Universidade do Contestado | www.unc.br/periodicos ISSNe: 2237-9029
DRd | Desenvolvimento Regional em debate
Revista eletrônica do Programa de Mestrado em Desenvolvimento Regional da Universidade do Contestado

REFINAMENTO DE MÉTODOS ASSINTÓTICOS NOS MODELOS DE REGRESSÃO LINDLEY-UNITÁRIA: UMA APLICAÇÃO EM DADOS DE CRESCIMENTO ECONÔMICO

IMPROVEMENT OF ASYMPTOTIC METHODS IN THE UNIT-LINDLEY REGRESSION MODELS: AN APPLICATION IN ECONOMIC GROWTH DATA

REFINAMIENTO DE MÉTODOS ASINTÓTICOS EN MODELOS DE REGRESIÓN UNITARIA DE LINDLEY: UNA APLICACIÓN A LOS DATOS DE CRECIMIENTO ECONÓMICO

Pedro Ricelly Gama de Oliveira[1]
Tatiane Ferreira do Nascimento Melo da Silva[2]
Tiago Moreira Vargas[3]
Michelli Karinne Barros da Silva[4]

RESUMO

Modelos de regressão são amplamente utilizados em Economia, principalmente quando os dados envolvidos são taxas e proporções. O modelo de regressão Lindley-Unitária é definido para dados restritos ao intervalo (0,1). Em problemas regulares, a inferência baseada na teoria assintótica pode não ser confiável, quando a amostra é pequena. É o caso da estimativa de máxima verossimilhança e do teste de Wald. Correções de vieses dos estimadores de máxima verossimilhança e ajustes feitos nas estatísticas de teste são uma forma amplamente utilizada para resolver tais problemas. Neste artigo, obtemos uma expressão para corrigir o viés e uma fórmula para a matriz de covariância de segunda ordem para os estimadores de máxima verossimilhança no modelo de regressão Lindley-Unitária. Evidências numéricas mostram que os estimadores corrigidos têm vieses menores e que o teste de Wald baseado em covariância de segunda ordem é mais preciso. Por fim, é apresentada uma aplicação a dados econômicos, em que a taxa de Crescimento do PIB per capita é modelada em função da variável de abertura a preços constantes.

Palavras-chave: correção de viés; teste de Wald modificado; matriz de covariância de segunda ordem; regressão Lindley-unitária.

[1]Mestre em Economia. Faculdade de Administração Ciências Contábeis e Ciências Econômicas, Universidade Federal de Goiás, Goiânia, Goiás, Brasil. E-mail: rgo.pedro@gmail.com. ORCID: https://orcid.org/0000-0001-7160-4432

[2]Doutora em Estatística. Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia, Goiás, Brasil. E-mail: *tmelo@ufg.br.* ORCID: https://orcid.org/0000-0001-5035-2280

[3]Doutor em Estatística. Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia, Goiás, Brasil. E-mail: vargas@ufg.br. ORCID: https://orcid.org/0000-0003-3697-5101

[4]Doutora em Estatística. Departamento de Estatística, Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil. E-mail: michelli.karinne@gmail.com. ORCID: https://orcid.org/0000-0002-3224-4152

ABSTRACT

Regression models are widely used in Economics, particularly when the data involved are rates and proportions. The Unit-Lindley regression model is defined for data restricted to the (0,1) range. In regular problems, inference based on asymptotic theory can be unreliable when the sample is small. This is the case of the maximum likelihood estimation and the Wald test. Corrections of biases in the maximum likelihood estimators and adjustments made in the test statistics are a widely used way to solve such problems. In this article, we obtain an expression to the correct the bias and a formula for the second-order covariance matrix for the maximum likelihood estimators in the Unit-Lindley regression model. Numerical evidence shows that the corrected estimators are less biased and that the Wald test based on second-order covariance is more accurate. Finally, an application to economic data is presented, in which the Growth Rate of Real GDP per capita is modeled as a function of openness in constant prices.

Keywords: bias correction; modified Wald test; second-order covariance matrix. Unit-Lindley regression.

RESUMEN

Modelos de regresión son ampliamente utilizados en Economía, principalmente cuando los datos involucrados son tasas y proporciones. El modelo de regresión Lindley-Unitaria está definido para datos restringidos al intervalo (0,1). En problemas regulares, la inferencia basada en la teoría asintótica puede no ser confiable cuando la muestra es pequeña. Este es el caso de la estimación de máxima verosimilitud y la prueba de Wald. Las correcciones de sesgo de los estimadores de máxima verosimilitud y los ajustes realizados en las estadísticas de prueba son una forma ampliamente utilizada para resolver tales problemas. En este artículo, obtenemos una expresión para corregir el sesgo y una fórmula para la matriz de covarianza de segundo orden para los estimadores de máxima verosimilitud en el modelo de regresión Lindley-Unitaria. Evidencia numérica muestra que los estimadores corregidos tienen sesgos más pequeños y que la prueba de Wald basada en la covarianza de segundo orden es más precisa. Por último, se presenta una aplicación a datos económicos, en la que se modela la Tasa de Crecimiento del PIB Real per cápita en función de la apertura en precios constantes.

Palavras clave: Corrección de sesgo. Prueba de Wald modificada. Matriz de covarianza de segundo orden. Regresión Lindley-Unitaria.

Como citar este artigo: OLIVEIRA, Pedro Ricelly Gama de *et al*. Refinamento de métodos assintóticos nos modelos de regressão Lindley-unitária: uma aplicação em dados de crescimento econômico. **DRd – Desenvolvimento Regional em debate**, v. 15, p. 427-445, 27 jun. 2025. Doi: https://doi.org/10.24302/drd.v15.5431.

Artigo recebido em: 15/05/2024
Artigo aprovado em: 25/03/2025
Artigo publicado em: 27/06/2025

1 INTRODUCTION

In practice, we often need to model proportion and percentage data restricted to the (0,1) range. In economics, a key example is the rate of economic growth, which measures the relative change in GDP over time. Accurately modeling such data is essential for understanding macroeconomic dynamics, assessing the effects of policy interventions, and forecasting economic performance. Traditional models, such as linear regression, may not be suitable due to the bounded nature of the dependent variable, making specialized distributions particularly useful.

