Computational Statistics and Data Analysis 56 (2012) 2442–2453

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bayesian inference for the correlation coefficient in two seemingly unrelated regressions

Min Wang, Xiaoqian Sun *

Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA

ARTICLE INFO

Article history:
Received 29 May 2011
Received in revised form 25 January 2012
Accepted 26 January 2012
Available online 4 February 2012

Keywords:
Seemingly unrelated regressions
Reference priors
Intrinsic discrepancy loss
Bayesian reference criterion
Hypothesis testing
Point estimation

ABSTRACT

We study the problems of hypothesis testing and point estimation for the correlation coefficient between the disturbances in the system of two seemingly unrelated regression equations. An objective Bayesian solution to each problem is proposed based on combined use of the invariant loss function and the objective prior distribution for the unknown model parameters. It is shown that this new solution possesses an invariance property under monotonic reparameterization of the quantity of interest. The performance of the proposed solution is examined through a simulation study. Furthermore, the solution is illustrated by an application to the real annual data for analyzing the investment model.

1. Introduction

Consider a system of two seemingly unrelated regression (SUR) equations

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \tag{1}$$

where in the ith equation $\mathbf{y}_i$ is an $n \times 1$ vector of observations, $\mathbf{X}_i$ is an $n \times p_i$ regressor matrix with rank p_i, $\boldsymbol{\beta}_i$ is a $p_i \times 1$ vector of unknown regression coefficients and $\boldsymbol{\varepsilon}_i$ is an $n \times 1$ vector of disturbances. The rows of the matrix $[\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2]$ are assumed to be independently distributed, each having a bivariate normal distribution with mean zero and unknown covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

where $\sigma_1 > 0$, $\sigma_2 > 0$ and $-1 < \rho < 1$. For ease of exposition, and without loss of generality, the system (1) can be written in compact form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

by letting $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$, $\mathbf{X} = \mathrm{diag}(\mathbf{X}_1, \mathbf{X}_2)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2')'$, where diag represents a block diagonal matrix. Here, $\boldsymbol{\varepsilon}$ follows a $2n$-dimensional normal distribution $N(\mathbf{0}_{2n}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, where $\mathbf{0}_{2n}$ is a $2n \times 1$ vector of zeros, $\mathbf{I}_n$ is an n-dimensional identity matrix and $\otimes$ denotes the Kronecker product operator. The parameter of interest in this paper is ρ, the correlation coefficient between the two equations. As we can see, $\rho = 0$ implies that the two equations are independent.

Such a system, first introduced by Zellner (1962), has been widely applied in many fields of economics, industry, and other sciences. This paper mainly focuses on two SUR equations, because this two-equation system is simple and has received

* Corresponding author.
E-mail address: xsun@clemson.edu (X. Sun).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.01.021

much attention in the past few decades. Zellner (1963) pointed out that Zellner's estimator of the regression coefficients is more efficient than the least squared estimator (LSE) when the disturbances are correlated between the equations. Nevertheless, as described in Kariya (1981), the LSE will be preferred when ρ is close to zero, since an estimated covariance matrix is involved in Zellner's estimator. Accordingly, the problem of testing the independence of the two equations becomes important and is entertained by Kariya (1981) and Kurata (2004) and the references therein.

There has been considerable interest in improving upon estimation of the regression coefficients of each equation. See Revankar (1974), Liu (2000, 2002a,b), Wang (2005) and Wang et al. (2011) and others. It is of interest to note that the efficiency of many proposed estimators often rests on the unknown parameter ρ. For example, the improved estimator proposed by Revankar (1974) is superior to the LSE only if $\rho^2 > 1/(n-r-1)$, where r is the rank of the matrix $[\mathbf{X}_1, \mathbf{X}_2]$. One further improved estimator derived by Liu (2002a) dominates both the estimator of Revankar (1974) and the LSE in terms of the mean square error matrix (MSEM) under the condition that $\rho^2 > \big(3/(n-r-4)\big)^{3/4}$. We can see that the point estimation of ρ is often required but has received less attention in the literature.

