and σ based on $\mathbf{X}^{(1)}$ and update the parameters $\boldsymbol{\Sigma}$, $\boldsymbol{\nu}$, and $\boldsymbol{\Delta}$ (see, for example, Fuller 1996). If a least squares approach is used we use the following estimates,

$$\begin{aligned}\hat{\mu}^{(1)} &= n^{-1}\sum_{t=1}^{n} X_t^{(1)},\\ \hat{\rho}^{(1)} &= \left\{\sum_{t=2}^{n}(X_{t-1}^{(1)} - \overline{X}_{-n}^{(1)})^2\right\}^{-1}\left\{\sum_{t=2}^{n}(X_t^{(1)} - \overline{X}_{-1}^{(1)})(X_{t-1}^{(1)} - \overline{X}_{-n}^{(1)})\right\},\\ \{\hat{\sigma}^{(1)}\}^2 &= (n-3)^{-1}\sum_{t=2}^{n}\left[X_t^{(1)} - \hat{\mu}^{(1)} - \hat{\rho}^{(1)}(X_{t-1}^{(1)} - \hat{\mu}^{(1)})\right]^2,\end{aligned}$$

where $\overline{X}_{-n} = (n-1)^{-1}\sum_{t=1}^{n-1} X_t$ and $\overline{X}_{-1} = (n-1)^{-1}\sum_{t=2}^{n} X_t$.

Step 7. Repeat **Step 3** $\sim$ **Step 6** until the parameter estimates converge. We use the following convergence rule,

$$\frac{(\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)})^T(\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)})}{\{\hat{\boldsymbol{\theta}}^{(k)}\}^T\hat{\boldsymbol{\theta}}^{(k)}} < \text{tol},$$

where $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\rho})^T$. In our simulations, tol=0.001 was used.

The imputation method works by maximizing an approximate full likelihood, which is obtained iteratively based on simulations of the censored part of the data. One of the benefits of the imputation method is that we are not limited to a specific procedure for the estimation step. In other words, once we augment the data, any suitable method (such as Yule–Walker, least squares, or maximum likelihood methods) can be used for the parameter estimation.

4. EFFECT OF THE CENSORING RATE

4.1. Fisher information matrix for correlated observations.

This section introduces the calculation of the Fisher information matrix of a stationary time series model to give insight into the effect of censoring. The joint probability density of a stationary Gaussian time series of size n can always be represented via a multivariate normal distribution. The log-likelihood of a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}\,|\,\mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by (3) and $\mathbf{y}$ is a realization of a time series $\{Y_t,\ t = 1, \dots, n\}$. Let $\boldsymbol{\theta}$ denote the vector consisting of all parameters in the model. Then the expected Fisher information matrix can be obtained by

$$I_n(\boldsymbol{\theta}) = \mathrm{E}\left[-\frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right],$$

where the expectation is taken with respect to the $\mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if we had complete data without censoring. However, the computation of the expectation becomes very complicated in the presence of censoring, where X_t are collected instead of Y_t, because multiple incomplete integrals are involved. As the expectation is difficult to calculate, we use the observed Fisher information,

$$\hat{I}_n(\boldsymbol{\theta}) = -\frac{\partial^2 L(\boldsymbol{\theta}\,|\,\mathbf{x})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}, \tag{6}$$

which converges in probability to the expected Fisher information (Hogg & Craig 1994, § 8.3). For example, for an AR(1) process the conditional distribution of Y_t given Y_{t-1} is independent of Y_s, $s < t - 1$. Thus, in a CENAR(1) model, we need to consider four scenarios to compute (6), see the Appendix.

4.2. Fisher information matrix for Gaussian CENAR(1).

We derive the observed Fisher information matrix based on a realization from a Gaussian CENAR(1) process and examine how the information changes as a function of the censoring rate. The details of the calculation are described in the Appendix. Let $I_{n,C}(\mu, \sigma, \rho)$ be the Fisher information matrix, where C stands for "censored". Then,

$$I_{n,C}(\mu, \sigma, \rho) \;=\; \begin{bmatrix} D_{\mu^2} & D_{\mu,\sigma} & D_{\mu,\rho} \\ D_{\mu,\sigma} & D_{\sigma^2} & D_{\sigma,\rho} \\ D_{\mu,\rho} & D_{\sigma,\rho} & D_{\rho^2} \end{bmatrix}, \tag{7}$$

where

$$D_{\mu,\sigma} = \frac{\partial^2 \log L(\mu, \sigma, \rho\,|\,\mathbf{x})}{\partial\mu\partial\sigma},$$

and similarly for the other entries. It is not analytically feasible to calculate the expected Fisher information matrix since the expectations of the second derivatives with respect to the censoring part involve complicated incomplete integrals. However, it is relatively simple to calculate the observed Fisher information matrix since we can just plug-in sample values in the derivatives.