Regression models for data restricted to the (0,1) range allow for more precise estimation and inference, allowing economists to identify key growth drivers, such as trade openness, investment rates, and institutional quality, while accounting for the inherent constraints of the data. One of the most widely used models for rate and proportion data is Beta regression, proposed by Ferrari and Cribari-Neto (2004), which adopts a parameterization based on the mean and dispersion parameters. However, despite its popularity, the Beta distribution has limitations, particularly in handling extreme skewness or in cases where the variance structure does not adequately capture the data characteristics. To address these challenges, alternative distributions have been proposed to enhance flexibility and improve empirical fit.

Recent studies have introduced several distributions as alternatives to the Beta model for data in the (0,1) range, e.g., Gómez-Déniz, Sordo and Calderín-Ojeda (2014), Jodrá and Jiménez-Gamero (2016), Altun and Cordeiro (2020), Guedes, Cribari-Neto and Espinheira (2020). One notable example is the Unit-Lindley distribution, proposed by Mazucheli, Menezes, and Chakraborty (2019), which was derived from a transformation of the Lindley distribution (Lindley, 1958). The authors applied the Unit-Lindley model to analyze data on inadequate access to water and sewage in Brazilian households. The Lindley distribution has desirable statistical properties, including greater flexibility in skewness and kurtosis, making it a promising alternative to the commonly used exponential distribution (Ghitany, Atieh, & Nadarajah, 2008).

From an inferential perspective, maximum likelihood estimators (MLEs) are generally biased when the sample is small. This bias is usually of order n^{-1}, with n being the sample size. Thus, a correction in the bias is necessary when the amount of data is small. The formulation of expressions that allow bias correction makes it possible to obtain more accurate estimators. Cox and Snell (1968) proposed a general expression for the order bias of MLEs. Since then, the formula proposed by these authors has been applied to various models. **Bias reduction techniques have gained significant attention in the literature due to their ability to improve inference, particularly in small-sample scenarios where standard estimators perform poorly.** Bias correction methods for nonlinear regression models with *t*-Student errors can be found in Vasconcelos and Cordeiro (2000). More recently, bias correction formulas have been developed for Unit-Gamma distribution parameters (Mazucheli; Menezes; Dey, 2018), elliptical models with general parameterization (Melo; Ferrari; Patriota, 2018), and multivariate Dirichlet regression models (Melo *et al.*, 2020).

In addition to bias, first-order asymptotic approximations for the covariance matrix of MLEs may be inaccurate in small samples. A common approach to mitigate this issue is to employ high-order asymptotic theory, which refines inferential procedures to enhance accuracy. Second-order covariance matrix approximations have been studied in various contexts (Shenton; Bowman, 1977; Peers; Iqbal, 1985), offering more precise confidence intervals and hypothesis testing results. Several studies have addressed second-order covariance corrections, including Rocha, Simas, and Cordeiro (2010), Lemonte (2011), Cordeiro *et al.* (2014), Magalhães (2016), and Lemonte (2020).

This study follows two key statistical inference approaches for the Unit-Lindley regression model. The first focuses on deriving the bias correction expression for the maximum likelihood estimators (MLEs) in this model, aiming to reduce bias in small samples. While Mazucheli, Menezes, and Chakraborty (2019) obtained bias corrections for the parameter of the Unit-Lindley distribution, our study extends this to the regression model parameters. The second approach aims to derive a formula for the second-order covariance matrix of the MLEs. Using this refined covariance matrix, we propose a modified Wald test for Unit-Lindley models, addressing the well-known issue that the standard Wald test, under small samples, relies on first-order asymptotic approximations that may lack accuracy.

The accuracy of statistical inference is particularly crucial in economic studies, where small or moderate sample sizes are common. Examples include analyses of economic growth, international trade, and sectoral performance, where biased estimators and imprecise confidence intervals can lead to misleading conclusions about policy effectiveness and key economic determinants. Therefore, bias correction techniques and higher-order asymptotic refinements enhance inference reliability, reducing errors in hypothesis testing and effect estimation. In the context of the Unit-Lindley regression model, the asymptotic refinements proposed in this study improve estimation precision, making the model more robust for applications in economics, such as growth rate analysis and trade modeling.

In addition to providing a compact and accessible framework for these asymptotic refinements, we also present Monte Carlo simulation studies and an empirical application to real data. Our application focuses on Brazil's economic growth rate and international trade. Recent studies have examined the determinants of economic growth in various regions, including OECD member countries (De La Fuente-Mella, Vallina-Hernandez, & Fuentes-Solís, 2020), Africa (Oyebowale & Algarhi, 2020), and the Eurozone (Pegkas, Staikouras, & Tsamadias, 2020). Moreover, Oliveira and Lima (2024) analyzed the convergence between economic sector growth and socio-economic development indicators in the intermediate regions of the Brazilian border zone, using data from 2005 to 2017. In developing countries, macroeconomic factors such as foreign aid, foreign direct investment, fiscal and monetary policies, international trade, physical and human capital, natural resources, and geopolitical and financial conditions play a significant role in shaping economic growth (Chirwa & Odhiambo, 2016). Episodes of accelerated growth— characterized by a per capita growth increase of 2% or more—are particularly noteworthy (Hausmann, Prichett, & Rodrik, 2005), as i exemplified by Brazil's rapid economic expansion in the 1960s and 1970s.

By bridging statistical methodology and economic application, this study provides methodological advancements that enhance inference in models used for economic growth analysis, offering more reliable tools for policymakers and researchers.

The rest of the paper is organized as follows. In Section 2 we present the Unit-Lindley regression model. In Section 3 we derive improved MLEs, the second-order covariance matrix, and an improved Wald test in the Unit-Lindley regression model. Monte Carlo simulation results are presented and discussed in Section 4, and Section 5 outlines an application for illustrative purposes. Finally, in Section 6, we present the conclusions reached during the course of the study.

