Improved Inference for First Order Autocorrelation using Likelihood Analysis

M. Rekkas* Y. Sun[†] A. Wong[‡]

Abstract

Testing for first-order autocorrelation in small samples using the standard asymptotic test can be seriously misleading. Recent methods in likelihood asymptotics are used to derive more accurate p-value approximations for testing the autocorrelation parameter in a regression model. The methods are based on conditional evaluations and are thus specific to the particular data obtained. A numerical example and three simulations are provided to show that this new likelihood method provides higher order improvements and is superior in terms of central coverage even for autocorrelation parameter values close to unity.

Keywords: Likelihood analysis; p-value; Autocorrelation

*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

[†]Graduate student. Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, ON M3J 1P3

[‡]Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, ON M3J 1P3

We are indebted to D.A.S. Fraser for helpful discussions and to an anonymous referee for many valuable suggestions. Rekkas and Wong gratefully acknowledge the support of the National Sciences and Engineering Research Council of Canada.

1 Introduction

Testing for first-order autocorrelation in small samples using the standard asymptotic test can be seriously misleading, especially for (absolute) values of the autocorrelation parameter close to one. While the past two decades have seen a tremendous advancement in the theory of small-sample likelihood asymptotic inference methods, their practical implementation has significantly lagged behind despite their exceptionally high accuracy compared to traditional first-order asymptotic methods. In this paper, inference for the autocorrelation parameter of a first-order model is considered. Recent developments in likelihood asymptotic theory are used to obtain p-values that more accurately assess the parameter of interest.

Consider the multiple linear regression model

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + \varepsilon_t, \quad t = 1, 2, \dots, n \quad (1)$$

with an autoregressive error structure of order one [AR(1)]

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t. \quad (2)$$

The random variables, v_t , are independently normally distributed with $E[v_t] = 0$ and $E[v_t^2] = \sigma^2$. Throughout this paper the process for ε_t is assumed to be stationary so that the condition $|\rho| < 1$ holds. Further, the independent variables in the model are considered to be strictly exogenous.

An alternative way to present the multiple linear regression model with AR(1) Gaussian error structure is as follows:

$$y = X\beta + \sigma\varepsilon, \quad \varepsilon \sim N(0, \Omega),$$

where

$$\Omega = ((\omega_{ij})), \quad \omega_{ij} = \frac{\rho^{|i-j|}}{1 - \rho^2} \quad i, j = 1, 2, \dots, n$$

and

$$\begin{aligned} y &= (y_1, y_2, \dots, y_n)^T \\ X &= \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \\ \beta &= (\beta_0, \beta_1, \dots, \beta_k)^T \\ \varepsilon &= (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T. \end{aligned}$$

It is well known that in the presence of autocorrelation, the ordinary least squares estimator (OLS) for

β , defined as $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$, is not the best linear unbiased estimator of β . To determine whether autocorrelation exists in time series data, the null hypothesis of $\rho = 0$ is tested against a two-sided or one-sided alternative. If the null hypothesis cannot be rejected at conventional statistical levels, the estimation of the unknown parameters is carried through using OLS, if, on the other hand, the null hypothesis can be rejected alternative estimation techniques must be used.

Two common tests appearing in standard textbooks for assessing the autocorrelation parameter, ρ , are an asymptotic test and the Durbin-Watson test (see for example (Wooldridge, 2006)).¹ The asymptotic test uses the OLS residuals from regression model (1),

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)^T = y - X\hat{\beta}_{OLS}$$

to estimate ρ from the regression

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + \tilde{\nu}_t.$$

The standardized test statistic for testing $\rho = \rho_0$ is constructed as

$$z = \frac{\hat{\rho} - \rho_0}{\sqrt{(1 - \rho_0^2)/n}}. \quad (3)$$

This random variable is distributed asymptotically as standard normal. The Durbin-Watson test, for testing the hypothesis $\rho = 0$, uses the same OLS residuals to construct another test statistic

$$d = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}. \quad (4)$$

The distribution of d under the null hypothesis depends on the design matrix; formal critical bounds have been tabulated by Durbin and Watson (1951). However, as it can be shown that the value of d is bound from below by zero and from above by four, a value of d close to two does not suggest the presence of autocorrelation, while a value close to zero suggests positive autocorrelation and a value close to four suggests negative autocorrelation. The test has an inconclusive region for both alternative hypotheses. As the Durbin-Watson test is restricted to testing an autocorrelation parameter equal to zero in AR(1) models, this statistic will not be the focus in this paper. Distortions of this statistic in small samples however, have been noted (see Belsley (1997)).

Asymptotic inference for ρ can also be obtained from some simple likelihood-based asymptotic methods. See for example Hamilton (1994). For the Gaussian AR(1) model, two different likelihood functions for $\theta = (\beta, \rho, \sigma^2)^T$ can be constructed depending on the assumptions about the first observation. If the first response, y_1 , is treated as fixed (non-random), the corresponding conditional log-likelihood function is given

¹The literature on testing this parameter is vast, a survey is presented in King (1987). Applications using marginal likelihood for time series models have been developed and applied by Levenbach (1972), Tunnicliffe-Wilson (1989), Cheang and Reinsel (2000), and Reinsel and Cheang (2003). These authors have shown that the use of REML in the time series context has been successful.

by

$$\log \prod_{t=2}^n f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \beta, \rho, \sigma^2), \quad (5)$$

where the conditional distribution of $f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \theta)$ is normal. If, on the other hand, the first response is treated as a random variable, the exact log-likelihood function is given as

$$\log[f_{Y_1}(y_1; \beta, \rho, \sigma^2) \cdot \prod_{t=2}^n f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \beta, \rho, \sigma^2)], \quad (6)$$

where $f_{Y_1}(y_1; \theta)$ is the normal density of the first observation. Notice in comparison, the conditional log-likelihood function given in (5) uses only $(n - 1)$ observations. Using the log-likelihood functions defined in (5) and (6), standard large sample theory can be applied to obtain test statistics for conducting inference on ρ .

