Third-order inference for autocorrelation in nonlinear regression models

P. E. Nguimkeu∗ M. Rekkas†

Abstract

We propose third-order likelihood-based methods to derive highly accurate p-value approximations for testing autocorrelated disturbances in nonlinear regression models. The proposed methods are particularly accurate for small- and medium-sized samples whereas commonly used first-order methods like the signed log-likelihood ratio test, the Kobayashi (1991) test, and the standardized test can be seriously misleading in these cases. Two Monte Carlo simulations are provided to show how the proposed methods outperform the above first-order methods. An empirical example applied to US population census data is also provided to illustrate the implementation of the proposed method and its usefulness in practice.

Keywords: Nonlinear regression models; Autocorrelation; Likelihood Analysis; p-value

∗Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, email: npnl@sfu.ca.
†Corresponding Author. Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, email: mrekkas@sfu.ca, phone: (778) 782-6793, fax: (778) 782-5944
1 Introduction

Testing for autocorrelation is a prominent tool for performing model adequacy checks in applied time series analysis. Falsely assuming independent errors in an estimated model yields inefficient estimators which biases hypothesis testing toward invalid conclusions of significance (see for example, Poirier and Ruud (1988) and Gourieroux et al. (1984)). Methods developed to test autocorrelation have therefore been thoroughly explored in the literature, most of which have been tailored to the case of linear regression models. Well-known examples of such tests are the Durbin and Watson (1950) test for first-order autocorrelation and standardized tests based on simple likelihood asymptotic methods (see, e.g. Hamilton (1994)).

However, many time series econometric models are frequently defined in the form of nonlinear regressions and applying the above mentioned techniques to handle autocorrelation in nonlinear cases could be seriously misleading (see White (1992)). Apart from a few authors, little attention has been paid to developing testing procedures for autocorrelation of disturbances in nonlinear regression models. Using a quasi-maximum likelihood procedure, Kobayashi (1991) proposed an extended iterated Cochrane-Orcutt estimator for the nonlinear case. The criterion is a bias-corrected test statistic from the standardized one. White (1992) approximated the exact distribution of the Durbin-Watson test statistic for first-order autocorrelation in nonlinear models. His technique is based on a linearization of the model by first-order Taylor series expansion of the regression function around the nonlinear least square estimator of the model. A common feature of these methods is that like the Durbin-Watson test, they have bounds under the null and an inconclusive region where a critical value needs to be computed. Moreover, these tests are asymptotic and are often not suitable for small sample models. The latter are frequent when modeling annual time series variables like national accounts aggregates or decennial series like most population census data.

We propose a new testing procedure for first-order autocorrelation in nonlinear regression that uses small-sample likelihood asymptotic inference methods. These methods were developed by Fraser and Reid (1995) and have been proven to possess high accuracy compared to the traditional first-order asymptotic methods described above. Two Monte Carlo simulations are provided to show the superiority of this procedure over existing ones and an empirical study is presented to show its usefulness in practice.

The rest of the paper is organized as follows. Section 2 presents the background of the third-order inference method. Section 3 describes the model and the proposed tests. Monte Carlo simulations results are provided in Section 4 followed by an empirical study in Section 5. Some concluding remarks are given in Section 6.

2 Background

2.1 First-order approximations

Let $Y_1, \ldots, Y_n$ be independently and identically distributed from a parametric model with density function given by $f(y; \theta)$. Assume the full parameter $\theta = (\psi, \lambda)'$ is $p$ dimensional. The variable $\psi$ represents the
scalar component interest parameter and the vector $\lambda$ is the $p-1$ dimensional nuisance parameter. The likelihood function is proportional to the joint density function and is given by

$$L(\theta; y) = c \Pi f(y; \theta),$$

where $c$ is an arbitrary multiplicative constant. The log-likelihood function for $\theta$ for a sample $y = (y_1, y_2, \ldots, y_n)'$ is then given by

$$l(\theta) = l(\theta; y) = \sum_{i=1}^{n} \log f(y_i; \theta) + a. \tag{1}$$

Maximization of this function with respect to $\theta$ yields the maximum likelihood estimate $\hat{\theta} = (\hat{\psi}, \hat{\lambda}')$. The constrained maximum likelihood estimate denoted by $\hat{\theta}_{\psi} = (\hat{\psi}, \hat{\lambda}_{\psi}')$ is obtained by maximizing the log-likelihood function over $\lambda$ for a fixed value of $\psi$.

Based on the log-likelihood function, two familiar statistics can be derived for testing $\psi = \psi(\theta) = \psi_0$. These statistics are the Wald statistic ($q$)

$$q = (\hat{\psi} - \psi_0)\{j_{\psi\psi}(\hat{\theta})\}^{-1/2} \tag{2}$$

and the signed log-likelihood ratio statistic ($r$)

$$r = sgn(\hat{\psi} - \psi_0)[2\{l(\hat{\theta}) - l(\hat{\theta}_{\psi_0})\}]^{1/2}. \tag{3}$$

In terms of notation $j_{\psi\psi}(\hat{\theta})$ is represented in the estimated asymptotic variance of $\hat{\theta}$:

$$j_{\theta\theta}(\hat{\theta}) = \{j_{\theta\theta}(\hat{\theta})\}^{-1} = \begin{bmatrix} j_{\psi\psi}(\hat{\theta}) & j_{\psi\lambda}(\hat{\theta}) \\ j_{\lambda\psi}(\hat{\theta}) & j_{\lambda\lambda}(\hat{\theta}) \end{bmatrix}. \tag{4}$$

The matrix $j_{\theta\theta}(\theta)$ represents the information matrix which contains the second derivatives of the log-likelihood function:

$$j_{\theta\theta}(\theta) = \begin{bmatrix} -l_{\psi\psi}(\theta) & -l_{\psi\lambda}(\theta) \\ -l_{\lambda\psi}(\theta) & -l_{\lambda\lambda}(\theta) \end{bmatrix} = \begin{bmatrix} j_{\psi\psi}(\theta) & j_{\psi\lambda}(\theta) \\ j_{\lambda\psi}(\theta) & j_{\lambda\lambda}(\theta) \end{bmatrix}. \tag{5}$$

