COMMUNICATIONS IN MATHEMATICAL RESEARCH
**29**(3)(2013), 271–279

# Efficient Mean Estimation in Log-normal Linear Models with First-order Correlated Errors

ZHANG SONG[1,2] AND WANG DE-HUI[2,*]

(*1. Dalian Commodity Exchange, Dalian, Liaoning, 116023*)

(*2. School of Mathematics, Jilin University, Changchun, 130012*)

**Abstract:** In this paper, we propose a log-normal linear model whose errors are first-order correlated, and suggest a two-stage method for the efficient estimation of the conditional mean of the response variable at the original scale. We obtain two estimators which minimize the asymptotic mean squared error (MM) and the asymptotic bias (MB), respectively. Both the estimators are very easy to implement, and simulation studies show that they are perform better.

**Key words:** log-normal, first-order correlated, maximum likelihood, two-stage estimation, mean squared error

**2000 MR subject classification:** 62J12, 62M10, 62F10

**Document code:** A

**Article ID:** 1674-5647(2013)03-0271-09

## 1 Introduction

Log-normality is widely found in many fields from biology, medicine, insurance (see [1–3]), to geology, hydrology, environmentalology (see [4–6]), and so on. In these fields, researchers discover that linear models are often fitted to the logarithmic transformed response variables very well, and these are the ordinary log-normal linear models, whose errors are independently and identically subject to $N(0,\ \sigma^2)$. The efficient mean estimation in the ordinary log-normal linear models has been considered by numbers of authors in the literature. Bradu and Mundlak[7] derived the uniformly minimum variance unbiased (UMVU) estimator and its variance. The maximum likelihood (ML) estimator and the restricted maximum likelihood (REML) estimator have also been used frequently in practice. A general discussion can be found in [8]. Though the UMVU estimator has the smallest mean squared error (MSE) among all unbiased estimators, it may not have a smaller MSE than a biased estimator.

**Received date:** Nov. 22, 2011.
**Foundation item:** The NSF (11271155) of China and Research Fund (20070183023) for the Doctoral Program of Higher Education.
* **Corresponding author.**
**E-mail address:** sngchng@gmail.com (Zhang S), wangdh@jlu.edu.cn (Wang D H).

Zhou[9] showed the fact that a biased conditionally minimal MSE estimator had smaller MSE than the UMVU estimator. El-shaarawi and Viveros[10] proposed a bias-corrected REML estimator, which was termed the EV estimator. More recently, Shen and Zhu[11] developed two estimators which minimize the asymptotic MSE and the asymptotic bias, respectively.

The ordinary log-normal linear models assume that the errors are i.i.d. However, in many practical cases, because of the time or spacial continuity of the response variables, the errors are correlated, which violates the i.i.d. assumption. If people ignore the violation and stick to use the ordinary log-normal linear models, it would result in large bias, and even wrong inference. Suppose that $\boldsymbol{Z} = (Z_1, \cdots, Z_n)^{\mathrm{T}}$ is the response vector, and $\boldsymbol{x}_i = (1, x_{i1}, \cdots, x_{ip})^{\mathrm{T}}$ is the covariate vector for observation $i$. As first-order correlation is the most common phenomena, we propose a log-normal linear model with first-order correlated errors as follows:

$$\boldsymbol{Y} = \log(\boldsymbol{Z}) = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{X} = (x_1, \cdots, x_n)^{\mathrm{T}}, \qquad \boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^{\mathrm{T}}, \qquad \boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^{\mathrm{T}}$$

with

$$\varepsilon_i = \rho\eta_{i-1} + \eta_i, \qquad \eta_0 = 0,\ \eta_i \overset{\text{i.i.d}}{\sim} N(0, \sigma^2),\ i = 1, \cdots, n,\ |\rho| < 1.$$

Then

$$\varepsilon_i \sim N(0, \sigma^2(1+\rho^2)), \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\boldsymbol{\Sigma}), \qquad \boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1+\rho^2 & \rho & \cdots & 0 \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ 0 & \cdots & \rho & 1+\rho^2 \end{pmatrix}.$$

Apparently, if $\rho = 0$, the model degenerates into the ordinary log-normal linear model.

In this paper, we focus on the efficient estimation of the conditional mean of $Z_0$ given $\boldsymbol{x}_0$,

$$\mu(\boldsymbol{x}_0) = E(Z_0 \mid \boldsymbol{x}_0) = \exp\left\{\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{\beta} + \frac{\sigma^2(1+\rho^2)}{2}\right\},$$

where $\boldsymbol{x}_0$ is a new set of covariate values,

$$Z_0 = \exp\{\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_0\}$$

is the response variable at the original scale and $\varepsilon_0$ is the normal error with mean zero and variance $\sigma^2(1+\rho^2)$. In Section 2, we derive the estimators of $\mu(\boldsymbol{x}_0)$ and their MSE and bias when $\rho$ is known. In Section 3, we suggest a moment method to estimate $\rho$ and present its iterative algorithm, and thus, the estimators of $\mu(\boldsymbol{x}_0)$ when $\rho$ is unknown are obtained. In Section 4, we compare the MSE and bias of the estimators by simulation studies.