

# Censored Time Series Analysis with Autoregressive Moving Average Models

Jung Wook Park<sup>1</sup>, Marc G. Genton<sup>2</sup> and Sujit K. Ghosh<sup>3</sup>

Institute of Statistics Mimeo Series# 2578

November 4, 2005

## Abstract

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

**Some key words:** Censored time Series; Fisher Information Matrix; Gibbs Sampling; Imputation Method; Truncated Multivariate Normal.

**Short title:** Censored Time Series Analysis

---

<sup>1</sup>Clinical Pharmacology Statistics and Programming, GlaxoSmithKline, PO Box 13398, Five Moore Drive, Research Triangle Park, NC 27709-3398, USA.  
E-mail: jung.w.park@gsk.com

<sup>2</sup>Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.  
E-mail: genton@stat.tamu.edu

<sup>3</sup>Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA.  
E-mail: sghosh@stat.ncsu.edu

# 1 Introduction

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

$$\tilde{Y}_t = \rho_1 \tilde{Y}_{t-1} + \dots + \rho_p \tilde{Y}_{t-p} + \epsilon_t - \psi_1 \epsilon_{t-1} - \dots - \psi_q \epsilon_{t-q}, \quad (1)$$

where  $\tilde{Y}_t = Y_t - \mu$ ,  $\mu$  is the mean parameter,  $\rho_k, k = 1, \dots, p$  are the autoregressive parameters, and  $\psi_l, l = 1, \dots, q$  are the moving average parameters. The error process  $\epsilon_t$  is assumed to be a white noise with mean 0 and variance  $\sigma^2$ . This model is reduced to an AR( $p$ ) model when  $q = 0$  or 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 happens in various situations such as physical science, business, and economics. The measurement of rainfall is often limited due to the size of the gauge and signals may intentionally be made limited for convenience in storage or processing (for further examples, see e.g. Robinson, 1980). Our motivation is the analysis of cloud ceiling height, which is an important variable accounting for weather-related accidents and flight delays, provided by the National Center for Atmospheric Research (NCAR), measured at San Francisco. Due to an upper detection limit of the device, many observations are censored and this will potentially lead to biased estimates when we implement classical analysis tools that ignore censoring.

There may be two naive approaches to handle 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 Section 5.2. Thus it is necessary to use an alternative approach to improve the performance of the parameter estimation for the 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 highly depend on

the strength of such model assumptions. Thus the key issue for 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 subgrouped 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 and Brookmeyer (1986) suggested a full likelihood estimation and approximate method for an autoregressive time series model. However, the authors have pointed in their article that the method may not be feasible when the censoring rate is very high (Zeger and Brookmeyer, 1986, p.728). To overcome this limitation the authors have suggested the use of pseudolikelihood estimation. Hopke *et al.* (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 is regarded as a realization from a multivariate normal distribution. Next, we impute the censored values with a random vector from the conditional multivariate normal distribution given the observed part.

In Section 2, we define a censored time series model and its multivariate normal representation when the white noise is Gaussian. Then we introduce the imputation method to estimate the parameters of censored time series models. In Section 3, an algorithm is described to generate 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 the 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 we illustrate its use on a meteorological time series of cloud ceiling height, measured with a detection limit, using an ARMA(1,1) model in Section 6. 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 an ARMA( $p, q$ ) model defined in (1). 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,

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

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

We will call the process  $\{X_t\}$  a censored autoregressive and moving average model and denote it by CENARMA( $p, q$ ). The models corresponding to AR( $p$ ) and MA( $q$ ) will be 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 the  $Y_t$ 's.

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

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

$$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (3)$$

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 we can re-arrange the order of the data so that we can partition the data vector into an observed part  $\mathbf{Y}_O$  and a censored part  $\mathbf{Y}_C$  given by

$$\mathbf{P}\mathbf{Y} = \begin{pmatrix} \mathbf{P}_O \\ \mathbf{P}_C \end{pmatrix} \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_O \\ \mathbf{Y}_C \end{pmatrix}. \quad (4)$$

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

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

It follows (see e.g. Anderson, 1984) that the conditional distribution of  $\mathbf{Y}_C$  given  $\mathbf{Y}_O$  is also a multivariate normal distribution 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  $\mathbf{X} = (X_1, \dots, X_n)^T$ , we have

$$\mathbf{P}\mathbf{X} = \begin{pmatrix} \mathbf{P}_O \\ \mathbf{P}_C \end{pmatrix} \mathbf{X} = \begin{pmatrix} \mathbf{X}_O \\ \mathbf{X}_C \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \mathbf{Y}_O \\ \mathbf{X}_C \end{pmatrix}, \quad (6)$$

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  $\mathbf{X}_C$ , a truncated multivariate normal distribution,

$$\mathbf{Y}_C | \mathbf{X}_O, \mathbf{X}_C \in D_C \sim TN_{n_C}(\boldsymbol{\nu}, \boldsymbol{\Delta}, D_C),$$

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

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. Let us suppose we have an AR(1) time series process with mean  $\mu$ , variance  $\sigma^2$ , and autocorrelation  $\rho$ . We can consider the data as a random vector from a multivariate Gaussian distribution as in (3) where  $\boldsymbol{\mu} = \mu \mathbf{1}_n$  and  $\{\boldsymbol{\Sigma}\}_{ij} = \frac{\sigma^2}{1-\rho^2} \rho^{|i-j|}$ ,  $i, j = 1, \dots, n$ . 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 variable defined on the  $n$ -dimensional Euclidean space,  $\mathbb{R}^n$ , and let

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

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

- (0) Set  $\mathbf{x}^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})^T$ .
- (1) For,  $k = 1, 2, \dots$ , sample
 
$$x_1^{(k)} \sim g\left(x_1|\mathbf{x}_{-1}^{(k-1)}\right) I_{A_1}(x_1)$$

$$x_2^{(k)} \sim g\left(x_2|x_1^{(k)}, x_3^{(k-1)}, \dots, x_n^{(k-1)}\right) I_{A_2}(x_2)$$

$$\vdots$$

$$x_n^{(k)} \sim g\left(x_n|\mathbf{x}_{-n}^{(k)}\right) I_{A_n}(x_n)$$

- (2) Repeat until the samples follow the stationary distribution with density  $f(\mathbf{x})$ .