The statistics given in (2) and (3) have the standard normal as their limiting distribution and have an $O(n^{-1/2})$ rate of convergence. These approximations are thus accordingly known as first-order approximations. Tail probabilities for testing a particular value of $\psi$ can be approximated using either of these statistics with the cumulative standard normal distribution function $\Phi(\cdot)$, i.e. $\Phi(q)$ and $\Phi(r)$. The accuracy of these first-order methods suffers from the typical drawbacks of requiring a large sample size and an original distribution that is close to normal.
2.2 Higher-order approximations

Third-order tail probability approximations for testing a particular value of $\psi$ are derived by Barndorff-Nielsen (1991) and Lugannani and Rice (1980) using saddlepoint methods:

$$\Phi(r^*)$$, where

$$r^* = r - \frac{1}{r} \log \left( \frac{r}{Q} \right)$$

(4)

$$\Phi(r) + \phi(r) \left( \frac{1}{r} - \frac{1}{Q} \right)$$.

(5)

Approximation (4) is due to Barndorff-Nielsen (BN) and approximation (5) is due to Lugannani and Rice (LR). The statistic $r^*$ is known as the modified signed log-likelihood ratio statistic and the statistic $r$ is the signed log-likelihood ratio statistic defined in (3). The statistic $Q$ is a standardized maximum likelihood departure whose expression depends on the type of information available. Approximations (4) and (5) have an $O(n^{-3/2})$ rate of convergence and are thus referred to as third-order approximations. The function $\phi(\cdot)$ represents the density function of a standard normal.

Barndorff-Nielsen (1991) defined $Q$ for a suitably chosen ancillary statistic. However, outside of special classes of models, it may be very difficult to construct an appropriate ancillary statistic, if one even exists. And further, if an ancillary exists, it may not be unique. Fraser and Reid (1995) formally introduced a general third-order approach to determine an appropriate $Q$ in (4) and (5) in a general model context. Their approach can be used for inference on any one-dimensional parameter in any continuous model setting with standard asymptotic properties without the requirement of an explicit form for the ancillary statistic. Rather than requiring the existence of an ancillary statistic, their methodology requires only ancillary directions. Ancillary directions are defined as vectors tangent to an ancillary surface at the data point. The ancillary directions $V$ can be obtained as follows:

$$V = \left. \frac{\partial y}{\partial \theta'} \right|_{\hat{\theta}} = - \left. \left( \frac{\partial z}{\partial y'} \right)^{-1} \left( \frac{\partial z}{\partial \theta'} \right) \right|_{\hat{\theta}}$$.

(6)

The variable $z$ represents a pivotal quantity of the model whose distribution is independent of $\theta$. For instance, in the standard linear regression model, $y = X\beta + \sigma \epsilon$, the pivotal quantity is given by $z = (y - X\beta)/\sigma$.

These ancillary directions are used to calculate a locally defined canonical parameter, $\varphi$:

$$\varphi'(\theta) = \left[ \frac{\partial l(\theta; y)}{\partial y} \right] V.$$ 

(7)

Given this new reparameterization, the original parameter of interest must thus be recalibrated. This recalibration results in the new parameter $\chi$:

$$\chi(\theta) = \frac{\psi_{\varphi'}(\hat{\theta}_\psi)}{\psi_{\varphi'}(\hat{\theta}_\psi)} \varphi(\theta),$$

(8)

where $\psi_{\varphi'}(\theta) = \partial \psi(\theta)/\partial \varphi' = (\partial \psi(\theta)/\partial \theta')(\partial \varphi(\theta)/\partial \theta')^{-1}$. The modified maximum likelihood departure is
then constructed in the $\varphi$ space. The expression for $Q$ is given by

$$Q = \text{sgn}(\hat{\psi} - \psi) \left| \chi(\hat{\varphi}) - \chi(\varphi) \right| \left\{ \frac{|\hat{j}_{\varphi\varphi'}(\hat{\varphi})|}{|\hat{j}_{\lambda\lambda'}(\hat{\lambda})|} \right\}^{1/2},$$

where $\hat{j}_{\varphi\varphi'}$ and $\hat{j}_{\lambda\lambda'}$ are the observed information matrix evaluated at $\hat{\varphi}$ and observed nuisance information matrix evaluated at $\hat{\lambda}$, respectively, calculated in terms of the new $\varphi(\theta)$ reparameterization. The determinants can be computed as follows:

$$|\hat{j}_{\varphi\varphi'}(\hat{\varphi})| = |\hat{j}_{\theta\theta'}(\hat{\varphi})|^2$$
$$|\hat{j}_{\lambda\lambda'}(\hat{\lambda})| = |\hat{j}_{\lambda\lambda'}(\hat{\lambda})|^2.$$

The interested reader is directed to Fraser and Reid (1995) and Fraser, Reid and Wu (1999) for the mathematical details of the methodology. Given the overall objective was to obtain highly accurate tail probabilities for testing a specific value of the interest parameter $\psi$, we can use $Q$ and $r$ in formulas (4) and (5) to obtain two such probabilities. Finally, the p-value function, $p(\psi)$ obtained from (4) or (5) can be inverted to build a $(1 - \alpha)100\%$ confidence interval for $\psi$, defined by

$$\left( \min\{p^{-1}(\alpha/2), p^{-1}(1 - \alpha/2)\}, \max\{p^{-1}(\alpha/2), p^{-1}(1 - \alpha/2)\} \right)$$

The following section shows how the nonlinear regression model can be transformed to fit in the above framework and how the above techniques can then be applied to test first-order autocorrelation in the model disturbance terms.

### 3 Model Description and Inference

Consider the nonlinear regression model where the observations are taken through time

$$y_t = g(x_t; \beta) + u_t, \quad t = 1, \ldots, n$$

The disturbance term $u_t$ is assumed to follow a first-order autoregressive process (AR(1)):

$$u_t = \rho u_{t-1} + \epsilon_t, \quad t = 2, \ldots, n$$

The dependent variable $y_t$ is assumed to be strictly exogenous. The explanatory variables are denoted by $x_t = (x_{t1}, \ldots, x_{tk})'$ and the unknown coefficients are $\beta = (\beta_1, \ldots, \beta_k)'$. The regression function $g(\cdot)$ is a known real-valued measurable function. We assume that the regression function and parameters satisfy the regularity conditions for nonlinear regression models which are extensively discussed in the literature (see Amemiya (1985), pp.127-135).