For simulation purposes, we need to set up the cutoff point c which is obtained by solving $\Pr(Y_t > c) = \alpha$, where $\{Y_t\}$ is a Gaussian AR(1) process and α is the censoring probability. We call $\alpha \times 100\%$ the average censoring rate. For the Gaussian CENAR(1) model, the cutoff point c can be derived easily:

$$c \;=\; \mu + \sigma\frac{\Phi^{-1}(1-\alpha)}{\sqrt{1-\rho^2}}\sqrt{1-\rho^{2(n+1)}}, \tag{8}$$

where $\Phi(\,\cdot\,)$ is the standard normal distribution function and $\sqrt{1-\rho^{2(n+1)}}$ is a correction factor for a finite sample size. Another possible criterion for finding the cutoff point is to calculate c for every size of the time series and summarize it with an average or median:

$$\begin{aligned} c_a \;&=\; \mu + \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{1-\rho^2}}\left\{\frac{1}{n}\sum_{j=1}^{n}\sqrt{1-\rho^{2(j+1)}}\right\}, \quad \text{or} \\ c_m \;&=\; \mu + \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{1-\rho^2}} \times \text{Median}\left\{\sqrt{1-\rho^{2(j+1)}},\ j = 1, \dots, n\right\}. \end{aligned}$$

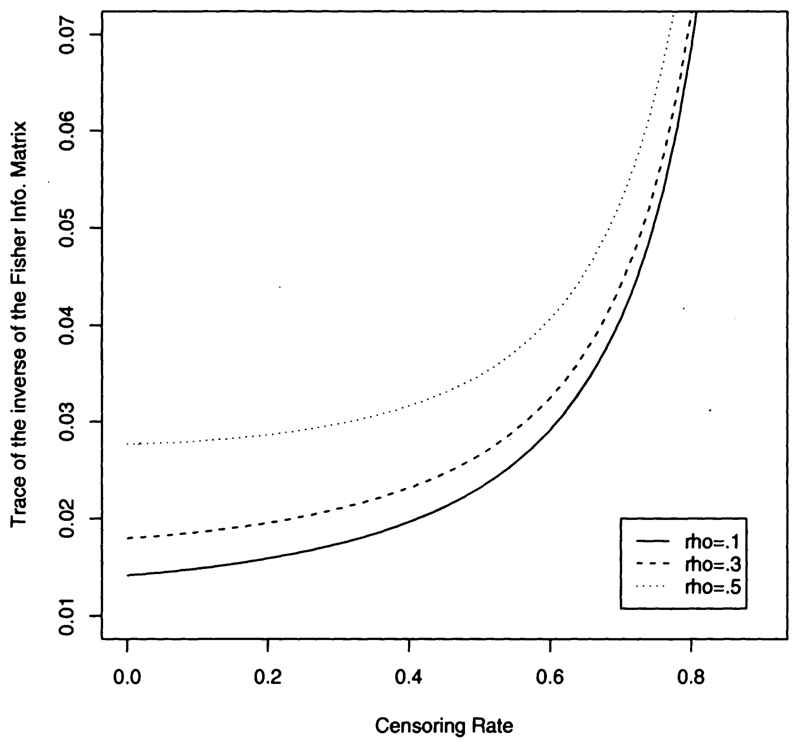

FIGURE 2: Plot of the median of the trace of the inverse of the observed Fisher information matrix as a function of the censoring rate ranging from 0% to 80% censoring. Samples of size $n = 200$ with $\mu = 0$, $\sigma = 1$, and $\rho = .1, .3, .5$ were used for the simulation with $N = 500$ replicates.

A simulation was conducted in order to illustrate the change of the Fisher information as a function of the censoring rate. A sample of size $n = 200$ and $N = 500$ simulation replicates were used to study the Fisher information. These parameters were set: $\mu = 0$, $\sigma = 1$, and $\rho = .1, .3, .5$. We changed the censoring rate from 0% to 100% to examine the behaviour of information as a function of the increasing rate of censoring. Figure 2 presents the trace of the inverse of the observed Fisher information matrix. Notice that the trace is the sum of the variances of the estimates of the parameters. Thus high values of the trace indicates high uncertainty and hence less informative estimates. Each point represents the median of 500 traces. We used the median instead of the mean because for a few samples an indefinite information matrix was observed when the censoring rate was very high. It appears that the curve increases steadily up to 40% censoring and then increases rapidly after 50% censoring. Overall, the trace appears to increase exponentially as the censoring rate increases. The figure was truncated at the censoring rate 80% since the trace values increase very rapidly for censoring rates larger than 80%, which suggests that the parameters cannot be estimated with reliable accuracy beyond 80% censoring.

Let $I_n(\mu, \sigma, \rho)$ represent the Fisher information matrix for the completely observed case. It is not analytically shown that $I_{n,C}(\mu, \sigma, \rho)$ is positive definite. However, our simulation study suggests that the observed Fisher information is positive definite. If $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices, then $\mathbf{A} \geq \mathbf{B}$ if and only if $\mathbf{A}^{-1} \leq \mathbf{B}^{-1}$, where $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a non-negative definite matrix (see Horn 1991). And if $\mathbf{A} \geq \mathbf{B}$, then $\det \mathbf{A} \geq \det \mathbf{B}$ and $\operatorname{tr}(\mathbf{A}) \geq \operatorname{tr}(\mathbf{B})$. Numerically we checked that $I_n(\mu, \sigma, \rho) \geq I_{n,C}(\mu, \sigma, \rho)$. Notice that this inequality quantifies the loss of information due to censoring.

5. SIMULATION STUDY

5.1. Data generation.

The data generating process is the following:
Step 1. Generate time series data from a Gaussian AR(1) process using a traditional method (e.g. arima.sim in the statistical software R).
Step 2. Construct the censored time series $X_t = Y_t \cdot (1 - I(Y_t > c)) + c \cdot I(Y_t > c)$.

