

A Kullback-Leibler divergence based comparison of approximate Bayesian estimations of ARMA models

Ayman A Amin^{1,a}

^aDepartment of Statistics, Mathematics, and Insurance, Faculty of Commerce,
Menoufia University, Egypt

Abstract

Autoregressive moving average (ARMA) models involve nonlinearity in the model coefficients because of unobserved lagged errors, which complicates the likelihood function and makes the posterior density analytically intractable. In order to overcome this problem of posterior analysis, some approximation methods have been proposed in literature. In this paper we first review the main analytic approximations proposed to approximate the posterior density of ARMA models to be analytically tractable, which include Newbold, Zellner-Reynolds, and Broemeling-Shaarawy approximations. We then use the Kullback-Leibler divergence to study the relation between these three analytic approximations and to measure the distance between their derived approximate posteriors for ARMA models. In addition, we evaluate the impact of the approximate posteriors distance in Bayesian estimates of mean and precision of the model coefficients by generating a large number of Monte Carlo simulations from the approximate posteriors. Simulation study results show that the approximate posteriors of Newbold and Zellner-Reynolds are very close to each other, and their estimates have higher precision compared to those of Broemeling-Shaarawy approximation. Same results are obtained from the application to real-world time series datasets.

Keywords: approximate posteriors distance, Kullback-Leibler calibration, multivariate t distribution, Jeffreys' prior, natural conjugate prior

1. Introduction

Autoregressive moving average (ARMA) modeling of time series has been successfully applied in different fields such as economics, finance, and engineering. Bayesian analysis of ARMA models is difficult due to the nonlinearity in the coefficients of moving average (MA) part, which complicates the likelihood function and makes the posterior density analytically intractable (Amin, 2019a). Accordingly, the exact posterior analysis of ARMA models requires the use of numerical integration, which is computationally expensive. In order to overcome this problem, three well-known analytic approximations were proposed by Newbold (1973), Zellner and Reynolds (1978), and Broemeling and Shaarawy (1984, 1988) to approximate the posterior density to be an analytically tractable closed-form distribution.

Newbold (1973) proposed the first analytic approximation by expanding unobserved errors as a linear function in the model coefficients using the first order Taylor's expansion. In addition, Zellner and Reynolds (1978) proposed the second analytic approximation, denoted by Zellner-Reynolds, by

¹ Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Menoufia University, Shibin El Kom, Al Minufya, Al Minufya, Egypt. E-mail: ayman.aamin@gmail.com

expanding the errors sum of squares, rather than the errors, of ARMA model as a quadratic function in the model coefficients using the second order Taylor's expansion. Moreover, Broemeling and Shaarawy (1984, 1988) proposed the third analytic approximation, denoted by Broemeling-Shaarawy, by simply replacing unobserved lagged errors with their estimates to linearize the errors as a function in the model coefficients.

Because of the simplicity of Broemeling-Shaarawy analytic approximation compared to other proposed approximations, it has been widely extended to the Bayesian analysis of many time series models such as seasonal ARMA models (Amin, 2009; Ismail and Amin, 2014), multivariate MA models (Shaarawy and Ali, 2012), double seasonal MA models (Amin, 2017b), and double seasonal ARMA models (Amin, 2017a, 2018). However, few work have been introduced to evaluate and compare the accuracy of these three analytic approximations. Ismail (1994) used a small scale of simulation study to investigate the accuracy of the three approximations in the case of MA model of order one, and Ali (1998) and Soliman (1999) extended his work to the MA model of order two and to ARMA and seasonal ARMA of order one, respectively. Therefore, there is a need in the Bayesian time series analysis to study how these analytic approximations are mathematically related and to comprehensively evaluate their accuracy, aiming to help researchers to make a trade-off between the simplicity and accuracy of these approximations. As a motivation to this work, the outcome of these analytic approximations for complicated time series models can be used as inputs to advanced methods such as Markov chain Monte Carlo (MCMC) methods (Amin and Ismail, 2015; Amin, 2020, 2022). For example, in the case of seasonal ARMA models where the posterior density is nonlinear function in the model coefficients and thus analytically intractable. So, these analytic approximations can be used to approximate the posterior density, which in turn is used to derive the full conditional posteriors as a main requirement to apply MCMC methods to introduce the Bayesian estimation and prediction for the underlying seasonal ARMA models.

In order to fill this gap, we first review these three analytic approximations and then we use the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) to study mathematically the relation between them and to measure the distance between their derived approximate posteriors for ARMA models. In addition, we evaluate the impact of the approximate posteriors distance in the Bayesian estimates of the model coefficients and their precision by generating a large number of Monte Carlo simulations from the approximate posteriors. Therefore, we can summarize our work in this paper as follows. First, we derive the KL divergence between Newbold, Zellner-Reynolds, and Broemeling-Shaarawy approximate posteriors and compute its calibration to study mathematically the relation between them and to measure their distances. Second, we use a large number of Monte Carlo simulations from the approximate posteriors to evaluate the impact of the posteriors distance in the Bayesian estimates of mean and precision of the model coefficients. Finally, we use real-world time series datasets to illustrate the use of the KL divergence to measure the distance between the approximate posteriors and show the impact of this distance in the Bayesian estimates.

The remainder of this paper is organized as follows. In Section 2 we present the background of the ARMA models and related Bayesian concepts, and in Section 3 we summarize the analytic approximations. In Section 4 we derive the KL divergence between the approximate posteriors and its calibration. In Section 5 we present the simulation study and real-world time series datasets to measure the distance between the approximate posteriors and evaluate the impact of this distance in the Bayesian estimates. Finally, we give the conclusions in Section 6.

2. ARMA models and related Bayesian concepts

Time series $\{y_t\}$ can be modeled by an autoregressive moving average (ARMA) model of order (p, q) ,

simply denoted by ARMA(p, q), and written as (Box *et al.*, 2015):

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t, \quad (2.1)$$

where $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ and $\theta_q(B) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$ are the autoregressive and moving average polynomials with orders p and q respectively, and B is a backshift operator defined as $B^k z_t = z_{t-k}$. $\{\varepsilon_t\}$ is a sequence of independent normal errors with zero mean and variance $\sigma^2 < \infty$. The ARMA model (2.1) is stationary if all the roots of $\phi_p(B) = 0$ lie outside the unit circle, and if all the roots of $\theta_q(B) = 0$ lie outside the unit circle this model is invertible. For more details about the stationarity and invertibility properties of time series models see (Box *et al.*, 2015). The model (2.1) can be simplified as:

$$\begin{aligned} y_t &= \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \\ &= X_t \beta + \varepsilon_t, \end{aligned} \quad (2.2)$$

where $X_t = (y_{t-1}, \dots, y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q})$ is the t^{th} row of the design matrix X of order $n \times m$, i.e. $m = p + q$, and $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T$ is the model coefficients.

For ARMA model with normally distributed errors, the likelihood function is given by

$$\begin{aligned} L(\beta, \sigma^2 | \mathbf{y}) &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) \right\} \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{Q_n}{2\sigma^2} \right\}, \end{aligned} \quad (2.3)$$

where $Q_n = \varepsilon^T \varepsilon$ is the errors sum of squares, and the natural conjugate prior is normal-gamma distribution. So, suppose $\beta \sim N_m(\mu_\beta, \sigma^2 \Sigma_\beta)$ and $\sigma^2 \sim \text{IG}(\nu/2, \lambda/2)$, the joint natural conjugate prior distribution of β and σ^2 can be written as:

$$\zeta_n(\beta, \sigma^2) \propto (\sigma^2)^{-\left(\frac{\nu+m}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma^2} \left[\lambda + (\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) \right] \right\}, \quad (2.4)$$

where $\mu_\beta, \Sigma_\beta, \nu$ and λ are hyperparameters need to be estimated.