The process $u_t$ is assumed to be stationary with $|\rho| < 1$. The random variable $\epsilon_t$ is independently and identically distributed with mean 0 and variance $\sigma^2$. In the presence of undetected autocorrelation, the
nonlinear least squares estimator or the nonlinear quasi maximum likelihood estimator of $\beta$ is inefficient (see Amemiya (1977); Gallant and Holly (1980)). In this framework, we are interested in determining whether autocorrelation exists in the disturbances in (12). We are therefore mainly testing the null hypothesis that the disturbance $u_t$ is serially uncorrelated, that is, $\rho = 0$, against the alternative hypothesis $\rho \neq 0$ or $\rho > 0$. Our framework however, allows us to test for general hypothesis of the form $\rho = \rho_0$ against $\rho \neq \rho_0$.

The nonlinear regression model (12) with error terms following an AR(1) process defined by (13), can be rewritten as follows:

$$y_t = g(x_t; \beta) + \rho(y_{t-1} - g(x_{t-1}, \beta)) + \epsilon_t, \quad \epsilon_t \sim \text{iid}(0, \sigma^2) \quad t = 2, \ldots, n$$

(14)

Conditional upon the first observation and under the assumption that $\epsilon_t$ is normally distributed, the conditional probability density function of the model is

$$f(y_2, \ldots, y_n | y_1; \rho, \beta, \sigma^2) = \prod_{t=2}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} \left\{\left\{ y_t - g(x_t; \beta) \right\} - \rho \left\{ y_{t-1} - g(x_{t-1}; \beta) \right\}\right]^2 \right] \right\}. \quad (15)$$

Using the fact that $\epsilon_t$’s are uncorrelated each other we can assume $u_1 = \frac{1}{\sqrt{1-\rho^2}} \epsilon_1$, where $\epsilon_1 \sim N(0, \sigma^2)$. This allows us to define the joint density function of the model by

$$f(y_1, \ldots, y_n; \rho, \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=2}^{n} \left\{ \left\{ y_t - g(x_t; \beta) \right\} - \rho \left\{ y_{t-1} - g(x_{t-1}; \beta) \right\}\right]^2 \right] \times \exp \left[ -\frac{1}{2\sigma^2} (1 - \rho^2) \left\{ y_1 - g(x_1; \beta) \right\}^2 \right].$$

This density function is used to construct the log-likelihood function:

$$l(\theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left\{ (1 - \rho^2) \left\{ y_1 - g(x_1; \beta) \right\}^2 + \sum_{t=2}^{n} \left\{ \left\{ y_t - g(x_t; \beta) \right\} - \rho \left\{ y_{t-1} - g(x_{t-1}; \beta) \right\}\right]^2 \right\},$$

(16)

where $\theta = (\rho, \beta, \sigma^2)$ denotes the parameter vector of interest. Note that the additive constant has been dropped from the log-likelihood function.

For a clearer understanding of our testing procedure, it is useful to rewrite our model (14) and the likelihood function (16) in reduced matrix formulations. The following notation is used for this purpose: $y = (y_1, \ldots, y_n)'$ is the vector of observations of the dependent variable $y_t$, $g(x, \beta) = (g(x_1, \beta), \ldots, g(x_n, \beta))'$ is the regression vector evaluated at $x = (x_1, \cdots, x_n)'$, and $\epsilon = (\epsilon_1, \cdots, \epsilon_n)'$ represents the vector of innovations. For $t = 2, \cdots, n$, we know from (14), that

$$y_t - \rho y_{t-1} = g(x_t; \beta) - \rho g(x_{t-1}, \beta) + \epsilon_t.$$  

(17)

\footnote{Note that $u_t - \rho u_{t-1} = \epsilon_t$ implies $u_t = (1 - \rho L)^{-1} \epsilon_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}$ Thus, by independence of $\epsilon_i$’s, $E u_1 = E \epsilon_1 = 0$ and $\text{Var} u_1 = \frac{1}{1 - \rho^2} \text{Var} \epsilon_1 = \frac{\sigma^2}{1 - \rho^2}$.}
This allows us to construct a matrix $B$ such that model (14) can be written as

$$By = Bg(x, \beta) + \epsilon,$$  

where for all $t \geq 2$, the $t^{th}$ row of $B$ has 1 in the $t^{th}$ position, $-\rho$ in the $(t-1)^{st}$ position and 0’s everywhere else. To account for the first observation, we set the first row of $B$ to have $\sqrt{1-\rho^2}$ in the first position, and 0’s everywhere else. Using this notation, the model can be rewritten in the reduced matrix form

$$y = g(x, \beta) + B^{-1}\epsilon,$$  

where $B$ is defined as the lower triangular matrix

$$B(\rho) = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix} \quad \text{and} \quad \epsilon \sim N(0, \sigma^2 I).$$

We can then obtain a matrix formulation for the log-likelihood function. To do so, we use model (18) to write $\epsilon = B\{y - g(x, \beta)\}$. We then have a new expression for the log-likelihood function (with the additive constant dropped)

$$l(\theta; y) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \{y - g(x, \beta)\}'A\{y - g(x, \beta)\},$$  

where $A$ is the tridiagonal positive definite matrix defined by

$$A(\rho) = B'B = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$  

The notations $B(\rho)$ and $A(\rho)$ are made to emphasize the dependence of the matrices $B$ and $A$ on the unknown parameter $\rho$.