In this paper, inference concerning the autocorrelation parameter is examined from the viewpoint of recent likelihood asymptotics. The general theory developed by Fraser and Reid (1995) will be used to obtain p-values for testing particular values of ρ that have known $O(n^{-3/2})$ distributional accuracy. This theory will be discussed in some detail in Section 2.2 below. The focus of this paper is on comparing the results from this approach to the asymptotic test given in (3) and to the signed log-likelihood departure derived from the unconditional log-likelihood function given in (6). A numerical example and three simulations will be provided to show the extreme accuracy of this new likelihood method even for (absolute) values of autocorrelation parameter close to one.

The structure of the paper is as follows. Likelihood asymptotics are presented in Section 2. Third-order inference for the first-order autocorrelation model is given in Section 3. Simulations and examples are recorded in Section 4. Section 5 concludes and gives suggestions for further research.

2 Likelihood Asymptotics

Background likelihood asymptotics are provided in this section as well as the general theory from Fraser and Reid (1995). For a sample $y = (y_1, y_2, \dots, y_n)^T$, the log-likelihood function for $\theta = (\psi, \lambda^T)^T$, where ψ is the one-dimensional component parameter of interest and λ is the $p - 1$ dimensional nuisance parameter, is denoted as $l(\theta) = l(\theta; y)$. The maximum likelihood estimate, $\hat{\theta} = (\hat{\psi}, \hat{\lambda}^T)^T$, is obtained by maximizing the exact log-likelihood with respect to θ and is characterized by the score equation

$$l_{\theta}(\hat{\theta}) = l_{\theta}(\hat{\theta}; y) = \left. \frac{\partial l(\theta)}{\partial \theta} \right|_{\hat{\theta}} = 0.$$

The constrained maximum likelihood estimate, $\hat{\theta}_{\psi} = (\psi, \hat{\lambda}_{\psi}^T)^T$, is obtained by maximizing the log-likelihood

with respect to λ while holding ψ fixed. The information matrix is given by

$$j_{\theta\theta^T}(\theta) = \begin{bmatrix} -\frac{\partial^2 l(\theta)}{\partial \psi \partial \psi} & -\frac{\partial^2 l(\theta)}{\partial \psi \partial \lambda^T} \\ -\frac{\partial^2 l(\theta)}{\partial \psi \partial \lambda^T} & -\frac{\partial^2 l(\theta)}{\partial \lambda \partial \lambda^T} \end{bmatrix} = \begin{bmatrix} -l_{\psi\psi}(\theta) & -l_{\psi\lambda^T}(\theta) \\ -l_{\psi\lambda^T}(\theta) & -l_{\lambda\lambda^T}(\theta) \end{bmatrix} = \begin{bmatrix} j_{\psi\psi}(\theta) & j_{\psi\lambda^T}(\theta) \\ j_{\psi\lambda^T}(\theta) & j_{\lambda\lambda^T}(\theta) \end{bmatrix}.$$

The observed information matrix evaluated at $\hat{\theta}$ is denoted as $j_{\theta\theta^T}(\hat{\theta})$. The estimated asymptotic variance of $\hat{\theta}$ is then given by

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

2.1 Large Sample Likelihood-Based Asymptotic Methods

Using the above notation, the two familiar likelihood-based methods that are used for testing the scalar component interest parameter $\psi = \psi(\theta) = \psi_0$ are the Wald departure and the signed log-likelihood departure:

$$q = (\hat{\psi} - \psi_0)\{j^{\psi\psi}(\hat{\theta})\}^{-1/2} \quad (7)$$

$$r = \text{sgn}(\hat{\psi} - \psi_0)[2\{l(\hat{\theta}) - l(\hat{\theta}_{\psi_0})\}]^{1/2}. \quad (8)$$

The limiting distribution of q and r is the standard normal. The corresponding p-values, $p(\psi_0)$ can be approximated by $\Phi(q)$ and $\Phi(r)$, where $\Phi(\cdot)$ is the standard normal distribution function. These methods are well known to have order of convergence $O(n^{-1/2})$ and are generally referred to as first-order methods. Note that $p(\psi)$ gives the p-value for any chosen value of ψ and thus is referred to as the significance function. Hence, a $(1 - \alpha)100\%$ confidence interval for ψ , (ψ_L, ψ_U) , can be obtained by inverting $p(\psi)$ such that

$$\psi_L = \min(p^{-1}(\alpha/2), p^{-1}(1 - \alpha/2))$$

$$\psi_U = \max(p^{-1}(\alpha/2), p^{-1}(1 - \alpha/2)).$$

2.2 Small Sample Likelihood-Based Asymptotic Method

Many methods exist in the literature that achieve improvements to the accuracy of the signed log-likelihood departure. See Reid (1996) and Severini (2000) for a detailed overview of this development. The approach developed by Fraser and Reid (1995) to more accurately approximate p-values will be the focus of this paper. Fraser and Reid (1995) show that this method achieves a known $O(n^{-3/2})$ rate of convergence and is referred to more generally as a third-order method.