In case of little or no information is available about the model parameters, Jeffreys' prior can be employed, and it is given as

$$\zeta_j(\beta, \sigma^2) \propto (\sigma^2)^{-1}, \quad \sigma^2 > 0. \quad (2.5)$$

Multiplying the likelihood function (2.3) by each one of these two joint prior distributions results in the following joint posteriors. For the natural conjugate prior, the joint posterior of β and σ^2 is:

$$\zeta_n(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left(\frac{n+\nu+m}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma^2} \left[\lambda + (\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) + Q_n \right] \right\}. \quad (2.6)$$

For Jeffreys' prior, the joint posterior of β and σ^2 is:

$$\zeta_j(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp \left\{ -\frac{Q_n}{2\sigma^2} \right\}. \quad (2.7)$$

It is worth observing that the unknown lagged errors are part of the design matrices X which complicates the likelihood function and makes the joint posteriors of β and σ^2 analytically intractable. As a result, analytic approximations were introduced to approximate these joint posteriors as we review in the following section.

3. Analytic approximations

3.1. Newbold approximation

Newbold (1973) expanded the errors as a linear function in the model coefficients β around their non-linear least squares estimates (NLSE) by using the first order Taylor's expansion. So, the approximate errors for Newbold (ε_t^N) can be calculated using the following formula:

$$\varepsilon_t^N = \hat{\varepsilon}_t + U_t (\beta - \hat{\beta}), \quad (3.1)$$

where $\hat{\varepsilon}_t$ are the errors estimated recursively using NLSE $\hat{\beta}$ of the model coefficients as:

$$\hat{\varepsilon}_t = y_t - \hat{X}_t \hat{\beta} \quad (3.2)$$

and \hat{X}_t is the vector of the explanatory variables ($y_{t-1}, y_{t-2}, \dots, y_{t-p}, \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \dots, \hat{\varepsilon}_{t-q}$) after replacing the error terms by their NLSE's. U_t is a derivative vector that can be defined as:

$$\begin{aligned} U_t &= (U_{1t}, U_{2t}, \dots, U_{mt}) \\ &= \left(\frac{\partial \varepsilon_t}{\partial \beta_1}, \frac{\partial \varepsilon_t}{\partial \beta_2}, \dots, \frac{\partial \varepsilon_t}{\partial \beta_m} \right) \Bigg|_{\beta=\hat{\beta}}. \end{aligned} \quad (3.3)$$

Accordingly, the Newbold approximate errors sum of squares Q_n^N can be obtained as:

$$Q_n^N = \hat{Q}_n + (\beta - \hat{\beta})^T U^T U (\beta - \hat{\beta}), \quad (3.4)$$

where $\hat{Q}_n = \hat{\varepsilon}^T \hat{\varepsilon}$ and $\hat{\varepsilon}_t$ is computed as in Equation (3.2), and U is a matrix with the t^{th} row is U_t defined in Equation (3.3). Therefore, the Newbold approximate likelihood function is given by

$$L^N(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\hat{Q}_n + (\beta - \hat{\beta})^T U^T U (\beta - \hat{\beta}) \right] \right\} \quad (3.5)$$

and the Newbold approximate joint posterior of β and σ^2 , in case of natural conjugate prior, can be obtained as:

$$\zeta_n^N(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left(\frac{n+\nu+m}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma^2} \left[\lambda + (\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) + \hat{Q}_n + (\beta - \hat{\beta})^T U^T U (\beta - \hat{\beta}) \right] \right\}. \quad (3.6)$$

Using some mathematical manipulations, we can prove that the marginal approximate posterior of the model coefficients β is a multivariate t distribution with degrees of freedom $\nu_n = (n + \nu)$ and location vector and dispersion matrix are respectively:

$$\mu_n^N = A_n^{-1} B_n \quad \text{and} \quad V_n^N = \frac{1}{\nu_n - 2} A_n^{-1} \left[\hat{\beta}^T U^T U \hat{\beta} + \hat{Q}_n + \lambda + \mu_\beta^T \Sigma_\beta^{-1} \mu_\beta - B_n^T A_n^{-1} B_n \right], \quad (3.7)$$

where $A_n^{-1} = (U^T U + \Sigma_\beta^{-1})^{-1}$ and $B_n = (U^T U \hat{\beta} + \Sigma_\beta^{-1} \mu_\beta)$. By setting $\lambda = 0$, $\Sigma_\beta^{-1} = \mathbf{0}$, and $\nu = -m$, we can obtain the Newbold approximate marginal posterior of β in the case of Jeffreys' prior as a multivariate t distribution with degrees of freedom $\nu_j = (n - m)$ and location vector and dispersion matrix are respectively:

$$\mu_j^N = \hat{\beta} \quad \text{and} \quad V_j^N = \frac{\hat{Q}_n}{\nu_j - 2} (U^T U)^{-1}. \quad (3.8)$$

3.2. Zellner-Reynolds approximation

Zellner and Reynolds (1978) expanded the errors sum of squares Q_n , instead of the errors, as a quadratic function in the model coefficients using the second order of Taylor's expansion. The Zellner-Reynolds approximation of the errors sum of squares can be calculated as:

$$Q_n^{\text{ZR}} = \hat{Q}_n + \frac{1}{2} (\beta - \hat{\beta})^T R (\beta - \hat{\beta}), \quad (3.9)$$

where R is an $m \times m$ matrix of the second derivatives that can be defined as:

$$R = \begin{bmatrix} \frac{\partial^2 Q_n}{\partial \beta_1^2} & \frac{\partial^2 Q_n}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 Q_n}{\partial \beta_1 \partial \beta_m} \\ \frac{\partial^2 Q_n}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 Q_n}{\partial \beta_2^2} & \cdots & \frac{\partial^2 Q_n}{\partial \beta_2 \partial \beta_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 Q_n}{\partial \beta_m \partial \beta_1} & \frac{\partial^2 Q_n}{\partial \beta_m \partial \beta_2} & \cdots & \frac{\partial^2 Q_n}{\partial \beta_m^2} \end{bmatrix}. \quad (3.10)$$

Accordingly, the Zellner-Reynolds approximate likelihood function is given by

$$L^{\text{ZR}}(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\hat{Q}_n + \frac{1}{2} (\beta - \hat{\beta})^T R (\beta - \hat{\beta}) \right] \right\} \quad (3.11)$$

and by employing the natural conjugate prior we can obtain the approximate joint posterior of β and σ^2 as:

$$\zeta_n^N(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left(\frac{n+\nu+m}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma^2} \left[\lambda + (\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) + \hat{Q}_n + \frac{1}{2} (\beta - \hat{\beta})^T R (\beta - \hat{\beta}) \right] \right\}. \quad (3.12)$$

With some manipulations we can simply derive the marginal approximate posterior of the model coefficients β as a multivariate t distribution with degrees of freedom $v_n = (n + \nu)$ and location vector and dispersion matrix are respectively:

$$\mu_n^{\text{ZR}} = A_z^{-1} B_z \quad \text{and} \quad V_n^{\text{ZR}} = \frac{1}{v_n - 2} A_z^{-1} \left[\frac{1}{2} \hat{\beta}^T R \hat{\beta} + \hat{Q}_n + \lambda + \mu_\beta^T \Sigma_\beta^{-1} \mu_\beta - B_z^T A_z^{-1} B_z \right], \quad (3.13)$$

where $A_z^{-1} = ((1/2)R + \Sigma_\beta^{-1})^{-1}$ and $B_z = ((1/2)R\hat{\beta} + \Sigma_\beta^{-1}\mu_\beta)$. By setting $\lambda = 0$, $\Sigma_\beta^{-1} = \mathbf{0}$, and $\nu = -m$, we can obtain the approximate marginal posterior of β in the case of Jeffreys' prior as a multivariate t distribution with degrees of freedom $v_j = (n - m)$ and location vector and dispersion matrix are respectively:

$$\mu_j^{\text{ZR}} = \hat{\beta} \quad \text{and} \quad V_j^{\text{ZR}} = \frac{\hat{Q}_n}{v_j - 2} \left(\frac{1}{2} R \right)^{-1}. \quad (3.14)$$