The maximum likelihood estimator $\hat{\theta} = (\hat{\rho}, \hat{\beta}, \hat{\sigma}^2)$ of $\theta$ can be calculated by solving simultaneously the first-order conditions $l_\theta(\hat{\theta}) = 0$ using some iterative procedure. We denote by $\nabla g(x, \beta) = (\nabla_1 g(x_1, \beta), \ldots, \nabla_k g(x_n, \beta))$ the $k \times n$ matrix of partial derivatives of $g(x, \beta)$, where $\nabla_{i\beta} g(x_t, \beta) = \left(\frac{\partial g(x_t, \beta)}{\partial \beta_1}, \ldots, \frac{\partial g(x_t, \beta)}{\partial \beta_k}\right)'$ is the
\( k \times 1 \) gradient of \( g(x_t; \beta) \). The first-order conditions for the MLE \( \hat{\theta} \) are given by:

\[
\begin{align*}
l_\rho(\hat{\theta}; y) &= -\frac{1}{2\sigma^2} \{ y - g(x, \hat{\beta}) \}' \hat{A}_\rho \{ y - g(x, \hat{\beta}) \} = 0 \\
l_\beta(\hat{\theta}; y) &= \frac{1}{\sigma^2} \nabla_\beta g(x, \hat{\beta}) \hat{A} \{ y - g(x, \hat{\beta}) \} = 0 \\
l_\sigma^2(\hat{\theta}; y) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \{ y - g(x, \hat{\beta}) \}' \hat{A} \{ y - g(x, \hat{\beta}) \} = 0,
\end{align*}
\]

(21)

where \( \hat{B} = B(\hat{\rho}), \hat{A} = A(\hat{\rho}) \) and \( \hat{A}_\rho = \left. \frac{\partial A(\rho)}{\partial \rho} \right|_{\rho=\hat{\rho}} \).

The most commonly used statistic for testing the hypothesis \( \rho = \rho_0 \) is the standardized test statistic given by (2) which in this context can be rewritten as

\[
STS = n^{1/2}(\hat{\rho} - \rho_0)(1 - \hat{\rho}^2)^{-1/2}.
\]

(22)

The statistic proposed by Kobayashi (1991) is a bias-corrected test statistic from the standardized one and is given by

\[
K = n^{1/2}(\hat{\rho} - \rho_0) + n^{-1/2}\text{trace}(\hat{J}^{-1} \hat{H}),
\]

(23)

where \( \hat{J} \) and \( \hat{H} \) are the \( k \times k \) matrices defined by

\[
\hat{J} = \sum_{t=1}^{n} \nabla_\beta \hat{B} g(x_t, \hat{\beta}) \nabla_\beta' \hat{B} g(x_t, \hat{\beta}) \text{ and } \hat{H} = \sum_{t=1}^{n} \nabla_\beta \hat{B} g(x_t, \hat{\beta}) \nabla_\beta' \hat{g}(x_{t-1}, \hat{\beta}).
\]

In practice, \( \hat{J}^{-1} \hat{H} \) can be easily obtained by regressing the columns of the \( n \times k \) matrix \( \nabla_\beta \hat{B} g(x_t, \hat{\beta}) \) upon the columns of the \( n \times k \) matrix \( \nabla_\beta' \hat{B} g(x_t, \hat{\beta}) \).\(^2\)

Let \( \nabla_{\beta \beta'} g(x_t, \beta) \) denotes the \( k \times k \) hessian matrix of \( g(x_t, \beta) \), \( t = 1, \cdots, n \). The information matrix \( j_{\theta \theta'}(\hat{\theta}) \) of our model can be obtained by calculating the second-order derivatives of the log-likelihood function:

\[
\begin{align*}
l_{\rho \rho}(\hat{\theta}; y) &= -\frac{1}{2\sigma^2} \{ y - g(x, \hat{\beta}) \}' \hat{A}_{\rho \rho} \{ y - g(x, \hat{\beta}) \} \\
l_{\rho \beta}(\hat{\theta}; y) &= \frac{1}{\sigma^2} \nabla_\beta g(x, \hat{\beta}) \hat{A}_{\rho} \{ y - g(x, \hat{\beta}) \} \\
l_{\rho \sigma^2}(\hat{\theta}; y) &= -\frac{1}{2\sigma^2} \{ y - g(x, \hat{\beta}) \}' \hat{A}_{\rho} \{ y - g(x, \hat{\beta}) \} \\
l_{\beta \beta}(\hat{\theta}; y) &= -\frac{1}{\sigma^2} \nabla_\beta \hat{g}(x, \hat{\beta}) \hat{A} \nabla_\beta \hat{g}(x, \hat{\beta}) + \frac{1}{\sigma^4} \sum_{t=1}^{n} \nabla_{\beta \beta'} g(x_t, \hat{\beta}) \hat{\omega}_t \\
l_{\beta \sigma^2}(\hat{\theta}; y) &= -\frac{1}{2\sigma^4} \nabla_\beta \hat{g}(x, \hat{\beta}) \hat{A} \{ y - g(x, \hat{\beta}) \} \\
l_{\sigma^2 \sigma^2}(\hat{\theta}; y) &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^4} \{ y - g(x, \hat{\beta}) \}' \hat{A} \{ y - g(x, \hat{\beta}) \},
\end{align*}
\]

where \( \hat{\omega}_t \) is the \( t^{th} \) element of the \( n \)-vector \( \hat{\omega} = \hat{A} \{ y - g(x, \hat{\beta}) \} \).

\(^2\)Note that in its original formulation the Kobayashi (1991) test was designed to test for \( \rho = 0 \); we adapted it in this framework to test for a more general null hypothesis of the form \( \rho = \rho_0 \).
For the reparametrization \( \varphi(\theta) \) given in (7) specific to the data point \( y = (y_1, \ldots, y_n)' \), the sample space gradient of the likelihood is required and obtained as follows:

\[
\frac{\partial l(\theta, y)}{\partial y} = -\frac{1}{\sigma^2} \{y - g(x, \beta)\}' A. \tag{24}
\]

To obtain the ancillary directions \( V \) in (6) we consider the following full dimensional pivotal quantity:

\[
z(y; \theta) = B \{y - g(x, \beta)\} / \sigma. \tag{25}
\]