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

$$x_j = \nu_j + \tau_j \Phi^{-1}\left[U_j \left\{1 - \Phi\left(\frac{c_j - \nu_j}{\tau_j}\right)\right\} + \Phi\left(\frac{c_j - \nu_j}{\tau_j}\right)\right], \quad (7)$$

where  $\nu_j = \mu_j - \boldsymbol{\sigma}_j^T \boldsymbol{\Sigma}_{-j}^{-1}(\mathbf{x}_{-j} - \boldsymbol{\mu}_{-j})$ ,  $\tau_j^2 = \sigma_{jj} - \boldsymbol{\sigma}_j^T \boldsymbol{\Sigma}_{-j}^{-1} \boldsymbol{\sigma}_j$ ,

$\boldsymbol{\mu}_{-j} = (\mu_1, \dots, \mu_{j-1}, \mu_{j+1}, \dots, \mu_n)^T$ ,  $\boldsymbol{\sigma}_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{j-1,j}, \sigma_{j+1,j}, \dots, \sigma_{nj})^T$ , and  $\boldsymbol{\Sigma}_{-j}$  is the matrix  $\boldsymbol{\Sigma}$  with  $j$ -th 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  $\{\hat{\sigma}^{(0)}\}^2$  (e.g. see Step 6). Then construct the following mean vector and covariance matrix  $\hat{\Sigma}^{(0)}$ ,

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

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

$$\begin{aligned}\hat{\boldsymbol{\nu}}^{(0)} &= \hat{\boldsymbol{\mu}}_C^{(0)} + \hat{\Sigma}_{CO}^{(0)} \left( \hat{\Sigma}_{OO}^{(0)} \right)^{-1} \left( \mathbf{x}_O - \hat{\boldsymbol{\mu}}_O^{(0)} \right) \\ \hat{\Delta}^{(0)} &= \hat{\Sigma}_{CC}^{(0)} - \hat{\Sigma}_{CO}^{(0)} \left( \hat{\Sigma}_{OO}^{(0)} \right)^{-1} \hat{\Sigma}_{OC}^{(0)},\end{aligned}\tag{8}$$

where the covariances are defined in (5).

**Step 4.** Generate a random sample  $\mathbf{x}_C^{(1)}$  from  $TN_{n_c} \left( \boldsymbol{\nu}^{(0)}, \Delta^{(0)}, D_C \right)$ , where  $D_C = (c, \infty)^{n_c} = (c, \infty) \times (c, \infty) \times \dots \times (c, \infty)$ .

**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  $\mu$ ,  $\rho$ , and  $\sigma$  based on  $\mathbf{X}^{(1)}$  and update the parameters  $\Sigma$ ,  $\boldsymbol{\nu}$ , and  $\Delta$  (see e.g. Fuller, 1996). If a least squares approach is used we use the following

estimates,

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

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

**Step 7.** Repeat **Step 3** ~ **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  $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\rho}, \hat{\sigma})^T$ . In our simulation,  $\text{tol}=0.001$  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 method, least squares, maximum likelihood, etc.) 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 more insight on 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,

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



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

$$\hat{I}_n(\boldsymbol{\theta}) = \left[ -\frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right], \quad (9)$$

since it converges in probability to the expected Fisher information (Hogg and Craig, 1995, Sec 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 (See the Appendix) to compute (9).

## 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 changes. 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,

$$I_{n,C}(\mu, \sigma, \rho) = \begin{bmatrix} D_{\mu^2} & D_{\mu,\sigma} & D_{\mu,\rho} \\ D_{\mu,\sigma} & D_{\sigma^2} & D_{\sigma,\rho} \\ D_{\mu,\rho} & D_{\sigma,\rho} & D_{\rho^2} \end{bmatrix}, \quad (10)$$

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

A simulation was conducted in order to illustrate the change of the Fisher information as a function of the censoring rate. A sample size of  $n = 200$  and a simulation of size  $N = 500$  were used to study the Fisher information. The parameters were set to  $\mu = 0, \sigma = 1$ , and  $\rho = 0.1, 0.3, 0.5$ . We changed the censoring rate from 0% to 100% to examine the behavior of information as a function of increasing rate of censoring. Figure 1 presents the trace of the inverse of the 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 few samples a negative definite information matrix was observed especially when the censoring rate was very high. It appears that the curve increases steadily up to 40% of censoring and then increases rapidly after 50% of censoring. Overall, the trace appears to increase exponentially as 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% of 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, the simulation 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  $\text{tr}(\mathbf{A}) \geq \text{tr}(\mathbf{B})$ . 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 A Simulation Study

We present the data generation for the Gaussian CENAR(1) and the results of our simulation.

