A matrix formula for the skewness of maximum likelihood estimators

Alexandre G. Patriota[a,*], Gauss M. Cordeiro[b]

[a] *Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010, São Paulo/SP, 05508-090, Brazil*
[b] *Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Brazil*

Abstract

We give a general matrix formula for computing the second-order skewness of maximum likelihood estimators. The formula was firstly presented in a tensorial version by Bowman and Shenton (1998). Our matrix formulation has numerical advantages, since it requires only simple operations on matrices and vectors. We apply the second-order skewness formula to a normal model with a generalized parametrization and to an ARMA model.

Key words: Asymptotic expansion, matrix operation, maximum likelihood estimator, skewness.

1. Introduction

A distribution is usually characterized by the mean, variance, skewness and kurtosis. The mean and variance represent the location and dispersion, respectively, whereas the skewness and kurtosis refer to the shape of the distribution. The latter two measures can be used to verify a departure from normality of the variable being analyzed. In this article, we shall concentrate only on the skewness measure in the context of maximum likelihood estimation. The most commonly used measure of skewness is the standardized third cumulant defined by $\gamma = \kappa_3/\kappa_2^{3/2}$, where κ_r is the rth cumulant of the distribution. When $\gamma > 0$ ($\gamma < 0$) the distribution has a positive (negative) skew. It is well known that $\gamma = 0$ for all symmetric distributions (e.g., normal, Student-t, power exponential, type I and type II logistics and so forth).

Under standard regular conditions, the maximum likelihood estimators (MLEs) are asymptotically normally distributed, and then, asymptotically, their skewness are equal to zero. However, for finite (small) sample sizes, the exact distribution of the MLEs may be very different from the normal one and, under this context, we can look at the skewness of the MLE to verify the departure from normality. The far away from zero is the skewness estimate, the farther might be the exact distribution of the MLEs from the normal distribution. Provided that the MLEs are asymptotically normally distributed, a first-order approximation for the skewness of the MLE is zero. Nevertheless, this gives us no information about the symmetry of this estimator under finite sample sizes. In view of that, Bowman and Shenton (1998) computed the second-order skewness of the MLE which can be seen as a more accurate estimate of its exact skewness. In addition, this skewness estimate can be used as a guide for computing the sample size as mentioned by Bowman and Shenton (2005). We just fix a value for γ, say $\gamma = 0.1$, and choose a sample size for which this skewness holds. Naturally, we have to take a previous sample to estimate unknown parameters in γ.

*Corresponding author
Email addresses: patriota.alexandre@gmail.com (Alexandre G. Patriota), gauss@deinfo.ufrpe.br (Gauss M. Cordeiro)

Let $\ell(\boldsymbol{\theta})$ be the total log-likelihood for some $p \times 1$ vector $\boldsymbol{\theta}$ of unknown parameters and let $\widehat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$. We assume that $\ell(\boldsymbol{\theta})$ is regular with respect to all $\boldsymbol{\theta}$ derivatives up to and including those of third order. We shall use the tensorial notation for joint cumulants of log-likelihood derivatives: $\kappa_{r,s} = E\{\partial\ell(\boldsymbol{\theta})/\partial\theta_r \partial\ell(\boldsymbol{\theta})/\partial\theta_s\}$, $\kappa_{rs} = E\{\partial^2\ell(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s\}$, $\kappa_{rst} = E\{\partial^3\ell(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s\partial\theta_t\}$ $\kappa_{rs,t} = E\{\partial^2\ell(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s \partial\ell(\boldsymbol{\theta})/\partial\theta_t\}$ and $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\theta_t$. All κ's are assumed to be of order $\mathcal{O}(n)$, where n is the sample size. These cumulants satisfy certain equations called Bartlett regularity conditions such as $\kappa_{r,s} = -\kappa_{rs}$, $\kappa_{r,st} = \kappa_{rst} - \kappa_{rs}^{(t)}$, $\kappa_{r,s,t} = -\kappa_{rst} - \kappa_{r,st} - \kappa_{s,rt} - \kappa_{t,rs}$ which usually facilitate their calculations.

The total expected information matrix is given by $\boldsymbol{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \{-\kappa_{rs}\}$. Let $\kappa^{r,s}$ be the corresponding (r,s)th element of the inverse information matrix, say $\boldsymbol{K}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} = \{\kappa^{r,s}\}$. The third central moment of the MLE $\widehat{\theta}_a$ is $\kappa_3(\widehat{\theta}_a) = E[(\widehat{\theta}_a - \theta_a)^3]$ for $a = 1, \ldots, p$. Using Taylor series expansion, Bowman and Shenton (1998) derived an approximation of order $\mathcal{O}(n^{-2})$ for this third central moment given by

$$\kappa_3(\widehat{\theta}_a) = \sum_{r,s,t=1}^{p} \kappa^{a,r}\kappa^{a,s}\kappa^{a,t} m_{rs}^{(t)}, \tag{1}$$

where $m_{rs}^{(t)} = 5\kappa_{rs}^{(t)} - (\kappa_{st}^{(r)} + \kappa_{rt}^{(s)} + \kappa_{rst})$. Hence, the second-order skewness of $\widehat{\theta}_a$ can be written as $\gamma(\widehat{\theta}_a) = \kappa_3(\widehat{\theta}_a)/(\kappa^{a,a})^{3/2}$ for $a = 1, \ldots, p$. The quantities $\kappa_3(\widehat{\theta}_a)$ and $(\kappa^{a,a})^{3/2}$ are of orders $\mathcal{O}(n^{-2})$ and $\mathcal{O}(n^{-3/2})$, respectively, and then the standardized third cumulant $\gamma(\widehat{\theta}_a)$ is of order $\mathcal{O}(n^{-1/2})$.

There are some recent works in the statistical literature regarding the skewness of the MLEs. Cysneiros et al. (2001) computed the second-order skewness and kurtosis for one-parameter exponential family. Cordeiro and Cordeiro (2001) applied formula (1) to obtain a matrix expression for the second-order skewness of the MLEs of the location and dispersion parameters in generalized linear models. Bowman and Shenton (2005) implemented a MAPLE script to compute the derivatives involved in (1) for two-parameter gamma and three-parameter Poisson mixture distributions. More recently, Cavalcanti et al. (2009) studied the second-order skewness of the MLEs in exponential family nonlinear models.