3.3. Broemeling-Shaarawy approximation

Broemeling and Shaarawy (1984, 1988) approximated the unobserved errors by simply replacing the unobserved lagged errors with their NLSE's to be a linear function in the model coefficients as follows:

$$\begin{aligned} \varepsilon_t^{\text{BS}} &= y_t - \hat{X}_t \beta \\ &= \hat{\varepsilon}_t - \hat{X}_t (\beta - \hat{\beta}), \end{aligned} \quad (3.15)$$

where \hat{X}_t is t^{th} row of the design matrix after replacing the unobserved lagged error terms by their NLSE's. Therefore, the Broemeling-Shaarawy approximate errors sum of squares Q_n^{BS} can be obtained as:

$$Q_n^{\text{BS}} = \hat{Q}_n + (\beta - \hat{\beta})^T \hat{X}^T \hat{X} (\beta - \hat{\beta}) - 2(\beta - \hat{\beta})^T \hat{X}^T \hat{\varepsilon} \quad (3.16)$$

and the approximate likelihood function is given by

$$L^{\text{BS}}(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\hat{Q}_n + (\beta - \hat{\beta})^T \hat{X}^T \hat{X} (\beta - \hat{\beta}) - 2(\beta - \hat{\beta})^T \hat{X}^T \hat{\varepsilon} \right] \right\} \quad (3.17)$$

and the Broemeling-Shaarawy approximate joint posterior of β and σ^2 , in case of natural conjugate prior, can be obtained as:

$$\begin{aligned} \zeta_n^{\text{BS}}(\beta, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-(\frac{n+\nu+m}{2}+1)} \exp \left\{ -\frac{1}{2\sigma^2} \left[\lambda + (\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) + \hat{Q}_n \right. \right. \\ \left. \left. + (\beta - \hat{\beta})^T \hat{X}^T \hat{X} (\beta - \hat{\beta}) - 2(\beta - \hat{\beta})^T \hat{X}^T \hat{\varepsilon} \right] \right\}. \end{aligned} \quad (3.18)$$

Using some manipulations, we can derive the marginal approximate posterior of the model coefficients β to be a multivariate t distribution with degrees of freedom $\nu_n = (n + \nu)$ and location vector and dispersion matrix are respectively:

$$\mu_n^{\text{BS}} = A_b^{-1} B_b \quad \text{and} \quad V_n^{\text{BS}} = \frac{1}{\nu_n - 2} A_b^{-1} \left[\mathbf{y}^T \mathbf{y} + \lambda + \mu_\beta^T \Sigma_\beta^{-1} \mu_\beta - B_b^T A_b^{-1} B_b \right], \quad (3.19)$$

where $A_b^{-1} = (\hat{X}^T \hat{X} + \Sigma_\beta^{-1})^{-1}$ and $B_b = (\hat{X}^T \mathbf{y} + \Sigma_\beta^{-1} \mu_\beta)$. By setting $\lambda = 0$, $\Sigma_\beta^{-1} = \mathbf{0}$, and $\nu = -m$, the approximate marginal posterior of β in the case of Jeffreys' prior can be obtained as a multivariate t distribution with degrees of freedom $\nu_j = (n - m)$ and location vector and dispersion matrix are respectively:

$$\mu_j^{\text{BS}} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T \mathbf{y} \quad \text{and} \quad V_j^{\text{BS}} = \frac{\hat{Q}_n}{\nu_j - 2} (\hat{X}^T \hat{X})^{-1}. \quad (3.20)$$

4. Kullback-Leibler divergence of approximate posteriors

The Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) can be used to measure the distance between any two of the approximate posteriors of the ARMA model coefficients introduced in Section 3. To explain the idea, suppose we have two multivariate t posteriors for β : the first is $\zeta_1(\beta | \mathbf{y})$ with parameters mean μ_1 , dispersion matrix V_1 , and degrees of freedom ν_1 , and the second is $\zeta_2(\beta | \mathbf{y})$ with parameters mean μ_2 , dispersion matrix V_2 , and degrees of freedom ν_2 . The KL divergence between these two posteriors can be written as:

$$\text{KL}[\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})] = \int \ln \left(\frac{\zeta_1(\beta | \mathbf{y})}{\zeta_2(\beta | \mathbf{y})} \right) \zeta_1(\beta | \mathbf{y}) d\beta. \quad (4.1)$$

This KL divergence is not symmetric and always non-negative, and it equals to zero when the two posteriors are the same. This implies that the smaller its value, the closer are the two posteriors. We

can simplify the KL divergence (4.1) as:

$$\begin{aligned} \text{KL} [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})] &= \int [-\ln(\zeta_2(\beta | \mathbf{y}))] \zeta_1(\beta | \mathbf{y}) d\beta - \int [-\ln(\zeta_1(\beta | \mathbf{y}))] \zeta_2(\beta | \mathbf{y}) d\beta \\ &= \text{CH} [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})] - H [\zeta_1(\beta | \mathbf{y})], \end{aligned} \quad (4.2)$$

where $\text{CH} [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})]$ and $H [\zeta_1(\beta | \mathbf{y})]$ are known as the cross-entropy and Shannon entropy respectively (Amin, 2017c). In our previous work (Amin, 2019b), we derived the approximate KL divergence between the two multivariate t posteriors $\zeta_1(\beta | \mathbf{y})$ and $\zeta_2(\beta | \mathbf{y})$ as:

$$\text{KL}[\zeta_1(\beta|\mathbf{y}), \zeta_2(\beta|\mathbf{y})] \approx \frac{1}{2} \log \left(\frac{|V_2|}{|V_1|} \right) + \frac{1}{2} \left(\frac{v_2+m}{v_2} \right) \left[\left(\frac{v_1}{v_1-2} \right) \text{tr}(V_2^{-1}V_1) + (\mu_1 - \mu_2)^T V_2^{-1} (\mu_1 - \mu_2) \right] - \frac{m}{2}. \quad (4.3)$$

This derived approximate KL divergence is not symmetric; however, we can compute the symmetric distance $\text{KL}^* [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})]$ as the average of two KL divergences as follows:

$$\begin{aligned} \text{KL}^* [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})] &= \frac{1}{2} \{ \text{KL} [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})] + \text{KL} [\zeta_2(\beta | \mathbf{y}), \zeta_1(\beta | \mathbf{y})] \} \\ &\approx \frac{1}{4} \{ \text{tr}(V_1^{*-1} V_2 + V_2^{*-1} V_1) + (\mu_1 - \mu_2)^T V^{-1} (\mu_1 - \mu_2) - 2m \}, \end{aligned} \quad (4.4)$$

where $V_1^* = ((1 - 2/v_2)/(1 + m/v_1))V_1$, $V_2^* = ((1 - 2/v_1)/(1 + m/v_2))V_2$, and $V^{-1} = (1 + m/v_1)V_1^{-1} + (1 + m/v_2)V_2^{-1}$.

Now, we can use this derived approximate KL divergence in Equation (4.4) to study the relation between the analytic approximations summarized in Section 3 and to measure the distance between their approximate posteriors. To do that we have to observe two things. First, the derived approximate KL divergence depends on the difference between the approximate posteriors means and the ratio of the posteriors dispersions. Second, all the marginal approximate posteriors of β derived in Section 3 are multivariate t distribution with the same degrees of freedom but with different location vector and dispersion matrix. Aiming to simplify the comparison, we compute the KL divergence between the marginal approximate posteriors of β in the case of employong Jeffreys' prior.