### 5.1 Data generation

For the data generating step, 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  $\alpha$  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:

$$c = \mu + \sigma \frac{\Phi^{-1}(1 - \alpha)}{\sqrt{1 - \rho^2}} \sqrt{1 - \rho^{2(n+1)}}, \quad (11)$$

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 criteria to find the cutoff point is to calculate

$c$  for every size of the time series and summarize it with an average or median:

$$c_a = \mu + \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{1-\rho^2}} \left\{ \frac{1}{n} \sum_{j=1}^n \sqrt{1-\rho^{2(j+1)}} \right\}, \text{ or}$$

$$c_m = \mu + \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{1-\rho^2}} \times \text{Median} \left\{ \sqrt{1-\rho^{2(j+1)}}, j = 1, \dots, n \right\}.$$

Now we describe the data generating process.

**Step 1.** Generate a 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  $X_t = Y_t \cdot (1 - C_t) + c \cdot C_t$ , where  $C_t = I(Y_t > c)$ .

For comparison purpose, we calculated the parameter estimates from the original complete data, that is  $Y_t$ , which is obtained before censoring. Also, we calculated the parameter estimates from treating the censored values as observed, that is  $X_t$ . For convenience, we denote the former by  $\hat{\theta}_{\text{raw}}$  and the latter by  $\hat{\theta}_{\text{cen}}$ . The estimates from the imputation method are denoted by  $\hat{\theta}_{\text{imp}}$ .

## 5.2 Results of the simulation study

In this simulation study we set  $\mu = 0$  and  $\sigma = 1$ . The sample size is set to  $n = 200$  and the simulation size is set to  $N = 200$ . We repeated the simulation for  $\rho$  equal to 0.3 and 0.7, and an average censoring rate  $c$  of 20% and 40%. To maintain the targeted average censoring we used the cutoff point derived in (11). 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 treating the censored part as observed. For convenience, we call the latter method the naive method.

Table 1 displays the biases of the parameter estimates and the standard deviations from the simulation and Figure 2 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 square error (MSE). Table 1 suggests that some of the parameter estimates are biased when we use the naive method. For example,  $\hat{\mu}_{\text{cen}}$  becomes highly biased when the censoring rate is 40% while  $\hat{\mu}_{\text{imp}}$  seems unbiased. The estimates for  $\sigma$  are always biased low for both censoring rates of 20% and 40%. It seems that the inference on  $\sigma$  is more significantly affected by censoring than  $\mu$  and  $\rho$ .

The imputation method presents another benefit in terms of the relative efficiency. We use the ratio of the estimated MSE's to calculate the relative efficiency. The estimated MSE

for  $\mu$  is calculated by

$$\widehat{\text{MSE}}(\hat{\mu}_1, \dots, \hat{\mu}_N) = (\bar{\hat{\mu}} - \mu)^2 + \frac{1}{N-1} \sum_{i=1}^N (\hat{\mu}_i - \bar{\hat{\mu}})^2, \quad (12)$$

where  $\bar{\hat{\mu}} = N^{-1} \sum_{i=1}^N \hat{\mu}_i$  and  $N$  is the simulation size. Similarly, the estimated MSE's for  $\rho$  and  $\sigma$  are computed. The relative efficiency in Table 2 is the ratio of the estimated MSE's from the imputation method to those from the naive method. It follows from Table 2 that the efficiency for  $\hat{\rho}$  is almost the same when  $\rho = .3$  and the censoring rate is 20%. However, in other cases such as  $\rho = .7$  or 40% of censoring, the efficiency of the imputation method is better than the naive method, especially for  $\mu$  and  $\sigma$ .

Overall, Table 1 implies that we get biased estimators for  $\mu$  and  $\sigma$  using the naive estimates and Table 2 implies that we lose efficiency when the naive method is used. Clearly, the use of the imputation method reduces the biases and increases the efficiency.

The box-plots in Figure 2 depict 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 method are significantly biased compared to those from complete data, especially for  $\mu$  and  $\sigma$ . The dispersion of the parameter estimates from the naive method is smaller than those from the imputation method. However, it does not mean that estimates are more efficient since the estimates from the naive approach are significantly biased.

Next we compare the Fisher information based on the imputation method to that based on the naive method. Figure 3 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% of censoring, the results appear almost equivalent. Thus this suggests that the imputation method recovers the censored part well enough to mimic the estimation based on complete data.

## 6 Application to a meteorological time series

Ceiling Height (CH) is defined as the distance from the ground to the bottom of a cloud measured in hundreds of feet. The CH is one of the major factors contributing to weather-related accidents and one of the major causes of flight delays. Moreover, adverse ceiling

condition can seriously affect the general aviation pilot on account of the fact that it can lead to the risk of disorientation, loss of control, and flight into obstructed terrain. Thus an accurate determination of the CH is important.

The data on CH, collected by the National Center for Atmospheric Research (NCAR), is observed hourly in San Francisco and collected in the month of March 1989, consisting of  $n = 729$  observations. The observations have a detection limit at  $c = 12,000$  feet and hence the data can be considered as a right-censored time series data. The censoring rate is 41.62% and there are three missing observations. For our imputation method corresponding to the missing observations, we use  $A_j = [-\infty, \infty]$ .

The data was fitted by using three candidate models, AR(1), AR(2), and ARMA(1,1), and the results are displayed in the first panel of Table 3. The ARMA(1,1) model gives the minimum AIC and hence we choose the ARMA(1,1) as the best model based on the naive approach. Then we use the imputation method to analyze the data. The results are displayed in the second panel of Table 3.