The quantity \( z(y; \beta) \) coincides with \( \epsilon / \sigma \) which is clearly a vector of independent standard normal deviates. The ancillary directions are then obtained as follows:

\[
V = \left\{ \frac{\partial z(y; \theta)}{\partial y} \right\}^{-1} \left\{ \frac{\partial z(y; \theta)}{\partial \theta} \right\}_{\hat{\theta}}
= \left( \hat{B}^{-1} \hat{B}_\rho \{y - g(x; \hat{\beta})\}; -\nabla_\beta g(x; \hat{\beta}); -\frac{1}{\sigma} \{y - g(x, \hat{\beta})\} \right).
\]

We can now use (24) and (26) to obtain our new locally defined parameter

\[
\varphi(\theta)' = \frac{\partial l(\theta, y)}{\partial y} \cdot V = [\varphi_1(\theta); \varphi_2(\theta); \varphi_3(\theta)], \tag{27}
\]

with

\[
\begin{align*}
\varphi_1(\theta) &= -\frac{1}{\sigma^2} \{y - g(x, \beta)\}' A \hat{B}^{-1} \hat{B}_\rho \{y - g(x; \hat{\beta})\} \\
\varphi_2(\theta) &= \frac{1}{\sigma^2} \{y - g(x, \beta)\}' A \nabla_\beta g(x; \hat{\beta}) \\
\varphi_3(\theta) &= \frac{1}{\sigma \hat{\sigma}^2} \{y - g(x, \beta)\}' A \{y - g(x, \hat{\beta})\}.
\end{align*}
\]

The parameters \( \varphi_1(\theta) \), \( \varphi_2(\theta) \), and \( \varphi_3(\theta) \) have dimensions \((1 \times 1)\), \((1 \times k)\) and \((1 \times 1)\) respectively, so that the overall parameter \( \varphi(\theta)' \) is a \((1 \times (k + 2))\) - vector. This therefore achieves a dimension reduction from \( n \), the dimension of \( y \), to \( k + 2 \), the dimension of \( \theta \).

Denote \( \psi(\theta) = \rho \), and \( \theta' = (\rho, \beta', \sigma^2) = (\psi, \lambda') \). The corresponding nuisance parameter is then \( \lambda' = (\beta', \sigma^2) \). Our next step is to reduce the dimension of the problem from \( k + 2 \) to 1, the dimension of the parameter of interest \( \psi(\theta) \). For this purpose, we need to define a new scalar parameter \( \chi(\theta) \) which in turn involves the parameter vector \( \varphi(\theta) \) as well as the constrained maximum likelihood estimator \( \hat{\theta}_\psi \).

We derive the constrained MLE, \( \hat{\theta}_\psi = (\rho, \hat{\beta}_\psi, \hat{\sigma}^2_\psi) \), by maximizing the log-likelihood function given by (20) with respect to \( \beta \) and \( \sigma^2 \) under the condition that \( \psi = \rho = 0 \). This leads to the following expression for the constrained log-likelihood:

\[
l(\theta_\psi; y) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \{y - g(x, \beta)\}' \{y - g(x, \beta)\}. \tag{29}
\]
The corresponding first-order conditions are
\[
\begin{align*}
  l_{\beta,\psi}(\hat{\theta}; y) &= \frac{1}{\sigma^2_{\psi}} \nabla_{\beta} g(x, \hat{\beta}_\psi) \{ y - g(x, \hat{\beta}_\psi) \} = 0 \\
  l_{\sigma^2,\psi}(\hat{\theta}; y) &= -\frac{n}{2\sigma^2_{\psi}} + \frac{1}{2\sigma^2_{\psi}} \{ y - g(x, \hat{\beta}_\psi) \} \{ y - g(x, \hat{\beta}_\psi) \} = 0.
\end{align*}
\]

Similar to the overall likelihood case, we can construct the observed constrained information matrix \( j_{\lambda^\prime}(\hat{\theta}_\psi) \) using the following second-order derivatives:
\[
\begin{align*}
  l_{\beta,\beta}(\hat{\theta}; y) &= -\frac{1}{\sigma^2_{\psi}} \nabla_{\beta} g(x, \hat{\beta}_\psi) \nabla_{\beta} g(x, \hat{\beta}_\psi) + \frac{1}{\sigma^2_{\psi}} \sum_{t=1}^{n} \nabla_{\beta} g(x_t, \hat{\beta}_\psi) \cdot \{ y_t - g(x_t; \hat{\beta}_\psi) \} \\
  l_{\beta,\sigma^2}(\hat{\theta}; y) &= -\frac{1}{\sigma^2_{\psi}} \nabla_{\beta} g(x, \hat{\beta}_\psi) \{ y - g(x, \hat{\beta}_\psi) \} \\
  l_{\sigma^2,\sigma^2}(\hat{\theta}; y) &= \frac{n}{2\sigma^4_{\psi}} \{ y - g(x, \hat{\beta}_\psi) \} \{ y - g(x, \hat{\beta}_\psi) \}.
\end{align*}
\]

Note that when testing for an arbitrary fixed value for \( \rho \), say \( \rho = \rho_0 \) (instead of \( \rho = 0 \) as above), the last two equations of system (21) give the constrained maximum likelihood estimates of \( \beta \) and \( \sigma^2 \), with the appropriate change of \( \hat{A} \) to \( A(\rho_0) \).

To construct \( \chi(\theta) \) given in (8), in addition to \( \varphi(\theta) \) we require the quantity \( \psi_{\varphi'}(\hat{\theta}_\psi) \). We can write
\[
\psi_{\varphi'}(\hat{\theta}_\psi) = \left\{ \frac{\partial \varphi(\theta)}{\partial \theta'} \right\} \left\{ \frac{\partial \varphi(\theta)}{\partial \theta} \right\}^{-1} \Big|_{\hat{\theta}_\psi}.
\]