In this article, we obtain a matrix formula for the tensorial equation (1) under a general framework. The main result is derived in Section 2. In Section 3, we apply the proposed matrix formula to a multivariate normal nonlinear model following a generalized parametrization. Section 4 is devoted to the skewness of the MLEs of the parameters in an ARMA model. Section 5 provides some simulation results. Section 6 presents an application and a R script algorithm (R Development Core Team, 2007) to calculate the skewness. Finally, we offer concluding remarks in Section 7.

2. General matrix formula

Here, we develop from (1) a matrix formula for the third central moment of the MLE under any regular statistical model. The following matrix equations will be intensively used in the algebraic development discussed in the article. Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ be general matrices of appropriate dimensions. We have

$$\text{tr}(\boldsymbol{AB}) = \text{vec}(\boldsymbol{A}^\top)^\top\text{vec}(\boldsymbol{B}),\ \text{vec}(\boldsymbol{AB}) = (\boldsymbol{I}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{B}) = (\boldsymbol{B}^\top\otimes\boldsymbol{I})\text{vec}(\boldsymbol{A}) \tag{2}$$

and

$$\boldsymbol{A}^\top\boldsymbol{C}\boldsymbol{D} = \{\boldsymbol{a}_r^\top\boldsymbol{C}\boldsymbol{d}_s\},\ \text{tr}\{\boldsymbol{A}^\top\boldsymbol{C}\boldsymbol{D}\boldsymbol{B}^\top\} = \text{vec}(\boldsymbol{A})^\top(\boldsymbol{B}\otimes\boldsymbol{C})\text{vec}(\boldsymbol{D}),\ \text{vec}(\boldsymbol{ACB}) = (\boldsymbol{B}^\top\otimes\boldsymbol{A})\text{vec}(\boldsymbol{C}), \tag{3}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, $\boldsymbol{D} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_m)$, $\text{vec}(\cdot)$ is the vec operator, which transforms a matrix into a vector by stacking the columns of the matrix one underneath the other and

"⊗" denotes the Kronecker product. These results and other methods in matrix differential calculus can be studied in Magnus and Neudecker (2007). To simplify the presentation of the matrix formula, we consider the following matrix operation: if $\boldsymbol{E} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p)$ is a $p \times p$ matrix, where $\boldsymbol{e}_j$ is a $p \times 1$ vector, we define the $p^2 \times p^2$ matrix $Q(\boldsymbol{E}) = \text{block–diag}(\boldsymbol{e}_1^\top, \ldots, \boldsymbol{e}_p^\top)$. We also write the inverse of the information matrix as $\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1} = (\boldsymbol{\kappa}^{(1)}, \ldots, \boldsymbol{\kappa}^{(p)})$, where $\boldsymbol{\kappa}^{(a)} = (\kappa^{1,a}, \ldots, \kappa^{p,a})^\top$. Further, we define the $p \times p^2$ matrix $\boldsymbol{M} = (\boldsymbol{M}^{(1)} \ldots \boldsymbol{M}^{(p)})$ based on the kernel quantity $m_{rs}^{(t)}$ in equation (1), namely

$$\boldsymbol{M}^{(t)} = \begin{pmatrix} m_{11}^{(t)} & \ldots & m_{1p}^{(t)} \\ \vdots & \ddots & \vdots \\ m_{p1}^{(t)} & \ldots & m_{pp}^{(t)} \end{pmatrix}.$$

First, we note that

$$\sum_{r,s,t=1}^{p} \kappa^{a,r}\kappa^{a,s}\kappa^{a,t} m_{rs}^{(t)} = \boldsymbol{\kappa}^{(a)\top}(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p)(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_{p^2})\text{vec}(\boldsymbol{M}), \tag{4}$$

where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. The matrix $\boldsymbol{M}$ may be defined in different ways. For example, $\boldsymbol{M} = (\boldsymbol{M}_{(1)}, \ldots, \boldsymbol{M}_{(p)})$ and $\boldsymbol{M} = (\boldsymbol{M}_{(1)}^*, \ldots, \boldsymbol{M}_{(p)}^*)$, where

$$\boldsymbol{M}_{(r)} = \begin{pmatrix} m_{r1}^{(1)} & \ldots & m_{r1}^{(p)} \\ \vdots & \ddots & \vdots \\ m_{rp}^{(1)} & \ldots & m_{rp}^{(p)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{M}_{(s)}^* = \begin{pmatrix} m_{1s}^{(1)} & \ldots & m_{1s}^{(p)} \\ \vdots & \ddots & \vdots \\ m_{ps}^{(1)} & \ldots & m_{ps}^{(p)} \end{pmatrix},$$

generate the same equation (4).

Using the identities $\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}) = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C}$ and $\boldsymbol{I}_{p^2} = \boldsymbol{I}_p \otimes \boldsymbol{I}_p$, we obtain

$$(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p)(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_{p^2}) = (\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p)((\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p) \otimes \boldsymbol{I}_p) = (\boldsymbol{\kappa}^{(a)\top}(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p)) \otimes \boldsymbol{I}_p.$$

Hence,

$$\sum_{r,s,t=1}^{p} \kappa^{a,r}\kappa^{a,s}\kappa^{a,t} m_{rs}^{(t)} = \boldsymbol{\kappa}^{(a)\top}[(\boldsymbol{\kappa}^{(a)\top}(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p)) \otimes \boldsymbol{I}_p]\text{vec}(\boldsymbol{M}) = \text{vec}(\boldsymbol{M})^\top \left[\left((\boldsymbol{\kappa}^{(a)} \otimes \boldsymbol{I}_p)\boldsymbol{\kappa}^{(a)}\right) \otimes \boldsymbol{I}_p\right]\boldsymbol{\kappa}^{(a)}.$$