Motivated by these factors, we propose an objective Bayesian solution for the problems of hypothesis testing and point estimation for the correlation coefficient. The resulting solution has several appealing properties. For example, (i) the solution is invariant under one-to-one reparameterization of either the data or the parameter ρ; (ii) the solution can be used to test for any value of $\rho = \rho_0 \in (-1, 1)$; (iii) the solution depends only on the sampling model and the available data.

The paper is organized as follows. In Section 2, we derive several objective priors based on the Fisher information matrix. Furthermore, some appropriate sampling algorithms are proposed to generate samples from the posterior distributions under these priors. An objective Bayesian solution for the problems of hypothesis testing and point estimation for ρ is proposed in Section 3. In Section 4, the resulting solution is illustrated in the context of simulated and real data. The concluding remarks are given in Section 5, with additional proofs given in the Appendix.

2. Objective Bayesian analysis

2.1. Under the Jeffreys independent prior

In the absence of prior knowledge, the noninformative priors are often adopted for unknown parameters. One of the most popular choices in the literature is the Jeffreys independent prior

$$\pi_{IJ}(\boldsymbol{\beta}, \sigma_1, \sigma_2, \rho) = \pi(\boldsymbol{\beta})\pi(\sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{3/2}}, \tag{3}$$

which is derived in detail through using the Fisher information matrix in the Appendix. Under the prior $\pi_{IJ}(\boldsymbol{\beta}, \sigma_1, \sigma_2, \rho)$, the full conditional posterior distributions $f_1(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \mathbf{y})$ and $f_2(\boldsymbol{\Sigma} \mid \boldsymbol{\beta}, \mathbf{y})$ are respectively given by

$$\boldsymbol{\beta} \mid (\boldsymbol{\Sigma}, \mathbf{y}) \sim N\big(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Psi}\big) \tag{4}$$

and

$$\boldsymbol{\Sigma} \mid (\boldsymbol{\beta}, \mathbf{y}) \sim IW\big(A, n\big), \tag{5}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{y}$, $\boldsymbol{\Psi} = (\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{X})^{-1}$, $A = [a_{ij}] = [(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}_j)]$ $(i, j = 1, 2)$ and IW stands for the inverse Wishart distribution. The following Gibbs sampling algorithm can be employed to iteratively sample $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ from the full conditional posterior distributions in (4) and (5):

Step 1. (Initialization). Specify an initial value $\boldsymbol{\beta}^{(0)}$;
Step 2. Repeat for $k = 1, 2, \ldots, N$;
- (a): Simulate $(\sigma_1, \sigma_2, \rho)$ from the conditional distribution of $\boldsymbol{\Sigma}$ given $(\boldsymbol{\beta}^{(k-1)}, \mathbf{y})$ in (5) and report $\big(\sigma_1^{(k)}, \sigma_2^{(k)}, \rho^{(k)}\big) = (\sigma_1, \sigma_2, \rho)$;
- (b): Simulate $\boldsymbol{\beta}$ from the conditional distribution of $\boldsymbol{\beta}$ given $(\boldsymbol{\Sigma}^{(k)}, \mathbf{y})$ in (4) and report $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}$;

Step 3. Return the values $\big\{\sigma_1^{(1)}, \sigma_2^{(1)}, \rho^{(1)}, \boldsymbol{\beta}^{(1)}, \ldots, \sigma_1^{(N)}, \sigma_2^{(N)}, \rho^{(N)}, \boldsymbol{\beta}^{(N)}\big\}$.

2.2. Under the reference priors

Since the seminal paper of Bernardo (1979), the reference priors have been illustrated as one of the most useful tools for developing noninformative priors. This idea was pursued further in Berger and Bernardo (1992a,b). It is well known that reference priors in multiparameter problems rest on the grouped ordering of the parameters. In our problem, the regression coefficients $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ can be put in any position of the parameter ordering, because the Fisher information matrix of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1, \sigma_2, \rho)$ does not depend on them. Thus, given that the parameter ρ is of interest in this paper, we consider four grouped orderings: $\{\rho, \sigma_1, \sigma_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$, $\{\rho, \sigma_2, \sigma_1, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$, $\{\rho, (\sigma_1, \sigma_2), \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$ and $\{\rho, (\sigma_1, \sigma_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\}$. Following the same idea of Berger and Sun (2008) in dealing with reference priors for the bivariate normal model, we derive the following two types of reference priors.