On the one hand we have
\[
\frac{\partial \varphi(\theta)}{\partial \theta'} = \varphi_{\theta'}(\theta) = \begin{pmatrix}
\frac{\partial \varphi_1(\theta)}{\partial \rho} & \frac{\partial \varphi_1(\theta)}{\partial \beta} & \frac{\partial \varphi_1(\theta)}{\partial \sigma^2} \\
\frac{\partial \varphi_2(\theta)}{\partial \rho} & \frac{\partial \varphi_2(\theta)}{\partial \beta} & \frac{\partial \varphi_2(\theta)}{\partial \sigma^2} \\
\frac{\partial \varphi_3(\theta)}{\partial \rho} & \frac{\partial \varphi_3(\theta)}{\partial \beta} & \frac{\partial \varphi_3(\theta)}{\partial \sigma^2}
\end{pmatrix},
\]

and on the other hand we have
\[
\frac{\partial \psi(\theta)}{\partial \theta'} = \left[ \frac{\partial \psi(\theta)}{\partial \rho}; \frac{\partial \psi(\theta)}{\partial \beta}; \frac{\partial \psi(\theta)}{\partial \sigma^2} \right] = [1; 0_{1 \times k}; 0].
\]

The surrogate parameter replacing \( \rho \) is then
\[
\chi(\theta) = \frac{[1; 0, \ldots, 0] \varphi_{\theta'}^{-1}(\hat{\theta}_\psi)}{||1; 0, \ldots, 0|| \varphi_{\theta'}^{-1}(\hat{\theta}_\psi)} \cdot \varphi(\theta).
\]

Note that the first term on the right hand side of equation (31) is \( \psi_{\varphi'}(\theta_\psi)/|\psi_{\varphi'}(\theta_\psi)| \) which is a \((1 \times (k + 2))\)-vector, so that \( \chi(\theta) \) is a scalar, since \( \varphi(\theta) \) is of dimension \((k + 2) \times 1\).

With these calculations, the modified maximum likelihood departure measure \( Q \) can then be obtained as
in (9) with determinants given in (10). The p-value of the test is given by combining $Q$ and the standard signed log-likelihood ratio statistic in (3) using the Barndorff-Nielsen (1991) or the asymptotically equivalent Lugannani and Rice (1980) expressions given by (4) and (5).

4 Monte Carlo simulation results

In this section, we perform two Monte Carlo simulation studies to gain a practical understanding of the p-values obtained by our testing procedure. The focus is to compare the results from the the third-order inference approach to the standardized asymptotic test statistic (STS) given by (2) or (22), the signed log-likelihood departure ($r$) derived from the log-likelihood function given in (3), and the Kobayashi (1991) test statistic (K) given in (23). The Barndorff-Nielsen (1991) and the Lugannani and Rice (1980) methods are respectively denoted BN and LR.

The accuracy of the different methods are evaluated by computing confidence intervals for $\rho$ in each case and using the following criteria:

- Coverage Probability: The proportion of the true parameter value falling within the confidence intervals.
- Coverage Error: The absolute difference between the nominal level and the coverage probability.
- Upper probability Error: The proportion of the true parameter value falling above the upper limit of the confidence interval.
- Lower Probability Error: The proportion of the true parameter value falling below the lower limit of the confidence interval.
- Average Bias: The average of the absolute difference between the upper and the lower probability errors and their nominal levels.

The above criteria were also considered by Rekkas et al. (2008) and Chang and Wong (2010).

4.1 Simulation study 1

Our first model is the one considered by Kobayashi (1991). The dependent variable $y_t$ is generated by the model:

$$y_t = \exp(x_t \beta) + u_t, \quad t = 1, \ldots, n$$

The disturbances $u_t$ are generated by an AR(1) process of the form

$$u_t = \rho u_{t-1} + \epsilon_t, \quad t = 2, \ldots, n$$

where the $\epsilon_t$ are distributed independently as standard normal. The sample sizes are set at $n = 10$, $n = 20$ and $n = 40$. We consider various values for the autocorrelation coefficient $\rho \in \{-0.9; -0.5; 0; 0.5; 0.9\}$ and we
compute 90%, 95% and 99% confidence intervals for $\rho$. We report only the results for a sample size of $n = 20$ and 95% confidence intervals.\(^3\) The explanatory variable $x_t$ is the realization of the first-order autoregressive process with the autocorrelation coefficient 0.9 and the variance $1/(1 - 0.9^2)$. The coefficient $\beta$ is set at 1/3.

For each sample size, 10,000 simulations are generated to compute the p-values for the testing methods: $r$, BN, STS, LR and K. The coefficients of the model are estimated by numerical maximization of log-likelihood functions.

Table 1: Results for Simulation Study 1 for $n=20$, 95% CI

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Method</th>
<th>Coverage Probability</th>
<th>Coverage Error</th>
<th>Upper Probability</th>
<th>Lower Probability</th>
<th>Average Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>$r$</td>
<td>0.9359</td>
<td>0.0141</td>
<td>0.0147</td>
<td>0.0494</td>
<td>0.0173</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9452</td>
<td>0.0048</td>
<td>0.0264</td>
<td>0.0284</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.9029</td>
<td>0.0471</td>
<td>0.0970</td>
<td>0.0001</td>
<td>0.0485</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9410</td>
<td>0.0090</td>
<td>0.0301</td>
<td>0.0289</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.8891</td>
<td>0.0609</td>
<td>0.1108</td>
<td>0.0001</td>
<td>0.0554</td>
</tr>
<tr>
<td>-0.5</td>
<td>$r$</td>
<td>0.9375</td>
<td>0.0125</td>
<td>0.0199</td>
<td>0.0426</td>
<td>0.0113</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9438</td>
<td>0.0062</td>
<td>0.0278</td>
<td>0.0284</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.9395</td>
<td>0.0105</td>
<td>0.0471</td>
<td>0.0134</td>
<td>0.0168</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9411</td>
<td>0.0089</td>
<td>0.0296</td>
<td>0.0293</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.9335</td>
<td>0.0165</td>
<td>0.0602</td>
<td>0.0063</td>
<td>0.0269</td>
</tr>
<tr>
<td>0</td>
<td>$r$</td>
<td>0.9377</td>
<td>0.0123</td>
<td>0.0212</td>
<td>0.0411</td>
<td>0.0100</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9496</td>
<td>0.0004</td>
<td>0.0250</td>
<td>0.0254</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.9261</td>
<td>0.0239</td>
<td>0.0279</td>
<td>0.0460</td>
<td>0.0119</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9460</td>
<td>0.0040</td>
<td>0.0277</td>
<td>0.0263</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.9293</td>
<td>0.0207</td>
<td>0.0388</td>
<td>0.0319</td>
<td>0.0103</td>
</tr>
<tr>
<td>0.5</td>
<td>$r$</td>
<td>0.9341</td>
<td>0.0159</td>
<td>0.0274</td>
<td>0.0385</td>
<td>0.0079</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9453</td>
<td>0.0047</td>
<td>0.0283</td>
<td>0.0264</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.8842</td>
<td>0.0658</td>
<td>0.0182</td>
<td>0.0976</td>
<td>0.0397</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9432</td>
<td>0.0068</td>
<td>0.0295</td>
<td>0.0273</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.8917</td>
<td>0.0583</td>
<td>0.0285</td>
<td>0.0798</td>
<td>0.0291</td>
</tr>
<tr>
<td>0.9</td>
<td>$r$</td>
<td>0.9342</td>
<td>0.0158</td>
<td>0.0409</td>
<td>0.0249</td>
<td>0.0080</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9459</td>
<td>0.0041</td>
<td>0.0262</td>
<td>0.0279</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.8694</td>
<td>0.0806</td>
<td>0.0001</td>
<td>0.1305</td>
<td>0.0653</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9428</td>
<td>0.0072</td>
<td>0.0287</td>
<td>0.0285</td>
<td>0.0036</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.8810</td>
<td>0.0690</td>
<td>0.0003</td>
<td>0.1187</td>
<td>0.0592</td>
</tr>
</tbody>
</table>