By application of (3), we have

$$\sum_{r,s,t=1}^{p} \kappa^{a,r}\kappa^{a,s}\kappa^{a,t} m_{rs}^{(t)} = \text{tr}\{\boldsymbol{M}^\top \boldsymbol{\kappa}^{(a)}\boldsymbol{\kappa}^{(a)\top}(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p)\} = \boldsymbol{\kappa}^{(a)\top}(\boldsymbol{\kappa}^{(a)\top} \otimes \boldsymbol{I}_p)\boldsymbol{M}^\top \boldsymbol{\kappa}^{(a)}.$$

We can express the vector $\boldsymbol{\kappa}_3(\widehat{\boldsymbol{\theta}}) = (\kappa_3(\widehat{\theta}_1), \ldots, \kappa_3(\widehat{\theta}_p))^\top$ as

$$\boldsymbol{\kappa}_3(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \boldsymbol{\kappa}^{(1)\top}(\boldsymbol{\kappa}^{(1)\top} \otimes \boldsymbol{I}_p)\boldsymbol{M}^\top \boldsymbol{\kappa}^{(1)} \\ \vdots \\ \boldsymbol{\kappa}^{(p)\top}(\boldsymbol{\kappa}^{(p)\top} \otimes \boldsymbol{I}_p)\boldsymbol{M}^\top \boldsymbol{\kappa}^{(p)} \end{pmatrix}.$$

Then, we obtain a matrix formula for the third central moment of the MLE

$$\boldsymbol{\kappa}_3(\widehat{\boldsymbol{\theta}}) = Q(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1})\big(Q(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1}) \otimes \boldsymbol{I}_p\big)\big(\boldsymbol{I}_p \otimes \boldsymbol{M}^\top\big)\text{vec}(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1}),$$

since

$$Q(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1})\big(Q(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1})\otimes \boldsymbol{I}_p\big) = \begin{pmatrix} \boldsymbol{\kappa}^{(1)\top}(\boldsymbol{\kappa}^{(1)\top}\otimes \boldsymbol{I}_p) & \dots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{\kappa}^{(p)\top}(\boldsymbol{\kappa}^{(p)\top}\otimes \boldsymbol{I}_p) \end{pmatrix}$$

and

$$\big(\boldsymbol{I}_p \otimes \boldsymbol{M}^\top\big)\mathrm{vec}(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1}) = \begin{pmatrix} \boldsymbol{M}^\top \boldsymbol{\kappa}^{(1)} \\ \vdots \\ \boldsymbol{M}^\top \boldsymbol{\kappa}^{(p)} \end{pmatrix}.$$

Finally, using the third identity in (3), we obtain

$$\boldsymbol{\kappa}_3(\widehat{\boldsymbol{\theta}}) = Q(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1})\big(Q(\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1})\otimes \boldsymbol{I}_p\big)\mathrm{vec}(\boldsymbol{M}^\top \boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1}). \tag{5}$$

The matrix formula (5) can be used to compute the second-order skewness of the MLEs in rather general regular statistical models, since it facilitates the numerical and algebraic computations involved. It depends only on the inverse information matrix $\boldsymbol{K}_{\boldsymbol{\theta\theta}}^{-1}$ and the basic matrix $\boldsymbol{M}$. For models with closed-form information matrix, it is possible to derive the skewness of the MLE in closed-form and then we can determine which aspects of the model contribute significantly to the skewness. In practice, equation (5) can be used without much difficulty, particularly if a suitable computerized algebra system is readily available. Although the tensorial expression (1) seems algebraically more appealing than (5), the latter is better suited for algebraic and numerical purposes in rather general parametric models, because it involves only trivial operations on matrices. Its main advantage over the tensorial formula (1) is that we can avoid computations involving higher-order arrays.

Equation (5) can be easily implemented in the `R` software (R Development Core Team, 2007) by using the package `Matrix` (see Section 6.1). This package gives a suitable treatment to sparse matrices which is much faster than the usual tensorial methods. It is worth mentioned that we need only the matrices $\boldsymbol{K}_{\boldsymbol{\theta\theta}}$ and $\boldsymbol{M}$ to compute the second-order skweness. On the one hand, the tensorial formula (1) needs four loops to compute the entire vector $\boldsymbol{\kappa}_3(\widehat{\boldsymbol{\theta}})$ which has complexity of order $O(p^4)$. On the other hand, although the matrix formula (5) requires the Kronecker product, it involves sparse and symmetric matrices and this turns the method computationally faster.

3. Multivariate normal model with a generalized parametrization

Now, we apply equation (5) to a multivariate normal nonlinear model with a generalized parametrization proposed by Patriota and Lemonte (2009). Let $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ be observed independent vectors for which the number of responses measured in the ith observation is q_i. The multivariate regression model can be written as

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}) + \boldsymbol{u}_i, \quad i = 1, \dots, n, \tag{6}$$

where $\boldsymbol{u}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$ and "$\overset{\text{ind}}{\sim}$" means "independently distributed as". Consequently, we have that $\boldsymbol{Y}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$, where $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{X}_i)$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{Z}_i)$ are known functional forms for the mean and the variance-covariance matrix, respectively. They are assumed to be three times continuously differentiable with respect to each element of $\boldsymbol{\theta}$. The model (6) admits that non-stochastic auxiliary variables $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ and $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ can also be observed, where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are $m_i \times 1$ and $k_i \times 1$ vectors, respectively, of known variables associated with the ith observed response $\boldsymbol{Y}_i$ which may have equal components. In addition, $\boldsymbol{\theta}$ is a p-vector of unknown parameters of interest (where $p < n$ and it is fixed). Since $\boldsymbol{\theta}$ must be identifiable in model (6), the functions $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ are defined to accomplish such restriction. Recently, Patriota et al. (2010) present some influence assessment procedures for this model, such as the local influence, total local influence of an individual and generalized leverage.

The large class of models (6) includes many important statistical models. We can mention, for instance, linear and nonlinear regression models, either homoscedastic or heteroscedastic. Heteroscedastic structural measurement error models can also be formulated within this class. These models were studied by de Castro et al. (2008), Patriota et al. (2009) and the references therein. Structural equation models (e.g., Bollen, 1989) represent a rich class of models with latent variables that can be written as a special case of equation (6). Simultaneous equations models (e.g., Magnus and Neudecker, 2007, Ch. 16) comprise endogenous and exogenous variables and, in the reduced form, they are special sub-models of the general model (6). The model (6) is perhaps the one with the highest degree of generality that can be considered in a multivariate normal set-up with independent observed vectors and therefore our list of examples is by no means exhaustive.

