Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Econometrics 127 (2005) 17–33
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Highly accurate likelihood analysis for the seemingly unrelated regression problem

D.A.S. Fraser[a], M. Rekkas[b,*], A. Wong[c]

[a]*Department of Statistics, University of Toronto, 100 St. George Street, Toronto, Ont., Canada M5S 3G3*
[b]*Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6*
[c]*SASIT, York University, 4700 Keele Street, Toronto, Ont., Canada M3J 1P3*

Accepted 30 June 2004
Available online 23 August 2004

Abstract

The linear and nonlinear seemingly unrelated regression problem with general error distribution is analyzed using recent likelihood theory that arguably provides the definitive distribution for assessing a scalar parameter; this involves implicit but well defined conditioning and marginalization for determining intrinsic measures of departure. Highly accurate *p*-values are obtained for the key difference between two regression coefficients of central interest. The *p*-value gives the statistical position of the data with respect to the key parameter. The theory and the results indicate that this methodology provides substantial improvement on first-order likelihood procedures, both in distributional accuracy, and in precise measurement of the key parameter.

JEL classification: C1; C3

Keywords: Asymptotics; Conditioning; Likelihood; *p*-value; SUR

*Corresponding author. Tel.: +1-604-2686793; fax: +1-604-2915944.
E-mail address: mrekkas@sfu.ca (M. Rekkas).

0304-4076/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2004.06.001

1. Introduction

In many areas of economics, relationships arise naturally that can be modeled in terms of several seemingly unrelated regression (SUR) equations. The foundational analysis of the SUR model was initiated by Zellner (1962, 1963) with an innovative use of generalized least squares, followed later by likelihood and Bayesian analyses. More recently, various other inference approaches have become available and accessible, such as the Bartlett corrected likelihood, the bootstrap, the Bayesian method of moments, and now the likelihood distributional analysis. The focus of this paper is on the likelihood distributional approach.

A prominent way of evaluating the various approaches is to examine the order of convergence of a candidate distribution to an objective. Traditional likelihood ratio analysis is typically $O(n^{-1/2})$ meaning that a candidate probability differs from a target by an error of the order $n^{-1/2}$, where n is the sample size or a related measure of the amount of relevant data. Also, for small and medium sized samples the likelihood ratio approach in particular is well known to have the potential for substantial inaccuracy. Bartlett adjustments have been developed and can provide a substantial increase in distributional accuracy; a Bartlett corrected likelihood interval can achieve distributional accuracy of order $O(n^{-2})$, but for directed or one-sided intervals, the order of accuracy drops to $O(n^{-1/2})$. For a discussion and application of Bartlett adjustments see, for example: Bartlett (1937), Barndorff-Nielsen and Cox (1994), Attfield (1995, 1998), and Rocke (1989). The bootstrap is initially $O(n^{-1/2})$, but by repetition can be raised to $O(n^{-1})$ or $O(n^{-3/2})$. Bootstrap methods have been applied and discussed in Rilstone and Veall (1996) and Rocke (1989). The Bayesian approach can attain arbitrary accuracy given a prior and with asymptotics can attain order $O(n^{-3/2})$. See Zellner (1971), Berry et al. (1996), and Mittelhammer et al. (2000) for a discussion of the Bayesian approach. Zellner (1997) proposed a Bayesian method of moments approach which does not require the use of a likelihood function to obtain point estimates and interval estimates; van der Merwe and Vilijoen (1988) applied this method to the SUR model.

The question of when these convergence rates manifest themselves, or more specifically how accurate the approximations are for very low values of n is an important one. Also of importance is the nature of the objective that is being approximated in each approach. With the Bartlett approach, departures on either side of a parameter value are treated in a compromised manner that trades off direction of departure against coverage; and if the direction of the parameter relative to the data point is of interest, as would typically be the case, then the $O(n^{-2})$ accuracy drops to just $O(n^{-1/2})$. For the bootstrap approach, the immediately targeted probabilities are those of a discrete distribution constructed from observed residuals. With the likelihood distributional approach, the full statistical model is examined in accord with local conditioning at the data point, and the corresponding p-value is arguably the appropriate probability assessment of the departure of data from the parameter (Fraser, 2004).

In this paper we use recent likelihood theory to derive p-values that more accurately assess scalar interest parameters for the linear and nonlinear SUR context

with normal or general error distribution. For this we apply the conditional likelihood-based method developed in Fraser and Reid (1995), which has $O(n^{-3/2})$ distributional accuracy and uses conditional techniques to accurately assess the departure of data from a hypothesized parameter value.

Within this intrinsic conditional framework the likelihood methods invoke distributional techniques closely related to the saddlepoint approach. As such they have been found to give exceptionally high accuracy even in the case of minimal sample size. A set of numerical examples and two simulations are used to illustrate that this new likelihood analysis provides substantial improvements on the usual likelihood ratio method and is superior in terms of central coverage even for very small sample sizes. This is important in the SUR context as economic data of this type may be quite limited. We also discuss briefly how this recent likelihood theory provides the definitive distribution for assessing a scalar parameter. This provides a benchmark for comparison when examining other inference procedures that might be applied.

In the present context we focus on the estimation, testing, and confidence evaluation for the case of a scalar interest parameter component of the full multi-dimensional model; the more general case with a vector parameter is accessible by sequencing through scalar parameter components.

The paper is organized as follows. Section 2 provides some background likelihood results including a description of the general theory from Fraser and Reid (1995). The applications and simulation results are given in Section 3. Conclusions are provided in Section 4.