We observe very different results between the naive approach and the imputation method. The estimates of  $\mu$  and  $\sigma$  based on the imputation method are inflated compared to those based on the naive method. For instance, the average CH based on the imputation method is about 8,333 feet., whereas the average CH based on naive method is about 6,706 feet. This is consistent with results we observed in the simulation study. Thus the naive method will mistakenly estimate the average CH to be lower than what it should be. Consequently, under-estimation of  $\mu$  can lead to higher risk for the general aviation pilot. Using the naive approach, the ARMA(1,1) model seems to fit the data better than other candidate models in terms of AIC. However, using the imputation method, an AR(2) model seems to fit the data better.

Figure 4 displays the original data 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 the average. This figure depicts what CH observations would have been observed if the device would not have 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. In addition we note that the naive approach can also lead to selecting an incorrect model. In the CH data case, we have a difference in the decision for the model selection between the naive method and the imputation method. Since

the imputation method gives more consistent estimation result than the naive approach, it will be more reasonable to choose the AR(2) model as the best model for the data rather than the ARMA(1,1).

## 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 than those from complete data.

In order to use the imputation method, we have used the fact that a Gaussian ARMA model can be related to a multivariate Gaussian 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.

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 topic of future investigations.

# Appendix: Derivation of the Fisher Information matrix for the Gaussian CENAR(1) Process

## A.1 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  $\{\boldsymbol{\Sigma}\}_{ij} = \frac{\sigma^2}{1-\rho^2} \rho^{|i-j|}$ . Using the fact that,

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

The density of  $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  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  $f_0(x_1) = \frac{\sqrt{1-\rho^2}}{\sigma} \phi\left(\frac{(x_1-\mu)\sqrt{1-\rho^2}}{\sigma}\right)$ ,  $f(x_t|x_{t-1}) = \frac{1}{\sigma} \phi\left(\frac{(x_t-\mu)-\rho(x_{t-1}-\mu)}{\sigma}\right)$ , and  $\phi(\cdot)$  denotes the standard normal density function (Wei and Reilly, 1990). Since 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

$$f(x_t|x_{t-1}) = \frac{1}{\sigma} \phi\left(\frac{(x_t - \mu) - \rho(x_{t-1} - \mu)}{\sigma}\right); \quad (\text{A.1})$$

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

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

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

$$f(x_t > c|x_{t-1}) = 1 - \Phi\left(\frac{(c - \mu) - \rho(y - \mu)}{\sigma}\right); \quad (\text{A.3})$$

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

$$\begin{aligned}
f(x_t > c | x_{t-1} > c) &= \frac{\int_c^\infty \int_c^\infty \frac{1-\rho^2}{\sigma^2} \phi\left(\frac{(x-\mu)\sqrt{1-\rho^2}}{\sigma}, \frac{(y-\mu)\sqrt{1-\rho^2}}{\sigma}; \rho\right) dx dy}{\int_c^\infty \int_{-\infty}^\infty \frac{1-\rho^2}{\sigma^2} \phi\left(\frac{(x-\mu)\sqrt{1-\rho^2}}{\sigma}, \frac{(y-\mu)\sqrt{1-\rho^2}}{\sigma}; \rho\right) dx dy} \\
&= \frac{K}{1 - \Phi\left(\frac{(c-\mu)\sqrt{1-\rho^2}}{\sigma}\right)}, \tag{A.4}
\end{aligned}$$

where

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

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

$$\begin{aligned}
\log L &= \log f_0(x_1)I(d_1 = 1) + \log f_0(x_1 > c)I(d_1 = 0) \\
&+ \sum_{t=2}^n \{\log f(x_t | x_{t-1})\}I(d_t = 1, d_{t-1} = 1) \\
&+ \sum_{t=2}^n \{\log f(x_t | x_{t-1} > c)\}I(d_t = 1, d_{t-1} = 0) \\
&+ \sum_{t=2}^n \{\log f(x_t > c | x_{t-1})\}I(d_t = 0, d_{t-1} = 1) \\
&+ \sum_{t=2}^n \{\log f(x_t > c | x_{t-1} > c)\}I(d_t = 0, d_{t-1} = 0). \tag{A.5}
\end{aligned}$$

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

## A.2 Derivatives calculation 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,

$$\frac{\partial \log\{1 - \Phi(z)\}}{\partial z} = -\lambda(z), \text{ and} \tag{A.6}$$

$$\frac{\partial \lambda(z)}{\partial z} = \lambda(z)\{\lambda(z) - z\}. \tag{A.7}$$

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



(1) Derivatives corresponding to  $X_1$

If  $X_1$  is observed, then we have

$$\frac{\partial^2 \log f_0(x_1)}{\partial \mu^2} = -\frac{1 - \rho^2}{\sigma^2}. \quad (\text{A.8})$$

If  $X_1$  is censored then, using (A.6) and (A.7), the second derivative with respect to  $\mu$  is

$$\frac{\partial^2 \log f_0(x_1 > c)}{\partial \mu^2} = -\frac{1 - \rho^2}{\sigma^2} \lambda(c_\mu) \{ \lambda(c_\mu) - c_\mu \}, \quad (\text{A.9})$$

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 (A.1) and the second derivatives are obtained similarly to the previous case. For example, the second derivative with respect to  $\mu$  is

$$\frac{\partial^2 \log f(x_t | x_{t-1})}{\partial \mu^2} = -\frac{(1 - \rho)^2}{\sigma^2}. \quad (\text{A.10})$$