2 REVIEW OF THE LITERATURE

The Lindley distribution was proposed by Lindley (1958) and its probability density function (p.d.f.) and cumulative distribution function (c.d.f.) are given, respectively, by:

$$f(x;\theta)=\frac{\theta^2}{\theta+1}(x+1)e^{-x\theta}, x>0, \theta\geq 0,$$

and

$$F(x;\theta)=1-(1+\frac{\theta x}{1+\theta})e^{-x\theta}. \quad (1)$$

Mazucheli, Menezes and Chakraborty (2019) proposed the Unit-Lindley distribution. These authors used the transformation $Y=X/(1+X)$ in (1), which resulted in the following c.d.f. and p.d.f, respectively:

$$F(y;\theta)=1-[1-\frac{\theta y}{(1+\theta)(y-1)}]\exp(\frac{-\theta y}{1-y}),$$

and

$$f(y;\theta)=\frac{\theta^2}{1+\theta}(1-y)^{-3}\exp(\frac{-\theta y}{1-y}), 0<y<1, \theta>0. \quad (2)$$

Moreover, Mazucheli, Menezes and Chakraborty (2019) reparameterized (2) in terms of the mean $\mu=1/(1+\theta)$ obtaining

$$f(y;\mu)=\frac{(1-\mu)^2}{\mu(1-y)^3}\exp[-\frac{y(1-\mu)}{\mu(1-y)}], 0<y<1, 0<\mu<1. \quad (3)$$

Let Y_1, Y_2, …, Y_n random sample the Unit-Lindley distribution with p.d.f. (3), i.e., $Y_i \sim UL(\mu_i)$, $i=1,2,\ldots,n$. The regression model is defined by

$$g(\mu_i)=X_i^T\beta=\eta_i \quad (4)$$

with $i=1,2,\ldots,n$, $\beta=(\beta_1,\beta_2,\ldots,\beta_p)^T$ being the vector of unknown parameters $(p\times 1)$, η_i the linear predictor and $X_i^T=(x_{i1},\ldots,x_{ip})$ the observations on *p* known covariates. The $g:(0,1)\rightarrow \mathbb{R}$ link function is known, and is strictly monotonous and twice differentiable.

Possible link function choices $g(\mu_i)$ are: logit $g(\mu_i)=\log(\frac{\mu_i}{1-\mu_i})$, complementary log-log $g(\mu_i)=\log[-\log(1-\mu_i)]$ and probit $g(\mu_i)=\Phi^{-1}(\mu_i)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Here, we used logit. The log-likelihood function is given by

$$l(\beta)=\sum_{i=1}^{n}[2\log(1-\mu_i)-\log(\mu_i)-3\log(1-y_i)-\frac{y_i(1-\mu_i)}{\mu_i(1-y_i)}].$$

The score vector, the observed and expected information matrix are respectively, given by

$$U_\beta(\beta)=X^T v, J(\beta)=X^T MX \text{ and } I(\beta)=X^T QX \qquad (5)$$

where X is the covariate matrix, the i-th element of the v vector is

$$v_i=[\frac{y_i(1-\mu_i)-\mu_i(1+\mu_i)(1-y_i)}{\mu_i^2(1-\mu_i)(1-y_i)}]h'(\eta_i),$$

the i-th elements of the diagonal matrices $M=diag\{m_1,m_2,\dots,m_n\}$ and $Q=diag\{q_1,q_2,\dots,q_n\}$ are given, respectively, by

$$m_i=[\frac{2}{(1-\mu_i)^2}-\frac{1}{\mu_i^2}+\frac{2y_i}{\mu_i^3(1-y_i)}]h'(\eta_i)^2+[\frac{2}{1-\mu_i}+\frac{1}{\mu_i}-\frac{y_i}{\mu_i^2(1-y_i)}]h''(\eta_i),$$

$$q_i=[\frac{1+2\mu_i-\mu_i^2}{\mu_i^2(1-\mu_i)^2}]h'(\eta_i)^2.$$

We denote $h'(\eta_i)$ and $h''(\eta_i)$ as the first- and second-order derivatives, respectively, of $h(\eta_i)=g^{-1}(\eta_i)$ with respect to β_j, or $i = 1, 2,.., n$ and $j = 1, 2, \dots, p$. When the link function is logit we have

$$h'(\eta_i)=\frac{e^{\eta_i}}{(1+e^{\eta_i})^2} \text{ and } h''(\eta_i)=\frac{e^{\eta_i}(1-e^{\eta_i})}{(1+e^{\eta_i})^3}.$$

The maximum likelihood estimator (MLE) for vector β, $\hat{\beta}=(\hat{\beta}_1,\hat{\beta}_2,\dots,\hat{\beta}_p)$, can be obtained by equating $U_\beta(\beta)$ to zero and solving the resulting system. However, such a system does not have closed-form solution and the use of iterative methods, such as nonlinear optimization algorithms are necessary.

3 METHODOLOGY

In this section, we present the second-order bias correction and an expression for the second-order covariance matrix for the maximum likelihood estimator of the unknown parameter vector β in the Unit-Lindley regression model. In addition, we present an improved Wald test.

3.1 IMPROVED ML ESTIMATORS

The general expression for the second-order bias correction of the MLE, proposed by Cox and Snell (1968), was obtained in matrix form by Cordeiro and Klein (1994):

$$B_{\hat{\beta}}(\beta)=I(\beta)^{-1}W_{vec}[I(\beta)^{-1}],$$

where *vec*(.) is the *vec* operator, which transforms a matrix into a vector by stacking the columns of the matrix one under the other, $I(\beta)$ is Fisher's information matrix defined in (5) and $W=(W^{(1)},W^{(2)},\dots,W^{(p)})$ is an array with dimension $p\times p^2$ where each $W^{(r)}$ has the -th element given by

$$w_{ts}^{(r)}=\kappa_{ts}^{(r)}-\frac{1}{2}\kappa_{tsr},$$

where $\kappa_{ts}^{(r)}$ and κ_{ts} are tensor notations for the cumulative log-likelihood derivatives introduced by Lawley (1956): $\kappa_{ts}^{(r)}=\partial\kappa_{ts}/\partial\beta_r$, $\kappa_{tsr}=E[\partial^3 l(\beta)/\partial\beta_t\partial\beta_s\partial\beta_r]$ and $-\kappa_{ts}$ (*t*,*s*)-th element of Fisher's information matrix $I(\beta)$. For the Unit-Lindley regression model we have

$$w_{ts}{}^{(r)}=\sum_{i=1}^{n}\{\frac{-2}{\mu_i^2(1-\mu_i)^3}h'(\eta_i)^3+[\frac{\mu_i^2-2\mu_i-1}{2\mu_i^2(1-\mu_i)^2}]h'(\eta_i)h''(\eta_i)\}x_{ir}x_{is}x_{it}, \tag{6}$$

with x_{ir} being -th element of the *X* matrix. The corrected maximum likelihood estimator is given by

$$\hat{\beta}_{BC}=\hat{\beta}-\hat{B}_{\hat{\beta}}(\beta),$$

$\hat{B}_{\hat{\beta}}(\beta)$ denotes the value of $B_{\hat{\beta}}(\beta)$ valued at $\hat{\beta}$. Note that the second order bias vector $B_{\hat{\beta}}(\beta)$ involves only simple algebraic operations between matrices. This expression can be easily calculated using software with support for matrix operations, such as R (R Core Team, 2022).