The theory developed by Fraser and Reid applies to the general case, where the dimension of the variable y is greater than the dimension of the parameter θ . In order to use existing statistical methods however, their theory requires an initial dimension reduction. In particular, the dimension of the variable y must be reduced to the dimension of the parameter θ . If this reduction is possible using sufficiency or ancillarity then third-order p-value approximations have previously been available.² See for example: Lugannani and

²A statistic $T(X)$ whose distribution is a function of θ , for data X , is a sufficient statistic if the conditional distribution of

Rice (1980), DiCiccio et al. (1990), Barndorff-Nielsen (1991), Fraser and Reid (1993), Skovgaard (1987). If reduction is not possible using either of these methods, then approximate ancillarity seems to be required. This latter case is the focus of the Fraser and Reid methodology. A subsequent dimension reduction from the parameter θ to the scalar parameter of interest ψ is required.

These reductions are achieved through two key reparameterizations. The first dimension reduction is done through a reparameterization from θ to φ and the second from the reparameterization from φ to χ . The construction of φ represents a very special new parameterization. The idea in this step is to obtain a local canonical parameter of an approximating exponential model. This is done so that existing saddlepoint approximations can be used. The parameterization to χ is simply a re-casting of the parameter of interest ψ in the new φ parameter space.

This new variable φ is obtained by taking the sample space gradient at the observed data point y^o calculated in the directions given by a set of vectors V :

$$\varphi^T(\theta) = \left. \frac{\partial}{\partial y} l(\theta; y) \right|_{y^o} \cdot V \quad (9)$$

$$V = \left. \frac{\partial y}{\partial \theta^T} \right|_{(y^o, \hat{\theta})}. \quad (10)$$

The set of vectors in V are referred to as ancillary directions or sensitivity directions and capture how the data is influenced by parameter change near the maximum likelihood value. The differentiation in (10) is taken for fixed values of a full-dimensional pivotal quantity and is defined from the total differentiation of this pivotal. A pivotal statistic $z(\theta, y)$ is a function of the variable y and the parameter θ that has a fixed distribution (independent of θ) and is a required component of the methodology. The expression in (10) can be rewritten in terms of the pivotal quantity:

$$V = \left\{ \frac{\partial z(y, \theta)}{\partial y^T} \right\}^{-1} \left\{ \frac{\partial z(y, \theta)}{\partial \theta^T} \right\} \Big|_{\hat{\theta}}. \quad (11)$$

Implicit in (9) is the necessary conditioning that reduces the dimension of the problem from n to p . This is done through the vectors in V which are based on the pivotal quantity $z(\theta, y)$ which in (9) serve to condition on an approximate ancillary statistic. This is a very technical point and the reader is referred to Fraser and Reid (1995) for full technical details.

The second step involves reducing the dimension of the problem from p to 1, with 1 being the dimension of the interest parameter $\psi(\theta)$. This step is achieved through the elimination of the nuisance parameters using a marginalization procedure.³ This procedure leads to the new parameter $\chi(\theta)$ which replaces $\psi(\theta)$

$$\chi(\theta) = \frac{\psi_{\varphi^T}(\hat{\theta}_{\psi})}{\left| \psi_{\varphi^T}(\hat{\theta}_{\psi}) \right|} \varphi(\theta), \quad (12)$$

X given T does not depend on θ . An ancillary statistic is a statistic whose distribution does not depend on θ .

³This is done through a marginal distribution obtained from integrating a conditional distribution based on nuisance parameters. See Fraser(2003).

where $\psi_{\varphi^T}(\theta) = \partial\psi(\theta)/\partial\varphi^T = (\partial\psi(\theta)/\partial\theta^T)(\partial\varphi(\theta)/\partial\theta^T)^{-1}$. This new variable $\chi(\theta)$ is simply the parameter of interest $\psi(\theta)$ recalibrated in the new parameterization.

Given this new reparameterization, the departure measure Q can be defined:

$$Q = \text{sgn}(\hat{\psi} - \psi) |\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)| \left\{ \frac{|\hat{j}_{\varphi\varphi^T}(\hat{\theta})|}{|\hat{j}_{(\lambda\lambda^T)}(\hat{\theta}_\psi)|} \right\}^{1/2}, \quad (13)$$

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

$$|\hat{j}_{\varphi\varphi^T}(\hat{\theta})| = |\hat{j}_{\theta\theta^T}(\hat{\theta})| |\varphi_{\theta^T}(\hat{\theta})|^{-2} \quad (14)$$

$$|\hat{j}_{(\lambda\lambda^T)}(\hat{\theta}_\psi)| = |\hat{j}_{\lambda\lambda^T}(\hat{\theta}_\psi)| |\varphi_{\lambda^T}^T(\hat{\theta}_\psi) \varphi_{\lambda^T}(\hat{\theta}_\psi)|^{-1}. \quad (15)$$

The expression in (13) is a maximum likelihood departure adjusted for nuisance parameters. The term in (13) involving the Jacobians, specifically,

$$\frac{|\hat{j}_{\varphi\varphi^T}(\hat{\theta})|}{|\hat{j}_{(\lambda\lambda^T)}(\hat{\theta}_\psi)|} \quad (16)$$

reflects the estimated variance of $|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|$. More precisely, the reciprocal of this term is an estimate of the variance of $|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|$.

Third-order accurate p-value approximations can be obtained by combining the signed log-likelihood ratio given in (8) and the new maximum likelihood departure from (13) using the expression

$$\Phi(r^*) = \Phi \left(r - r^{-1} \log \frac{r}{Q} \right) \quad (17)$$

due to Barndorff-Nielsen (1991). That is, for a null hypothesis of interest, $\psi = \psi_0$, use the observed data to compute the usual log-likelihood departure given in (8) as well as the maximum likelihood departure given in (13) and plug these quantities into the righthand side of (17) to obtain the observed p-value for testing $\psi = \psi_0$. An asymptotically equivalent expression to the Barndorff-Nielsen one is given by

$$\Phi(r) + \phi(r) \left\{ \frac{1}{r} - \frac{1}{Q} \right\}, \quad (18)$$

where ϕ is the standard normal density. This version is due to Lugannani and Rice (1980). P-values from both these approximations will be reported in the analyses below.