For comparison, we calculated the parameter estimates from the original complete data Y_t, which are obtained before censoring. Also, we calculated the parameter estimates from treating the censored values as observed, that is X_t. We applied two naive approaches as the candidate data analysis methods. One is to treat the censored values as observed and the other is to consider the censored observations as missing values. We will denote the former by NM1 and the latter by NM2, where NM stands for *naive method*. For convenience, we denote the estimates from the complete data analysis by $\hat{\theta}_{\text{CDA}}$ and those from NM1 and NM2 by $\hat{\theta}_{\text{NM1}}$ and $\hat{\theta}_{\text{NM2}}$, respectively. The estimates from the imputation method are denoted by $\hat{\theta}_{\text{IMP}}$.

TABLE 1: Bias and standard deviation (in parentheses) from the simulation results for a Gaussian CENAR(1). The results NM1 and NM2 are based on naive methods and the results IMP are based on the imputation method. The column 0% is for all observed data. The columns 20% and 40% correspond to the censoring rate. The autocorrelations used for the simulation are $\rho = .3$ (top) and $\rho = .7$ (bottom).

$\rho = .3$				
Method	Parameter	0% Censoring	20% Censoring	40% Censoring
	ρ	−.018 (.069)	−.040 (.068)	−.067 (.071)
NM1	μ	−.003 (.107)	−.089 (.063)	−.230 (.052)
	σ	.000 (.047)	−.163 (.041)	−.324 (.047)
	ρ	−.018 (.069)	−.129 (.084)	−.170 (.100)
NM2	μ	−.003 (.107)	−.355 (.073)	−.661 (.070)
	σ	.000 (.047)	−.221 (.047)	−.336 (.051)
	ρ	−.018 (.069)	−.022 (.071)	−.027 (.080)
IMP	μ	−.003 (.107)	−.008 (.078)	−.013 (.077)
	σ	.000 (.047)	−.013 (.059)	−.020 (.075)

$\rho = .7$				
Method	Parameter	0% Censoring	20% Censoring	40% Censoring
	ρ	−.018 (.052)	−.043 (.054)	−.075 (.063)
NM1	μ	−.001 (.251)	−.055 (.071)	−.149 (.058)
	σ	.003 (.049)	−.145 (.052)	−.291 (.065)
	ρ	−.018 (.052)	−.134 (.070)	−.233 (.092)
NM2	μ	−.001 (.251)	−.379 (.168)	−.780 (.137)
	σ	.003 (.049)	−.141 (.055)	−.231 (.067)
	ρ	−.018 (.052)	−.023 (.053)	−.024 (.058)
IMP	μ	−.001 (.251)	−.007 (.082)	−.011 (.082)
	σ	.003 (.049)	−.013 (.064)	−.021 (.080)

In this simulation study we set $\mu = 0$ and $\sigma = 1$. The sample size is set to $n = 200$ and we use $N = 200$ simulation replicates. We repeated the simulation for ρ equal to .3 and .7, and an average censoring rate of 20% and 40%. To maintain the targeted average censoring, we used the cutoff point derived in (8).

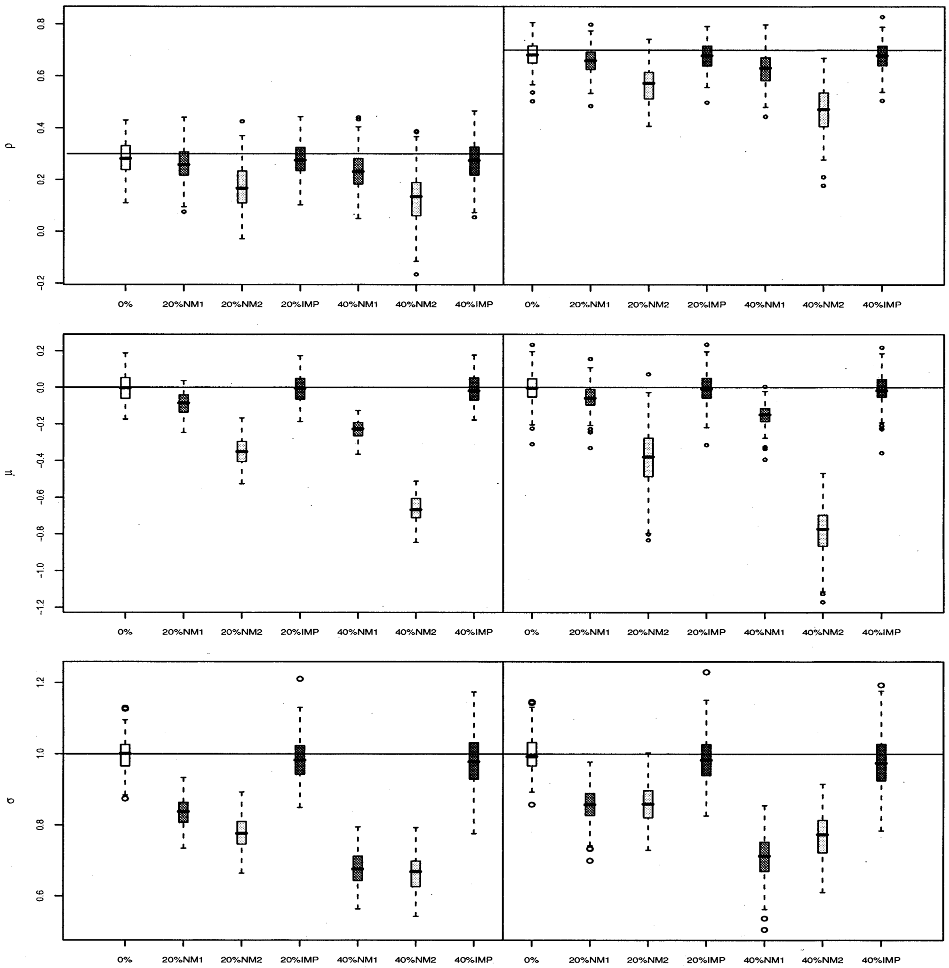

FIGURE 3: Box-plots of the estimation for ρ (top), μ (middle), and σ (bottom). The reference lines in the plots are the true values of the parameters. The results NM1 and NM2 are based on the naive methods and the results IMP are based on the imputation method.