Using the location vectors and dispersion matrices given in Equations (3.8) and (3.14), we can simplify the KL divergence between the Newbold and Zellner-Reynolds approximate posteriors of β as:

$$\text{KL}^* [\zeta_j^N(\beta | \mathbf{y}), \zeta_j^{\text{ZR}}(\beta | \mathbf{y})] \approx \frac{1}{4} \left\{ \frac{n}{n-m-2} \text{tr} \left(U^T U \left(\frac{1}{2} R \right)^{-1} + \frac{1}{2} R \left(U^T U \right)^{-1} \right) - 2m \right\}, \quad (4.5)$$

which indicates that the distance between the Newbold and Zellner-Reynolds approximate posteriors depends mainly on the relation between the two matrices $U^T U$ and $(1/2)R$. To investigate this relation, we can redefine the R matrix in Equation (3.10) by letting its ij^{th} element to be defined as:

$$\begin{aligned} R_{ij} &= \frac{d^2 Q}{d\beta_i d\beta_j} \\ &= 2 \left\{ \sum_{t=1}^n \left(\varepsilon_t \frac{\partial^2 \varepsilon_t}{\partial \beta_i \partial \beta_j} \right) + \sum_{t=1}^n \left(\frac{\partial \varepsilon_t}{\partial \beta_i} \times \frac{\partial \varepsilon_t}{\partial \beta_j} \right) \right\} \\ &= 2 \left\{ D_{ij} + (U^T U)_{ij} \right\}, \end{aligned} \quad (4.6)$$

which means $(1/2)R = D + U^T U$ and implies that $U^T U$ is a part of $(1/2)R$. Therefore, the distance between the Newbold and Zellner-Reynolds approximate posteriors is expected to be very small, i.e. they are very close to each other, especially when the matrix D closes to $\mathbf{0}$, which is evaluated by the simulation study in the next section. Similarly, we use the location vectors and dispersion matrices given in Equations (3.8) and (3.20) to simplify the KL divergence between the Newbold and Broemeling-Shaarawy approximate posteriors of β as:

$$\text{KL}^* [\zeta_j^N(\beta | \mathbf{y}), \zeta_j^{\text{BS}}(\beta | \mathbf{y})] \approx \frac{1}{4} \left\{ \frac{n}{n-m-2} \text{tr} \left(\hat{X}^T \hat{X} (U^T U)^{-1} + U^T U (\hat{X}^T \hat{X})^{-1} \right) + \frac{n(n-m-2)}{(n-m) \hat{Q}_n} (\hat{\beta} - \mu_j^{\text{BS}})^T \left[\hat{X}^T \hat{X} + U^T U \right] (\hat{\beta} - \mu_j^{\text{BS}}) - 2m \right\}, \quad (4.7)$$

which depends on the difference between the two approximate posteriors means and the ratio of the two matrices $\hat{X}^T \hat{X}$ and $U^T U$. Since the matrix U contains more information about the unobserved errors compared to the matrix \hat{X}^T , it is expected that the distance between the Newbold and Broemeling-Shaarawy approximate posteriors is very large, i.e. they are strongly different, which we evaluate by the simulation study in the next section. We can obtain similar conclusion about the distance between the Zellner-Reynolds and Broemeling-Shaarawy approximate posteriors, since their KL divergence can be simplified as:

$$\text{KL}^* [\zeta_j^{\text{ZR}}(\beta | \mathbf{y}), \zeta_j^{\text{BS}}(\beta | \mathbf{y})] \approx \frac{1}{4} \left\{ \frac{n}{n-m-2} \text{tr} \left(\hat{X}^T \hat{X} \left(\frac{1}{2} R \right)^{-1} + \frac{1}{2} R (\hat{X}^T \hat{X})^{-1} \right) + \frac{n(n-m-2)}{(n-m) \hat{Q}_n} (\hat{\beta} - \mu_j^{\text{BS}})^T \left[\hat{X}^T \hat{X} + \frac{1}{2} R \right] (\hat{\beta} - \mu_j^{\text{BS}}) - 2m \right\}. \quad (4.8)$$

It has to be noted that possible values of the approximate KL divergence are non-negative with no maximum limit, however, we can calibrate the values to be in the interval (0.5,1.0) to be able to decide about the distance between the approximate posteriors. When the calibration value closes to 0.5, it implies the two approximate posteriors are almost the same; and when its value closes to 1.0, the two approximate posteriors are strongly different. Suppose $\text{KL} [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})] = k$, using the idea of calibration proposed by McCulloch (1989) we can compute the calibration of KL divergence (KLC) between the two approximate posteriors as:

$$\text{KLC} [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})] = \frac{1}{2} \{ 1 - \exp(-2\text{KL} [\zeta_1(\beta | \mathbf{y}), \zeta_2(\beta | \mathbf{y})]) \}. \quad (4.9)$$

All the derived approximate KL divergences of the approximate posteriors and their calibration will be used in the next section to measure the distance between these posteriors and the impact of this distance in the Bayesian estimates using simulated and real-world time series datasets.

5. Simulation study and real application

We have two parts of work in this section. First, we present a simulation study to evaluate the KL divergence (and its calibration) between the considered approximate posteriors of the ARMA model coefficients and evaluate the impact of posteriors divergence in the Bayesian estimates. Second, we present two applications of our work to real world time series datasets.

Table 1: Results of KL divergence and its calibration (KLC) for MA(1)

θ		Sample size (n)					
		50	100	150	200	250	300
0.4	KL[N,ZR]*	0.067	0.030	0.019	0.014	0.011	0.009
	KLC[N,ZR]	0.661	0.614	0.593	0.580	0.571	0.565
	KL[N,BS]	0.373	0.136	0.123	0.111	0.105	0.102
	KLC[N,BS]	0.725	0.703	0.695	0.689	0.685	0.683
	KL[ZR,BS]	0.465	0.139	0.127	0.116	0.108	0.105
	KLC[ZR,BS]	0.738	0.708	0.699	0.692	0.686	0.684
0.7	KL[N,ZR]	0.098	0.151	0.027	0.021	0.016	0.013
	KLC[N,ZR]	0.679	0.628	0.604	0.592	0.582	0.574
	KL[N,BS]	1.244	0.703	0.546	0.554	0.545	0.538
	KLC[N,BS]	0.864	0.856	0.852	0.853	0.851	0.850
	KL[ZR,BS]	1.406	0.601	0.545	0.545	0.532	0.531
	KLC[ZR,BS]	0.860	0.852	0.848	0.850	0.848	0.848

* N, ZR and BS refer to Newbold, Zellner-Reynolds and Broemeling-Shaarawy approximate posteriors, respectively.

5.1. Simulation study

In this simulation study, we have two objectives: (1) evaluating the KL divergence (and its calibration) between the considered approximate posteriors of the ARMA model coefficients, and (2) evaluating the impact of this posteriors divergence in the Bayesian estimates for several simulated time series data with different sample sizes, different ARMA model orders, and different values of the ARMA model coefficients. In the following we discuss the simulation study design and results for each objective.

