Improved maximum likelihood estimators in a heteroskedastic errors-in-variables model

Alexandre G. Patriota · Artur J. Lemonte · Heleno Bolfarine

Received: 17 March 2009 / Revised: 15 June 2009 / Published online: 1 July 2009
© Springer-Verlag 2009

Abstract This paper develops a bias correction scheme for a multivariate heteroskedastic errors-in-variables model. The applicability of this model is justified in areas such as astrophysics, epidemiology and analytical chemistry, where the variables are subject to measurement errors and the variances vary with the observations. We conduct Monte Carlo simulations to investigate the performance of the corrected estimators. The numerical results show that the bias correction scheme yields nearly unbiased estimates. We also give an application to a real data set.

Keywords Bias correction · Errors-in-variables model · Heteroskedastic model · Maximum-likelihood estimation

1 Introduction

Heteroskedastic errors-in-variables (or measurement error) models have been extensively studied in the statistical literature and widely applied in astrophysics (to explain relationships between black hole masses and some variates of luminosities), epidemiology (to model the cardiovascular event with its risk factors), analytical chemistry (to compare different types of measurement instruments). The applicability of this model abound mainly in the astronomy literature where all quantities are subject to measurement errors (Akritas and Bershady 1996).

It is well-known that, when the measurement errors are ignored in the estimation process, the maximum-likelihood estimators (MLEs) become inconsistent. More specifically, the estimation of the slope parameter of a simple linear model is attenuated...
(Fuller 1987). When variables are subject to measurement errors, a special inference treatment must be carried out for the model parameters in order to avoid inconsistent estimators. Usually, a measurement equation is added to the model to capture the measurement error effect and then the MLEs from this approach are consistent, efficient and asymptotically normally distributed. A careful and deep exposition on the inferential process in errors-in-variables models can be seen in Fuller (1987) and the references therein.

Although consistent, asymptotically efficient and asymptotically normally distributed, the MLEs are oftentimes biased and point inference can be misleading. This is not a serious problem for relatively large sample sizes, since bias is typically of order $O(n^{-1})$, while the asymptotic standard errors are of order $O(n^{-1/2})$. However, for small or even moderate values of the sample size $n$, bias can constitute a problem. Bias adjustment has been extensively studied in the statistical literature. For example, Cook et al. (1986), Cordeiro (1993), Cordeiro and Vasconcellos (1997), Vasconcellos and Cordeiro (1997) and, more recently, Cordeiro (2008). Additionally, Patriota and Lemonte (2009) obtained general matrix formulae for the second-order biases of the maximum-likelihood estimators in a very general model which includes all previous works aforementioned. The model presented by the authors considers that the mean vector and the variance–covariance matrix of the observed variable have parameters in common. This approach includes the heteroskedastic measurement error model that we are going to study in this paper.

The main goal of this article is to define bias-corrected estimators using the general second-order bias expression derived in Patriota and Lemonte (2009) assuming that the model defined by (1) and (2) holds. Additionally, we compare the performance of bias-corrected estimators with the MLEs in small samples via Monte Carlo simulations. The numerical results show that the bias correction is effective in small samples and leads to estimates that are nearly unbiased and display superior finite-sample behavior.

The rest of the paper is as follows. Section 2 presents the multivariate heteroskedastic errors-in-variables model. Using general results from Patriota and Lemonte (2009), we derive in Sect. 3 the second-order biases of the MLEs of the parameters. The result is used to define bias-corrected estimates. In Sect. 4 the $O(n^{-1})$ biases of the estimates $\hat{\mu}_i$ and $\hat{\Sigma}_i$ are given. Monte Carlo simulation results are presented and discussed in Sect. 5. Section 6 gives an application. Finally, concluding remarks are offered in Sect. 7.

2 The model

The multivariate model assumed throughout this paper is

$$y_i = \beta_0 + \beta_1 x_i + q_i, \quad i = 1, \ldots, n,$$

where $y_i$ is a $(v \times 1)$ latent response vector, $x_i$ is a $(m \times 1)$ latent vector of covariates, $\beta_0$ is a $(v \times 1)$ vector of intercepts, $\beta_1$ is a $(v \times m)$ matrix, the elements of which are inclinations and $q_i$ is the equation error having a multivariate normal