5.2. Results of the simulation study.

The main findings from the simulation are summarized numerically and graphically. The tables and figures are based on the comparison between the imputation method and the two naive methods. Table 1 displays the biases of the parameter estimates and the standard deviations from the simulation and Figure 3 shows the box-plots of the parameter estimates. The results suggest that the imputation method improves the performance of the parameter estimation in terms of bias and mean squared error. Table 1 suggests that some of the parameter estimates are biased when we use the naive methods. For example, $\hat{\mu}_{\mathrm{NM1}}$ and $\hat{\mu}_{\mathrm{NM2}}$ are highly biased when the censoring rate is 40% while $\hat{\mu}_{\mathrm{IMP}}$ appears to be unbiased. The estimates for σ are always negatively biased for both censoring rates of 20% and 40%. It seems that the inference on σ is more significantly affected by censoring than the inference on μ and ρ.

The imputation method presents another benefit in terms of the relative efficiency. We use the ratio of the estimated mean squared errors to calculate the relative efficiency. The estimated mean squared error (MSE) for μ is calculated by

$$\hat{\text{MSE}}(\hat{\mu}_1,\ldots,\hat{\mu}_N) = (\bar{\hat{\mu}} - \mu)^2 + \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\mu}_i - \bar{\hat{\mu}})^2, \tag{9}$$

where $\bar{\hat{\mu}} = N^{-1}\sum_{i=1}^{N}\hat{\mu}_i$ and N is the number of simulation replicates. The estimated mean squared errors for ρ and σ are computed similarly. The relative efficiency in Table 2 is the ratio of the estimated mean squared errors from the imputation method to those from the naive methods. It follows from Table 2 that the efficiency gain for $\hat{\rho}$ is only about 11%, compared to the NM1 when $\rho = .3$ and the censoring rate is 20%. However, in other cases such as $\rho = .7$ or 40% censoring, the efficiency of the imputation method is much better than the naive methods, especially for μ and σ. Although both NM1 and NM2 produce biased and inefficient estimates compared to the IMP method, it appears that among the naive approaches, the performance of NM2 is the worst.

Overall, Table 1 shows that we get biased estimators for μ and σ using the naive estimates, and Table 2 shows that we lose efficiency when the naive methods are used. Clearly, the use of the imputation method reduces the biases and increases the efficiency.

The box-plots in Figure 3 show that the distribution of the parameter estimates from the imputation method are similar to those obtained from complete data. However, the distributions of the parameter estimates from the naive methods are significantly biased compared to those from complete data, especially for μ and σ. The dispersion of the parameter estimates from the naive methods is smaller than those from the imputation method. However, it does not mean that estimates are more efficient since the estimates from the naive approaches are significantly biased.

Next we compare the Fisher information based on the imputation method to that based on the naive methods. Figure 4 presents the comparison through box-plots, between the observed Fisher information and the plug-in version, denoted by $\hat{I}_n(\hat{\theta})$, of the observed Fisher information. The plot suggests that the distributions of the trace of the inverse of the Fisher information matrix from the imputation method are very similar to those from the complete data case. Especially for 20% censoring, the results appear almost equivalent. This suggests that the imputation method recovers the censored part well enough to mimic the estimation based on complete data. Note that we did not report the box-plots of the NM2 with 40% censoring because the bias was too large in that case.

TABLE 2: Relative efficiency based on the ratio of mean squared errors defined in (9) from the simulation.

		NM1		NM2	
	Parameter	20% Censoring	40% Censoring	20% Censoring	40% Censoring
	ρ	.89	.75	.24	.18
$\rho = .3$	μ	.51	.11	.05	.01
	σ	.13	.06	.07	.05
	ρ	.71	.41	.15	.06
$\rho = .7$	μ	.84	.27	.04	.01
	σ	.18	.08	.18	.12

6. APPLICATION TO A METEOROLOGICAL TIME SERIES

We apply our methodology to the meteorological time series of cloud ceiling height described in the Introduction. There were three missing observations and we used $A_j = (-\infty, \infty)$ for our imputation method corresponding to the missing observations.