3.2 SECOND-ORDER COVARIANCE MATRIX

Following Lawley (1956) notation to denote cumulants of the log-likelihood function, we have

$$\kappa_{bs,d}=E[(\partial^2 l/\partial\beta_b\partial\beta_s)(\partial l/\partial\beta_d)],\kappa_{ab,cd}=E[(\partial^2 l/\partial\beta_a\partial\beta_b)(\partial^2 l/\partial\beta_c\partial\beta_d)]-\kappa_{ab}\kappa_{cd},$$

and its derivatives $\kappa_{ad}^{(bc)}=\partial^2\kappa_{bc}/\partial\beta_a\partial\beta_d$, and $\kappa_{bcd}^{(a)}=\partial\kappa_{bcd}/\partial\beta_a$. In addition, $\kappa_{b,sd}=\kappa_{sd}^{(b)}-\kappa_{bsd}$.

A formula for a second-order covariance, $Cov_2(\hat{\beta})$ is provided by Peers and Iqbal (1985). From Magalhães (2016) we have the matrix form

$$Cov_2(\hat{\beta})=I(\beta)^{-1}+I(\beta)^{-1}[\Gamma+\Gamma^T]I(\beta)^{-1}$$

where the (a, b)-th element of Γ is given by $\gamma_{ab}=\frac{-1}{2}\gamma_{ab}^{(1)}+\frac{1}{4}\gamma_{ab}^{(2)}+\frac{1}{2}\gamma_{ab}^{(3)}$, and

$$\gamma_{ab}^{(1)}=\sum_{c,d=1}^{p}\kappa^{cd}\{2\kappa_{bc}^{(ad)}-\kappa_{bcd}^{(a)}+\kappa_{ac,bd}\},$$

$$\gamma_{ab}^{(2)}=\sum_{r,s,c,d=1}^{p}\kappa^{rs}\kappa^{cd}\{\kappa_{arc}[3\kappa_{bsd}+2\kappa_{b,sd}+8\kappa_{bs,d}]+2\kappa_{ar,c}[2\kappa_{b,sd}+\kappa_{bd,s}]\},$$

$$\gamma_{ab}^{(3)}=\sum_{r,s,c,d=1}^{p}\kappa^{rs}\kappa^{cd}\{\kappa_{bc}^{(a)}(\kappa_{sd}^{(r)}+\kappa_{r,sd})\}.$$

For the Unit-Lindley model, after performing some algebra, it can be expressed as

$$\Gamma=X^T[\frac{1}{2}TZ_D+\frac{1}{4}(\Lambda\odot Z\odot Z)]X+\frac{1}{2}X^T[(\Delta\odot ZZ_D)Y\odot X], \tag{7}$$

where "$\odot$" denotes the direct product of matrices (or Hadamard product), $Z=XI(\beta)^{-1}X^T=\{z_{ij}\}$, $Z_D=diag\{z_{11},z_{22},\dots,z_{nn}\}$, Y is a matrix ($n\times p$) of ones, $T=diag\{t_1,t_2,\dots,t_n\}$, with $t_i=2\xi_i-\lambda_i+\phi_i$ for $i=1,2,\dots,n$. Furthermore, we have that the -th elements of the matrices Λ and Δ are, respectively, $\nu_i(3\nu_j+10\tau_j)+6\tau_i\tau_j$, and $\delta_i(\delta_j+\tau_j)$, for $j=1,2,\dots,n$, where,

$$\xi_i=2[\frac{3\mu_i^4-12\mu_i^3-2\mu_i^2+8\mu_i-3}{\mu_i^4(1-\mu_i)^4}]h'(\eta_i)^4+10[\frac{\mu_i^3-3\mu_i^2-\mu_i+1}{\mu_i^3(1-\mu_i)^3}]h'(\eta_i)^2h''(\eta_i)$$

$$+2[\frac{\mu_i^2-2\mu_i-1}{\mu_i^2(1-\mu_i)^2}][h''(\eta_i)^2+h'(\eta_i)h'''(\eta_i)],$$

$$\lambda_i=[\frac{12(\mu_i^4-4\mu_i^3+\mu_i^2+2\mu_i-1)}{\mu_i^4(1-\mu_i)^4}]h'(\eta_i)^4+[\frac{6(3\mu_i^3-9\mu_i^2-\mu_i+3)}{\mu_i^3(1-\mu_i)^3}]h'(\eta_i)^2h''(\eta_i)$$

$$+3[\frac{\mu_i^2-2\mu_i-1}{\mu_i^2(1-\mu_i)^2}][h''(\eta_i)^2+h'(\eta_i)h'''(\eta_i)],$$

$$\phi_i=\frac{(1+2\mu_i-\mu_i^2)}{\mu_i^2(1-\mu_i)^2}[\frac{2}{\mu_i}h'(\eta_i)^2-h''(\eta_i)]^2,$$

$$\nu_i=\frac{4(\mu_i^3-3\mu_i^2+1)}{\mu_i^3(1-\mu_i)^3}h'(\eta_i)^3+\frac{3\mu_i^2-2\mu_i-1}{\mu_i^2(1-\mu_i)^2}h'(\eta_i)h''(\eta_i),$$

$$\delta_i=\frac{2(\mu_i^3-3\mu_i^2-\mu_i+1)}{\mu_i^3(1-\mu_i)^3}h'(\eta_i)^3+\frac{2(\mu_i^2-2\mu_i-1)}{\mu_i^2(1-\mu_i)^2}h'(\eta_i)h''(\eta_i),$$

$$\tau_i=\frac{\mu_i^2-2\mu_i-1}{\mu_i^2(1-\mu_i)^2}[\frac{2}{\mu_i}h'(\eta_i)^2-h''(\eta_i)]h'(\eta_i).$$

The expression given for Γ in (7) also involves only simple operations between matrices. Therefore, it can be easily computed using the R software.