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

The log-likelihood follows from (A.2). Using (A.6) and (A.7), the second derivatives are obtained. For example, the second derivative with respect to  $\mu$  is

$$\begin{aligned} \frac{\partial^2 \log f(x_t | x_{t-1} > c)}{\partial \mu^2} &= -\frac{1 - \rho^2}{\sigma^2} + \frac{1 - \rho^2}{\sigma^2} \lambda(c_\mu) \{ \lambda(c_\mu) - c_\mu \} \\ &\quad - \frac{(1 - \rho)^2}{\sigma^2} \lambda(c_\mu^*) \{ \lambda(c_\mu^*) - c_\mu^* \}, \end{aligned} \quad (\text{A.11})$$

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 (A.3). Using (A.6) and (A.7), the second derivatives are obtained. For example, the second derivative with respect to  $\mu$  is

$$\frac{\partial^2 \log f(x_t > c | x_{t-1})}{\partial \mu^2} = -\left(\frac{1 - \rho}{\sigma}\right)^2 \lambda(c_\mu^{**}) \{ \lambda(c_\mu^{**}) - c_\mu^{**} \}, \quad (\text{A.12})$$

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 (A.4) and we need to calculate the derivatives of  $K$  with respect to the parameters. Using Leibnitz' rule, the first derivative with respect to  $\mu$  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 derivatives is obtained by

$$\frac{\partial^2 K}{\partial \mu^2} = \frac{2(1-\rho^2)}{\sigma^2} c_\mu \phi(c_\mu) \{1 - \Phi(c_\mu^\dagger)\} + \frac{2(1-\rho)\sqrt{1-\rho^2}}{\sigma^2} \phi(c_\mu) \phi(c_\mu^\dagger). \quad (\text{A.13})$$

Finally the second derivatives with respect to  $\mu$  is

$$\frac{\partial^2 \log f(x_t > c | x_{t-1} > c)}{\partial \mu^2} = K^{-1} \left\{ \frac{\partial^2 K}{\partial \mu^2} \right\} - K^{-2} \left\{ \frac{\partial K}{\partial \mu} \right\}^2 - \frac{\partial^2 \log \{1 - \Phi(c_\mu)\}}{\partial \mu^2}. \quad (\text{A.14})$$

We also confirmed all algebraic calculations by using the software Mathematica<sup>®</sup>.

### A.3 Observed Fisher information Matrix

We combine all the results from the previous section to construct the observed Fisher information matrix as in (10). For example,  $D_{\mu^2}$  is the sum of the second derivatives as in (A.8), (A.9), (A.10), (A.11), (A.12), and (A.14) corresponding to one of four cases.

## References

- Anderson, T.W (1984). *An Introduction to Multivariate Statistical Analysis*, Wiley.
- Box, G. E. P., Jenkins, G. M. (1970). *Time Series Analysis : Forecasting and Control.*, Holden-Day, San Francisco.
- Fuller, W. A. (1996). *Introduction to Statistical Time Series*, Wiley.
- Gelfand, A. E., Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398-409.
- Helsel, D. R. (1990). Less than obvious: statistical treatment of data below the detection limit, *Environmental Science and Technology*, **24(12)**, 1766-1774.
- Hogg, R. V., Craig, A. T. (1995). *Introduction to mathematical statistics*, Prentice Hall.
- Hopke, P.K., Liu, C., Rubin, D. B. (2001). Multiple imputation for multivariate data with missing and below-threshold measurements: time-series concentrations of pollutants in the Arctic., *Biometrics*, **57(1)**, 22-33.
- Horn, R. A. (1991). *Topics in Matrix Analysis*, Cambridge University Press.
- 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>
- Robert, C. P. (1995). Simulation of truncated normal variables, *Statistics and Computing*, **5**, 121-125.
- Robinson, R. M. (1980). Estimation and forecasting for time series containing censored or missing observations, *Time Series*, North-Holland Publishing Company, 167-182.
- Wei, W.S, Reilly, D.P. (1990). *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley
- Zeger, S. L., Brookmeyer, R. (1986). Regression analysis with censored autocorrelated data, *Journal of the American Statistical Association*, **81**, 722-729.