2. Background asymptotics

2.1. First-order results

Let $y = (y_1, \ldots, y_n)$ be an array of independent variables and y_i have density $f_i(y_i; \theta)$, where the full parameter in the model $\theta = (\psi, \lambda')'$ has dimension p, the parameter of interest ψ has dimension 1 and the nuisance parameter λ has dimension $p - 1$.

The log likelihood function is

$$l(\theta; y) = \sum_{i=1}^{n} l^i(\theta; y_i) = \sum_{i=1}^{n} \log f_i(y_i; \theta) \tag{1}$$

and with observed data y^o we obtain the observed likelihood function $l(\theta) = \sum_{i=1}^{n} l^i(\theta)$, where $l^i(\theta) = l^i(\theta; y_i^o)$. Likelihood analyses will typically use the maximum likelihood value $\hat{\theta} = (\hat{\psi}, \hat{\lambda}')' = \text{argmax}_\theta\, l(\theta; y)$ and the constrained maximum likelihood value $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)' = \text{argmax}_\lambda\, l(\theta; y)$ for each tested ψ value. In cases where there is not an explicit nuisance parameter, the constrained maximum value may often require Lagrange multiplier methods where

$$(\hat{\theta}_\psi, \hat{\alpha}) = \underset{\theta,\alpha}{\text{argmax}}\{l(\theta; y) + \alpha[\psi(\theta) - \psi]\}$$

is calculated by setting the gradient of the right-hand expression with respect to θ and α equal to zero.

When $\theta = \psi$ is a scalar parameter of interest there are two familiar first-order departure measures, the Wald departure and the signed likelihood ratio departure:

$$q = (\hat{\theta} - \theta)|\hat{j}_{\theta\theta}|^{1/2}, \tag{2}$$

$$r = \text{sgn}(\hat{\theta} - \theta)[2\{l(\hat{\theta}; y) - l(\theta; y)\}]^{1/2}, \tag{3}$$

where $\hat{j}_{\theta\theta} = -\partial^2 l(\theta; y)/\partial\theta\partial\theta|_{\hat{\theta}}$ is the observed information. The corresponding first-order p-values are $\Phi(q)$ and $\Phi(r)$, where Φ is the standard normal distribution function. We do not mention here the score measure of departure as it seems not to be a contender in most contexts. With small sample sizes or very nonnormal distributions, these quantities themselves can be very misleading as the following example illustrates.

Example 1. Consider a single observation from the exponential distribution

$$f(y; \theta) = \theta e^{-\theta y} \quad y > 0.$$

For the data value, say $y = 1$, consider the assessment of the parameter value $\theta = 8$. We have $q = -7$ and $r = -3.137055$ leading to the approximate p-values $\Phi(q) = 1.27981 \times 10^{-12}$ and $\Phi(r) = 0.000853$. The exact p-value for $\hat{\theta}$ is easily calculated and given by $P(y \geqslant 1; \theta = 8) = 0.000335$; the approximations can be highly inaccurate.

When $\theta = (\psi, \lambda')'$ with ψ a scalar interest parameter and λ a vector of nuisance parameters, there are simple analogs of the q and r above:

$$q = (\hat{\psi} - \psi)\frac{|\hat{j}_{\theta\theta'}|^{1/2}}{|\tilde{j}_{\lambda\lambda'}|^{1/2}}, \tag{4}$$

$$r = \text{sgn}(\hat{\psi} - \psi)[2\{l(\hat{\theta}; y) - l(\hat{\theta}_\psi; y)\}]^{1/2}, \tag{5}$$

where $\tilde{j}_{\lambda\lambda'} = -\partial^2 l(\theta; y)/\partial\lambda\partial\lambda'|_{\hat{\theta}_\psi}$ is the observed nuisance information. Approximate p-values based on first-order likelihood theory are given by $\Phi(q)$ and $\Phi(r)$. These values can likewise be misleading.

2.2. Recent higher-order results

For p-values that more accurately measure departure and more reliably have a null distribution that is uniform (0,1), some model-data information beyond the observed likelihood is essential. For this, recent likelihood theory focuses on the use of a nominal canonical reparameterization

$$\varphi'(\theta) = \frac{\partial}{\partial V} l(\theta; y)\bigg|_{y^o} = \frac{\partial}{\partial y'} l(\theta; y)\bigg|_{y^o} V, \tag{6}$$

where $\partial l(\theta; y)/\partial V$ denotes directional derivatives with respect to $V = (v_1, \ldots, v_p)$. The vectors in V indicate how y responds to a change in the parameter θ and can be

obtained by using a full n dimensional pivotal quantity $k = k(y, \theta) = (k_1, \ldots, k_n)'$ and calculating the gradient of y with respect to θ at the data for fixed pivotal:

$$V = \frac{\partial y}{\partial \theta'}\bigg|_{(y^o,\hat{\theta})} = \left\{\frac{\partial k(y,\theta)}{\partial y'}\right\}^{-1}\left\{\frac{\partial k(y,\theta)}{\partial \theta'}\right\}\bigg|_{(y^o,\hat{\theta})}. \tag{7}$$