3.3 IMPROVED WALD TEST

Let the parameter vector be partitioned as $\beta=(\psi^\top,\omega^\top)$, where $\psi=(\beta_1,\ldots,\beta_q)^\top$ is the vector of parameters of interest and $\omega=(\beta_{q+1},\ldots,\beta_p)^\top$ is the vector of nuisance parameters. Here, the null and alternative hypotheses are, respectively: $H_0:\psi=\psi^{(0)}$ and $H_1:\psi\neq\psi^{(0)}$, where $\psi^{(0)}$ is a known q-vector. The unrestricted maximum likelihood estimator of β is denoted by $\hat{\beta}=(\hat{\psi}^\top,\hat{\omega}^\top)^\top$. We use "^" for matrices and vectors to indicate that they are computed at $\hat{\beta}$.

The Wald statistic for testing H_0 is

$$W=(\hat{\psi}-\psi^{(0)})^\top[\hat{I}^{\psi\psi}(\hat{\beta})]^{-1}(\hat{\psi}-\psi^{(0)}),$$

where $I^{\psi\psi}$ is the upper left submatrix of $I(\beta)^{-1}$ for the parameters of interest.

The improved Wald test is obtained by replacing $\hat{I}^{\psi\psi}(\hat{\beta})$ with the second-order covariance matrix $\widehat{Cov}_{2_{\psi\psi}}(\hat{\beta})$, which denotes the upper left submatrix of $\widehat{Cov}_2(\hat{\beta})$ for the parameters of interest. That is,

$$W_c=(\hat{\psi}-\psi^{(0)})^\top[\widehat{Cov}_{2_{\psi\psi}}(\hat{\beta})]^{-1}(\hat{\psi}-\psi^{(0)}).$$

We emphasize here that the W_c statistics does not change the order of convergence of the Wald test (Magalhães, 2016).

4 RESULTS AND DISCUSSION

4.1 MONTE CARLO SIMULATION

In this section, we present the results of Monte Carlo simulation experiments in which we evaluate the finite sample performance of the original MLEs and their bias-corrected version. In addition, we compare, via Monte Carlo simulation, the first-order covariance matrix $I(\beta)^{-1}$ and the second-order covariance matrix $Cov_2(\hat{\beta})$ in relation to the sample (empirical) covariance matrix for small sized samples. Finally, we evaluate the finite sample performance of the Wald test (*W*) and its corrected version (W_c).

The simulations are based on the Unit-Lindley regression model (4). All simulations are performed using the R software (R Core Team, 2022). The number of Monte Carlo replications is 10,000 (ten thousand) with sample sizes $n=10,20,30,50$. The tests are carried out at the following nominal levels: α = 1%, 5% and 10%. The values of the covariates are taken as random draws from the uniform distribution *U* (-0.5,0.5). In this simulation study we consider the logit link function, that is, $g(\mu)=\eta=\mu/(1-\mu)$. We consider the model with two to six covariates and we tested one and two parameters (*q*=1,2). The true values set of the parameters are taken as $\beta_1=0$, $\beta_2=-2$, $\beta_3=1,$ $\beta_4=\beta_5=0.5$ and $\beta_6=1.5$ for *q* = 1, whereas *q* = 2, $\beta_1=\beta_2=0$, $\beta_3=-2$, $\beta_4=1$ and $\beta_5=\beta_6=0.5$.

Tables 1-2 present the bias, the bias-corrected estimates, and corresponding estimated root mean squared errors (RMSE) of the MLE for different sample sizes and covariates. In general, we observe that the bias-corrected estimates are less biased than the original MLE. For instance, when $n=10$, $p=5$ (see Table 1) the estimated biases of $\hat{\beta}_2$ are -0,75 (MLE), and -0.02 (bias-corrected). As *n* increases, both the bias and the root mean squared error of all the estimators decrease, as expected.

In Table 3 we have the sample variances (Var_A), second-order variances (Var_2), and first-order variances Var_1 of $\hat{\beta}$. Remember that $Var_1(\hat{\beta})=I(\hat{\beta})^{-1}$. We observe in this table that the second-order variances, obtained in (3.2), are much closer to the sample variances than the first-order variances, when the sample is small. For instance, for $\hat{\beta}_5$ with $p=5$ and $n=10,$ the sample and second-order variances are equal to 2.14 while the first-order variance is 1.59. We also observe that as the number of covariates (*p*) increases, the difference between the first-order variance and the sample variances also increases. Furthermore, as expected, when the increase the sample size, these variances become are closer to each other.

The null rejection rates of the Wald (*W*) and Corrected Wald tests (W_c) are displayed in Tables 4 and 5 for different values of *n* and *p;* entries are given as percentages. We present the rejection rates of $H_0:\psi=\psi^{(0)}$. When *q* = 1 (Table 4) we consider $\psi^{(0)}$= 0 and *q* = 2 for (Table

5) we have $\psi^{(0)}=(0,0)^T$. All tests were carried out at the following nominal levels: α= 1%, 5% and α=10%. The Wald test is markedly liberal when the sample size is small. For instance, for the case where p = 6 and q = 2, the Wald test displays rejection rates equal to 7.5% (α = 1%), 16.8% (α= 5%), and 24.5% (α= 10%). The test based on the corrected Wald statistic shows better performance than the test based on the original Wald statistic in all cases, with rejection rates close to the nominal levels. For instance, when p = 5 , q = 2, n = 20, and α = 1%, the rejection rates are 1.2%$ (W_c) and 2.0% (W). When p = 3 , q = 1, n = 10, the rejection rates are 1.0%, 5.1%, 10.1% (α = 1%, 5%, 10%$) for W_c and 2.1%, 7.2%, 13.2%$ (α= 1%, 5%, 10%) for W. Overall, the best performing test is the one that employs W_c as the test statistic.

