STATISTICA, anno LXXXIV, n. 1, 2024

SHRINKAGE ESTIMATION OF LINEAR REGRESSION MODELS WITH ARIMA ERRORS AND APPLICATION TO CANADIAN CRIME RATES DATA

Sharandeep Pandher
Department of Mathematics and Statistical Sciences, University of Alberta, Edmonton, AB, Canada
Shakhawat Hossain [1]
Department of Mathematics and Statistics, University of Winnipeg, Winnipeg, MB, Canada
Andrei Volodin
Department of Mathematics and Statistics, University of Regina, Regina, SK, Canada

SUMMARY

Shrinkage methods for estimating the parameters of a regression model with autoregressive integrated moving average (ARIMA) errors are presented when some regression parameters are restricted to a subspace. The estimates are obtained by maximizing the likelihood function with and without restrictions, yielding the unrestricted and restricted estimators, respectively. Shrinkage estimators optimally combine these two estimators. To demonstrate the optimality of these estimators, we use metrics such as asymptotic distributional bias (ADB) and asymptotic distributional risk (ADR), aiming to minimize both quantities. We show that the relative efficiency of the shrinkage estimator is superior to that of the unrestricted estimator when the shrinkage dimension exceeds two. Our large-sample theory and simulation study demonstrate that shrinkage estimators dominate the unrestricted estimator across the entire parameter space. An empirical example using Canadian crime rate data is also provided.

Keywords: Linear regression model; ARIMA; Monte Carlo simulation; Shrinkage estimators; Asymptotic biases and risks.

1. INTRODUCTION

In the domain of financial time series analysis, the effectiveness of time series regression models incorporating economic variables can be compromised by the frequent presence of non-homoscedastic residuals. Hossain and Ghahramani (2016) considered this model with GARCH errors for addressing this issue. However, in certain scenarios,

[1] Corresponding Author. E-mail: sh.hossain@uwinnipeg.ca

there might be significant serial dependence among the observations of the response variable, which violates the assumption of independence. In such cases, it is appropriate to consider a time series regression model with ARIMA errors. When we extend pure ARIMA models by including additional covariates that are not part of the time series being modeled, these covariates can improve the accuracy of the forecast by capturing additional information relevant to the response. This combination of regression and ARIMA modeling is a versatile tool for handling a wide range of time series patterns. While the covariates can account for trend and other non-temporal influences on the response variable, the ARIMA component captures temporal dependence in the errors.

Regression models with ARIMA errors (RegARIMA) are frequently employed in the analysis of real-world data. Chen *et al.* (2022) used an ARIMA model in conjunction with infectious disease reports from January 1, 2013, to December 31, 2020, to analyze epidemic characteristics and forecast incidence trends in Anhui province, China. Xu and Qin (2021) proposed a novel hybrid model for interval-valued time series by integrating ARIMA and regression tree models. Johansen *et al.* (2012) introduced C_p statistics for regression models with stationary and non-stationary ARIMA errors, providing a detailed discussion of the asymptotic properties of maximum likelihood (ML) estimators and accompanying simulation studies. William (2011) presented a notable implementation of linear regression models with ARIMA errors in Fortran, examining the computational and theoretical aspects of Gaussian ML estimation using three real-world datasets. Bianco *et al.* (2001) employed a robust estimation technique for outlier detection in regression models with ARIMA errors, comparing simulation results with classical methods based on ML estimates and Kalman filtering. Davis and Dunsmuir (1997) utilized least absolute deviation (LAD) estimation in a linear regression model with autoregressive moving average (ARMA) errors under general conditions, establishing the asymptotic properties of the LAD estimator through functional limit theorems. Otto *et al.* (1987) considered an iterative general least squares approach to ML estimation of regression models with ARIMA errors.Wincek and Reinsel (1986) considered the exact ML estimation of the RegARMA model with possibly consecutive and nonconsecutive time series data.

In this paper, we examine the James-Stein shrinkage estimation procedure for RegARIMA models in situations where numerous potential covariates are being considered. Including a large number of insignificant covariates can significantly hinder the model's forecasting performance. Researchers face a common challenge in model building: balancing the need for accurate predictions of forecasting with the desire to use only the most relevant variables. Selecting only statistically significant covariates can lead to underfitting, while including too many variables results in overfitting. To overcome this dilemma, researchers aim to maximize predictive power while minimizing the number of covariates. James-Stein shrinkage estimation provides a solution by leveraging information from insignificant covariates to achieve this balance. This method effectively shrinks the coefficients of less impactful variables, reducing their influence on the model while preserving the contributions of truly relevant ones.