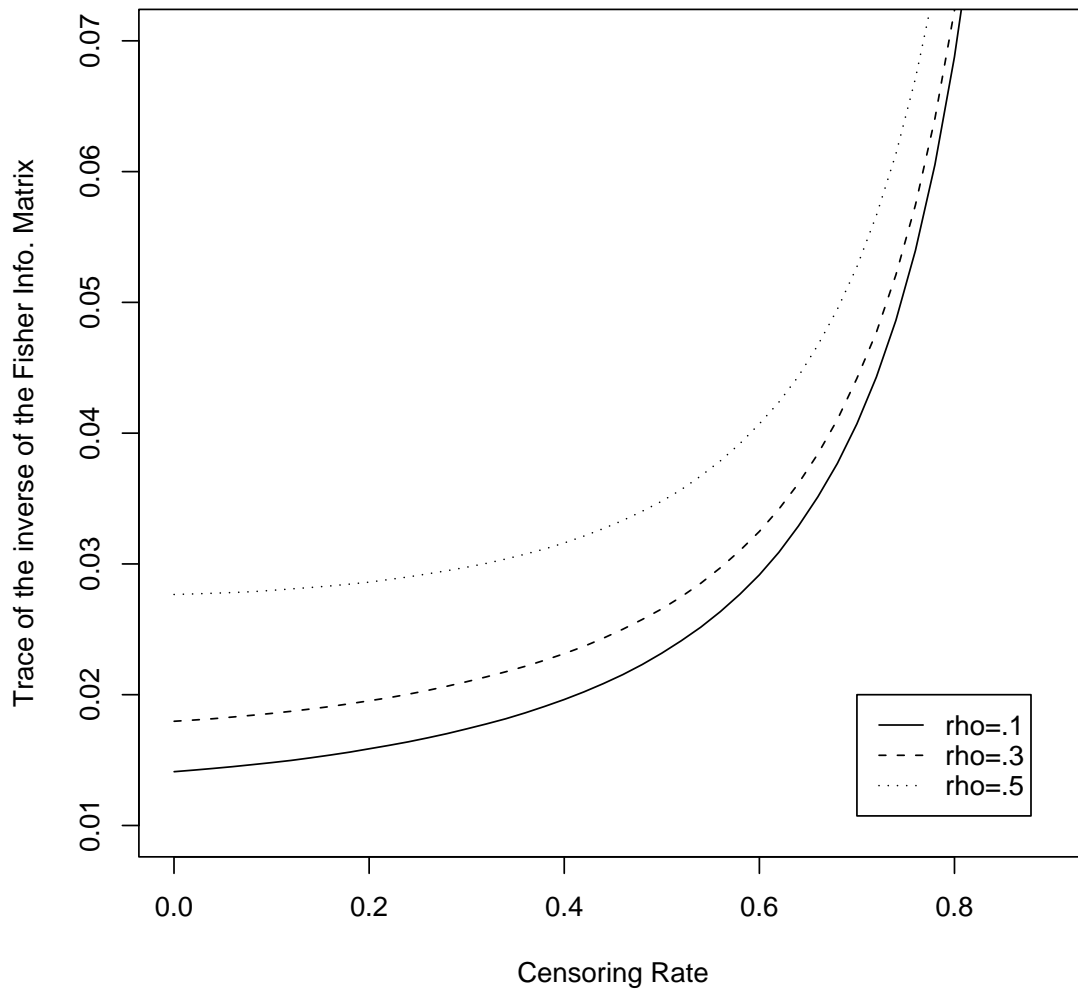


Figure 1: Plot of the median of the trace of the inverse of the observed Fisher information matrix by censoring rate from 0% to 80% of censoring.  $\rho = 0.1, 0.3, 0.5$ ,  $\mu = 0$ , and  $\sigma = 1$  were used for the simulation with 500 replications.