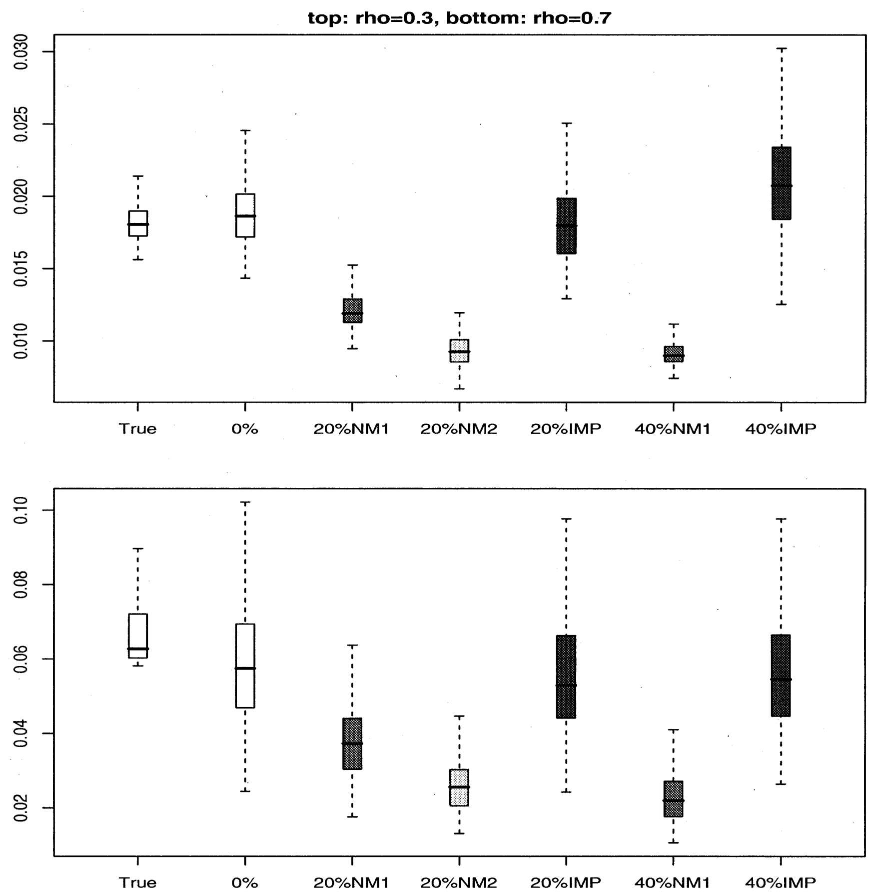

FIGURE 4: Box-plots of the trace of the inverse of the observed Fisher information matrix and the plug-in version of the observed Fisher information matrix, by censoring rate. The far left box-plots show the trace of the inverse of the observed Fisher information matrix evaluated with the true parameter.

The results from three estimation methods, NM1, NM2, and Hopke–Liu–Rubin, were compared to those from the imputation (IMP) method. The data were fitted by using three candidate models, AR(1), AR(2), and ARMA(1,1), after log-transformation and the results are displayed in Table 3. We implemented the Hopke–Liu–Rubin method using the Bayesian statistical software WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/).

We observe very different results among the IMP/Hopke–Liu–Rubin methods and the naive methods based on three models. The estimates of μ and σ based on the imputation method and the Hopke–Liu–Rubin one are inflated compared to those based on NM1 and NM2. For instance, the average log-transformed ceiling height values based on the imputation method and on the Hopke–Liu–Rubin method are slightly greater than 4, whereas those based on NM1 and NM2 are about 3.7 and 3.0, respectively. This is consistent with the results we observed in the simulation study. Thus the naive methods will mistakenly estimate the average cloud ceiling height to be lower than what it should be. Consequently, underestimation of μ can lead to higher risk for the general aviation pilot. In NM1, the ARMA(1,1) model gives the minimum Akaike information criterion (AIC) and hence we choose the ARMA(1,1) as the best model based on NM1. In NM2, AR(1) was selected due to the minimum AIC. For Hopke–Liu–Rubin, the AR(2) model was selected as the best model based on the deviance information criteria (DIC), see Spiegelhalter, Best, Carlin & Van der Linde (2003). In the imputation method, the AR(2) was chosen as best.

Figure 5 displays the original data in log-scale and the augmented data based on the parameter estimates of a CENAR(2) model using the imputation method. To obtain the augmented data, we generate 20 realizations of the time series for the censored and missing parts and calculate their average. This figure depicts the cloud ceiling heights that would have been observed had the device not had a detection limit.

From the simulation study, we observed that the parameter estimates may be biased when we treat the censored part as observed or missing. In addition we note that the naive approaches could also lead to selecting an incorrect model. In the ceiling height data case, the decision is different for the model selection between the naive methods and the imputation method. Since the imputation method gives a more consistent estimation result than the naive approach, it is reasonable to choose the AR(2) model as the best among the three models for the data.

TABLE 3: Parameter estimates for the log-transformed cloud ceiling height data. Three candidate models were used and AIC (DIC for the Hopke–Liu–Rubin (HLR) method) is displayed as a criterion for the model selection. Bold fonts represent the "best" model.

			NM1			
Model	$\hat{\mu}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\psi}$	$\hat{\sigma}$	AIC
AR(1)	3.720 (.123)	.778 (.025)	NA	NA	.736	1594
AR(2)	3.707 (.149)	.638 (.038)	.182 (.038)	NA	.724	1574
ARMA(1,1)	**3.704 (.159)**	**.872 (.024)**	**NA**	**.243 (.046)**	**.723**	**1572**
			NM2			
Model	$\hat{\mu}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\psi}$	$\hat{\sigma}$	AIC
AR(1)	**3.004 (.147)**	**.841 (.027)**	**NA**	**NA**	**.562**	**743**
AR(2)	3.001 (.152)	.801 (.055)	.047 (.056)	NA	.562	744
ARMA(1,1)	3.000 (.153)	.857 (.031)	NA	.055 (.066)	.562	744
			HLR			
Model	$\hat{\mu}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\psi}$	$\hat{\sigma}$	DIC
AR(1)	4.297 (.347)	.884 (.024)	NA	NA	1.048	1472
AR(2)	**4.194 (.407)**	**.740 (.049)**	**.159 (.049)**	**NA**	**1.035**	**1465**
ARMA(1,1)	4.195 (.404)	0.913 (.021)	NA	.171 (.053)	1.035	1466
			IMP			
Model	$\hat{\mu}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\psi}$	$\hat{\sigma}$	AIC
AR(1)	4.147 (.196)	.827 (.022)	NA	NA	.910	1904
AR(2)	**4.129 (.236)**	**.689 (.038)**	**.173 (.038)**	**NA**	**.877**	**1852**
ARMA(1,1)	4.211 (.264)	.898 (.020)	NA	.214 (.044)	.930	1938