Table 1 – Bias and RMSE of the ML estimates and their bias adjusted version; n = 10 and 20

		ML		**Bias corrected**		**ML**		**Bias corrected**	
p	β	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
	β_1	0.23	1.26	0.01	1.24	−0.04	0.75	−0.01	0.75
3	β_2	−0.38	1.26	−0.02	1.18	−0.04	0.7	−0.01	0.70
	β_3	−0.34	1.27	−0.01	1.21	0.06	0.74	0.02	0.74
	β_1	0.30	1.31	0.02	1.28	0.00	0.74	0.00	0.75
4	β_2	−0.49	1.35	−0.02	1.23	−0.04	0.71	0.00	0.71
	β_3	−0.40	1.34	−0.02	1.26	0.06	0.79	0.01	0.79
	β_4	−0.01	1.10	0.00	1.10	−0.03	0.68	−0.01	0.68
	β_1	0.18	1.38	−0.03	1.37	0.00	0.78	−0.01	0.79
	β_2	−0.75	1.58	−0.02	1.38	−0.04	0.71	0.00	0.71
5	β_3	−0.42	1.35	−0.02	1.28	0.05	0.85	0.00	0.86
	β_4	0.15	01.09	0.00	01.08	−0.01	0.67	−0.01	0.67
	β_5	−0.47	1.54	−0.01	1.45	0.06	0.76	0.01	0.77
	β_1	0.89	02.08	0.06	1.82	−0.10	0.99	−0.03	0.98
	β_2	−1.03	1.79	−0.08	1.42	−0.08	0.75	−0.02	0.75
6	β_3	−1.08	2.17	−0.07	1.82	0.15	0.94	0.03	0.93
	β_4	0.18	1.28	0.02	1.23	−0.01	0.68	−0.01	0.69
	β_5	0.05	1.70	0.01	1.64	−0.09	0.86	−0.03	0.85
	β_6	0.73	1.90	0.06	1.70	−0.16	0.93	−0.04	0.92

Table 2 – Bias and RMSE of the ML estimates and their bias adjusted version; *n*=30 and 50

		ML			**Bias corrected**			**ML**			**Bias corrected**	
p	β	**Bias**	**RMSE**		**Bias**	**RMSE**		**Bias**	**RMSE**		**Bias**	**RMSE**
	β_1	−0.02	0.58		−0.01	0.58		−0.02	0.43		0.00	0.43
3	β_2	0.02	0.54		0.00	0.54		−0.01	0.40		0.00	0.40
	β_3	0.04	0.54		0.01	0.54		0.02	0.41		0.00	0.41
	β_1	−0.01	0.57		0.00	0.57		−0.01	0.43		0.00	0.43
4	β_2	0.03	0.55		0.00	0.55		0.00	0.41		0.00	0.41
	β_3	0.04	0.57		0.00	0.57		0.02	0.42		0.00	0.42
	β_4	−0.01	0.54		−0.01	0.54		0.00	0.39		−0.01	0.39
	β_1	−0.01	0.61		−0.01	0.61		−0.01	0.45		0.00	0.45
	β_2	0.04	0.56		0.02	0.56		0.00	0.42		0.01	0.42
5	β_3	0.03	0.60		−0.01	0.60		0.01	0.45		−0.01	0.45
	β_4	−0.01	0.55		−0.01	0.55		0.00	0.41		−0.01	0.41
	β_5	0.06	0.62		0.01	0.62		0.02	0.46		0.01	0.46
	β_1	0.01	0.61		0.00	0.61		−0.01	0.44		0.00	0.44
	β_2	0.04	0.57		0.01	0.57		0.00	0.42		0.00	0.42
6	β_3	0.03	0.60		0.01	0.60		0.02	0.43		0.00	0.43
	β_4	−0.01	0.55		−0.02	0.55		0.00	0.40		−0.01	0.40
	β_5	0.04	0.60		−0.01	0.61		0.00	0.44		0.00	0.44
	β_6	0.04	0.60		0.00	0.60		0.00	0.43		0.00	0.43

Table 3 – Sample variances, second-order variances, and first-order variances of $\widehat{\beta}$.

	$p=3$										
		$n=10$				$n=20$				$n=30$		
		Var_A	Var_2	Var_1		Var_A	Var_2	Var_1		Var_A	Var_2	Var_1
β_1		1.53	1.55	1.30		0.56	0.54	0.49		0.34	0.33	0.31
β_2		1.44	1.50	1.21		0.49	0.50	0.46		0.29	0.30	0.28
β_3		1.50	1.54	1.29		0.55	0.55	0.54		0.29	0.29	0.28
$p=4$												
		$n=10$				$n=20$				$n=30$		
β_1		1.64	1.68	1.37		0.55	0.56	0.50		0.33	0.33	0.31
β_2		1.58	1.63	1.26		0.50	0.51	0.46		0.30	0.31	0.28
β_3		1.63	1.69	1.38		0.62	0.64	0.57		0.32	0.33	0.30
β_4		1.21	1.26	1.00		0.46	0.45	0.40		0.29	0.29	0.27
$p=5$												
		$n=10$				$n=20$				$n=30$		
β_1		1.88	1.91	1.52		0.62	0.61	0.54		0.37	0.36	0.33
β_2		1.93	1.97	1.52		0.50	0.52	0.45		0.31	0.32	0.29
β_3		1.64	1.70	1.37		0.73	0.74	0.65		0.36	0.35	0.32
β_4		1.17	1.24	0.96		0.45	0.46	0.40		0.31	0.31	0.28
β_5		2.14	2.14	1.59		0.58	0.59	0.51		0.38	0.38	0.35

Table 4 – Null rejection rates q = 1 and different values of n and p; entries are percentages.

			$\alpha=1\%$			$\alpha=5\%$			$\alpha=10\%$	
p	n		W	W_c		W	W_c		W	W_c
2	10		1.5	0.9		7.0	4.9		12.4	9.6
	20		1.4	1.1		6.2	5.3		11.2	10.1
	30		1.2	1.1		5.7	5.1		11.0	10.2
	50		1.1	1.0		5.9	5.4		10.6	10.2
3	10		2.1	1.0		7.2	5.1		13.2	10.1
	20		1.5	1.1		6.3	5.1		12.1	10.4
	30		1.5	1.2		6.3	5.5		11.5	10.4
	50		1.3	1.1		6.1	5.6		11.0	10.4
4	10		2.4	1.2		7.6	5.2		13.5	9.9
	20		1.4	1.0		6.0	4.8		11.7	9.7
	30		1.3	1.1		5.7	4.7		11.0	9.8
	50		1.1	0.9		5.4	4.9		10.7	10.0
5	10		2.3	1.0		8.2	5.1		14.4	10.3
	20		1.6	1.0		6.4	5.0		12.3	9.8
	30		1.4	1.0		6.5	5.5		12.1	10.7
	50		1.4	1.2		5.9	5.2		11.0	10.2
6	10		4.6	2.0		11.6	7.0		18.0	12.0
	20		1.7	1.1		7.2	5.2		12.6	10.2
	30		1.6	1.0		6.2	4.9		11.7	9.9
	50		1.3	1.0		5.9	5.1		10.8	9.8

Table 5 – Null rejection rates: q = 2 and different values of n and p; entries are percentages.