5.1.1. Evaluating the Kullback-Leibler divergence and its calibration

In order to evaluate the KL divergence (and its calibration) between the considered three approximate posteriors of the ARMA model coefficients, we generate 1,000 time series of size n (from 50 to 300 with an increment value of 50) from different models: MA(1) with $\theta = 0.4$ and 0.7 , MA(2) with $(\theta_1, \theta_2) = (0.4, 0.3)$ and $(0.3, 0.6)$, ARMA(1, 1) with $(\phi, \theta) = (0.5, 0.3)$ and $(0.4, 0.6)$, ARMA(1, 2) with $(\phi, \theta_1, \theta_2) = (0.4, 0.5, 0.4)$ and $(0.7, 0.3, 0.4)$, and ARMA(2, 2) with $(\phi_1, \phi_2, \theta_1, \theta_2) = (0.4, -0.3, 0.5, 0.4)$ and $(1.3, -0.6, 0.3, 0.6)$; and in all these models we set the model variance $\sigma^2 = 1$. For each time series, we compute the location vectors and dispersion matrices given in Equations (3.8), (3.14) and (3.20) for the Newbold, Zellner-Reynolds and Broemeling-Shaarawy approximate posteriors of the ARMA model coefficients β . We then use these location vectors and dispersion matrices to compute the KL divergence and its calibration between the corresponding approximate posteriors. The average of the KL divergence and its calibration are computed and presented in Tables 1–5.

From these simulation results, we can observe some important remarks.

- First, the KL divergence between the Newbold and Zellner-Reynolds approximate posteriors is very small especially for simple models and large sample size, for example their KL calibration values are between 0.57 and 0.71 in most of the cases for $n = 300$. This confirms that the Newbold and Zellner-Reynolds approximate posteriors are very close to each other. On contrary, the KL divergence between the Newbold and Broemeling-Shaarawy approximate posteriors is very large, since the KL calibration values are between 0.77 and 0.97 in most of the cases for $n = 300$, which confirms that these approximate posteriors are strongly different. Similar conclusion can be drawn about Zellner-Reynolds and Broemeling-Shaarawy approximate posteriors.
- Second, when the sample size increases the approximate posteriors divergences decreases, since

Table 2: Results of KL divergence and its calibration (KLC) for MA(2)

(θ_1, θ_2)		Sample size (n)					
		50	100	150	200	250	300
(0.4, 0.3)	KL[N,ZR]	0.188	0.083	0.053	0.039	0.031	0.025
	KLC[N,ZR]	0.763	0.688	0.653	0.632	0.618	0.607
	KL[N,BS]	0.618	0.274	0.236	0.217	0.211	0.205
	KLC[N,BS]	0.833	0.798	0.784	0.775	0.772	0.769
	KL[ZR,BS]	0.726	0.301	0.253	0.229	0.221	0.213
	KLC[ZR,BS]	0.851	0.808	0.790	0.780	0.776	0.773
(0.3, 0.6)	KL[N,ZR]	0.482	0.116	0.068	0.049	0.039	0.032
	KLC[N,ZR]	0.789	0.708	0.669	0.645	0.631	0.619
	KL[N,BS]	2.176	0.805	0.693	0.641	0.625	0.610
	KLC[N,BS]	0.910	0.895	0.891	0.886	0.883	0.882
	KL[ZR,BS]	2.230	0.822	0.698	0.645	0.628	0.611
	KLC[ZR,BS]	0.918	0.898	0.893	0.887	0.884	0.883

Table 3: Results of KL divergence and its calibration (KLC) for ARMA(1,1)

(ϕ, θ)		Sample size (n)					
		50	100	150	200	250	300
(0.5, 0.3)	KL[N,ZR]	0.145	0.073	0.044	0.033	0.026	0.022
	KLC[N,ZR]	0.738	0.673	0.640	0.623	0.610	0.600
	KL[N,BS]	0.625	0.204	0.164	0.143	0.131	0.128
	KLC[N,BS]	0.807	0.761	0.742	0.732	0.726	0.723
	KL[ZR,BS]	0.795	0.231	0.180	0.156	0.142	0.137
	KLC[ZR,BS]	0.825	0.775	0.752	0.740	0.732	0.729
(0.4, 0.6)	KL[N,ZR]	0.184	0.074	0.046	0.035	0.027	0.022
	KLC[N,ZR]	0.743	0.676	0.642	0.624	0.611	0.601
	KL[N,BS]	1.526	0.748	0.576	0.550	0.544	0.540
	KLC[N,BS]	0.906	0.893	0.885	0.882	0.881	0.882
	KL[ZR,BS]	1.719	0.730	0.598	0.561	0.550	0.547
	KLC[ZR,BS]	0.911	0.894	0.885	0.882	0.881	0.882

Table 4: Results of KL divergence and its calibration (KLC) for ARMA(1,2)

$(\phi, \theta_1, \theta_2)$		Sample size (n)					
		50	100	150	200	250	300
(0.4, 0.5, 0.4)	KL[N,ZR]	0.440	0.198	0.107	0.080	0.062	0.048
	KLC[N,ZR]	0.841	0.755	0.710	0.683	0.662	0.647
	KL[N,BS]	1.991	0.873	0.722	0.666	0.649	0.628
	KLC[N,BS]	0.937	0.923	0.917	0.912	0.910	0.909
	KL[ZR,BS]	2.236	0.930	0.758	0.695	0.672	0.648
	KLC[ZR,BS]	0.948	0.927	0.921	0.915	0.913	0.911
(0.7, 0.3, 0.4)	KL[N,ZR]	0.359	0.145	0.091	0.066	0.050	0.041
	KLC[N,ZR]	0.828	0.742	0.698	0.672	0.652	0.638
	KL[N,BS]	1.334	0.661	0.565	0.524	0.502	0.484
	KLC[N,BS]	0.927	0.902	0.894	0.886	0.883	0.879
	KL[ZR,BS]	1.467	0.701	0.591	0.545	0.517	0.498
	KLC[ZR,BS]	0.937	0.909	0.898	0.890	0.886	0.882

more information are provided by the observed time series enabling the analytic approximations to reduce the uncertainty about the model parameters. For example, in case of ARMA(1, 1) with $(\phi, \theta) = (0.5, 0.3)$, the KLC between the Newbold and Zellner-Reynolds posteriors is about 0.74 for $n = 50$ and it becomes about 0.60 for $n = 300$; and the KLC between the Newbold and Broemeling-Shaarawy posteriors is about 0.81 for $n = 50$ and it is reduced to be about 0.72 for $n = 500$.

Table 5: Results of KL divergence and its calibration (KLC) for ARMA(2,2)

$(\phi_1, \phi_2, \theta_1, \theta_2)$		Sample size (n)					
		50	100	150	200	250	300
$(0.4, -0.3, 0.5, 0.4)$	KL[N,ZR]	2.537	0.370	0.239	0.207	0.136	0.109
	KLC[N,ZR]	0.902	0.817	0.776	0.749	0.724	0.706
	KL[N,BS]	4.038	1.532	1.148	0.913	0.864	0.825
	KLC[N,BS]	0.957	0.928	0.925	0.921	0.920	0.919
	KL[ZR,BS]	6.679	1.696	1.270	1.030	0.915	0.876
	KLC[ZR,BS]	0.970	0.943	0.936	0.934	0.928	0.926
$(1.3, -0.6, 0.3, 0.6)$	KL[N,ZR]	0.595	0.257	0.152	0.111	0.085	0.070
	KLC[N,ZR]	0.887	0.798	0.747	0.715	0.692	0.676
	KL[N,BS]	4.310	1.691	1.496	1.425	1.334	1.297
	KLC[N,BS]	0.977	0.975	0.973	0.972	0.970	0.969
	KL[ZR,BS]	4.652	1.701	1.527	1.457	1.359	1.320
	KLC[ZR,BS]	0.981	0.977	0.975	0.973	0.971	0.969

- Third, the more complicated ARMA model the larger KL divergence between the approximate posteriors. For example, in case of $n = 100$, the KLC between the Newbold and Zellner-Reynolds and between the Newbold and Broemeling-Shaarawy posteriors are about 0.61 and 0.70 respectively for MA(1) with $\theta = 0.4$ and they are increased to be about 0.82 and 0.93 respectively for ARMA(2,2) with $(\phi_1, \phi_2, \theta_1, \theta_2) = (0.4, -0.3, 0.5, 0.4)$.