In order to follow the same notation introduced by Patriota and Lemonte (2009), we take the full quantities $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \ldots, \boldsymbol{Y}_n^\top)^\top$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1(\boldsymbol{\theta})^\top, \ldots, \boldsymbol{\mu}_n(\boldsymbol{\theta})^\top)^\top$, $\boldsymbol{\Sigma} = \text{block–diag}(\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{\Sigma}_1(\boldsymbol{\theta}))$ and $\boldsymbol{u} = \boldsymbol{Y} - \boldsymbol{\mu}$. The log-likelihood function associated with model (6), apart from a constant, is

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}. \tag{7}$$

In order to guarantee the asymptotic properties of the MLE $\widehat{\boldsymbol{\theta}}$, such as consistency, sufficiency and normality, we assume that the usual regularity conditions on $\ell(\boldsymbol{\theta})$ hold and also that the parameter $\boldsymbol{\theta}$ is an interior point of $\boldsymbol{\Theta}$ which is an open subset of $\mathbb{R}^p$. The quantities $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ are defined in such a way that the log-likelihood (7) becomes a regular function. In general, we assume the regularity conditions stated in Cox and Hinkley (1974, Ch. 9) on the behavior of $\ell(\boldsymbol{\theta})$ as $n \to \infty$.

We define the following quantities $(r, s, t = 1, 2, \ldots, p)$:

$$\boldsymbol{a}_r = \frac{\partial\boldsymbol{\mu}}{\partial\theta_r},\ \boldsymbol{a}_{sr} = \frac{\partial^2\boldsymbol{\mu}}{\partial\theta_s\partial\theta_r},\ \boldsymbol{C}_r = \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_r},\ \boldsymbol{C}_{sr} = \frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_s\partial\theta_r} \text{ and } \boldsymbol{A}_r = \frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta_r} = -\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_r\boldsymbol{\Sigma}^{-1}.$$

Let

$$\boldsymbol{F} = \begin{pmatrix}\boldsymbol{D}\\ \boldsymbol{V}\end{pmatrix},\quad \boldsymbol{H} = \begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & 2(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\end{bmatrix}^{-1} \quad\text{and}\quad \boldsymbol{u}^* = \begin{bmatrix}\boldsymbol{u}\\ -\text{vec}(\boldsymbol{\Sigma} - \boldsymbol{u}\boldsymbol{u}^\top)\end{bmatrix}, \tag{8}$$

where $\boldsymbol{D} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$ and $\boldsymbol{V} = (\text{vec}(\boldsymbol{C}_1), \ldots, \text{vec}(\boldsymbol{C}_p))$. Here, $\boldsymbol{F}$ is assumed to have full rank p. To compute the derivatives of $\ell(\boldsymbol{\theta})$, we can make use of matrix differential calculus methods (Magnus and Neudecker, 2007). The score function and the expected information are $\boldsymbol{U}_{\boldsymbol{\theta}} = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{u}^*$ and $\boldsymbol{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F}$, respectively. An iterative algorithm to calculate the MLE is given by

$$(\boldsymbol{F}^{(m)\top}\boldsymbol{H}^{(m)}\boldsymbol{F}^{(m)})\boldsymbol{\theta}^{(m+1)} = \boldsymbol{F}^{(m)\top}\boldsymbol{H}^{(m)}\boldsymbol{v}^{(m)},\quad m = 0, 1, \ldots, \tag{9}$$

where $\boldsymbol{v}^{(m)} = \boldsymbol{F}^{(m)}\boldsymbol{\theta}^{(m)} + \boldsymbol{u}^{*(m)}$ and m is the iteration counter. The cycle scheme (9) is an iterative re-weighted least squares algorithm and the iterations continue until convergence is achieved. Sometimes this algorithm does not converge, neither find the actual maximum of the likelihood function nor a relative maximum point which is an interior point of a restricted parametric space. In these cases, other numerical methods can be used such as the Gauss-Newton and Quasi-Newton methods.

For the general model (6), Patriota and Lemonte (2009) obtained the cumulants

$$\begin{aligned}\kappa_{tsr} &= \mathrm{tr}\{(\boldsymbol{A}_r\boldsymbol{\Sigma}\boldsymbol{A}_s + \boldsymbol{A}_s\boldsymbol{\Sigma}\boldsymbol{A}_r)\boldsymbol{C}_t\} + \frac{1}{2}\mathrm{tr}\{\boldsymbol{A}_s\boldsymbol{C}_{tr} + \boldsymbol{A}_r\boldsymbol{C}_{ts} + \boldsymbol{A}_t\boldsymbol{C}_{sr}\} \\ &\quad - (\boldsymbol{a}_t^\top\boldsymbol{A}_s\boldsymbol{a}_r + \boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_r + \boldsymbol{a}_s^\top\boldsymbol{A}_r\boldsymbol{a}_t + \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{sr} + \boldsymbol{a}_{ts}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_r + \boldsymbol{a}_s^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{tr})\end{aligned} \tag{10}$$

and

$$\begin{aligned}\kappa_{ts}^{(r)} &= \frac{1}{2}\mathrm{tr}\{(\boldsymbol{A}_r\boldsymbol{\Sigma}\boldsymbol{A}_s + \boldsymbol{A}_s\boldsymbol{\Sigma}\boldsymbol{A}_r)\boldsymbol{C}_t + \boldsymbol{A}_t\boldsymbol{C}_{rs} + \boldsymbol{A}_s\boldsymbol{C}_{rt}\} \\ &\quad - (\boldsymbol{a}_{rt}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_s + \boldsymbol{a}_t^\top\boldsymbol{A}_r\boldsymbol{a}_s + \boldsymbol{a}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_{rs}).\end{aligned} \tag{11}$$