Table 1 reports the results for the case of 95% confidence intervals and a sample size of $n = 20$. For a 95% CI, the nominal levels for coverage probability, coverage error, upper and lower probability errors, and average bias are given as 0.95, 0, 0.025, 0.025, and 0, respectively. Figure 1 gives a graphical illustration of the behavior of the tests. Based on our criteria, the standardized test statistic (STS) is the least satisfactory method. The signed log-likelihood method ($r$) gives better coverage probability than the Kobayashi method (K). The third-order methods BN and LR produce excellent results. They give upper and lower probabilities that are fairly symmetric with relatively small average biases. Unreported simulations show that for smaller sample sizes like $n = 10$, the BN and LR behave even better relative to the other methods. It can

\(^3\)Results for all simulations performed are available from the authors.
also be seen from Figure 1 that the proposed methods give results that are consistently close to the nominal values and are stable, while the other methods, particularly the STS and K statistics, are less satisfactory especially as the values of \( \rho \) get closer to the boundaries.

Figure 1: Graphical illustration of the results of Simulation Study 1

4.2 Simulation study 2

We now consider the model defined by

\[
y_t = \frac{\beta_1}{1 + e^{\beta_2 + \beta_3 x_t}} + u_t, \quad t = 1, \ldots, n
\]
The disturbances \( u_t \) are generated by an AR(1) process of the form

\[
\begin{align*}
u_t &= \rho u_{t-1} + \epsilon_t, & t = 2, \ldots, n
\end{align*}
\]

Like in the previous study, \( \epsilon_t \) is assumed to be distributed independently as standard normal and the explanatory variable \( x_t \) is a size \( n \) vector of deterministic integers ranging from 1 to \( n \). The parameter values are given as \( \beta_1 = 439, \beta_2 = 4.7, \) and \( \beta_3 = -0.3 \).

We consider 10,000 replications with sample sizes \( n = 10 \) and \( n = 20 \) where the autocorrelation coefficient takes values \( \rho \in \{-0.9; -0.5; 0; 0.5; 0.9\} \). We compute 90\%, 95\% and 99\% confidence intervals. Table 2 records only the simulation results for 95\% confidence interval and sample size \( n = 20 \) while Figure 2 graphically illustrates the behavior of the different methods.⁴

![Table 2](image)

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>Method</th>
<th>Coverage Probability</th>
<th>Coverage Error Probability</th>
<th>Upper Probability</th>
<th>Lower Probability</th>
<th>Average Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>r</td>
<td>0.9325</td>
<td>0.0175</td>
<td>0.0094</td>
<td>0.0581</td>
<td>0.0243</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9462</td>
<td>0.0038</td>
<td>0.0282</td>
<td>0.0256</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.9172</td>
<td>0.0328</td>
<td>0.0828</td>
<td>0.0000</td>
<td>0.0414</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9451</td>
<td>0.0049</td>
<td>0.0280</td>
<td>0.0269</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.889</td>
<td>0.0610</td>
<td>0.1110</td>
<td>0.0000</td>
<td>0.0555</td>
</tr>
<tr>
<td>-0.5</td>
<td>r</td>
<td>0.9380</td>
<td>0.0120</td>
<td>0.0110</td>
<td>0.0510</td>
<td>0.0200</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9449</td>
<td>0.0051</td>
<td>0.0283</td>
<td>0.0268</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.9550</td>
<td>0.0050</td>
<td>0.0260</td>
<td>0.0190</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9440</td>
<td>0.0060</td>
<td>0.0285</td>
<td>0.0275</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.9400</td>
<td>0.0100</td>
<td>0.0550</td>
<td>0.0050</td>
<td>0.0250</td>
</tr>
<tr>
<td>0</td>
<td>r</td>
<td>0.9150</td>
<td>0.0350</td>
<td>0.0060</td>
<td>0.0790</td>
<td>0.0365</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9450</td>
<td>0.0050</td>
<td>0.0274</td>
<td>0.0276</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.9030</td>
<td>0.0470</td>
<td>0.0120</td>
<td>0.0850</td>
<td>0.0365</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9440</td>
<td>0.0060</td>
<td>0.0279</td>
<td>0.0281</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.9250</td>
<td>0.0250</td>
<td>0.0350</td>
<td>0.0400</td>
<td>0.0125</td>
</tr>
<tr>
<td>0.5</td>
<td>r</td>
<td>0.9090</td>
<td>0.0410</td>
<td>0.0070</td>
<td>0.0840</td>
<td>0.0385</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9435</td>
<td>0.0065</td>
<td>0.0278</td>
<td>0.0287</td>
<td>0.0032</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.7820</td>
<td>0.1680</td>
<td>0.0110</td>
<td>0.2070</td>
<td>0.0980</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9420</td>
<td>0.0080</td>
<td>0.0285</td>
<td>0.0295</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.8460</td>
<td>0.1040</td>
<td>0.0240</td>
<td>0.1300</td>
<td>0.0530</td>
</tr>
<tr>
<td>0.9</td>
<td>r</td>
<td>0.8990</td>
<td>0.0510</td>
<td>0.0170</td>
<td>0.0840</td>
<td>0.0335</td>
</tr>
<tr>
<td></td>
<td>BN</td>
<td>0.9390</td>
<td>0.0110</td>
<td>0.0290</td>
<td>0.0320</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>STS</td>
<td>0.6760</td>
<td>0.2740</td>
<td>0.0000</td>
<td>0.3240</td>
<td>0.1620</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.9340</td>
<td>0.0160</td>
<td>0.0310</td>
<td>0.0350</td>
<td>0.0080</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.7130</td>
<td>0.2370</td>
<td>0.0010</td>
<td>0.2860</td>
<td>0.1425</td>
</tr>
</tbody>
</table>