Consider the study analyzing the impact of various macroeconomic indicators (covariates) — including inflation, interest rates, unemployment, exchange rates, debt levels, and political instability—on a country's GDP growth rate. Because GDP growth often exhibits autocorrelation, an ARIMA model will be employed to account for temporal dependencies. Not all macroeconomic indicators will significantly influence GDP growth. Therefore, the goal is to maximize the predictive ability of the GDP growth rate while minimizing the number of insignificant covariates in the time series regression model. We can identify insignificant covariates using classical model selection procedures or adaptive LASSO. James–Stein shrinkage estimation allows us to achieve this goal by utilizing information from the insignificant covariates. Specifically, let $\boldsymbol{\xi}$ be the $k \times 1$ regression and ARMA time series parameters, which are partitioned into two sub-vectors as $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)^\top$, where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are assumed to have dimensions $(k_1 + p + q + 1) \times 1$ and $k_2 \times 1$, respectively, such that $k = k_1 + k_2 + p + q + 1$. Here, $\boldsymbol{\xi}_1$ is the coefficient vector for significant covariates and ARMA time series parameters, and $\boldsymbol{\xi}_2$ is a coefficient vector for insignificant covariates. We are interested in the estimation of $\boldsymbol{\xi}_1$ using the auxiliary information about the parameter vector $\boldsymbol{\xi}_2$ when their values are near some specified value. Without loss of generality, we consider the hypothesis $H_0 : \boldsymbol{\xi}_2 = \mathbf{0}$, the $k_2 \times 1$ null vector. This parameter partition strategy was used by Hossain and Ghahramani (2016) for a linear regression model with GARCH errors and employed shrinkage techniques to estimate the regression parameters when other parameters were considered as nuisance.

Recent literature has investigated James Stein's shrinkage method to regression model suitable for stationary time series data, incorporating various error structures such as AR, ARMA, and GARCH. We extend these methods to effectively handle non-stationary time series by appropriately differencing the data within the framework of linear regression with ARIMA errors. In the recent literature, researchers have shown significant interest in the usefulness of James-Stein shrinkage method for parameter estimation in the linear regression model with AR, ARMA, and GARCH errors. Paolella (2019) worked on the application of shrinkage estimator for linear models with ARMA and GARCH errors. Hossain and Ghahramani (2016) considered shrinkage and positive shrinkage estimators in linear regression with GARCH error. Thomson *et al.* (2015) developed shrinkage estimation in a linear regression model with AR errors. In addition, the authors studied the asymptotic features of the estimators in the context of risks and biases. The utility of the proposed estimators was tested on Los Angeles pollution mortality data. Wu and Wang (2012) proposed a shrinkage procedure for a linear regression model with ARMA error. This procedure simultaneously estimates the parameters and selects the informative variables in the regression, autoregressive, and moving average components. Chan and Chen (2011) applied the adaptive LASSO method to a linear regression model with ARMA errors. This method achieves model selection consistency and produces asymptotically unbiased estimators for the nonzero coefficients. Wang *et al.* (2007) introduced a modified LASSO for regression models with AR errors and its superiority compared with a traditional LASSO in the context of the BIC. This was done through a simulation study and a practical example of an electricity demand data set.

The paper is structured as follows. In Section 2 we present the model specification, parameter estimation, score vector and Hessian matrix of the RegARIMA model. Section 3 reviews the asymptotic distributional biases and risks of the proposed estimators. In Section 4 we consider a Monte Carlo simulation study to evaluate the numerical performance of the proposed estimators with respect to the maximum likelihood estimator. An application of RegARIMA models is represented in Section 5. Section 6 closes the article with concluding remarks.

2. Models and estimation strategy

2.1. Linear regression model with ARIMA(p,d,q) errors

We assume a multiple linear regression model as

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + \eta_t, \quad t = 1,2,\ldots,N = n+d, \tag{1}$$

where y_t be the response, $\boldsymbol{x}_t = (x_{t1}, x_{t2}, \cdots, x_{tk})^\top$ be a $k \times 1$ predictor vector and $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)^\top$ is a $k \times 1$ vector of unknown regression coefficients.