			$\alpha=1\%$			$\alpha=5\%$			$\alpha=10\%$	
p	n		W	W_c		W	W_c		W	W_c
	10		2.6	1.4		8.6	5.3		14.7	10.2
3	20		1.4	1.0		6.5	5.3		12.0	9.9
	30		1.4	1.1		6.2	5.1		11.6	10.3
	50		1.2	1.0		5.4	4.9		11.1	10.2
	10		3.5	1.8		9.9	6.1		15.6	10.7
4	20		1.7	1.1		7.0	5.0		13.1	10.3
	30		1.4	1.1		6.3	5.2		12.0	10.1
	50		1.4	1.0		5.8	5.0		11.3	10.2
	10		4.7	2.2		12.4	7.6		18.8	13.0
5	20		2.0	1.2		7.3	5.1		13.2	10.0
	30		1.6	1.1		6.7	5.3		12.6	10.3
	50		1.4	1.1		6.0	5.0		11.2	9.9
	10		7.5	3.1		16.8	9.0		24.5	14.9
6	20		2.2	1.1		7.9	5.5		14.1	10.4
	30		1.8	1.1		7.0	5.1		12.7	10.2
	50		1.4	1.1		6.4	5.1		11.9	10.1

4.2 REAL ILLUSTRATION: ECONOMIC GROWTH DATA

The data used in this section were taken from Penn World Table 6.1 (Heston; Summers; Aten, 2002), whose goal is to model Growth Rate of Real GDP per capita as a function of openness in constant prices (OPENK) for Brazil from 1962 to 1973. We consider the following Unit-Lindley regression model to fit these data:

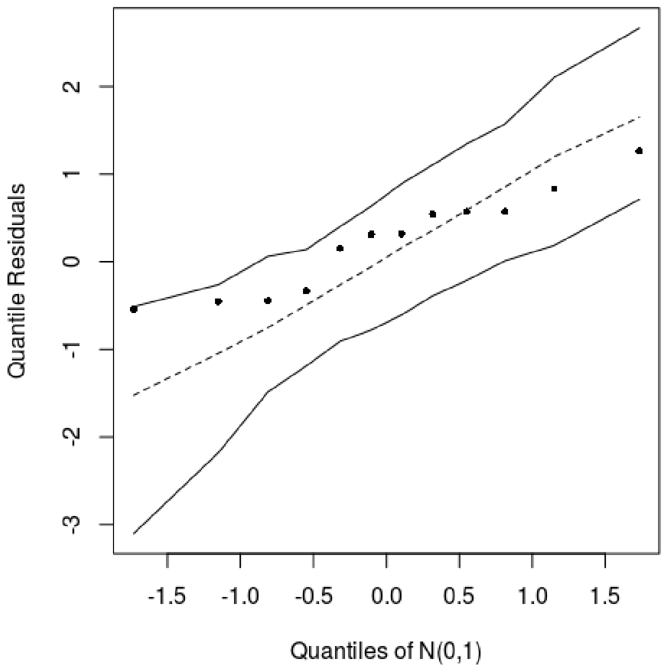
$$logit(\mu_i) = \beta_0 + \beta_1 OPENK_i,$$

where *logit*(.) is the logit link function, with i = 1, 2, ..., 12. In the simulated envelope (Figure 1), we observe that all residuals remain within the confidence bands, indicating that the fitted model adequately represents the data.

The ML estimates are $\widehat{\beta_0}$= 35.45 and $\widehat{\beta_1}$= -6.11. The bias adjusted estimates are $\widehat{\beta_0}$ = 36.64 and $\widehat{\beta_1}$= -6.13. Although the corrected estimates are close to the original estimates, the Akaike Information Criterion (AIC) yields a value of 18.10 for the case of the model whose estimates are uncorrected and -46.86 for the corrected version, indicating that the best fitted regression provides the corrected estimates.

The first- and second-order variance of $\widehat{\beta_0}$ are, respectively, $Var_1(\widehat{\beta}_0)$= 428.86 and $Var_2(\widehat{\beta}_0)$= 505.26. We observed a great difference between first- and second-order variances. For $\widehat{\beta_1}$ we have no significant difference, since $Var_1(\widehat{\beta}_1)$ = 3.29 and $Var_2(\widehat{\beta}_1)$= 3.85.

Figure 1 – The Quantile–Quantile (QQ) plot of the randomized quantile residuals for Unit-Lindley regression models.



Performing the test $H_0 : \beta_1 = 0 \times H_1 : \beta_1 \neq 0$ we obtain the following values for uncorrected and corrected Wald statistics, respectively, $W = 2.93$ (p-value = 0.087) and W_c = 2.49 (p-value = 0.115). At the 10% significance level, the corrected Wald test does not reject the null hypothesis, while the uncorrected Wald test rejects the null hypothesis.

The tests lead to different conclusions, and the conclusion obtained using the adjusted test is compatible with that achieved in Veloso, Villela and Giambiagi (2008). These authors argue that Brazil's accelerated growth between 1968 and 1973 was not directly driven by trade openness but rather by a favorable external environment, including factors such as abundant international liquidity and increasing demand for Brazilian exports. The non-significance of the variable OPENK suggests that, unlike in more recent periods, trade openness was not a key determinant of economic growth in Brazil at that time. This may be attributed to the country's economic model, which was characterized by strong state intervention, import substitution policies, and public investments in infrastructure and industry. Moreover, the comparison between the statistical tests highlights the importance of using robust techniques to avoid biased conclusions about the determinants of economic growth.

5 CONCLUSIONS AND RECOMMENDATIONS

In this paper, we use the Unit-Lindley regression model as an alternative for modeling data that take values in the (0,1) range, based on the Lindley distribution, which has good statistical properties. We presented two key theoretical contributions for small sample sizes: (i) the expression for correcting the bias of the maximum likelihood estimator and (ii) the second-order covariance matrix formula, from which we propose a modified Wald test. The results showed that the corrected estimators exhibit less bias than the uncorrected ones, and the Wald test based on the second-order covariance matrix yields rejection rates closer to the nominal levels, as evidenced by the Monte Carlo experiments.

In practical terms, we illustrated the application of these results using economic data, with the growth rate of real GDP per capita as the dependent variable and economic openness as the covariate. The application demonstrated that the corrected Wald test produces inferences that differ from those obtained using the original statistic. This approach has significant implications for economic data analysis, especially in small sample contexts.