3 Third-Order Inference for Autocorrelation

The third-order method outlined above is now applied to the Gaussian AR(1) model for inference concerning the autocorrelation parameter ρ (using earlier notation $\psi(\theta) = \rho$). For the parameter vector, $\theta = (\beta, \rho, \sigma^2)^T$,

the probability density function of ε is given by

$$f(y; \theta) = (2\pi)^{-n/2} |\Sigma^{-1}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (y - X\beta)^T \Sigma^{-1} (y - X\beta) \right],$$

where $\Sigma = \sigma^2 \Omega$, with

$$|\Omega| = \frac{1}{1 - \rho^2} \quad \text{and} \quad \Omega^{-1} = \begin{bmatrix} 1 & -\rho & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & \cdots & 0 & 0 \\ & & & \vdots & & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} \equiv A.$$

The log-likelihood function (with the constant dropped) is then given by

$$\begin{aligned} l(\theta) &= -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (y - X\beta)^T \Sigma^{-1} (y - X\beta) \\ &= -\frac{n}{2} \log \sigma^2 + \frac{1}{2} \log(1 - \rho^2) - \frac{1}{2\sigma^2} (y - X\beta)^T A (y - X\beta). \end{aligned} \quad (19)$$

This function is equivalent to the log-likelihood function given in (6).

The overall maximum likelihood estimate (MLE) of θ , denoted as $\hat{\theta}$, is obtained by simultaneously solving the first-order conditions $l_{\theta}(\hat{\theta}) = 0$:

$$\begin{aligned} l_{\beta}(\hat{\theta}) &= \frac{1}{\hat{\sigma}^2} X^T \hat{A} (y - X\hat{\beta}) = 0 \\ l_{\sigma^2}(\hat{\theta}) &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} (y - X\hat{\beta})^T \hat{A} (y - X\hat{\beta}) = 0 \\ l_{\rho}(\hat{\theta}) &= -\frac{\hat{\rho}}{1 - \hat{\rho}^2} - \frac{1}{2\hat{\sigma}^2} (y - X\hat{\beta})^T \hat{A}_{\rho} (y - X\hat{\beta}) = 0, \end{aligned}$$

where $A_{\rho} = \partial A / \partial \rho$. Solving these first-order conditions gives

$$\hat{\beta} = (X^T \hat{A} X)^{-1} X^T \hat{A} y \quad (20)$$

$$\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\beta})^T \hat{A} (y - X\hat{\beta}), \quad (21)$$

with $\hat{\rho}$ satisfying

$$-\frac{\hat{\rho}}{1 - \hat{\rho}^2} - \frac{1}{2\hat{\sigma}^2} (y - X\hat{\beta})^T \hat{A}_{\rho} (y - X\hat{\beta}) = 0.$$

This last equation is further simplified by defining $e = y - X\hat{\beta}$ and noting

$$e^T \hat{A}_{\rho} e = 2 \left(\rho \sum_{i=2}^{n-1} e_i^2 - \sum_{i=1}^{n-1} e_i e_{i+1} \right).$$

This then gives the condition that $\hat{\rho}$ must satisfy

$$-\frac{\hat{\rho}}{1-\hat{\rho}^2} - \frac{1}{\hat{\sigma}^2} \left(\hat{\rho} \sum_{i=2}^{n-1} e_i^2 - \sum_{i=1}^{n-1} e_i e_{i+1} \right) = 0,$$

which implies

$$\frac{\hat{\rho}^3}{\hat{\sigma}^2} \sum_{i=2}^{n-1} e_i^2 - \frac{\hat{\rho}^2}{\hat{\sigma}^2} \sum_{i=1}^{n-1} e_i e_{i+1} - \left(1 + \frac{1}{\hat{\sigma}^2} \sum_{i=2}^{n-1} e_i^2 \right) \hat{\rho} + \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n-1} e_i e_{i+1} = 0. \quad (22)$$

Given this information about the likelihood function and the overall maximum likelihood estimate, the quantities $\hat{\psi} = \hat{\rho}, l(\hat{\theta})$, and $j_{\theta\theta}(\hat{\theta})$ can be obtained. To construct the information matrix, recall, the second derivatives of the log-likelihood function are required:

$$\begin{aligned} l_{\beta\beta}(\theta) &= -\frac{1}{\sigma^2} X^T A X \\ l_{\beta\sigma^2}(\theta) &= -\frac{1}{\sigma^4} X^T A (y - X\beta) \\ l_{\beta\rho}(\theta) &= \frac{1}{\sigma^2} X^T A_{\rho} (y - X\beta) \\ l_{\sigma^2\sigma^2}(\theta) &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} (y - X\beta)^T A (y - X\beta) \\ l_{\sigma^2\rho}(\theta) &= \frac{1}{2\sigma^4} (y - X\beta)^T A_{\rho} (y - X\beta) \\ l_{\rho\rho}(\theta) &= -\frac{1}{2\sigma^4} (y - X\beta)^T A_{\rho\rho} (y - X\beta), \end{aligned}$$

where $A_{\rho\rho} = \partial^2 A / \partial \rho^2$.