This reparameterization is used for a Wald-type departure measure Q:

$$Q = \operatorname{sgn}(\hat{\psi} - \psi)|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|\left\{\frac{|\hat{j}_{\varphi\varphi'}(\hat{\theta})|}{|\hat{j}_{(\lambda\lambda')}(\hat{\theta}_\psi)|}\right\}^{1/2}, \tag{8}$$

where $\hat{j}_{\varphi\varphi'}$ and $\hat{j}_{(\lambda\lambda')}$ are the observed information matrix and observed nuisance information matrix, respectively, calculated in the nominal parameterization scale. More specifically, they can be obtained as follows:

$$|\hat{j}_{\varphi\varphi'}(\hat{\theta})| = |\hat{j}_{\theta\theta'}(\hat{\theta})||\varphi_{\theta'}(\hat{\theta})|^{-2}, \tag{9}$$

$$|\hat{j}_{(\lambda\lambda')}(\hat{\theta}_\psi)| = |\hat{j}_{\lambda\lambda'}(\hat{\theta}_\psi)||\varphi'_\lambda(\hat{\theta}_\psi)\varphi_{\lambda'}(\hat{\theta}_\psi)|^{-1}. \tag{10}$$

And also, $\chi(\theta)$ acts as a scalar canonical parameter for a one parameter marginal model for testing the interest parameter ψ

$$\chi(\theta) = \frac{\psi_{\varphi'}(\hat{\theta}_\psi)}{|\psi_{\varphi'}(\hat{\theta}_\psi)|}\,\varphi(\theta), \tag{11}$$

where $\psi_{\varphi'}(\theta) = \partial\psi(\theta)/\partial\varphi' = (\partial\psi(\theta)/\partial\theta')(\partial\varphi(\theta)/\partial\theta')^{-1}$. In the case with no explicit nuisance parameterization, the Lagrange method mentioned above is needed and a more general formula replacing (8) is recorded in Fraser et al. (1999a, b).

If the model is exponential then $\varphi(\theta)$ can be taken to be any version of the canonical parameter. More generally, $\varphi(\theta)$ is a gradient of the log likelihood taken in directions that conform to an approximate ancillary that describes the model structure locally. For most third-order likelihood analyses it suffices to view the initial full model as being exponential with log likelihood $l(\theta)$ and canonical parameter $\varphi(\theta)$:

$$f^*(s;\theta) = \exp\{l(\theta) + [\{\varphi(\theta) - \varphi(\hat{\theta})'\}s]\}|\hat{j}_{\varphi\varphi'}|^{-1/2},$$

where s is the score variable with observed value $s = 0$.

The highly accurate likelihood-based p-value $p(\psi)$ assessing a scalar $\psi(\theta) = \psi$ can then be obtained by combining the r from (3) with the Q from (8) and using either of the asymptotically equivalent expressions:

$$\Phi(r) + \phi(r)\left\{\frac{1}{r} - \frac{1}{Q}\right\} \tag{12}$$

and

$$\Phi\left(r - r^{-1}\log\frac{r}{Q}\right) = \Phi(r^*), \tag{13}$$

due respectively to Lugannani and Rice (1980) and Barndorff-Nielsen (1991). These p-values have a distribution that is uniform (0,1) to third order; ϕ is the standard normal density function.

Example 1 revisited. For the simple exponential life model with $y = 1$ and $\theta = 8$, we obtained $\Phi(r) = 0.000853$ with exact value 0.000335. Since the parameter θ is a canonical parameter as it stands, we have $\varphi(\theta) = \theta$. The corresponding φ from (6) in this scalar canonical parameter case is just the q from (4) recorded earlier as $q = -7$. The resulting p-values from (12) and (13) are 0.000341 and 0.000346. These values offer a significant improvement from the likelihood value toward the exact.

Example 2. Consider a sample $(y_1, \ldots, y_n)$ from the two-parameter gamma distribution with density

$$f(y_i; \theta) = \Gamma^{-1}(\psi)\lambda^{-\psi} \exp\{\psi \log y - y/\lambda\}/y \quad y, \psi, \lambda > 0.$$

This model is exponential with canonical parameter $\varphi(\theta) = (\psi, \lambda^{-1})$ and has a minimal sufficient statistic $(x_1, x_2) = (\sum y_i, \sum \log y_i)$. Consider a small sample size $n = 2$ and data say $(1, 4)$. If we are interested in obtaining the p-value for $\psi = 1$, from (5) we obtain $\Phi(r) = 0.7965$ and from (12) and (13) we obtain 0.4321 and 0.4812. These latter third-order values compare very well for this small sample size with the exact value of 0.4000 obtained by numerical integration. The first-order likelihood ratio value is off by a factor of 2, the third-order value by 15–20%; but then this is the smallest possible sample size here and does involve the elimination of a nuisance parameter. In this extreme context the method is clearly not breaking down and we can attribute this to the intrinsic use of likelihood.

3. Seemingly unrelated regressions

3.1. Linear equations

Consider the following M equation model:

$$y_m = X_m\beta_m + \varepsilon_m, \quad m = 1, \ldots, M, \tag{14}$$

where y_m is a $N \times 1$ dependent variable, X_m is a $N \times k_m$ full column rank matrix of nonstochastic independent variables, β_m is the $k_m \times 1$ vector of regression coefficients, and ε_m is the $N \times 1$ stochastic error vector independent of X_m. The usual error structure for the classical linear regression formulation for $m = 1, \ldots, M$ is

$$E[\varepsilon_m] = 0, \quad E[\varepsilon_m\varepsilon_m'] = \sigma_m^2 I_N.$$

The above set of equations can be stacked and represented as the system

$$y = X\beta + \varepsilon, \tag{15}$$

where y is $NM \times 1$, X is $NM \times K$, β is $K \times 1$, ε is $NM \times 1$, $K = \sum_{m=1}^{M} k_m$, and $E[\varepsilon] = 0$. If the errors across equations are contemporaneously correlated then we have

$$E[\varepsilon\varepsilon'] = \begin{bmatrix} \sigma_1^2 I_N & \sigma_{12} I_N & \dots & \sigma_{1M} I_N \\ \sigma_{21} I_N & \sigma_2^2 I_N & \dots & \sigma_{2M} I_N \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} I_N & \sigma_{M2} I_N & \dots & \sigma_M^2 I_N \end{bmatrix} = \Sigma \otimes I_N. \tag{16}$$