The results essentially bear the same features as in Simulation Study 1. Based on our comparative criteria, the performance of the third-order methods are again superior to all the other methods. They give results very close to the nominal values even for values of the autocorrelation coefficient close to the boundaries.

---

⁴Results for all simulations performed are available from the authors.
whereas the $r$, STS and K methods give less satisfactory results. On the whole, the two simulation studies show that significantly improved accuracy can be obtained for testing first-order autocorrelated disturbances in nonlinear regression models using third-order likelihood methods.\footnote{It is useful to note that both $q$ and $r$ in (2) and (3) are close to zero if $\hat{\rho}$ is near $\rho$. This generates a singularity point in (4) and (5) and poses obvious numerical difficulties to the evaluation of the BN and LR p-values when it arises. These singularities are however correctly handled in practice by implementing the bridging methods proposed by Fraser et al. (2003).} In the next section we apply these methods to obtain p-values and confidence intervals for autocorrelation parameter in an empirical example.

5 Empirical Example

The behavior of the third-order approximation tests for autocorrelation in nonlinear regression models has been simulated in the preceding sections to confirm theoretical accuracy. In this section an empirical example...
is presented to show the usefulness of the method. The example is a model of population growth common
in demographic studies. This example is taken from Fox (2002). The population size \(y_t\) is assumed to follow
a logistic model defined by

\[ y_t = \frac{\beta_1}{1 + e^{\beta_2 + \beta_3 t}} + u_t, \quad t = 1, \ldots, n \]  

(32)

The parameter \(\beta_1\) represents the asymptote toward which the population grows, the parameter \(\beta_2\) reflects
the relative size of the population at time \(t = 0\) and the parameter \(\beta_3\) accounts for the growth rate of the
population.

This model is the same as the one examined with simulated data in Simulation Study 2 above. We
estimate it using census population data for the United States from 1790 through 2010. To fit this population
growth model to the US data, we use numerical optimization that requires reasonable start values for the
parameters. A good way to choose those values use insightful information about the structure of the data.
The US Census Bureau projects the population to reach 439 millions by 2050.\(^6\) One can thus choose this
value to be the start value of the asymptotic population, that is, \(\beta_1^0 = 439\). Then using the same reasoning
as in Fox (2002) we can derive \(\beta_2^0 = 4.7\) and \(\beta_3^0 = -0.3\).

Table 3 reports the naive estimation of the logistic model. Figure 3(a) gives the plots of the population
over time where the line in the figure represents the fit of the logistic population growth model. A graphical
analysis of the residuals plot depicted by Figure 3(b) suggests the presence of positive autocorrelation in the
error terms of the model.

Table 3: Results for the logistic regression of model (32) with \(u_t \sim IID(0, \sigma_u^2)\)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_1)</td>
<td>484.7947</td>
<td>35.3013</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>4.0652</td>
<td>0.0641</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>-0.2084</td>
<td>0.0089</td>
</tr>
<tr>
<td>(\sigma_u^2)</td>
<td>22.7796</td>
<td>17.4921</td>
</tr>
</tbody>
</table>

To further investigate this observation, we assume that the disturbance term \(u_t\) is an AR(1) process, that is,

\[ u_t = \rho u_{t-1} + \epsilon_t, \quad t = 2, \ldots, n, \quad \text{with} \quad \epsilon_t \sim N(0, \sigma^2). \]  

(33)

We then compute the Maximum Likelihood Estimators of the model assuming autocorrelated errors. The
results of the estimation are given in Table 4. Figure 4(a) pictures the population together with the fitted
values whereas Figure 4(b) depicts the MLE residuals plotted.

It can be seen that while the estimators used to generate the results in Tables 3 and 4 produce somewhat compar-
able values, the maximum likelihood estimation produces smaller standard errors for the estimates. The

MLE estimates are then used to obtain the p-values and build the confidence intervals for the autocorrelation test as described in the previous sections.

Table 5 reports the 95% confidence intervals for $\rho$ obtained from our five methods. The resultant confidence intervals display considerable variation and could lead to different inferences about $\rho$. The first-order methods, STS, $r$, and $K$, give wider confidence intervals compared to the confidence intervals produced from the two third-order methods, BN and LR. The latter two intervals are theoretically more accurate. It is instructive to note however, that the residuals depicted in Figure 4(b) indicate that the nature of the autocorrelation structure in the model goes beyond first-order and suggests higher-order structures be considered. Developing higher-order inference methods for general AR processes in the nonlinear regression model context is left for future research endeavors.
6 Concluding remarks

This paper used third-order likelihood theory to derive highly accurate p-values for testing first-order autocorrelated disturbances in nonlinear regression models. Two Monte Carlo experiments were carried out to compare the performance of the proposed third-order methods BN and LR with those of the first-order methods: r, STS, and K. The results show that the proposed methods produce higher order improvements over the first-order methods and are superior in terms of the simulation criteria we considered. In particular, the results indicate that although the STS may be a commonly used method, applying it to small- and medium-sized samples could be seriously misleading. To illustrate the implementation of the proposed method an empirical example using the logistic population growth model applied to US population census data was provided.
References