5.1.2. Evaluating the impact of posteriors distance in Bayesian estimates

In order to evaluate the distance impact of the Newbold, Zellner-Reynolds and Broemeling-Shaarawy approximate posteriors in the Bayesian estimates of the ARMA model coefficients β , we use the Monte Carlo simulations to obtain samples of β from these approximate posteriors as follows.

- First, we generate 1,000 time series of size n (from 50 to 300 with an increment value of 50) from different ARMA models: MA(1) with $\theta = 0.4$ and 0.7 , MA(2) with $(\theta_1, \theta_2) = (0.3, 0.6)$, and ARMA(1, 1) with $(\phi, \theta) = (0.3, -0.7)$; and in all these models we set the model variance $\sigma^2 = 1$.
- Second, for each time series, we generate 1,000 values of β from the Newbold, Zellner-Reynolds and Broemeling-Shaarawy marginal approximate posteriors of β .
- Third, for each Monte Carlo simulation chain of β , we compute the Bayesian estimates of β which include mean, standard deviation, and quantiles 2.5% and 97.5% as an 95% credible interval.
- Fourth, we compute the mean absolute percentage errors (MAPE) to compare the Bayesian estimates of β , obtained by simulations from different approximate posteriors, to the true value used to generate the underlying time series.
- Fifth, we evaluate the variability of β obtained from each of the approximate posteriors by calculating the ratio of standard deviations of β from different posteriors, for each generated chain; and then we report the mean and quantiles 2.5% and 97.5% of these ratios. Results of these simulations are presented in Tables 6–11 and discussed in the following.

Results of the Bayesian estimates of the MA(1) coefficients are presented in Table 6 and their comparison criteria are presented in Table 7. From these results we can observe that the Bayesian estimates of the coefficients mean obtained from all the approximate posteriors are almost the same, however, the standard deviation estimates obtained from the Newbold and Zellner-Reynolds approximate posteriors are strongly different from those obtained from the Broemeling-Shaarawy approximate posterior.

Table 6: Bayesian estimates of the coefficients of MA(1)

θ	n	$\zeta_j^N(\beta \mathbf{y})$				$\zeta_j^{ZR}(\beta \mathbf{y})$				$\zeta_j^{BS}(\beta \mathbf{y})$			
		$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$
0.4	50	0.413	0.016	0.108	0.715	0.413	0.018	0.108	0.715	0.410	0.020	0.107	0.708
	100	0.406	0.008	0.210	0.607	0.406	0.009	0.210	0.607	0.402	0.010	0.201	0.583
	150	0.405	0.006	0.242	0.562	0.405	0.006	0.242	0.562	0.403	0.007	0.236	0.567
	200	0.403	0.004	0.266	0.533	0.403	0.004	0.266	0.533	0.402	0.005	0.254	0.540
	250	0.402	0.003	0.283	0.520	0.402	0.003	0.283	0.520	0.401	0.004	0.272	0.530
	300	0.403	0.003	0.301	0.506	0.403	0.003	0.301	0.506	0.402	0.003	0.291	0.517
0.7	50	0.723	0.010	0.483	1.000	0.723	0.011	0.483	1.000	0.715	0.020	0.405	1.007
	100	0.710	0.005	0.546	0.867	0.710	0.006	0.546	0.867	0.704	0.010	0.503	0.883
	150	0.707	0.003	0.576	0.831	0.707	0.004	0.576	0.831	0.704	0.007	0.536	0.862
	200	0.705	0.003	0.596	0.809	0.705	0.003	0.596	0.809	0.703	0.005	0.552	0.840
	250	0.704	0.002	0.614	0.795	0.704	0.002	0.614	0.795	0.702	0.004	0.574	0.828
	300	0.704	0.002	0.628	0.786	0.704	0.002	0.628	0.786	0.702	0.003	0.591	0.814

Table 7: Comparison criteria between estimates from different posteriors for MA(1).

θ	n	MAPE			$\hat{\sigma}_j^{ZR} / \hat{\sigma}_j^N$			$\hat{\sigma}_j^{BS} / \hat{\sigma}_j^N$		
		$\hat{\mu}_j^N$	$\hat{\mu}_j^{ZR}$	$\hat{\mu}_j^{BS}$	$\hat{\mu}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$Q_{2.5\%}$	$Q_{97.5\%}$
0.4	50	58.331	58.331	60.645	1.132	0.630	2.348	1.399	1.013	2.153
	100	39.391	39.391	40.614	1.066	0.708	1.735	1.223	1.042	1.570
	150	31.481	31.481	33.127	1.040	0.743	1.523	1.212	1.061	1.465
	200	27.033	27.033	29.372	1.030	0.772	1.421	1.205	1.074	1.395
	250	23.999	23.999	26.294	1.025	0.798	1.349	1.202	1.087	1.368
	300	21.747	21.747	23.791	1.022	0.809	1.322	1.200	1.099	1.342
0.7	50	28.388	28.388	34.865	1.276	0.549	3.199	3.520	1.261	19.866
	100	18.299	18.299	23.246	1.120	0.651	2.116	2.234	1.412	4.095
	150	14.496	14.496	18.943	1.066	0.686	1.752	2.101	1.485	3.187
	200	12.217	12.217	16.757	1.055	0.723	1.653	2.042	1.541	2.824
	250	10.532	10.532	14.989	1.047	0.742	1.591	2.025	1.601	2.702
	300	9.541	9.541	13.526	1.035	0.754	1.489	2.014	1.645	2.598

Table 8: Bayesian estimates of the coefficients of MA(2)

	n	$\zeta_j^N(\beta \mathbf{y})$				$\zeta_j^{ZR}(\beta \mathbf{y})$				$\zeta_j^{BS}(\beta \mathbf{y})$			
		$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$
$\theta_1 = 0.3$	50	0.310	0.012	0.049	0.577	0.310	0.017	0.049	0.577	0.310	0.021	0.002	0.622
	100	0.303	0.006	0.139	0.471	0.303	0.007	0.139	0.471	0.301	0.010	0.099	0.486
	150	0.303	0.004	0.163	0.438	0.303	0.005	0.163	0.438	0.302	0.007	0.139	0.459
	200	0.301	0.003	0.184	0.419	0.301	0.003	0.184	0.419	0.302	0.005	0.154	0.438
	250	0.301	0.003	0.193	0.404	0.301	0.003	0.193	0.404	0.301	0.004	0.173	0.427
	300	0.301	0.002	0.207	0.395	0.301	0.002	0.207	0.395	0.302	0.003	0.189	0.417
$\theta_2 = 0.6$	50	0.641	0.012	0.330	1.000	0.641	0.015	0.330	1.000	0.621	0.021	0.298	0.956
	100	0.611	0.006	0.419	0.808	0.611	0.008	0.419	0.808	0.603	0.010	0.387	0.823
	150	0.606	0.004	0.471	0.745	0.606	0.005	0.471	0.745	0.601	0.007	0.429	0.772
	200	0.603	0.003	0.481	0.725	0.603	0.003	0.481	0.725	0.600	0.005	0.459	0.744
	250	0.602	0.003	0.500	0.703	0.602	0.003	0.500	0.703	0.599	0.004	0.480	0.721
	300	0.601	0.002	0.514	0.690	0.601	0.002	0.514	0.690	0.598	0.003	0.491	0.708

For example, when $n = 50$ and $\theta = 0.7$, all the approximate posteriors give a mean estimate of about 0.72, and the standard deviation estimates are 0.010, 0.011, and 0.020 obtained from the Newbold, Zellner-Reynolds and Broemeling-Shaarawy approximate posteriors, respectively.