If Σ is known, parameter estimates can be obtained by using the generalized least squares (GLS) estimator $b_{\mathrm{GLS}} = [X'(\Sigma^{-1} \otimes I_N)X]^{-1} X'(\Sigma^{-1} \otimes I_N)y$. In practice however, Σ is rarely known and for this case feasible generalized least squares (FGLS) estimators have been proposed. The equation-by-equation ordinary least squares residuals can be used to consistently estimate Σ. Both these estimators are due to Zellner (1962, 1963). Iterating on this FGLS procedure produces maximum likelihood estimates with equivalence conditions given in Oberhofer and Kmenta (1974).

For illustration and without any loss of generality,[1] let us focus on the two equations regression system. The methodology can be applied to several sets of seemingly unrelated regression equations. Consider:

$$y_{1i} = \alpha_0 + \alpha_1 w_{1i} + \alpha_2 z_{1i} + \varepsilon_{1i}, \quad i = 1, \dots, N, \tag{17}$$

$$y_{2i} = \gamma_0 + \gamma_1 w_{2i} + \gamma_2 z_{2i} + \varepsilon_{2i}, \quad i = 1, \dots, N. \tag{18}$$

Alternatively, using matrix notation, we have

$$y = X\beta + \varepsilon \;\Leftrightarrow\; \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},$$

where y_1 and y_2 are $N \times 1$, X_1 and X_2 are $N \times 3$, β_1 and β_2 are 3×1, and ε_1 and ε_2 are $N \times 1$, with

$$\varepsilon \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 I_N & \rho\sigma_1\sigma_2 I_N \\ \rho\sigma_1\sigma_2 I_N & \sigma_2^2 I_N \end{bmatrix} = \Sigma \otimes I_N \right), \tag{19}$$

where $\rho = \sigma_{12}/\sigma_1\sigma_2$ is the correlation coefficient. We emphasize that as long as we know the distributional form of the random error, the methodology is applicable. If the random error is not normally distributed, the likelihood function will change, and the r and Q will then be defined accordingly based on this function. The imposition here of normality is for ease of illustration; see Fraser et al. (1999a, b) for details when errors are not normally distributed.

[1] The third-order likelihood method can be applied in the context of any number of regression equations provided interest is on a scalar component parameter of the full model.

The log likelihood function is obtained as

$$L = -N\log(2\pi) - \frac{N}{2}\log(|\Sigma|) - \frac{1}{2}(y - X\beta)'(\Sigma^{-1} \otimes I_N)(y - X\beta). \tag{20}$$

Maximization of the log likelihood function over the parameter space produces the maximum likelihood estimator $\hat{\beta}$ of β and $\hat{\Sigma}$ of Σ. These quantities are then used to construct the signed square root of the log likelihood ratio r given in (5). For the calculation of Q, we need the conditioning vectors V in (7) obtained from an appropriate pivotal quantity.[2] For this, consider the model expressed in the form

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + (C \otimes I_N) \begin{bmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \end{bmatrix},$$

where $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2)$ are standard normal, and C is the lower triangular matrix defined by the decomposition $\Sigma = CC'$, namely

$$C = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{bmatrix},$$

see Anderson (1984). Note that the order of triangular factorization may have a minor effect on the assessment of ρ; we examine this elsewhere. We thus obtain the pivotal quantity (k_{1i}, k_{2i}) for an observation (y_{1i}, y_{2i}):

$$k_{1i} = (y_{1i} - \alpha_0 - \alpha_1 w_{1i} - \alpha_2 z_{1i})/\sigma_1, \tag{21}$$

$$k_{2i} = \frac{(y_{2i} - \gamma_0 - \gamma_1 w_{2i} - \gamma_2 z_{2i})/\sigma_2 - \rho(y_{1i} - \alpha_0 - w_{1i}\alpha_1 - z_{1i}\alpha_2)/\sigma_1}{\sqrt{1-\rho^2}}. \tag{22}$$

Here, k_{1i} represents the first variable, y_{1i}, after location scale standardization, while k_{2i} represents y_{2i} after location (including regression on y_{1i}) and scale standardization. The third-order methods in Section 2.2 then give inference for the scalar parameter of interest. That is, V, $\varphi(\theta)$, $\chi(\theta)$ and Q are calculated directly from (6) to (11) and tail probabilities can be obtained by either the Lugannani and Rice formula or the Barndorff-Nielsen formula.

In what follows we record the results of two simulations to assess the performance of the third-order method and then apply the method to several examples.

Example 3. A simulation study is performed to assess the performance of the first- and third-order methods.[3] We consider the two commodity demand model:

$$\log q_1 = \alpha_0 + (\gamma - (1/2)\delta)\log p_1 + \varepsilon_1,$$

$$\log q_2 = \beta_0 + (\gamma + (1/2)\delta)\log p_2 + \varepsilon_2,$$

[2] As sufficiency does not reduce the dimension of the variable y to that of the parameter θ, we need to implicitly construct a pseudo full rank exponential model through conditioning on an approximate ancillary.

[3] Computations were done using Splus Version 2.1, Maple 5 Release 4, and a program written by Jianrong Wu (available at http://fisher.utstat.toronto.edu/dfraser). Source code for any of the third-order calculations used in the paper is available upon request.