Finally, we suggest that future research explore alternative methods of refinement in hypothesis testing, such as those based on the likelihood ratio statistic rather than the Wald statistic.

REFERENCES

ALTUN, E.; CORDEIRO, G.M. The unit-improved second-degree Lindley distribution: inference and regression modeling, **Computational Statistics**, v. 35, p. 259-279, 2020.

CHIRWA, T.G.; ODHIAMBO, N.M. Macroeconomic determinants of economic growth: A review of international literature, **South East European Journal of Economics and Business**, v. 11, p. 33–47, 2016.

CORDEIRO, G.M.; BOTTER, D.A.; Cavalcanti, A.B.; Barroso, L.P. Covariance matrix of the bias corrected maximum likelihood estimator in generalized linear models, **Statistical Papers**, v. 55, p. 43-652, 2014.

CORDEIRO, G.M.; KLEIN, R. Bias correction in ARMA models, **Statistics and Probability Letters**, v. 19, p. 169-176, 1994.

COX, D.R.; SNELL, E.J. A General Definition of Residuals, **Journal of the Royal Statistical Society**: Series B (Methodological), v. 30, p. 248-265, 1968.

DE LA FUENTE-MELLA, H.; VALLINA-HERNANDEZ, A.M.; FUENTES-SOLÍS, R. Stochastic analysis of the economic growth of OECD countries, **Economic Research-Ekonomska Istrazivanja**, v. 33, p. 2189- 2202, 2020.

FERRARI, S.L.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions, **Journal of Applied Statistics**, v. 31, p. 799-815, 2004.

GHITANY, M.E.; ATIEH, B.; NADARAJAH, S. Lindley distribution and its application, **Mathematics and Computers in Simulation**, v. 78, p. 493-506, 2008.

GÓMEZ-DÉNIZ, E.; SORDO, M.A.; CALDERÍN-OJEDA, E. The Log-Lindley distribution as an alternative to the beta regression model with applications in insurance, **Insurance:** Mathematics and Economics, v. 54, p. 49–57, 2014.

GUEDES, A.C.; CRIBARI-NETO, F.; ESPINHEIRA, P.L. Modified likelihood ratio tests for unit gamma regressions, **Journal of Applied Statistics**, v. 47, p. 1562-1586, 2020.

HAUSMANN, R.; PRICHETT, L.; RODRIK, D. Growth accelerations, **Journal of Economic Growth**, v. 10, p. 303-329, 2005.

HESTON, A.; SUMMERS, R.; ATEN, B. **Penn World Table Version 6.1**. Technical report, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, 2002.

JODRÁ, P.; JIMÉNEZ-GAMERO, M. D. A note on the Log-Lindley distribution, **Insurance:** Mathematics and Economics, v. 71, p. 189-194, 2016.

LAWLEY, D.N. A General Method for Approximating to the Distribution of Likelihood Ratio Criteria, **Biometrika**, v. 43, p. 295-303, 1956.

LEMONTE, A.J. Covariance matrix formula for Birnbaum-Saunders regression models, **Journal of Statistical Computation and Simulation**, v. 81, p. 899–908, 2011.

LEMONTE, A.J. Covariance matrix of maximum likelihood estimators in censored exponential regression models, **Communications in Statistics - Theory and Methods**, v.51, p. 1765-1777, 2020.

LINDLEY, D.V. Lindley, Fiducial Distributions and Bayes Theorem, **Journal of the Royal Statistical Society**: Series B (Methodological), v. 20, p. 102-107, 1958.

MAGALHÃES, T.M. **Refinamento de propriedades assintóticas do estimador de máxima verossimilhança**. PhD thesis, Universidade de São Paulo, 2016.

MAZUCHELI, J.; MENEZES, A.F.B., CHAKRABORTY, S. On the one parameter unit-Lindley distribution and its associated regression model for proportion data, **Journal of Applied Statistics**, v. 46, p. 700-714, 2019.

MAZUCHELI, J.; MENEZES, A.F.B.; DEY, S. Improved maximum-likelihood estimators for the parameters of the unit-gamma distribution, **Communications in Statistics-Theory and Methods**, v. 47, p. 3767-3778, 2018.

MELO, T.F.N.; FERRARI, S.L.P.; PATRIOTA, A.G. Improved estimation in a general multivariate elliptical model, **Brazilian Journal of Probability and Statistics**, v. 32, p. 44-68, 2018.

MELO, T.F.N.; VARGAS, T.M.; LEMONTE, A.J.; MORENO-ARENAS, G. On improved estimation in multivariate dirichlet regressions. **Communications in Statistics-Theory and Methods**, v. 49, p. 5765-5777, 2020.

OLIVEIRA, L. R.; LIMA, J. F. Crescimento e desenvolvimento econômico na fronteira brasileira: o que mostram os indicadores de convergência?. **DRd - Desenvolvimento Regional Em Debate**, v. 14, p. 97–120, 2024.

OYEBOWALE, A.Y.; ALGARHI, A.S. Macroeconomic determinants of economic growth in Africa, **International Review of Applied Economics**, v. 34, p. 839-857, 2020.

PEERS, H.W.; IQBAL, M. Asymptotic Expansions for Confidence Limits in the Presence of Nuisance Parameters, with Applications, **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 47, p. 547-554, 1985.

PEGKAS, P.; STAIKOURAS, C.; TSAMADIAS, C. On the determinants of economic growth: Empirical evidence from the Eurozone countries, **International Area Studies Review**, v. 23 p. 210-229, 2020.

ROCHA, A.V.; SIMAS, A.B.; CORDEIRO, G.M. Second-order asymptotic expressions for the covariance matrix of maximum likelihood estimators in dispersion models, **Statistics and Probability Letters**, v. 80, pp. 718–725, 2010.

SHENTON, L.; BOWMAN, K.O. **Maximum Likelihood Estimation in Small Samples** (Griffin's statistical monographs & courses. C. Griffin), 1977.

R CORE TEAM. **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria, 2022. Available in https://www.R-project.org/. Accessed on June 24, 2024.

VASCONCELOS, K. L.; CORDEIRO, G.M. Bias corrected estimates in multivariate student t regression models, **Communications in Statistics Part B: Simulation and Computation**, v. 29, p. 797-822, 2000.

VELOSO, F.A.; VILLELA, A.; GIAMBIAGI, F. Determinantes do "milagre" econômico brasileiro (1968-1973): uma análise empírica. **Revista Brasileira de Economia**, v. 62, p. 221-246, 2008.