The results of the comparison criteria show that the Newbold and Zellner-Reynolds approximate

Table 9: Comparison criteria between estimates from different posteriors for MA(2)

	n	MAPE			$\hat{\sigma}_j^{ZR} / \hat{\sigma}_j^N$			$\hat{\sigma}_j^{BS} / \hat{\sigma}_j^N$		
		$\hat{\mu}_j^N$	$\hat{\mu}_j^{ZR}$	$\hat{\mu}_j^{BS}$	$\hat{\mu}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$Q_{2.5\%}$	$Q_{97.5\%}$
$\theta_1 = 0.3$	50	71.962	71.962	80.500	1.730	0.284	3.258	2.820	1.084	17.998
	100	45.565	45.565	53.860	1.094	0.799	1.662	1.733	1.180	2.824
	150	36.249	36.249	43.715	1.059	0.828	1.408	1.614	1.270	2.226
	200	32.125	32.125	38.935	1.043	0.840	1.328	1.593	1.294	2.106
	250	28.718	28.718	34.831	1.037	0.857	1.300	1.584	1.327	1.978
	300	25.837	25.837	31.518	1.030	0.875	1.255	1.575	1.349	1.890
$\theta_2 = 0.6$	50	44.276	44.276	44.943	1.223	0.489	4.470	2.920	1.097	20.293
	100	25.730	25.730	29.330	1.179	0.704	2.301	1.754	1.198	2.874
	150	19.218	19.218	23.505	1.098	0.733	1.770	1.626	1.273	2.276
	200	16.030	16.030	19.674	1.066	0.763	1.566	1.602	1.300	2.116
	250	14.045	14.045	17.280	1.049	0.783	1.483	1.591	1.332	1.998
	300	12.431	12.431	15.695	1.039	0.792	1.395	1.581	1.351	1.904

Table 10: Bayesian estimates of the coefficients of ARMA(1, 1)

	n	$\zeta_j^N(\beta \mathbf{y})$				$\zeta_j^{ZR}(\beta \mathbf{y})$				$\zeta_j^{BS}(\beta \mathbf{y})$			
		$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$\hat{\sigma}$	$Q_{2.5\%}$	$Q_{97.5\%}$
$\phi = 0.3$	50	0.192	0.113	-0.733	0.645	0.192	0.088	-0.733	0.645	0.249	0.196	-0.858	0.957
	100	0.274	0.035	-0.205	0.615	0.274	0.036	-0.205	0.615	0.311	0.058	-0.226	0.762
	150	0.291	0.023	-0.071	0.584	0.291	0.032	-0.071	0.584	0.316	0.039	-0.116	0.711
	200	0.294	0.018	-0.008	0.541	0.294	0.019	-0.008	0.541	0.317	0.029	-0.061	0.672
	250	0.297	0.014	0.033	0.521	0.297	0.014	0.033	0.521	0.314	0.023	0.005	0.622
	300	0.299	0.012	0.059	0.513	0.299	0.012	0.059	0.513	0.317	0.019	0.046	0.601
$\theta = -0.7$	50	-0.628	0.091	-1.000	0.540	-0.628	0.073	-1.000	0.540	-0.674	0.218	-1.209	0.676
	100	-0.698	0.022	-1.000	-0.195	-0.698	0.025	-1.000	-0.195	-0.729	0.068	-1.062	-0.207
	150	-0.705	0.014	-0.924	-0.370	-0.705	0.020	-0.924	-0.370	-0.725	0.046	-1.011	-0.338
	200	-0.703	0.010	-0.887	-0.458	-0.703	0.011	-0.887	-0.458	-0.723	0.034	-0.982	-0.417
	250	-0.705	0.008	-0.867	-0.472	-0.705	0.008	-0.867	-0.472	-0.720	0.027	-0.944	-0.463
	300	-0.704	0.007	-0.864	-0.509	-0.704	0.007	-0.864	-0.509	-0.720	0.023	-0.931	-0.480

Table 11: Comparison criteria between estimates from different posteriors for ARMA(1, 1)

	n	MAPE			$\hat{\sigma}_j^{ZR} / \hat{\sigma}_j^N$			$\hat{\sigma}_j^{BS} / \hat{\sigma}_j^N$		
		$\hat{\mu}_j^N$	$\hat{\mu}_j^{ZR}$	$\hat{\mu}_j^{BS}$	$\hat{\mu}$	$Q_{2.5\%}$	$Q_{97.5\%}$	$\hat{\mu}$	$Q_{2.5\%}$	$Q_{97.5\%}$
$\phi = 0.3$	50	84.462	84.462	100.995	0.712	0.390	3.296	2.492	0.889	8.979
	100	55.174	55.174	67.819	0.968	0.595	2.942	2.081	0.931	5.954
	150	42.149	42.149	52.445	2.265	0.665	2.592	1.901	0.961	4.208
	200	36.603	36.603	45.429	1.083	0.678	2.081	1.777	1.018	3.595
	250	31.576	31.576	39.450	1.037	0.700	1.774	1.748	1.050	3.272
	300	29.338	29.338	36.662	1.028	0.724	1.710	1.724	1.097	3.065
$\theta = -0.7$	50	34.121	34.121	41.079	0.809	0.355	5.575	10.055	1.048	72.894
	100	19.429	19.429	24.556	0.911	0.445	4.614	9.783	1.200	70.867
	150	14.694	14.694	19.314	0.978	0.518	3.485	7.235	1.369	21.411
	200	11.755	11.755	15.667	1.141	0.558	2.443	4.385	1.611	13.234
	250	9.975	9.975	13.448	1.056	0.596	2.086	4.189	1.670	10.889
	300	9.274	9.274	12.541	1.055	0.614	1.960	4.122	1.779	10.158

posteriors have almost the same MAPE values that are relatively less than those of the Broemeling-Shaarawy posterior. In addition, the ratios of standard deviations show that the standard deviation estimates obtained from the Broemeling-Shaarawy approximate posterior are relatively very large compared to those obtained from the Newbold and Zellner-Reynolds posteriors, which are very close.

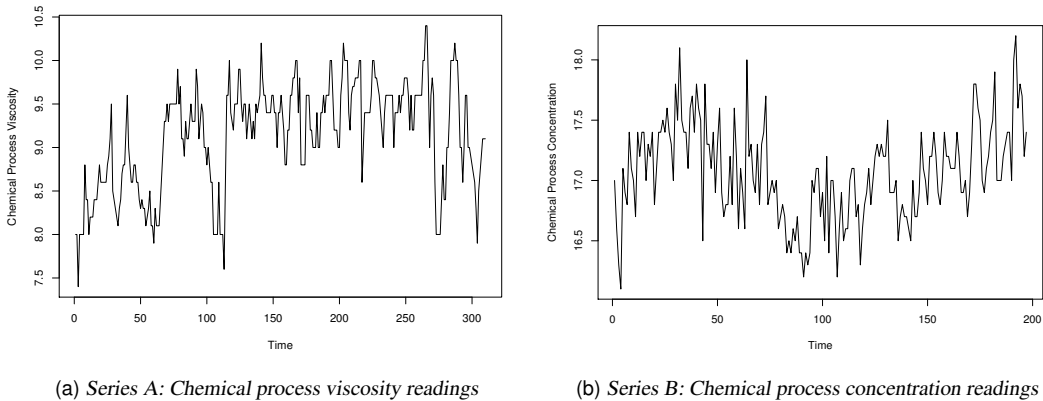


Figure 1: Time-plot of the real-world time series datasets.