Table 1
Central coverage and noncoverage percentages

Method	Central coverage (%)	% Outside lower limit	% Outside upper limit
Nominal	90	5	5
lr	85.22	6.78	8.00
Bootstrap	84.97	8.13	6.90
Percentile-t	87.88	5.33	6.79
Bartlett	88.69	4.76	6.55
LR	90.06	4.52	5.42
BN	90.04	4.54	5.42
Nominal	95	2.5	2.5
lr	91.60	4.02	4.38
Bootstrap	92.04	3.78	4.18
Percentile-t	93.57	3.16	3.27
Bartlett	93.91	2.72	3.37
LR	95.32	2.26	2.42
BN	95.26	2.28	2.46
Nominal	99	0.5	0.5
lr	97.72	1.20	1.08
Bootstrap	97.48	1.37	1.15
Percentile-t	98.26	0.82	0.92
Bartlett	99.15	0.45	0.40
LR	99.00	0.56	0.44
BN	98.98	0.56	0.46

where q_1, q_2 and p_1, p_2 represent the quantities and prices of the two goods, respectively. ε_1 and ε_2 represent the stochastic error terms and are assumed to be normally distributed. Our interest is in testing the equality of the price elasticities. Five thousand simulated samples each of size 15 were randomly generated from the above equations with the following true values: $\alpha_0 = 3.5509$, $\beta_0 = 3.1797$, $\gamma = -0.6469$, $\delta = 0$, $\sigma_1^2 = 0.3458$, $\sigma_2^2 = 0.3127$, and $\rho = 0.4430$. The values for $\log p_1$ are (2.3761, 2.5675, 2.2240, 1.5271, 2.5684, 2.0420, 2.0022, 2.0117, 2.1707, 2.6035, 1.5980, 2.1235, 1.7441, 1.9775, 1.8896), and the values for $\log p_2$ are (1.4983, 2.3829, 1.7675, 2.6398, 2.4351, 2.1696, 1.9902, 1.8500, 1.4322, 0.6911, 1.9835, 1.7646, 1.6409, 2.2132, 1.6162).[4]

Table 1 below provides coverage and noncoverage percentages for nominal 90%, 95%, and 99% confidence intervals for covering the true δ value of 0. The labels "lr", "LR", and "BN" represent the likelihood ratio method, the Lugannani and Rice formula, and the Barndorff-Nielsen formula, respectively.

For comparison purposes, bootstrapped confidence intervals were computed using the classical nonparametric bootstrap and the percentile-t method. A Bartlett adjustment was also computed using the bootstrap as proposed in Rocke (1989). These intervals are reported in Table 1 under the headings "bootstrap", "percentile-t", and "Bartlett", respectively. For each of the 5000 replications, bootstrap samples

[4]Data values for p_1 and p_2 are the first 15 observations in Table 11.3 from Judge et al. (1988, p. 460).

of size 200 were used. The percentile-t method of bootstrapping was chosen in addition to the standard bootstrap since it was found to be consistently more accurate than the various other bootstrapping methods considered in the simulation experiments conducted in Rilstone and Veall (1996). As can been seen from Table 1, the results obtained from the classical bootstrap are not significantly better than those obtained from the first-order likelihood ratio method. However, the percentile-t bootstrapped confidence intervals do provide an improvement on those obtained from the likelihood ratio method, though they are still substantially less accurate than those obtained from the third-order likelihood methods. These results for the bootstrap are consistent with those reported in Rilstone and Veall (1996). Moderate gains to the percentile-t method can be obtained from using the Bartlett adjustment procedure. Notice that the third-order methods give more symmetric intervals on the percentage scale than the other methods.

The standard errors for the nominal 90%, 95%, and 99% central coverages based on 5000 simulated samples can be obtained by using the standard Binomial formula and they are 0.42%, 0.31%, and 0.14%, respectively. Similarly, the standard errors for the nominal 5%, 2.5%, and 0.5% tail coverages based on 5000 simulated samples can again be obtained by using the Binomial type formula and are 0.31%, 0.22%, and 0.10%, respectively.

All our third-order results in Table 1 lie within two standard errors of their nominal levels. The values resulting from the likelihood ratio method, the classical bootstrap, and the percentile-t bootstrap are all greater than two standard errors from their nominal levels. A similar statement can be made for the results based on the Bartlett adjustment with the exception of the results produced for the case of 99% coverage.

Example 4. A second simulation is performed using a system of three equations. Consider the three commodity demand model given in Judge et al. (1988, p. 460):

$$\log q_1 = \alpha_0 + \alpha_1 \log p_1 + \alpha_2 \log y + \varepsilon_1,$$

$$\log q_2 = \gamma_0 + \gamma_1 \log p_2 + \gamma_2 \log y + \varepsilon_2,$$

$$\log q_3 = \zeta_0 + \zeta_1 \log p_3 + \zeta_2 \log y + \varepsilon_3,$$

where q_1, q_2, q_3 and p_1, p_2, p_3 represent the quantities and prices of the three goods, respectively, income is represented by y. ε_1, ε_2, and ε_3 represent the stochastic error terms and are assumed to be normally distributed. Using the price and income data given in Judge et al. (1988, p. 460), 5000 simulated samples of size 30 were randomly generated from the above equations. The following true values were set: $\alpha_0 = -4$, $\alpha_1 = -1$, $\alpha_2 = 1.5$, $\gamma_0 = -3$, $\gamma_1 = -1$, $\gamma_2 = 1$, $\zeta_0 = 0.5$, $\zeta_1 = -1$, $\zeta_2 = 1$, $\sigma_1^2 = 0.17$, $\sigma_2^2 = 0.21$, $\sigma_3^2 = 0.03$, $\rho_{12} = -0.07$, $\rho_{13} = -0.74$, and $\rho_{23} = -0.52$. Our interest is in testing $\alpha_1 = -1$.