Then, we have

$$\begin{aligned}m_{rs}^{(t)} &= 2\boldsymbol{a}_s^\top\boldsymbol{A}_r\boldsymbol{a}_t + 2\boldsymbol{a}_r^\top\boldsymbol{A}_s\boldsymbol{a}_t + \mathrm{tr}\{\boldsymbol{A}_t\boldsymbol{\Sigma}\boldsymbol{A}_s\boldsymbol{C}_r\} - 4\boldsymbol{a}_s^\top\boldsymbol{A}_t\boldsymbol{a}_r \\ &\quad - 3\boldsymbol{a}_{st}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_r - 3\boldsymbol{a}_{rt}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_s + \frac{3}{2}\mathrm{tr}\{\boldsymbol{A}_r\boldsymbol{C}_{st} + \boldsymbol{A}_s\boldsymbol{C}_{rt}\} \\ &\quad + 3\boldsymbol{a}_{rs}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_t - \frac{3}{2}\mathrm{tr}\{\boldsymbol{A}_t\boldsymbol{C}_{rs}\}\end{aligned}$$

and by using the identities in (2) systematically, the matrix $\boldsymbol{M}^{(t)}$ for this general model can be written as

$$\boldsymbol{M}^{(t)} = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{O}_t\boldsymbol{H}\boldsymbol{F} - 3\left(\boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F}_t + \boldsymbol{F}_t^\top\boldsymbol{H}\boldsymbol{F} - [\boldsymbol{v}_t^\top\boldsymbol{H}][\frac{\partial\boldsymbol{F}}{\partial\boldsymbol{\theta}}]\right),$$

where

$$\boldsymbol{O}_t = 4\begin{pmatrix}\boldsymbol{C}_t & -\boldsymbol{a}_t^\top\otimes\boldsymbol{\Sigma} \\ -\boldsymbol{a}_t\otimes\boldsymbol{\Sigma} & \boldsymbol{C}_t\otimes\boldsymbol{\Sigma}\end{pmatrix}, \quad \boldsymbol{F}_t = \frac{\partial\boldsymbol{F}}{\partial\theta_t} = \begin{pmatrix}\boldsymbol{D}_t \\ \boldsymbol{V}_t\end{pmatrix},$$

$\boldsymbol{D}_t = (\boldsymbol{a}_{1t}, \ldots, \boldsymbol{a}_{pt})$, $\boldsymbol{V}_t = (\mathrm{vec}(\boldsymbol{C}_{1t}), \ldots, \mathrm{vec}(\boldsymbol{C}_{pt}))$, $\boldsymbol{v}_t = (\boldsymbol{a}_t^\top, \mathrm{vec}(\boldsymbol{C}_t)^\top)^\top$ and $N = \sum_i q_i$. Here, $\partial\boldsymbol{F}/\partial\boldsymbol{\theta}$ is an array of dimension $N(N+1)\times p\times p$ and $[\cdot][\cdot]$ represents the bracket product of a matrix by an array as defined by Wei (1998, p. 188). The bracket product can also be written as

$$[\boldsymbol{v}_t^\top\boldsymbol{H}][\frac{\partial\boldsymbol{F}}{\partial\boldsymbol{\theta}}] = \big(\boldsymbol{I}_p\otimes(\boldsymbol{v}_t^\top\boldsymbol{H})\big)\boldsymbol{G} \tag{12}$$

where

$$\boldsymbol{G} = \begin{pmatrix}\boldsymbol{G}_{11} & \ldots & \boldsymbol{G}_{1p} \\ \vdots & \ldots & \vdots \\ \boldsymbol{G}_{p1} & \ldots & \boldsymbol{G}_{pp}\end{pmatrix}$$

with $\boldsymbol{G}_{rs} = (\boldsymbol{a}_{rs}^\top, \mathrm{vec}(\boldsymbol{C}_{rs})^\top)^\top$.

Hence, we have all ingredients for computing the second-order skewness

$$\boldsymbol{\kappa}_3(\widehat{\boldsymbol{\theta}}) = Q\big((\boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F})^{-1}\big)\big(Q\big((\boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F})^{-1}\big)\otimes\boldsymbol{I}_p\big)\mathrm{vec}(\boldsymbol{M}^\top(\boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F})^{-1}).$$

For models with closed-form information matrix $\boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F}$, it is possible to derive closed-form expressions for the skewness of the estimate $\widehat{\boldsymbol{\theta}}$. If the second derivatives of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are equal to zero (i.e., $\boldsymbol{F}_t = \boldsymbol{0}$ for all $t = 1, \ldots, p$), the matrix $\boldsymbol{M}^{(t)}$ reduces to $\boldsymbol{M}^{(t)} = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{O}_t\boldsymbol{H}\boldsymbol{F}$.

4. ARMA model

We consider an ARMA (p, q) model defined by

$$y_i = \alpha_1 y_{i-1} + \ldots + \alpha_{p_1} y_{i-p_1} + u_i - \beta_1 u_{i-1} - \ldots - \beta_{p_2} u_{i-p_2}, \tag{13}$$

where the $u_i's$ are independent random variables with mean zero and variance σ^2 and $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ is the observed time series of length n. Let $\boldsymbol{\tau} = (\alpha_1, \ldots, \alpha_{p_1}, \beta_1, \ldots, \beta_{p_2})^\top$ be the $b = p_1 + p_2$ vector of linear parameters. We have $p = b + 1$ unknown parameters to be estimated, namely those parameters in $\boldsymbol{\tau}$ and σ^2. The log-likelihood $\ell(\boldsymbol{\theta})$ for $\boldsymbol{\theta} = (\boldsymbol{\tau}^\top, \sigma^2)^\top$ given $\boldsymbol{y}$, apart from a constant, is

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y}, \tag{14}$$

where $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Omega}$ is the covariance matrix of $\boldsymbol{y}$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\tau})$. Notice that the log-likelihood function (14) has the same structure of the log-likelihood function (7), therefore we can use the same quantities already defined and derived for the general case. We remark that, in this case, the matrix $\boldsymbol{\Sigma}$ is not block diagonal. The same previous iterative procedure to attain the maximum likelihood estimates can be used here.