To obtain the new locally defined parameter $\varphi(\theta)$ given by (9), two components are required. The first is the sample space gradient evaluated at the data, that is the derivative of (19) with respect to y evaluated at y^o :

$$\frac{\partial l(\theta; y)}{\partial y} = -\frac{1}{\sigma^2} (y - X\beta)^T A. \quad (23)$$

And the second is the ancillary vectors V . To obtain V , given by (10), a full-dimensional pivotal quantity, $z(y, \theta)$, is required. The pivotal quantity for this problem is specified as the vector of independent standard normal deviates

$$z(y, \theta) = \frac{U(y - X\beta)}{\sigma}, \quad (24)$$

where U is defined as the lower triangular matrix

$$U = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 0 & 0 & \cdots & 0 & 0 \\ & & \vdots & & & & \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}. \quad (25)$$

This choice of pivotal quantity coincides with the standard quantity used to estimate the parameters of an AR(1) model in the literature (see for example Hamilton (1994)). Together with the overall MLE, the ancillary direction array can be constructed as follows

$$\begin{aligned} V &= \left\{ \frac{\partial z(y, \theta)}{\partial y^T} \right\}^{-1} \left\{ \frac{\partial z(y, \theta)}{\partial \theta^T} \right\} \Big|_{\hat{\theta}} \\ &= \left\{ -X, \quad U^{-1} \frac{\partial U}{\partial \rho} (y - X\beta), \quad -\frac{(y - X\beta)}{\sigma} \right\} \Big|_{\hat{\theta}} \\ &= \left\{ -X, \quad \hat{U}^{-1} \frac{\partial U}{\partial \rho} \Big|_{\hat{\theta}} (y - X\hat{\beta}), \quad -\frac{(y - X\hat{\beta})}{\hat{\sigma}} \right\}. \end{aligned} \quad (26)$$

Using (23) and (26), $\varphi^T(\theta)$ is obtained as

$$\varphi^T(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} (y - X\beta)^T A X & -\frac{1}{\sigma^2} (y - X\beta)^T A \hat{U}^{-1} \frac{\partial U}{\partial \rho} \Big|_{\hat{\theta}} (y - X\hat{\beta}) & \frac{1}{\sigma^2 \hat{\sigma}} (y - X\beta)^T A (y - X\hat{\beta}) \end{bmatrix}. \quad (27)$$

For convenience, define

$$\begin{aligned} \varphi_1(\theta) &= \frac{1}{\sigma^2} (y - X\beta)^T A X \\ \varphi_2(\theta) &= -\frac{1}{\sigma^2} (y - X\beta)^T A \hat{U}^{-1} \frac{\partial U}{\partial \rho} \Big|_{\hat{\theta}} (y - X\hat{\beta}) \\ \varphi_3(\theta) &= \frac{1}{\sigma^2 \hat{\sigma}} (y - X\beta)^T A (y - X\hat{\beta}). \end{aligned}$$

So that,

$$\varphi^T(\theta) = [\varphi_1(\theta) \quad \varphi_2(\theta) \quad \varphi_3(\theta)].$$

The dimension reduction from n , the dimension of the variable y , to $k+2$, the dimension of the parameter θ is evidenced from the expression for $\varphi^T(\theta)$. The dimension of $\varphi^T(\theta)$ given by (27) is $(1 \times k+2)$ with $\varphi_1(\theta)$ having dimension $1 \times k$.

To obtain the further reduction to the dimension of the interest parameter ρ , $\chi(\theta)$ is required. As can be seen by (12), the scalar parameter $\chi(\theta)$ involves $\varphi(\theta)$ as well as the constrained maximum likelihood estimate $\hat{\theta}_\psi$. To derive the constrained MLE, the log-likelihood function given by (19) must be maximized with respect to β and σ^2 while holding ρ fixed. Thus, for fixed ρ , (20) and (21) are the constrained maximum likelihood estimates for β and σ^2 (with appropriate changes to $\hat{A}$), respectively. The other required component of $\chi(\theta)$

is $\psi_{\varphi^T}(\theta) = (\partial\psi(\theta)/\partial\theta^T)(\partial\varphi(\theta)/\partial\theta^T)^{-1}$ evaluated at $\hat{\theta}_\psi$. The first term in parentheses is computed as

$$\partial\psi(\theta)/\partial\theta^T = [\partial\psi(\theta)/\partial\beta \quad \partial\psi(\theta)/\partial\rho \quad \partial\psi(\theta)/\partial\sigma^2] = [0 \quad 1 \quad 0],$$

where $\partial\psi(\theta)/\partial\beta$ is of dimension $1 \times k$. The second term in parentheses is a $(k+2 \times k+2)$ matrix and involves differentiation of $\varphi(\theta)$ with respect to θ and is calculated from

$$\partial\varphi(\theta)/\partial\theta^T = \begin{bmatrix} \partial\varphi_1(\theta)/\partial\beta & \partial\varphi_1(\theta)/\partial\rho & \partial\varphi_1(\theta)/\partial\sigma^2 \\ \partial\varphi_2(\theta)/\partial\beta & \partial\varphi_2(\theta)/\partial\rho & \partial\varphi_2(\theta)/\partial\sigma^2 \\ \partial\varphi_3(\theta)/\partial\beta & \partial\varphi_3(\theta)/\partial\rho & \partial\varphi_3(\theta)/\partial\sigma^2 \end{bmatrix}.$$

With the calculation of the determinants given in (14) and (15), the departure measure, Q , can then be calculated from (13). Third-order inference concerning ρ can be obtained from plugging into either (17) or (18). Unfortunately, an explicit formula is not available for Q as a closed form solution for the MLE does not exist. For the interested reader, Matlab code is available from the authors (for the example below) to help with the implementation of this method.

4 Numerical Illustrations

An example and a set of three simulations is considered in this section. For expositional clarity, confidence intervals are reported.