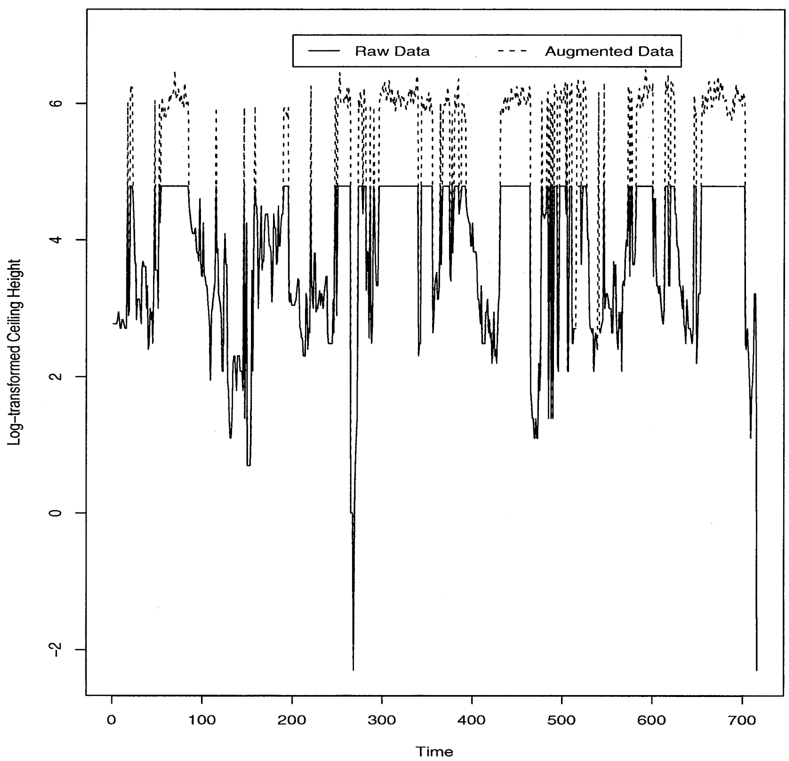

FIGURE 5: The super-imposed time plots of the log-transformed original series containing censored observations and the augmented series based on a Gaussian CENAR(2) model fit.

7. DISCUSSION

The main message of the imputation method is that we should account for the variability of the censored part of the data by mimicking the complete data. That is, we impute the incomplete part with a conditional random sample rather than the conditional expectation or certain constants. Simulation results suggest that the imputation method reduces the possible biases and has similar standard errors to those from complete data.

In order to use the imputation method, we have taken advantage of the fact that a Gaussian ARMA model can be related to a multivariate normal distribution. We use this property to characterize the variance-covariance matrix for time series models. It may be noticed that our imputation method is not limited to generating samples from truncated multivariate normal distributions. We can easily extend it to other multivariate distributions, e.g., multivariate t-distributions or even multivariate skew-elliptical distributions for which the conditional distribution is well known. See the book edited by Genton (2004) for a recent account.

Our method is obviously not restricted to ARMA models and can be directly extended to several other settings. For example, vector ARMA time series models can easily be fitted to censored data with our imputation approach. The analysis of spatial censored data by means of regression models with spatially correlated errors is also straightforward. Finally, parametric models of spatial or temporal covariances can be replaced by nonparametric estimators as long as the corresponding covariance matrix is guaranteed to be positive definite. Such extensions will be the topics of future investigations.

APPENDIX

A. Likelihood of the Gaussian CENAR(1) process.

We treat the time series data from an AR(1) process as a random vector from the multivariate normal distribution as in (3), where $\boldsymbol{\mu} = \mu \mathbf{1}_n$ and

$$\{\boldsymbol{\Sigma}\}_{ij} = \frac{\sigma^2}{1-\rho^2}\rho^{|i-j|}.$$

In this case, we have

$$\begin{aligned} |\boldsymbol{\Sigma}| &= \sigma^{2n}(1-\rho^2)^{-1}, \quad \text{and} \\ (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= (1-\rho^2)(x_1-\mu)^2 + \sum_{t=2}^{n}[(x_t-\mu)-\rho(x_{t-1}-\mu)]^2. \end{aligned}$$

The density of $\mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be written as

$$f(\mathbf{x}\,|\,\mu,\sigma^2,\rho) = f_0(x_1)\prod_{t=2}^{n} f(x_t\,|\,x_{t-1}),$$

where

$$f_0(x_1) = \frac{\sqrt{1-\rho}}{\sigma}\phi\left(\frac{(x_1-\mu)\sqrt{1-\rho^2}}{\sigma}\right), \quad f(x_t\,|\,x_{t-1}) = \frac{1}{\sigma}\phi\left(\frac{(x_t-\mu)-\rho(x_{t-1}-\mu)}{\sigma}\right),$$

and $\phi(\cdot)$ denotes the standard normal density function (Wei 1990). Because some observations are right censored at c, we need the last expression to be modified in order to set up the correct likelihood for the following four cases:

(1) Both X_t and X_{t-1} are observed:

$$f(x_t\,|\,x_{t-1}) = \frac{1}{\sigma}\phi\left(\frac{(x_t-\mu)-\rho(x_{t-1}-\mu)}{\sigma}\right); \tag{10}$$

(2) X_t is observed but X_{t-1} is censored:

$$\begin{aligned} f(x_t\,|\,x_{t-1}>c) &= \left\{1-\Phi\left(\frac{(c-\mu)\sqrt{1-\rho^2}}{\sigma}\right)\right\}^{-1}\frac{\sqrt{1-\rho^2}}{\sigma}\phi\left(\frac{(x_t-\mu)\sqrt{1-\rho^2}}{\sigma}\right) \\ &\quad \times\left\{1-\Phi\left(\frac{(c-\mu)-\rho(x_t-\mu)}{\sigma}\right)\right\}; \end{aligned} \tag{11}$$

(3) X_t is censored but X_{t-1} is observed:

$$f(x_t>c\,|\,x_{t-1}) = 1-\Phi\left(\frac{(c-\mu)-\rho(x_{t-1}-\mu)}{\sigma}\right); \tag{12}$$

(4) Both X_t and X_{t-1} are censored:

$$\begin{aligned} &f(x_t>c\,|\,x_{t-1}>c) \\ &= \frac{\displaystyle\int_c^\infty\int_c^\infty \frac{1-\rho^2}{\sigma^2}\phi_2\left(\frac{(x-\mu)\sqrt{1-\rho^2}}{\sigma},\frac{(y-\mu)\sqrt{1-\rho^2}}{\sigma};\rho\right)dx\,dy}{\displaystyle\int_c^\infty\int_{-\infty}^\infty \frac{1-\rho^2}{\sigma^2}\phi_2\left(\frac{(x-\mu)\sqrt{1-\rho^2}}{\sigma},\frac{(y-\mu)\sqrt{1-\rho^2}}{\sigma};\rho\right)dx\,dy} \\ &= \frac{K}{1-\Phi\left(\dfrac{(c-\mu)\sqrt{1-\rho^2}}{\sigma}\right)}, \end{aligned} \tag{13}$$

where

$$\phi_2(t,u;\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(t^2 - 2\rho t u + u^2)\right\}.$$

We also need to consider whether X_1 is observed or censored. Denote by $d_t = 1$ if $x_t \le c$ and 0 otherwise. Then the log-likelihood that accounts for censoring is represented by:

$$\begin{aligned}\log L &= \log f_0(x_1) I(d_1 = 1) + \log f_0(x_1 > c) I(d_1 = 0)\\ &+ \sum_{t=2}^{n} \{\log f(x_t \,|\, x_{t-1})\} I(d_t = 1, d_{t-1} = 1)\\ &+ \sum_{t=2}^{n} \{\log f(x_t \,|\, x_{t-1} > c)\} I(d_t = 1, d_{t-1} = 0)\\ &+ \sum_{t=2}^{n} \{\log f(x_t > c \,|\, x_{t-1})\} I(d_t = 0, d_{t-1} = 1)\\ &+ \sum_{t=2}^{n} \{\log f(x_t > c \,|\, x_{t-1} > c)\} I(d_t = 0, d_{t-1} = 0).\end{aligned}$$

The Fisher information is obtained from the separate second derivatives of each term.

B. Calculation of the derivatives based on truncated distributions.

It is straightforward to calculate the second derivatives for the observed case but it is necessary to use the following fact for the censored case. Let $\lambda(z) = \phi(z)/\{1 - \Phi(z)\}$. Then

$$\frac{\partial \log\{1 - \Phi(z)\}}{\partial z} = -\lambda(z), \quad \text{and} \tag{14}$$

$$\frac{\partial \lambda(z)}{\partial z} = \lambda(z)\{\lambda(z) - z\}. \tag{15}$$

Due to space limitation, we only display the second derivatives with respect to μ as an example. Other derivatives can be obtained similarly.

(1) Derivatives corresponding to X_1:

If X_1 is observed, then we have

$$\frac{\partial^2 \log f_0(x_1)}{\partial \mu^2} = -\frac{1-\rho^2}{\sigma^2}. \tag{16}$$

If X_1 is censored then, using (14) and (15), the second derivative with respect to μ is

$$\frac{\partial^2 \log f_0(x_1 > c)}{\partial \mu^2} = -\frac{1-\rho^2}{\sigma^2}\lambda(c_\mu)\{\lambda(c_\mu) - c_\mu\}, \tag{17}$$

where

$$c_\mu = \frac{(c-\mu)\sqrt{1-\rho^2}}{\sigma}.$$

(2) Both X_t and X_{t-1} are observed:

The log-likelihood follows from (10) and the second derivatives are obtained similarly to the previous case. For example, the second derivative with respect to μ is

$$\frac{\partial^2 \log f(x_t \,|\, x_{t-1})}{\partial \mu^2} = -\frac{(1-\rho)^2}{\sigma^2}. \tag{18}$$

(3) X_t is observed but X_{t-1} is censored:

The log-likelihood follows from (11). Using (14) and (15), the second derivatives are obtained. For example, the second derivative with respect to μ is

$$\begin{aligned}\frac{\partial^2 \log f(x_t \,|\, x_{t-1} > c)}{\partial \mu^2} &= -\frac{1-\rho^2}{\sigma^2} + \frac{1-\rho^2}{\sigma^2}\lambda(c_\mu)\{\lambda(c_\mu) - c_\mu\} \\ &\quad - 6\frac{(1-\rho)^2}{\sigma^2}\lambda(c_\mu^*)\{\lambda(c_\mu^*) - c_\mu^*\}, \qquad (19)\end{aligned}$$

where $c_\mu^* = \{(x_t - \mu) - \rho(c - \mu)\}/\sigma$.