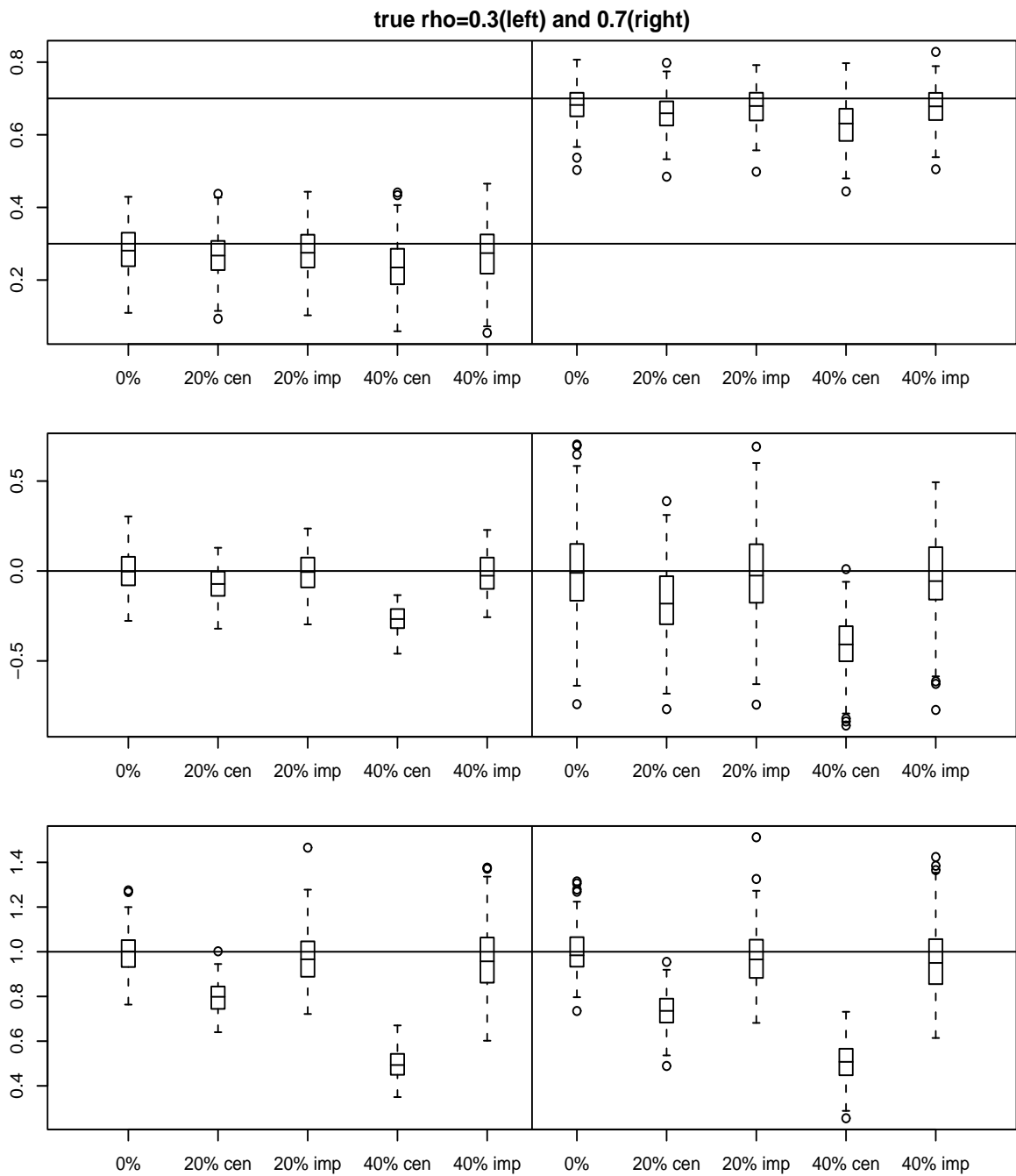


Figure 2: Box-plots of the estimation for  $\rho$  (top),  $\mu$  (middle), and  $\sigma$  (bottom). The reference lines in plots are the true values of the parameter. The index ‘cen’ means the results from the naive approach and ‘imp’ represents the results from the imputation method.

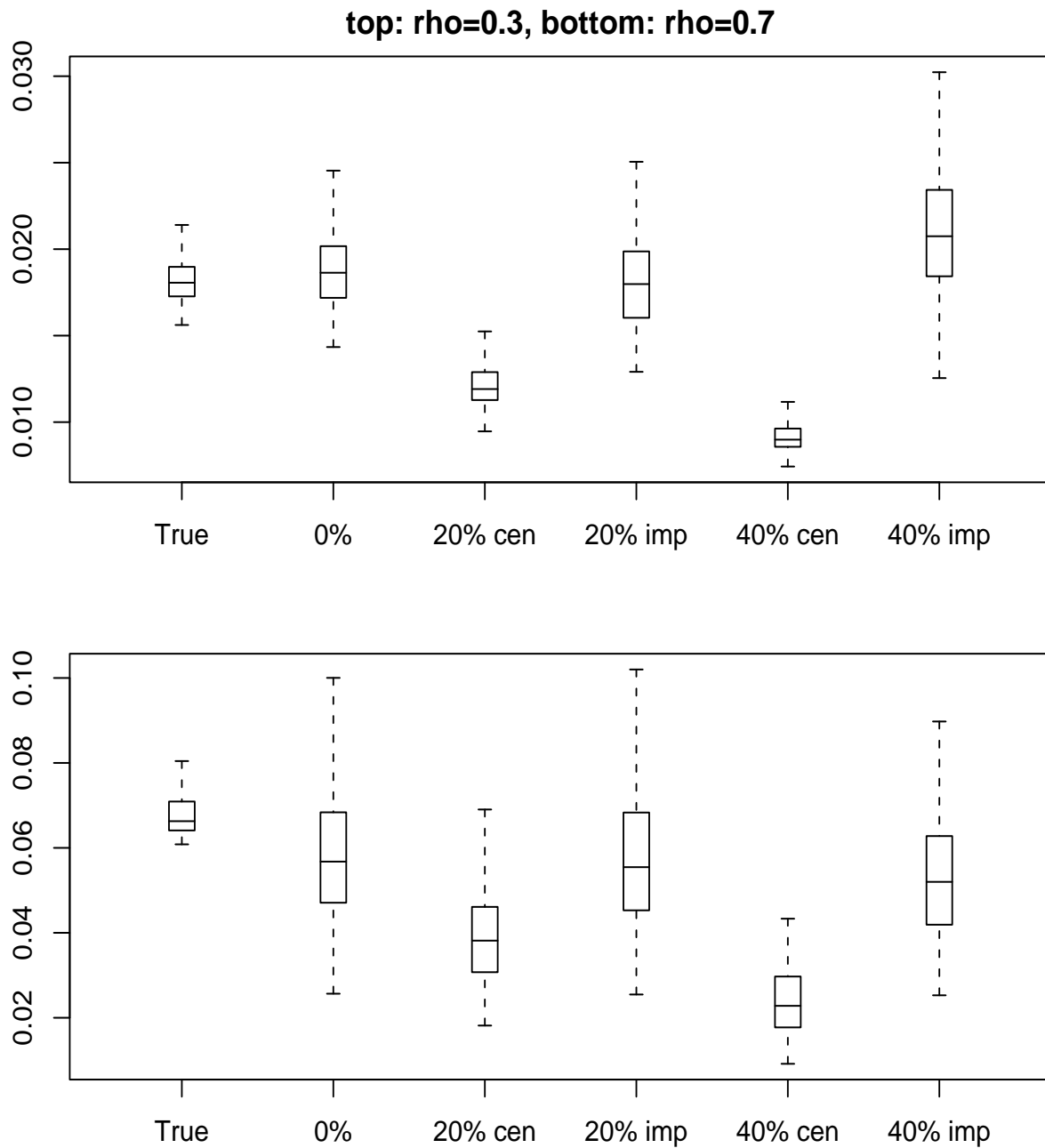


Figure 3: Box-plots of the trace of the inverse of the observed Fisher information matrix and the plug-in version of the observed Fisher information matrix by censoring rate. The far left box-plots show the trace of the inverse of the observed Fisher information matrix evaluated with the true parameter.