We get the same conclusions from the results of MA(2) model presented in Tables 8 and 9 and ARMA(1, 1) model presented in Tables 10 and 11 with some additional general remarks. First, when the sample size is getting larger, the Bayesian estimates become more accurate, and hence the MAPE values and the ratios of standard deviation estimates are highly reduced. Second, the larger coefficients values used to generate the time series the more accurate estimates obtained from the approximate posteriors. In general, the simulation results confirm that the Newbold and Zellner-Reynolds approximate posteriors are strongly different from the Broemeling-Shaarawy approximate posterior, and the impact of that can be observed in the Bayesian estimates of the coefficients standard deviation. Thus, the Newbold and Zellner-Reynolds approximations reduce the posterior estimate of the coefficients standard deviation by at least 17% in the case of simple models to 80% in the case of complicated models compared to the Broemeling-Shaarawy approximation, which is strongly reflected in the 95% credible intervals of the coefficients.

5.2. Application to real-world time series

We introduce two real-world time series to demonstrate how the KL divergence can be used in reality to measure the distance between the derived approximate posteriors and to show the impact of this distance in the Bayesian estimates of the model coefficients. These two real-world time series datasets represent viscosity and concentration outputs from two different chemical processes (Box *et al.*, 2015). These datasets were collected on full-scale processes to show the effect on these outputs of uncontrolled disturbances such as variations in feedstock. These two time series datasets are presented in Figure 1.

It can observe from Figure 1 that these two time series are not stationary, so we use the first difference to stationarize them; and we then identify the best suitable order of the ARMA model for the differenced time series datasets by using the corrected Akaike's information criterion (AIC_c) with assuming the maximum order is five for both autoregressive and moving average parts. We compute the AIC_c for all the combinations of ARMA models and select the best model with smallest AIC_c value, and accordingly we identify the best model for both the differenced datasets is ARMA(1, 1). Using the identified models for these two time series datasets and employing the Jeffreys' priors, we first obtain the Newbold, Zellner-Reynolds and Broemeling-Shaarawy approximate posteriors of the model coefficients. Second, we compute the KL divergence and its calibration between these approximate posteriors and present their results in Table 12. Finally, we compute the Bayesian estimates for

Table 12: Results of KL divergence and its calibration for real-world time series datasets

Dataset	KL[N, ZR]	KLC[N, ZR]	KL[N, BS]	KLC[N, BS]	KL[ZR, BS]	KLC[ZR, BS]
Series A	0.0248	0.6100	2.1950	0.9969	1.9265	0.9947
Series B	0.0273	0.6153	0.1145	0.7262	0.1088	0.7211

Table 13: Bayesian estimates of the coefficients for real-world time series datasets

Dataset	$\zeta_j^N(\beta y)$		$\zeta_j^{ZR}(\beta y)$		$\zeta_j^{BS}(\beta y)$		$\hat{\sigma}_j^{ZR} / \hat{\sigma}_j^N$	$\hat{\sigma}_j^{BS} / \hat{\sigma}_j^N$	
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$			
Series A	ϕ	0.2155	0.0092	0.2155	0.0092	0.1142	0.0138	1.0086	1.5027
	θ	-0.8193	0.0032	-0.8193	0.0035	-0.7267	0.0190	1.1136	5.9836
Series B	ϕ	-0.4613	0.1723	-0.4613	0.1953	-0.4082	0.2375	1.1338	1.3789
	θ	0.3246	0.1956	0.3246	0.2228	0.2666	0.2431	1.1390	1.2430

the coefficients of these models and present results in Table 13. It is worth noting that these results are consistent with our conclusions in the simulation study in previous subsection.

6. Conclusions

In this paper we first reviewed the Newbold, Zellner-Reynolds, and Broemeling-Shaarawy analytic approximations proposed in the literature to approximate the posterior density of ARMA models to be analytically tractable. We then derived the KL divergence between these approximate posteriors and compute its calibration to measure the distance between these posteriors. We used a large number of Monte Carlo simulations from the approximate posteriors to evaluate the impact of the posteriors distance in the Bayesian estimates of mean and precision of the model coefficients. Simulation results confirmed that the Newbold and Zellner-Reynolds approximate posteriors are very close to each other, and they are strongly different from the Broemeling-Shaarawy approximate posterior. In particular, the results demonstrated that the Newbold and Zellner-Reynolds approximations show higher precision of the model coefficients compared to the Broemeling-Shaarawy approximation, which is reflected in the coefficients credible intervals. We applied our work to real-world time series datasets and their results are consistent with those of the simulation study.

References

Amin AA (2009). *Bayesian inference for seasonal ARMA models: A Gibbs sampling approach* (Masters thesis), Cairo University, Egypt.

Amin AA (2017a). Gibbs sampling for double seasonal ARMA models. In *Proceedings of the 29th Annual International Conference on Statistics and Modeling in Human and Social Sciences*, Cairo, Egypt.

Amin AA (2017b). Bayesian inference for double seasonal moving average models: A Gibbs sampling approach, *Pakistan Journal of Statistics and Operation Research*, **13**, 483–499.

Amin AA (2017c). Sensitivity to prior specification in Bayesian identification of autoregressive time series models, *Pakistan Journal of Statistics and Operation Research*, **14**, 699–713.

Amin AA (2018). Bayesian inference for double SARMA models, *Communications in Statistics-Theory and Methods*, **47**, 5333–5345.

Amin AA (2019a). Gibbs sampling for Bayesian prediction of SARMA processes, *Pakistan Journal of Statistics and Operation Research*, **15**, 397–418.

Amin AA (2019b). Kullback-Leibler divergence to evaluate posterior sensitivity to different priors for

- autoregressive time series models, *Communications in Statistics-Simulation and Computation*, **48**, 1277–1291.
- Amin AA (2020). Bayesian analysis of double seasonal autoregressive models, *Sankhya B*, **82**, 328–352.
- Amin AA (2022). Gibbs sampling for Bayesian estimation of triple seasonal autoregressive models, *Communications in Statistics-Theory and Methods*. Available from: <https://doi.org/10.1080/03610926.2022.2043379>
- Amin AA and Ismail MA (2015). Gibbs sampling for double seasonal autoregressive models, *Communications for Statistical Applications and Methods*, **22**, 557–573.
- Ali SS (1998). *Approximations for posterior densities of moving average models: A comparative study* (Masters thesis), Cairo University, Egypt.
- Broemeling LD and Shaarawy S (1984). Bayesian inferences and forecasts with moving average processes, *Communications in Statistics*, **13**, 1871–1888.
- Broemeling LD and Shaarawy S (1988). Time series: A Bayesian analysis in the time domain, *Bayesian Analysis of Time Series and Dynamic Models*, (Ed. J. Spell, pp. 1–21), Marcel Dekker, New York.
- Box GEP, Jenkins GM, Reinsel GC, and Ljung GM (2015). *Time Series Analysis: Forecasting and Control* (5th ed), John Wiley & Sons, New York.
- Ismail MA (1994). *Bayesian forecasting for nonlinear time series* (Ph.D. thesis), University of Wales, UK.
- Ismail MA and Amin AA (2014). Gibbs sampling For SARMA models, *Pakistan Journal of Statistics*, **30**, 153–168.
- Kullback S and Leibler RA (1951). On information and sufficiency, *The Annals of Mathematical Statistics*, **22**, 79–86.
- McCulloch RE (1989). Local model influence, *Journal of the American Statistical Association*, **84**, 473–478.
- Newbold P (1973). Bayesian estimation of box Jenkins transfer function noise models, *Journal of the Royal Statistical Society: Series B*, **35**, 323–336.
- Shaarawy SM and Ali SS (2012). Bayesian model order selection of vector moving average processes, *Communications in Statistics - Theory and Methods*, **41**, 684–698.
- Soliman E (1999). *On Bayesian time series analysis* (Ph.D. thesis), Cairo University, Egypt.
- Zellner A and Reynolds R (1978). Bayesian analysis of ARMA models. In *Proceedings of the Sixteenth Seminar on Bayesian Inference in Econometrics*.

Received February 4, 2022; Revised April 1, 2022; Accepted April 6, 2022