(4) X_t is censored but X_{t-1} is observed:

The log-likelihood follows from (12). Using (14) and (15), the second derivatives are obtained. For example, the second derivative with respect to μ is

$$\frac{\partial^2 \log f(x_t > c \,|\, x_{t-1})}{\partial \mu^2} = -\left(\frac{1-\rho}{\sigma}\right)^2 \lambda(c_\mu^{**})\{\lambda(c_\mu^{**}) - c_\mu^{**}\}, \qquad (20)$$

where $c_\mu^{**} = \{(c - \mu) - \rho(x_{t-1} - \mu)\}/\sigma$.

(5) Both X_t and X_{t-1} are censored:

The log-likelihood follows from (13) and we need to calculate the derivatives of K with respect to the parameters. Using the Leibnitz rule, the first derivative with respect to μ is obtained by

$$\frac{\partial K}{\partial \mu} = \frac{2\sqrt{1-\rho^2}}{\sigma}\phi(c_\mu)\{1 - \Phi(c_\mu^\dagger)\},$$

where $c_\mu^\dagger = c_\mu(1-\rho)/\sqrt{1-\rho^2}$. Then the second derivative is obtained by

$$\frac{\partial^2 K}{\partial \mu^2} = \frac{2(1-\rho^2)}{\sigma^2}c_\mu\phi(c_\mu)\{1 - \Phi(c_\mu^\dagger)\} + \frac{2(1-\rho)\sqrt{1-\rho^2}}{\sigma^2}\phi(c_\mu)\phi(c_\mu^\dagger).$$

Finally the second derivative with respect to μ is

$$\frac{\partial^2 \log f(x_t > c \,|\, x_{t-1} > c)}{\partial \mu^2} = K^{-1}\left\{\frac{\partial^2 K}{\partial \mu^2}\right\} - K^{-2}\left\{\frac{\partial K}{\partial \mu}\right\}^2 - \frac{\partial^2 \log\{1 - \Phi(c_\mu)\}}{\partial \mu^2}. \qquad (21)$$

We also confirmed all algebraic calculations by using the software Mathematica®.

C. Observed Fisher information matrix.

We combine all the results from the previous section to construct the observed Fisher information matrix as in (7). For example, D_{μ^2} is the sum of the second derivatives as in (16), (17), (18), (19), (20), and (21) corresponding to one of four cases.

ACKNOWLEDGEMENTS

The authors thank the Editor, the Associate Editor, and an anonymous referee for insightful comments on the manuscript. The authors are grateful to the National Center for Atmospheric Research (Boulder, CO) for providing the data. The work of Genton was partially supported by a grant from the United States National Science Foundation.

REFERENCES

T. W. Anderson (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edition. Wiley, New York.

G. E. P. Box & G. M. Jenkins (1970). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

W. A. Fuller (1996). *Introduction to Statistical Time Series*. Wiley, New York.

A. E. Gelfand & A. F. M. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

M. G. Genton. ed. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman & Hall/CRC Press, Boca Raton, Florida.

D. R. Helsel (1990). Less than obvious: statistical treatment of data below the detection limit. *Environmental Science and Technology*, 24, 1766–1774.

R. V. Hogg & A. T. Craig (1994). *Introduction to Mathematical Statistics*, 5th edition. Prentice Hall, New York.

P. K. Hopke, C. Liu & D. B. Rubin (2001). Multiple imputation for multivariate data with missing and below-threshold measurements: time-series concentrations of pollutants in the Arctic. *Biometrics*, 57, 22–33.

R. A. Horn (1991). *Topics in Matrix Analysis*. Cambridge University Press, Cambridge.

R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

C. P. Robert (1995). Simulation of truncated normal variables. *Statistics and Computing*, 5, 121–125.

P. M. Robinson (1980). Estimation and forecasting for time series containing censored or missing observations. In *Time Series: Proceedings of the International Meeting held at Nottingham University, Nottingham, March 26–30, 1979* (Oliver D. Anderson, ed.), North-Holland, Amsterdam, pp. 167–182.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin & A. Van der Linde (2003). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B*, 64, 583–616.

W. S. Wei (1990). *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley, Redwood City, California.

S. L. Zeger & R. Brookmeyer (1986). Regression analysis with censored autocorrelated data. *Journal of the American Statistical Association*, 81, 722–729.

Received 4 November 2005
Accepted 11 September 2006

Jung Wook PARK: jung.w.park@gsk.com
Clinical Pharmacology Statistics and Programming
GlaxoSmithKline
Research Triangle Park, NC 27709-3398, USA

Marc G. GENTON: genton@stat.tamu.edu
Department of Statistics, Texas A&M University
College Station, TX 77843-3143, USA

Sujit K. GHOSH: sghosh@stat.ncsu.edu
Department of Statistics, North Carolina State University
Raleigh, NC 27695-8203, USA