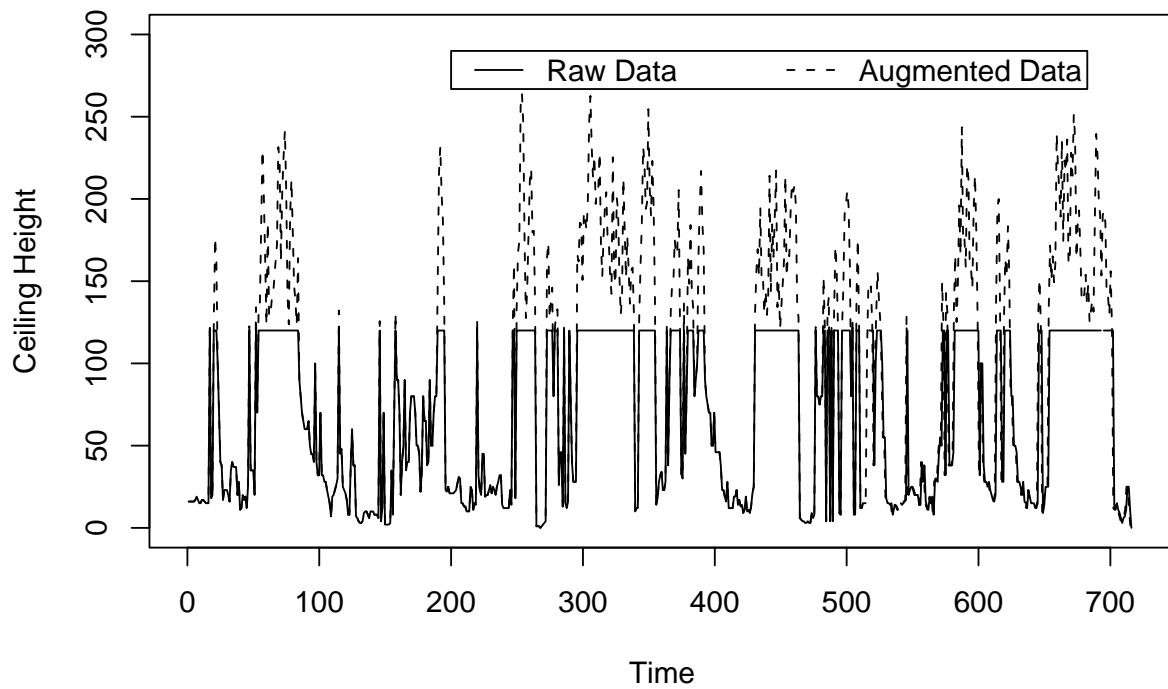


Figure 4: The super-imposed time plots of the original series containing censored observations and the augmented series based on a Gaussian CENAR(2) model fit.

Table 1: Bias and standard deviation (in the parenthesis) from the simulation results for a Gaussian CENAR(1). The results ‘naive’ are based on treating the censored values as observed and the results ‘imputed’ are based on the imputation method. The column ‘0%’ is for all observed data (complete-case analysis). The columns of ‘20%’ and ‘40%’ correspond to the censoring rate. The autocorrelations used for the simulation are 0.3 (top) and 0.7 (bottom).

		$\rho = .3$		
		0%	20%	40%
naive	$\rho$	-.018(.069)	-.032(.067)	-.062(.070)
	$\mu$	-.003(.107)	-.074(.093)	-.267(.070)
	$\sigma$	.002(.093)	-.203(.070)	-.504(.065)
imputed	$\rho$	-.018(.069)	-.022(.071)	-.027(.080)
	$\mu$	-.003(.107)	-.011(.107)	-.019(.106)
	$\sigma$	.002(.093)	-.022(.117)	-.034(.149)

		$\rho = .7$		
		0%	20%	40%
naive	$\rho$	-.018(.052)	-.043(.054)	-.075(.058)
	$\mu$	-.001(.251)	-.170(.204)	-.407(.159)
	$\sigma$	.003(.098)	-.266(.088)	-.493(.092)
imputed	$\rho$	-.018(.052)	-.023(.053)	-.024(.058)
	$\mu$	-.001(.251)	-.025(.246)	-.035(.239)
	$\sigma$	.003(.098)	-.022(.127)	-.035(.159)

Table 2: Relative efficiency based on the ratio of MSE defined in (12) from the simulation.

		20%	40%
$\rho = 0.3$	$\rho$	1.00	0.80
	$\mu$	0.83	0.15
	$\sigma$	0.31	0.09
$\rho = 0.7$	$\rho$	0.71	0.42
	$\mu$	0.87	0.31
	$\sigma$	0.21	0.10



Table 3: Parameter estimates for the cloud ceiling height data. Three candidate models were used and AIC is displayed as a criteria for the model selection (bold fonts represent the “best” model).

Results based on the naive approach

	$\hat{\mu}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\psi}$	$\hat{\sigma}$	AIC
AR(1)	67.06 (5.077)	0.78 (0.024)	NA	NA	30.47	6904
AR(2)	67.06 (6.110)	0.63 (0.037)	0.18 (0.037)	NA	29.95	6881
<b>ARMA(1,1)</b>	<b>67.06</b> <b>(6.85)</b>	<b>0.89</b> <b>(0.022)</b>	NA	<b>-0.29</b> <b>(0.048)</b>	<b>29.80</b>	<b>6874</b>

Results based on the imputation method

	$\hat{\mu}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\psi}$	$\hat{\sigma}$	AIC
AR(1)	93.76 (11.44)	0.86 (0.019)	NA	NA	42.39	7405
<b>AR(2)</b>	<b>83.33</b> <b>(10.57)</b>	<b>0.68</b> <b>(0.037)</b>	<b>0.19</b> <b>(0.037)</b>	NA	<b>37.07</b>	<b>7215</b>
ARMA(1,1)	93.13 (13.29)	0.90 (0.019)	NA	-0.19 (0.044)	42.05	7395