4.1 Example

Consider the simple example presented in Wooldridge (2006, page 445) for the estimation of the Phillips curve using 49 observations

$$\text{inf}_t = \beta_0 + \beta_1 \text{ue}_t + \varepsilon_t,$$

with an AR(1) error structure

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t.$$

The variables, inf_t and ue_t represent the CPI inflation rate and the civilian unemployment rate in the United States from 1948-1996. The random variables, v_t , are normally distributed with $E[v_t] = 0$ and $E[v_t^2] = \sigma^2$.

Table 1 reports the 90% confidence intervals for ρ obtained from the standardized test statistic given in equation (3), the signed log-likelihood ratio statistic given in equation (8), and the Lugannani and Rice and Barndorff-Nielsen approximations given in equations (11) and (12). These methods will henceforth be abbreviated as STS, r , LR, and BN, respectively. Given the interval results arising from the three methods are clearly very different, the accuracy of the first-order methods must be pursued. To this end, various simulation studies will be performed to compare results obtained by each of these three methods.

Table 1: 90% Confidence Interval for ρ

Method	90% Confidence Interval for ρ
STS	(0.6193, 0.9250)
r	(0.6043, 0.9153)
BN	(0.6566, 0.9942)
LR	(0.6592, 0.9999)

4.2 Simulation Studies

The superiority of the third-order method can be seen through the simplest possible simulations. Throughout the simulation studies, the accuracy of a method will be evaluated based on the following criteria:

1. Coverage probability: the percentage of a true parameter value falling within the intervals.
2. Coverage error: the absolute difference between the nominal level and coverage probability.
3. Upper (lower) error probability: the percentage of a true parameter value falling above (below) the intervals.
4. Average bias: the average of the absolute difference between the upper and lower error probabilities and their nominal levels.

The simulations include the results from both first-order methods (STS and r) as well as from both third-order approximations, Lugannani and Rice (LR) and Barndorff-Nielsen (BN).

Simulation Study 1:

The first simulation generates 10,000 random samples each of size 15, from the following Gaussian AR(1) model for various values of the autocorrelation parameter ρ , ranging from strong positive autocorrelation to strong negative autocorrelation:

$$\begin{aligned}
 y_t &= \varepsilon_t \\
 \varepsilon_t &= \rho\varepsilon_{t-1} + \nu_t, \quad t = 1, 2, \dots, 15.
 \end{aligned}$$

The variables, ν_t , are distributed as standard normal. Note carefully that the design matrix X is null for this simulation; y_t is equal to $\rho\varepsilon_{t-1} + \nu_t$. A 95% confidence interval for ρ is calculated using each of the three methods. The nominal coverage probability, coverage error, upper and lower error probabilities, and average bias are 0.95, 0, 0.025 (0.025), and 0, respectively. Table 2 records the results of this simulation study for selected values of ρ .

Based on a 95% confidence interval, 2.5% is expected in each tail. While the third-order methods produce upper and lower error probabilities that are relatively symmetric, with a tail probability totalling approximately 5%, those produced by the first-order methods are heavily skewed, and in the case of the

Table 2: Results for Simulation Study 1

ρ	Method	Coverage Probability	Coverage Error	Upper Probability	Lower Probability	Average Bias
-0.9	STS	0.8200	0.1300	0.1762	0.0038	0.0862
	r	0.9501	0.0001	0.0383	0.0116	0.0134
	BN	0.9545	0.0045	0.0215	0.0240	0.0023
	LR	0.9524	0.0024	0.0218	0.0258	0.0020
-0.5	STS	0.9277	0.0223	0.0710	0.0013	0.0348
	r	0.9519	0.0019	0.0318	0.0163	0.0077
	BN	0.9497	0.0003	0.0260	0.0243	0.0009
	LR	0.9497	0.0003	0.0259	0.0244	0.0008
0	STS	0.9581	0.0081	0.0192	0.0227	0.0040
	r	0.9505	0.0005	0.0231	0.0264	0.0016
	BN	0.9482	0.0018	0.0243	0.0275	0.0016
	LR	0.9481	0.0019	0.0244	0.0275	0.0016
0.5	STS	0.9280	0.0220	0.0027	0.0693	0.0333
	r	0.9489	0.0011	0.0198	0.0313	0.0057
	BN	0.9464	0.0036	0.0276	0.0260	0.0018
	LR	0.9463	0.0037	0.0277	0.0260	0.0019
0.9	STS	0.8190	0.1310	0.0033	0.1777	0.0872
	r	0.9460	0.0040	0.0111	0.0429	0.0159
	BN	0.9489	0.0011	0.0238	0.0273	0.0018
	LR	0.9466	0.0034	0.0254	0.0280	0.0017

standardized test statistic, the total error probability reaches as high as 18%. In terms of coverage error, the likelihood ratio performs well, however, due to the distortion in the tails, the average bias is never less than that achieved by either the Lugannani and Rice or Barndorff-Nielsen approximations. The signed log-likelihood ratio method is superior to the the standardized test statistic in all cases considered; notice the 82% coverage probability of the standardized test statistic for absolute values of ρ close to unity.

Simulation Study 2:

Consider the simple linear regression model:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 X_t + \varepsilon_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + \nu_t, \quad t = 1, 2, \dots, 50. \end{aligned}$$

The variables, ν_t , are distributed as $N(0, \sigma^2)$ and the design matrix is given in Table 3. 10,000 samples of size 50 were generated with parameter values given as: $\beta_0 = 2, \beta_1 = 1$, and $\sigma^2 = 1$ for various values of ρ .