For the defined model ARMA (13), we have $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\mu} = \boldsymbol{0}$. Thus, by using the same quantities defined in the previous section, $\boldsymbol{D} = \boldsymbol{D}_t = \boldsymbol{0}$ and the matrix $\boldsymbol{M}^{(t)}$, for $t = 1, \ldots, p$, becomes

$$\boldsymbol{M}^{(t)} = \boldsymbol{V}^\top \boldsymbol{H}_2 \boldsymbol{O}_{2t} \boldsymbol{H}_2 \boldsymbol{V} - 3 \left(\boldsymbol{V}^\top \boldsymbol{H}_2 \boldsymbol{V}_t + \boldsymbol{V}_t^\top \boldsymbol{H}_2 \boldsymbol{V} - \left[\boldsymbol{v}_{2t}^\top \boldsymbol{H}_2 \right] \left[\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\theta}} \right] \right), \tag{15}$$

where

$$\boldsymbol{H}_2 = \frac{1}{2} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}, \quad \boldsymbol{O}_{2t} = 4 \boldsymbol{C}_t \otimes \boldsymbol{\Sigma}, \quad \boldsymbol{V}_t = \frac{\partial \boldsymbol{V}}{\partial \theta_t},$$

$\boldsymbol{V}_t = (\mathrm{vec}(\boldsymbol{C}_{1t}), \ldots, \mathrm{vec}(\boldsymbol{C}_{pt}))$ and $\boldsymbol{v}_{2t} = \mathrm{vec}(\boldsymbol{C}_t)^\top$. In order to identify the contribution of the parameters in $\boldsymbol{\tau}$ and σ^2, we partition the matrix $\boldsymbol{M}^{(t)}$ as follows

$$\boldsymbol{M}^{(t)} = \begin{pmatrix} \boldsymbol{M}_{11}^{(t)} & \boldsymbol{M}_{12}^{(t)} \\ \boldsymbol{M}_{21}^{(t)} & M_{22}^{(t)} \end{pmatrix}$$

where $\boldsymbol{M}_{11}^{(t)} = \{m_{rs}^{(t)}\}$, for $r, s = 1 \ldots, b$, is a $b \times b$ matrix that is the contribution of the parameters in $\boldsymbol{\tau}$, $\boldsymbol{M}_{12}^{(t)} = \boldsymbol{M}_{21}^{(t)\top} = \{m_{rs}^{(t)}\}$, for $r = 1, \ldots, b$ and $s = p$, is a $b \times 1$ vector which is a type of cross-contribution of $\boldsymbol{\tau}$ and σ^2 and, finally, $M_{22}^{(t)} = m_{pp}^{(t)}$ is a real number that is the contribution of σ^2. These matrices can be formed by decomposing the matrices $\boldsymbol{V}$ and $\boldsymbol{V}_t$ as follows

$$\boldsymbol{V} = (\widetilde{\boldsymbol{V}}, \ \widetilde{\gamma}) \quad \text{and} \quad \boldsymbol{V}_t = (\widetilde{\boldsymbol{V}}_t, \ \widetilde{\gamma}_t)$$

where $\widetilde{\boldsymbol{V}} = (\mathrm{vec}(\boldsymbol{C}_1), \ldots, \mathrm{vec}(\boldsymbol{C}_b))$, $\widetilde{\gamma} = \mathrm{vec}(\boldsymbol{C}_p)$, $\widetilde{\boldsymbol{V}}_t = (\mathrm{vec}(\boldsymbol{C}_{1t}), \ldots, \mathrm{vec}(\boldsymbol{C}_{bt}))$ and $\widetilde{\gamma}_t = \mathrm{vec}(\boldsymbol{C}_{pt})$. Then, by representation (12), the matrix $\boldsymbol{G}$ can be partitioned as

$$\boldsymbol{G} = \begin{pmatrix} \widetilde{\boldsymbol{G}}_1 & \widetilde{\boldsymbol{G}}_2 \\ \widetilde{\boldsymbol{G}}_2^\top & \widetilde{\boldsymbol{G}}_3 \end{pmatrix},$$

where

$$\widetilde{\boldsymbol{G}}_1 = \begin{pmatrix} \boldsymbol{G}_{11} & \ldots & \boldsymbol{G}_{1b} \\ \vdots & \ldots & \vdots \\ \boldsymbol{G}_{b1} & \ldots & \boldsymbol{G}_{bb} \end{pmatrix}, \quad \widetilde{\boldsymbol{G}}_2 = \begin{pmatrix} \boldsymbol{G}_{1p} \\ \vdots \\ \boldsymbol{G}_{bp} \end{pmatrix} \quad \text{and} \quad \widetilde{\boldsymbol{G}}_3 = \boldsymbol{G}_{pp}.$$

Then,

$$M_{11}^{(t)} = \widetilde{V}^\top H_2 O_{2t} H_2 \widetilde{V} - 3\left(\widetilde{V}^\top H_2 \widetilde{V}_t + \widetilde{V}_t^\top H_2 \widetilde{V} - (I_b \otimes (v_{2t}^\top H_2))\widetilde{G}_1\right),$$

$$M_{12}^{(t)} = \widetilde{V}^\top H_2 O_{2t} H_2 \widetilde{\gamma} - 3\left(\widetilde{V}^\top H_2 \widetilde{\gamma}_t + \widetilde{V}_t^\top H_2 \widetilde{\gamma} - [I_b \otimes (v_{2t}^\top H_2)]\widetilde{G}_2\right)$$

and

$$M_{22}^{(t)} = \widetilde{\gamma}^\top H_2 O_{2t} H_2 \widetilde{\gamma} - 3\left(2\widetilde{\gamma}^\top H_2 \widetilde{\gamma}_t - v_{2t}^\top H_2 \widetilde{G}_3\right).$$

Also, the Fisher information is given by

$$K_{\theta\theta} = \begin{pmatrix} K_{\tau\tau} & K_{\tau\sigma^2} \\ K_{\sigma^2\tau} & K_{\sigma^2\sigma^2} \end{pmatrix} = \begin{pmatrix} \widetilde{V}^\top H_2 \widetilde{V} & \widetilde{V}^\top H_2 \widetilde{\gamma} \\ \widetilde{\gamma}^\top H_2 \widetilde{V} & \widetilde{\gamma}^\top H_2 \widetilde{\gamma} \end{pmatrix}. \quad (16)$$

The matrix M for the ARMA model is easily constructed from equation (15) and then the skewness of the MLEs can be calculated from the matrix M and the information matrix (16).