Table 2 below reports the simulation results with labels as given in Example 3. Again, for the bootstrap results, samples of size 200 were used for each replication. In terms of central coverage, the third-order methods (LR and BN) outperform all

Table 2
Central coverage and noncoverage percentages

Method	Central coverage (%)	% Outside lower limit	% Outside upper limit
Nominal	90	5	5
lr	86.17	7.00	6.83
Bootstrap	86.72	6.81	6.47
Percentile-t	87.02	6.35	6.63
Bartlett	88.43	5.89	5.68
LR	90.08	4.63	5.29
BN	90.05	4.68	5.27
Nominal	95	2.5	2.5
lr	91.59	3.85	4.56
Bootstrap	91.68	3.71	4.61
Percentile-t	92.62	2.98	4.40
Bartlett	93.68	2.82	3.50
LR	95.23	2.28	2.49
BN	95.22	2.30	2.48
Nominal	99	0.5	0.5
lr	97.67	1.29	1.04
Bootstrap	97.74	1.17	1.09
Percentile-t	98.27	0.85	0.88
Bartlett	98.63	0.65	0.72
LR	98.94	0.52	0.54
BN	98.98	0.54	0.48

the other methods considered. The computation time involved for these third-order methods was, however, significant for this simulation. The relative performance of the other methods is comparable to their performance in Example 3. The Bartlett adjustment procedure performs relatively well, producing better coverage than the likelihood ratio and the classical and percentile-t bootstraps. From the simulations examined, the superior coverage of the third-order method is clear.

As in the previous simulation experiment, all our third-order results in Table 2 lie within two standard errors of their nominal levels. The results produced using the likelihood ratio method, the classical bootstrap, and the percentile-t bootstrap are all at least two standard errors away from their nominal levels. The results based on the Bartlett adjustment lie in their respective intervals for the case of 99% coverage.

If interest lies in testing the equality of parameters across equations (e.g. $\alpha_1 = \gamma_1 = \zeta_1$) or on any multi-dimensional hypothesis, sequencing through scalar component parameters is necessary. The third-order methodology currently handles the testing of scalar component interest parameters.

Example 5. We consider the investment model and data discussed in Zellner (1962) and earlier in Boot and de Witt (1960). A firm's gross investment in period i, y_i, is modeled as a linear function of the firm's stock market value at the beginning of the period, w_i, and the firm's capital stock at the beginning of the period, z_i. Two US

firms are considered, General Electric and Westinghouse for a 20 year period. The investment equations for the two firms are given by (17) and (18), respectively.

Suppose we are interested in testing a common scalar component of the coefficient vectors. For example, suppose we are interested in one or other of the two hypotheses: $\gamma_1 - \alpha_1 = \delta$ and $\gamma_2 - \alpha_2 = \tilde{\delta}$. Given the way in which we have expressed the statistical model no explicit nuisance parameterization is obviously available for use to assess either of the above hypotheses. We therefore choose to reparameterize the model so as to provide an explicit nuisance parameterization. If for instance, we are interested in the parameter difference $\gamma_1 - \alpha_1 = \delta$, then we may reparameterize the model as follows:

$$y_{1i} = \alpha_0 + (\tilde{\gamma} + (1/2)\delta)w_{1i} + \alpha_2 z_{1i} + \varepsilon_{1i}, \quad i = 1, \ldots, N,$$

$$y_{2i} = \gamma_0 + (\tilde{\gamma} - (1/2)\delta)w_{2i} + \gamma_2 z_{2i} + \varepsilon_{2i}, \quad i = 1, \ldots, N$$

and assess the parameter δ. The parameter separation used is:

$$\theta' = (\delta, (\alpha_0, \tilde{\gamma}, \alpha_2, \gamma_0, \gamma_2, \sigma_1^2, \sigma_2^2, \rho)) = (\psi, \lambda)'.$$

Tables 3 and 4 provide 90%, 95%, and 99% confidence intervals for δ and $\tilde{\delta}$, respectively. We note that at the 90% acceptance level the likelihood ratio rejects the equality of the market value coefficients, whereas the opposite conclusion is reached using either of the third-order confidence intervals. This finding is consistent with the literature that finds that the standard first-order methods tend to over-reject tests.

Example 6. To illustrate the results for a system with greater than two equations, we consider Example 5 with the addition of a third firm, IBM. We also use Stata to produce conventional p-values for this example. Data for this firm is provided in Boot and de Witt (1960). The investment equations for General Electric, Westinghouse, and IBM are given below:

$$y_{1i} = \alpha_0 + \alpha_1 w_{1i} + \alpha_2 z_{1i} + \varepsilon_{1i}, \quad i = 1, \ldots, N,$$
$$y_{2i} = \gamma_0 + \gamma_1 w_{2i} + \gamma_2 z_{2i} + \varepsilon_{2i}, \quad i = 1, \ldots, N,$$
$$y_{3i} = \zeta_0 + \zeta_1 w_{3i} + \zeta_2 z_{3i} + \varepsilon_{3i}, \quad i = 1, \ldots, N.$$

Table 3
Confidence intervals for δ

Method	90% CI		95% CI		99% CI	
	Lower	Upper	Lower	Upper	Lower	Upper
lr	0.00151	0.03660	−0.00205	0.04038	−0.00953	0.04835
LR	−0.00030	0.03889	−0.00432	0.04323	−0.01278	0.05255
BN	−0.00029	0.03888	−0.00430	0.04320	−0.01275	0.05249