Table 3: Design Matrix for Simulation 2

t	X_0	X_1	t	X_0	X_1
1	1	2.3846	26	1	-0.6645
2	1	-0.2998	27	1	1.6852
3	1	0.7914	28	1	0.7722
4	1	-1.7490	29	1	-0.3484
5	1	-1.7062	30	1	2.0686
6	1	0.2231	31	1	0.6002
7	1	0.9711	32	1	-1.3160
8	1	-2.1686	33	1	0.6553
9	1	-1.0717	34	1	0.2159
10	1	0.4011	35	1	0.2730
11	1	-1.2638	36	1	-0.8858
12	1	0.5482	37	1	-0.7351
13	1	-1.2954	38	1	0.2834
14	1	0.6048	39	1	0.9487
15	1	0.4588	40	1	0.8860
16	1	-0.2083	41	1	0.9119
17	1	-0.9748	42	1	0.0747
18	1	0.5060	43	1	0.9844
19	1	-0.5636	44	1	1.0850
20	1	-1.7085	45	1	0.7950
21	1	0.9227	46	1	0.3001
22	1	0.0891	47	1	1.1010
23	1	0.1783	48	1	-1.1671
24	1	-0.9638	49	1	1.6045
25	1	-0.5363	50	1	-0.9970

Again 95% confidence intervals for ρ are obtained for each simulated sample. Table 4 records the simulation results only for $\rho = 0$. The superiority of the third-order methods is clear with coverage error equal to 0.0002 and average bias of 0.0003 for both methods. While the signed log-likelihood ratio method outperformed the standardized test statistic in the previous simulation, it is not the case here, the standardized test statistic achieves substantially lower coverage error and bias for $\rho = 0$. The asymmetry in the upper and lower tail probabilities for both these methods still persists however. This asymmetry can be evidenced further in Figure 1 where upper and lower error probabilities are plotted for various values of ρ used in the simulation. The discord between the first-order methods and the nominal value of 0.025 is very large,

especially for the standardized test statistic as the absolute value of ρ approaches unity. The average bias and coverage probability are provided in Figure 2. It can be seen from this figure that the proposed methods give results very close to the nominal values whereas the first-order methods give results that are less satisfactory especially for values of ρ close to one.

Table 4: Results for Simulation Study 2 for $\rho = 0$

ρ	Method	Coverage Probability	Coverage Error	Upper Probability	Lower Probability	Average Bias
0	STS	0.9509	0.0009	0.0317	0.0174	0.0072
	r	0.9431	0.0069	0.0373	0.0196	0.0089
	BN	0.9498	0.0002	0.0248	0.0254	0.0003
	LR	0.9498	0.0002	0.0248	0.0254	0.0003

Figure 1: Upper and lower error probabilities

Figure 2: Average bias and coverage probability

Simulation Study 3:

Consider the multiple linear regression model

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \nu_t, \quad t = 1, 2, \dots, 50.$$

The variables, ν_t , are distributed as $N(0, \sigma^2)$ and design matrix given in Table 5. 10,000 samples of size 50 were generated with parameter values given as: $\beta_0 = 2, \beta_1 = 1, \beta_2 = 1$, and $\sigma^2 = 1$ for various values of ρ .

Table 5: Design Matrix for Simulation 3

t	X_0	X_1	X_2	t	X_0	X_1	X_2
1	1	-0.4326	-1.0106	26	1	0.8580	-0.2340
2	1	-1.6656	0.6145	27	1	1.2540	0.1184
3	1	0.1253	0.5077	28	1	-1.5937	0.3148
4	1	0.2877	1.6924	29	1	-1.4410	1.4435
5	1	-1.1465	0.5913	30	1	0.5711	-0.3510
6	1	1.1909	-0.6436	31	1	-0.3999	0.6232
7	1	1.1892	0.3803	32	1	0.6900	0.7990
8	1	-0.0376	-1.0091	33	1	0.8156	0.9409
9	1	0.3273	-0.0195	34	1	0.7119	-0.9921
10	1	0.1746	-0.0482	35	1	1.2902	0.2120
11	1	-0.1867	0	36	1	0.6686	0.2379
12	1	0.7258	-0.3179	37	1	1.1908	-1.0078
13	1	-0.5883	1.0950	38	1	-1.2025	-0.7420
14	1	2.1832	-1.8740	39	1	-0.0198	1.0823
15	1	-0.1364	0.4282	40	1	-0.1567	-0.1315
16	1	0.1139	0.8956	41	1	-1.6041	0.3899
17	1	1.0668	0.7310	42	1	0.2573	0.0880
18	1	0.0593	0.5779	43	1	-1.0565	-0.6355
19	1	-0.0956	0.0403	44	1	1.4151	-0.5596
20	1	-0.8323	0.6771	45	1	-0.8051	0.4437
21	1	0.2944	0.5689	46	1	0.5287	-0.9499
22	1	-1.3362	-0.2556	47	1	0.2193	0.7812
23	1	0.7143	-0.3775	48	1	-0.9219	0.5690
24	1	1.6236	-0.2959	49	1	-2.1707	-0.8217
25	1	-0.6918	-1.4751	50	1	-0.0592	-0.2656

Again 95% confidence interval for ρ are obtained for each simulated sample. Simulation results for $\rho = 0$ are provided in Table 6. From this table, the superiority of the third-order method is evident, the standardized test statistic outperforms the signed log-likelihood ratio test for this particular value of ρ . The upper and lower error probabilities are plotted in Figure 3. The average bias and coverage probability are plotted in Figure 4. The conclusions from these graphs are similar to those reached from the previous simulation. The proposed method gives results very close to the nominal values even for values of the autocorrelation coefficient close to one whereas the first-order methods give results that are less satisfactory.