5. Simulation studies

The main goal of this section is to compare the sample and analytical skewness measures of the MLEs for an errors-in-variables model using Monte Carlo simulation. The sample sizes were taken as $n = 15, 25, 35, 50, 100$ and 1000 and the number N of Monte Carlo replications was 15,000. All simulations were performed using R Development Core Team (2007).

The simple errors-in-variables model considers that (Y_i, X_i) is a vector of random variables defined by

$$Y_i = \alpha + \beta x_i + e_i \quad \text{and} \quad X_i = x_i + u_i, \quad i = 1, \ldots, n, \quad (17)$$

where $x_i \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $e_i \sim \mathcal{N}(0, \sigma^2)$ and $u_i \sim \mathcal{N}(0, \sigma_u^2)$. Here, σ_u^2 is known and x_i, e_i and u_i are mutually uncorrelated for $i = 1, \ldots, n$. Let $\boldsymbol{Y}_i = (Y_i, X_i)^\top$ and $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma^2)^\top$, we have $\boldsymbol{Y}_i \sim \mathcal{N}_2(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$, where

$$\mu_i(\theta) = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{and} \quad \Sigma_i(\theta) = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

Hence,

$$\mu(\theta) = \mathbf{1}_n \otimes \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{and} \quad \Sigma(\theta) = I_n \otimes \begin{pmatrix} \beta^2\sigma_x^2 + \sigma^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

From the previous expressions, we immediately obtain

$$a_1 = \mathbf{1}_n \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad a_2 = \mathbf{1}_n \otimes \begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, \quad a_3 = \mathbf{1}_n \otimes \begin{pmatrix} \beta \\ 1 \end{pmatrix}, \quad a_4 = a_5 = \mathbf{0}$$

and $\boldsymbol{a}_{rs} = \mathbf{0}$ for all r, s except for

$$a_{23} = a_{32} = \mathbf{1}_n \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Further, $\boldsymbol{C}_1 = \boldsymbol{C}_3 = \mathbf{0}$ and

$$C_2 = I_n \otimes \begin{pmatrix} 2\beta\sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & 0 \end{pmatrix}, \; C_4 = I_n \otimes \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix} \quad \text{and} \quad C_5 = I_n \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Additionally, $\boldsymbol{C}_{rs} = \mathbf{0}$ for all r, s except for

$$\boldsymbol{C}_{22} = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\sigma_x^2 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } \boldsymbol{C}_{24} = \boldsymbol{C}_{42} = \boldsymbol{I}_n \otimes \begin{pmatrix} 2\beta & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus,

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{vec}(\boldsymbol{C}_2) & \mathbf{0} & \mathrm{vec}(\boldsymbol{C}_4) & \mathrm{vec}(\boldsymbol{C}_5) \end{pmatrix} \text{ and } \boldsymbol{F}_t = \begin{pmatrix} \boldsymbol{a}_{1t} & \boldsymbol{a}_{2t} & \boldsymbol{a}_{3t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{vec}(\boldsymbol{C}_{2t}) & \mathbf{0} & \mathrm{vec}(\boldsymbol{C}_{4t}) & \mathrm{vec}(\boldsymbol{C}_{5t}) \end{pmatrix}.$$

Therefore, the required quantities to determine $\boldsymbol{\kappa}_3(\widehat{\boldsymbol{\theta}})$ using expression (5) are given. The MLEs are $\widehat{\alpha} = \bar{Y} - \widehat{\beta}\bar{X}$, $\widehat{\beta} = S_{YX}/(S_X^2 - \sigma_u^2)$, $\widehat{\mu} = \bar{X}$, $\widehat{\sigma}_x^2 = S_X^2 - \sigma_u^2$ and $\widehat{\sigma}^2 = S_Y^2 - \widehat{\beta}^2\widehat{\sigma}_x^2$, where $\bar{Y}$ and $\bar{X}$ are the sample means and S_Y^2 and S_X^2 are the sample variances of Y and X, respectively, and S_{YX} is the sample covariance of (Y, X).

For comparison purposes, we compute (for each parameter and sample size) the sample moment ratio statistics $\rho = m_3/m_2^{3/2}$, where $m_k = N^{-1}\sum_{i=1}^{N}(\widehat{\theta}_j - \bar{\widehat{\theta}})^k$ for $k = 2, 3$ and $\bar{\widehat{\theta}} = N^{-1}\sum_{i=1}^{N}\widehat{\theta}_j$. Further, we also compute (for each parameter, sample size and Monte Carlo simulation) the asymptotic skewness $\gamma = \kappa_3/\kappa_2^{3/2}$ for $i = 1, \ldots, N$ and calculate the average $\bar{\gamma} = N^{-1}\sum_{i=1}^{N}\gamma_j$.

The true values of the regression parameters were set at $\alpha = 67$, $\beta = 0.42$, $\mu_x = 70$, $\sigma_x^2 = 247$ and $\sigma^2 = 43$. The parameter values were selected in order to represent the data (the yields of corn on Marshall soil in Iowa) described in Fuller (1987, p. 18). The known measurement error variance is $\sigma_u^2 = 57$ (which was attained through a previous experiment).

Table 1 lists the asymptotic and sample skewness of the MLEs. The figures in this table confirm that the asymptotic skewness and the sample skewness are generally in good agreement and they both converge to zero when n increases (as expected).

We remark that, there is a problem with the sample moment ratio ρ for the MLE of β (and consequently for the MLE of α), since for finite samples sizes the expectation of $\widehat{\beta}$ is not defined (see Fuller, 1987, pag 28, exercise 13(b)). For this reason, this measure may not be reliable to estimate the skewness of $\widehat{\alpha}$ and $\widehat{\beta}$. However, as there is no other reliable measure for the skewness we keep it in the simulations.