Table 4
Confidence intervals for $\tilde{\delta}$

Method	90% CI		95% CI		99% CI	
	Lower	Upper	Lower	Upper	Lower	Upper
lr	−0.15308	−0.00844	−0.16813	0.00614	−0.19919	0.03623
LR	−0.15517	0.00451	−0.17204	0.02114	−0.20724	0.05594
BN	−0.15517	0.00436	−0.17203	0.02089	−0.20721	0.05550

Table 5
p-values for testing either α_1 or γ_2

Method	$H_0 : \alpha_1 = 0$ vs. $H_1 : \alpha_1 \neq 0$	$H_0 : \gamma_2 = 0$ vs. $H_1 : \gamma_2 \neq 0$
mle	0.0108	0.0567
Bootstrap	0.0325	0.1204
lr	0.0159	0.0835
LR	0.0147	0.0741
BN	0.0147	0.0741

Suppose we are interested in testing a scalar component parameter of the model, say $\alpha_1 = 0$ or $\gamma_2 = 0$ against a two-sided alternative. Table 5 provides first- and third-order p-values in addition to standard bootstrap results for testing either of these two hypotheses. For the bootstrap results, 10,000 samples were used. The label "mle" refers to the p-value associated with the maximum likelihood Wald departure. The p-values for the "mle" and "bootstrap" cases are reported directly from Stata.

As evidenced from Table 5, these two standard quantities produce very discordant results. In fact, they can lead to different conclusions with respect to hypothesis testing. The p-values produced from the third-order likelihood methods are closer to the p-values obtained from the likelihood ratio method.

3.2. Nonlinear equations

Consider the following M equation model:

$$y_m = f_m(x_m, \beta_m) + \varepsilon_m, \quad m = 1, \ldots, M,$$

where y_m is the $N \times 1$ response vector, x_m is the k_m set of $N \times 1$ design vectors in equation m, β_m is the $k_m \times 1$ vector of unknown parameters, and the ε_m's are the $N \times 1$ stochastic error components each identically and independently distributed with mean zero and covariance matrix Σ. The above M equations can be written as

$$y = f(\beta) + \varepsilon,$$

where $y = (y_1', \ldots, y_M')'$ is $NM \times 1$, $f(\beta) = (f_1'(x_1, \beta_1), \ldots, f_M'(x_M, \beta_M))'$ is $NM \times 1$, and $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_M')'$ is $NM \times 1$.

With unknown Σ matrix one can proceed with FGLS, iterated FGLS provides maximum likelihood estimates. The equation-by-equation nonlinear least squares residuals can be used to consistently estimate Σ. To obtain the Q derived from higher order likelihood, the third-order method of Section 2.2 is applied directly.

Example 7. Consider the familiar Cobb–Douglas production function

$$Y = f(L, K) = AL^{\alpha}K^{\beta} \quad \alpha, \beta \geqslant 0,$$

where Y is output, L is labor, and K is capital. We consider two firms and for estimation purposes we assume additive contemporaneously correlated errors. We further assume that each firm's production technology is governed by the same production function. We estimate this model with 10 observations per firm. Assuming the random error has a normal distribution, Table 6 provides the 90%, 95%, and 99% confidence intervals for testing the parameter α. Table 7 provides the 90%, 95%, and 99% confidence intervals for testing the parameter γ, where γ is equal to $\alpha + \beta$.

The results consistently show that the likelihood ratio method produces shorter confidence intervals, viewed here as inappropriate. The present third-order likelihood approach is applicable to a wide spectrum of nonlinear problems in production theory.

Example 8. We now turn to a nonlinear problem with more nuisance parameters and more data. We consider the linear expenditure system discussed in Stone (1954). The specific model we consider is from Judge et al. (1988). Suppose a consumer with

Table 6
Confidence intervals for α

Method	90% CI		95% CI		99% CI	
	Lower	Upper	Lower	Upper	Lower	Upper
lr	−0.00728	0.33127	−0.04641	0.36508	−0.13327	0.43731
LR	−0.02810	0.35970	−0.07470	0.40127	−0.18500	0.49641
BN	−0.02820	0.35937	−0.07449	0.40063	−0.18414	0.49460

Table 7
Confidence intervals for γ

Method	90% CI		95% CI		99% CI	
	Lower	Upper	Lower	Upper	Lower	Upper
lr	0.71890	0.83852	0.70114	0.84857	0.65761	0.86923
LR	0.70515	0.84514	0.68215	0.85681	0.62229	0.88141
BN	0.70531	0.84509	0.68242	0.85673	0.61820	0.88120