Table 6: Results for Simulation Study 3 for $\rho = 0$

ρ	Method	Coverage Probability	Coverage Error	Upper Probability	Lower Probability	Average Bias
0	STS	0.9527	0.0027	0.0287	0.0186	0.0051
	r	0.9395	0.0105	0.0372	0.0233	0.0069
	BN	0.9491	0.0009	0.0250	0.0259	0.0004
	LR	0.9492	0.0008	0.0250	0.0258	0.0004

Figure 3: Upper and lower error probabilities

Figure 4: Average bias and coverage probability

The simulation studies have shown the improved accuracy that can be obtained for testing the autocorrelation parameter in first-order autoregressive models. The proposed method can be applied to obtain either a p-value or a confidence interval for testing the autocorrelation parameter in an AR(1) model. The third-order methods produce results which are remarkably close to nominal levels, with superior coverage and symmetric upper and lower error probabilities compared to the results from the first-order methods. It is recommended that third-order methods be employed for reliable and improved inference for small- and medium-sized samples; if first-order methods are used, they should be used with caution and viewed with some skepticism.

5 Conclusion

Recently developed third-order likelihood theory was used to obtain highly accurate p-values for testing the autocorrelation parameter in a first-order model. The simulation results indicate that significantly improved inferences can be made by using third-order likelihood methods. The method was found to outperform the standardized test statistic in every case and across all criteria considered. The method further outperformed the signed log-likelihood ratio method in terms of average bias and produced symmetric tail error probabilities. As the proposed method relies on familiar likelihood quantities and given its ease of computational implementation, it is a highly tractable and viable alternative to conventional methods. Further, with appropriately defined pivotal quantity, the proposed method can readily be extended to models of higher order autocorrelation. Extensions to this line of research include the consideration of fully dynamic models with lagged regressors as well as conducting inference directly for the regression parameters. For this latter case, Veall (1986) provides Monte Carlo evidence to show that the standard bootstrap does not improve inference in a regression model with highly autocorrelated errors and a strongly trended design matrix.

References

- [1] Barndorff-Nielsen, O., 1991, Modified Signed Log-Likelihood Ratio, *Biometrika* 78, 557-563.
- [2] Belsley, D., 1997, A Small-Sample Correction for Testing for g -Order Serial Correlation with Artificial Regressions, *Computational Economics* 10, 197-229.
- [3] Cheang, W., Reinsel, G., 2000, Bias Reduction of Autoregressive Estimates in Time Series Regression Model through Restricted Maximum Likelihood, *Journal of the American Statistical Association* 95, 1173-1184.
- [4] DiCiccio, T., Field, C., Fraser, D., 1990, Approximation of Marginal Tail Probabilities and Inference for Scalar Parameters, *Biometrika* 77, 77-95.
- [5] Durbin, J., Watson, G., 1951, Testing for Serial Correlation in Least Squares, II, *Biometrika* 38, 159-178.
- [6] Fraser, D., 2003, Likelihood for Component Parameters, *Biometrika* 90, 327-339.
- [7] Fraser, D., Reid, N., 1993, Third Order Asymptotic Models: Likelihood Functions Leading to Accurate Approximations for Distribution Functions, *Statistica Sinica* 3, 67-82.
- [8] Fraser, D., Reid, N., 1995, Ancillaries and Third Order Significance, *Utilitas Mathematica* 47, 33-53.
- [9] Fraser, D., Reid, N., Wu, J., 1999, A Simple General Formula for Tail Probabilities for Frequentist and Bayesian Inference, *Biometrika* 86, 249-264.
- [10] Fraser, D., Wong, A., Wu, J., 1999, Regression Analysis, Nonlinear or Nonnormal: Simple and Accurate p Values From Likelihood Analysis, *Journal of the American Statistical Association* 94(448), 1286-1295.
- [11] Hamilton, J.D., 1994, *Time Series Analysis* (Princeton University Press, New Jersey).
- [12] King, M., 1987, Testing for Autocorrelation in Linear Regression Models: A Survey, Chapter 3 in King, M. and Giles, D. (eds), *Specification Analysis in the Linear Models: Essays in Honour of Donald Cochrane* (Routledge, London).
- [13] Levenbach, H., 1972, Estimation of Autoregressive Parameters from a Marginal Likelihood Function, *Biometrika* 59, 61-71.
- [14] Lugannani, R., Rice, S., 1980, Saddlepoint Approximation for the Distribution of the Sums of Independent Random Variables, *Advances in Applied Probability* 12, 475-490.
- [15] Reid, N., 1996, Higher Order Asymptotics and Likelihood: A Review and Annotated Bibliography, *Canadian Journal of Statistics* 24, 141-166.

- [16] Reinsel, G., Cheang, W., 2003, Approximate ML and REML Estimation for Regression Models with Spatial or Time Series AR(1) Noise, *Statistics & Probability Letters* 62, 123-135.
- [17] Skovgaard, I., 1987, Saddlepoint Expansion for Conditional Distributions, *Journal of Applied Probability* 24, 875-887.
- [18] Severini, T., 2000, *Likelihood Methods in Statistics* (Oxford University Press, Oxford).
- [19] Tunnicliffe-Wilson, G., 1989, On the use of Marginal Likelihood in Time Series Model Estimation, *Journal of the Royal Statistical Society Series B* 51, 15-27.
- [20] Veall, M., 1986, Bootstrapping Regression Estimators under First-Order Serial Correlation, *Economics Letters* 21, 41-44.
- [21] Wooldridge, J., 2006, *Introductory Econometrics: A Modern Approach* (Thomson South-Western, USA).