Table 1: Asymptotic (γ) and sample (ρ) skewness of the MLEs

		α	β	μ	σ_x^2	σ^2
$n = 15$	$\bar{\gamma}$	-0.6445	0.6550	0.0000	0.7303	0.4879
	ρ	-1.1479	1.2014	-0.0050	0.8207	0.6839
$n = 25$	$\bar{\gamma}$	-0.4189	0.4271	0.0000	0.5657	0.4370
	ρ	-0.6038	0.6089	0.0068	0.5717	0.5319
$n = 35$	$\bar{\gamma}$	-0.3249	0.3318	0.0000	0.4781	0.3870
	ρ	-0.4438	0.4729	0.0016	0.4453	0.4022
$n = 50$	$\bar{\gamma}$	-0.2574	0.2631	0.0000	0.4000	0.3321
	ρ	-0.2992	0.3093	-0.0032	0.4223	0.3105
$n = 100$	$\bar{\gamma}$	-0.1707	0.1748	0.0000	0.2828	0.2409
	ρ	-0.2388	0.2325	0.0449	0.2700	0.2454
$n = 1000$	$\bar{\gamma}$	-0.0512	0.0525	0.0000	0.0894	0.0776
	ρ	-0.0720	0.0832	-0.0006	0.0487	0.0447

The estimated skewness for $\widehat{\mu}_x = \bar{X}$ is zero as expected since the distribution of $\bar{X}$ is symmetric.

6. An application

We consider an application to a small data set given by Fuller (1987, p. 18). Table 2 presents the data which are yields of corn and determinations of available soil nitrogen collected at 11 sites on Marshall soil in Iowa. Following Fuller (1987, p. 18), the estimates of soil nitrogen contain measurement errors arising from two sources. First, only a small sample of soil is selected from each plot and, as a result, there is the sampling error associated with the use of the sample to represent the whole population. Second, there is a measurement error associated with the chemical analysis used to determined the level of nitrogen in the soil sampled. The variance arising from these two sources is estimated as $\sigma_u^2 = 57$. According to Fuller (1987, p. 18), the model (17) holds.

Table 2: Yields of corn on Marshall soil in Iowa.

Site	Yield (Y)	Soil Nitrogen (X)	Site	Yield (Y)	Soil Nitrogen (X)
1	86	70	7	99	50
2	115	97	8	96	70
3	90	53	9	99	94
4	86	64	10	104	69
5	110	95	11	96	51
6	91	64			

Table 3: MLEs (standard error) and asymptotic skewness.

Parameter	MLEs	γ (skewness)
α	66.8606 (11.73)	-0.5670
β	0.4331 (0.16)	0.5793
μ_x	70.6364 (5.02)	0.0000
σ_x^2	220.1405 (118.17)	0.8528
σ^2	38.4058 (20.94)	0.7085

The MLEs, the corresponding standard errors and the estimated skewness are listed in Table 3. The figures in this table show that the skewness of the MLEs of the variances σ_x^2 and σ^2 are high (0.8528 and 0.7085, respectively) which indicates a departure from normality.

6.1. R code

We present below an R code to calculate the second-order skewness of the MLEs. The user just need to give the expected information matrix $\boldsymbol{K_{\theta\theta}}$ and the matrix $\boldsymbol{M}$

```
require("Matrix")

K3.corrected<-function(K,M){
  K<-solve(K)
```

```
  l<-ncol(K)
  aux<-function(j)
    matrix(kronecker(diag(l)[,j], K1[,j]))
  D.K<-Matrix(sapply(1:l, aux),sparse=TRUE)
  Kurtosis.corrected<-t(D.K)%*%kronecker(t(D.K),Diagonal(l))%*%as.vector(t(M)%*%K)
  return(Kurtosis.corrected)
}
```

7. Conclusions and Remarks

We present a general matrix formula for computing the second-order skewness of the maximum likelihood estimators (when their third central moments exist). The matrix formula was applied to a multivariate normal nonlinear model and to an ARMA model. It can be easily implemented into a computer algebra system such as Mathematica or Maple, or into a programming language with support for matrix operations, such as Ox or R. In practice, the skewness can be applied to verify a departure from normality of these estimators for finite sample sizes. We also provide an R code to compute the skewness.

Acknowledgements

We gratefully acknowledge grants from FAPESP and CNPQ (Brazil).

References

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Wiley, New York.

Bowman, K.O. and Shenton, L.R. (1998). *Asymptotic skewness and the distribution of maximum likelihood estimators*, *Communications in Statistics - Theory and Methods*, **27**, 2743–2760.

Bowman, K.O. and Shenton, L.R. (2005). *The asymptotic variance and skewness of maximum likelihood estimators using Maple, Journal of Statistical Computation and Simulation*, **75**, 975–986.

Cavalcanti, A.B., Cordeiro, G.M., Botter, D.A. and Barroso, L. (2009). *Asymptotic Skewness in Exponential Family Nonlinear Models*, *Communications in Statistics - Theory and Methods*, **38**, 2275–2287.

Cordeiro, K.O. and Cordeiro, L.R. (2001). *Skewness for parameters in generalized linear models*, *Communications in Statistics - Theory and Methods*, **30**, 1317–1334.

Cox, D.R., Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cysneiros, F.J., Santos, S.J., Cordeiro, G.M. (2001). Skewness and kurtosis for maximum likelihood estimator in one parameter exponential family models. *Brazilian Journal of Probability and Statistics*, **15**, 85–105.

de Castro, M., Galea, M., Bolfarine, H. (2008). Hypothesis testing in an errors-in-variables model with heteroscedastic measurement errors. *Statistics in Medicine* **27**, 5217–5234.

Fuller, W. (1987). *Measurement Error Models*. Wiley: Chichester.

Magnus, J.R., Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester, third edition.

Patriota, A.G., Lemonte, A.J. (2009). Bias correction in a multivariate normal regression model with general parameterization. *Statistics and Probability Letters*, **79**, 1655–1662

Patriota, A.G., Bolfarine, H., de Castro, M. (2009). A heteroscedastic structural errors-in-variables model with equation error. *Statistical Methodology*, **6**, 408–423

Patriota, A.G., Lemonte, A.J., de Castro M. (2010). Influence diagnostics in a multivariate normal regression model with general parameterization. *Statistical Methodology*, **7**, 644–654

R Development Core Team (2007). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria

Wei, B.C. (1998). *Exponential Family Nonlinear Models*. Singapore: Springer.