Table 8
Confidence intervals for β_1

Method	90% CI		95% CI		99% CI	
	Lower	Upper	Lower	Upper	Lower	Upper
lr	0.14117	0.26245	0.12897	0.27456	0.10459	0.29934
LR	0.13731	0.26741	0.12378	0.28018	0.09692	0.30670
BN	0.13731	0.26739	0.12379	0.28016	0.09697	0.30666

utility function

$$U(q_1, q_2, q_3) = \beta_1 \log(q_1 - \gamma_1) + \beta_2 \log(q_2 - \gamma_2) + \beta_3 \log(q_3 - \gamma_3)$$

spends her income y on three goods q_1, q_2, and q_3 with respective prices p_1, p_2, and p_3. The individual's demand system is then characterized by the first-order conditions of the utility maximization problem

$$\max_{\{q_1,q_2,q_3\}} U(q_1, q_2, q_3) = \beta_1 \log(q_1 - \gamma_1) + \beta_2 \log(q_2 - \gamma_2) + \beta_3 \log(q_3 - \gamma_3)$$

subject to the budget constraint $y = p_1q_1 + p_2q_2 + p_3q_3$. The resulting demand functions (with $\beta_1 + \beta_2 + \beta_3 = 1$) are:

$$p_1q_1 = p_1\gamma_1 + \beta_1(y - p_1\gamma_1 - p_2\gamma_2 - p_3\gamma_3),$$
$$p_2q_2 = p_2\gamma_2 + \beta_2(y - p_1\gamma_1 - p_2\gamma_2 - p_3\gamma_3),$$
$$p_3q_3 = p_3\gamma_3 + \beta_3(y - p_1\gamma_1 - p_2\gamma_2 - p_3\gamma_3).$$

For estimation purposes we assume additive errors for each of the above demand functions and further assume that these errors are contemporaneously correlated. We do not impose the usual constraints: $q_1 > \gamma_1$, $q_2 > \gamma_2$, $q_3 > \gamma_3$, $0 < \beta_1 < 1$, $0 < \beta_2 < 1$, $0 < \beta_3 < 1$; we could however assess the validity of these constraints from our estimates. As one of the equations must be dropped for estimation purposes, we estimate the first two equations. We use the data given in Judge et al. (1988, p. 460). Table 8 records the 90%, 95%, and 99% confidence intervals for testing the parameter β_1.

Again, the tendency of the likelihood ratio method to over reject is evidenced by the narrower confidence intervals produced.

4. Conclusion

Third-order likelihood theory was applied to obtain highly accurate p-values for testing scalar interest components of a multi-dimensional parameter in the seemingly unrelated regression equations context. The results indicate that improved inferences can be made using the third-order likelihood method and that this method outperforms the likelihood ratio method, the classical bootstrap, the percentile-t

bootstrap, and the Bartlett bootstrap in terms of central coverage. The performance of this method with small sample sizes and the ease of computational implementation makes it a highly attractive and tractable alternative. As further exploration, we intend to examine SUR models for stochastically dependent errors, such as those that arise with time series data.

Acknowledgements

We would like to thank the editor, the associate editor and the two anonymous referees for helpful comments and suggestions. Fraser and Wong gratefully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada.

References

Anderson, T., 1984. An Introduction to Multivariate Statistical Analysis. Wiley, New York.
Attfield, C., 1995. A Bartlett adjustment to the likelihood ratio test for a system of equations. Journal of Econometrics 66, 207–223.
Attfield, C., 1998. Bartlett adjustments for systems of linear equations with linear restrictions. Economic Letters 60, 277–283.
Barndorff-Nielsen, O., 1991. Modified signed log-likelihood ratio. Biometrika 78, 557–563.
Barndorff-Nielsen, O., Cox, D.R., 1994. Inference and Asymptotics. Chapman & Hall, London.
Bartlett, M., 1937. Properties of sufficiency and statistical tests. Proceedings of the Society of London Series A 160, 259–268.
Berry, D., Chaloner, K., Geweke, J., 1996. Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner. Wiley, New York.
Boot, J., de Witt, G., 1960. Investment demand: an empirical contribution to the aggregation problem. International Economic Review 1, 3–30.
Fraser, D., 2004. Ancillaries and conditional inference. Statistical Science, forthcoming.
Fraser, D., Reid, N., 1995. Ancillaries and third order significance. Utilitas Mathematica 47, 33–53.
Fraser, D., Reid, N., Wu, J., 1999a. A simple general formula for tail probabilities for frequentist and Bayesian inference. Biometrika 86, 249–264.
Fraser, D., Wong, A., Wu, J., 1999b. Regression analysis, nonlinear or nonnormal: simple and accurate p values from likelihood analysis. Journal of the American Statistical Association 94 (448), 1286–1295.
Judge, G., Hill, R., Griffiths, W., Lutkepohl, H., Lee, T., 1988. Introduction to the Theory and Practice of Econometrics. Wiley, New York.
Lugannani, R., Rice, S., 1980. Saddlepoint approximation for the distribution of the sums of independent random variables. Advances in Applied Probability 12, 475–490.
Mittelhammer, R., Judge, G., Miller, D., 2000. Econometric Foundations. Cambridge University Press, Cambridge.
Oberhofer, W., Kmenta, J., 1974. A general procedure for obtaining maximum likelihood estimates in generalized regression models. Econometrica 43 (3), 579–590.
Rilstone, P., Veall, M., 1996. Using bootstrapped confidence intervals for improved inferences with seemingly unrelated regression equations. Econometric Theory 12, 569–580.
Rocke, D., 1989. Bootstrap Bartlett adjustment in seemingly unrelated regression. Journal of the American Statistical Association 84 (406), 598–601.
Stone, R., 1954. Linear expenditure systems and demand analysis: an application to the pattern of British demand. Economic Journal 64 (255), 511–527.

van der Merwe, A., Viljoen, C., 1988. Bayesian analysis of the seemingly unrelated regression model. Manuscript, Department of Mathematical Statistics, University of the Free State, Bloemfontein, SA.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. Journal of the American Statistical Association 57, 348–368.

Zellner, A., 1963. Estimators for seemingly unrelated regression equations: some exact finite sample results. Journal of the American Statistical Association 58, 977–992.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

Zellner, A., 1997. The Bayesian method of moments (BMOM): theory and applications. Advances in Econometrics 12, 85–105.