We assume that error terms η_t in Eq. (1) is generated by an ARIMA model of order (p,d,q), where p is the order of autoregressive part, d is the degree of differences, and q is the order of moving average part. From η_t, we can modify error term ω_t of $(n+d)-d$ differences $\omega_1, \omega_2, \cdots, \omega_n$. where $\omega_t = \nabla^d \eta_t$. While the differences are appeared in η_t of Eq. (1), all corresponding series (both of the dependent and the explanatory variables) should occurs the difference (Pankratz, 1991). Differencing is very useful technique because it helps address specific types of mean non-stationarity. Thus, we applied d differences to the error term in Eq. (1) to make it follows an ARMA(p, q) process. So the multiple linear regression model in Eq. (1) can be re-written as

$$y_t^* = \boldsymbol{x}_t^{*\top} \boldsymbol{\beta} + w_t, \quad t = 1,2,\ldots,n, \tag{2}$$

where $y_t{}^* = \nabla^d y_t$, $\boldsymbol{x}_t^{*\top} = \nabla^d \boldsymbol{x}_t^\top$, $\omega_t = \nabla^d \eta_t$, and $\nabla = 1-B$, with B being the backward shift operator. George *et al.* (2008) mentioned that the general problem of fitting the parameters $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_q)^\top$ of the ARIMA error (η_t) is equivalent to fitting ω_t as stationary ARMA(p,q) error which can be written as

$$\omega_t - \phi_1 \omega_{t-1} - \phi_2 \omega_{t-2} - \ldots - \phi_p \omega_{t-p} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}, \tag{3}$$

or, in short, $\phi(B)\omega_t = \theta(B) a_t$, where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ is the AR polynomial, and $\theta(z) = 1 - \theta_1 z - \cdots - \theta_q z^q$ is the MA polynomial. We assume that $\{a_t\}$ is a sequence of i.i.d. random variables with mean 0, common variance σ^2, and finite fourth moment. We also assume that the polynomials $\phi(z)$ and $\theta(z)$ have no common roots and that all their roots lie outside the unit circle in the complex plane. Moreover, we assume that $\{\boldsymbol{x}_t\}$ is a strictly stationary and ergodic process with finite second moment. Hence, linear regression with ARIMA(p,d,q) errors in Eq. (1) is equivalent to linear regression with stationary ARMA(p,q) errors in Eq. (2).

2.2. *Estimation of parameters*

In the past, researchers studied the linear regression models with ARMA and ARIMA errors using maximum likelihood estimation. For instance, Furno (1996) proposed an information matrix approach to study the linear regression model with ARMA errors, maximizing the likelihood function with respect to regression parameters $\boldsymbol{\beta}$ and unknown variance parameters $\boldsymbol{\eta} = (\boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, \sigma^2)^\top$. Wincek and Reinsel (1986) introduced an exact maximum likelihood method for estimating the parameters of regression model with ARMA errors. We use the same method for estimating the parameters. In matrix form, the model in Eq. (2) can be written as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}, \tag{4}$$

where $\boldsymbol{Y} = (\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_n^*)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*)^\top$, and $\boldsymbol{W} = (w_1, \ldots w_n)^\top$ and the vector $\boldsymbol{W}$ has a zero mean and covariance matrix $\boldsymbol{\Gamma}$ with (t, t')th element $\text{cov}(w_t, w_{t'}) = \gamma(t - t')$, where the autocovariances are functions of unknown parameters $\boldsymbol{\eta}$. We want to estimate the regression parameter vector $\boldsymbol{\xi}$ by maximizing the Gaussian log-likelihood function in Eq. (5). The log-likelihood function of Eq. (4) can be written as

$$l(\boldsymbol{\xi}) = -\frac{1}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\boldsymbol{W}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{W}, \tag{5}$$

where $\boldsymbol{W} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$ is a covariance matrix of $\boldsymbol{W}$ and $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \boldsymbol{\eta}^\top)^\top$ be the $(k + p + q + 1) \times 1$ vector of unknown parameters. The log-likelihood function in Eq. (5) is a complicated function of the unknown parameters. To maximize this function, the iterative Newton-Raphson method is needed, requiring the evaluation of the score vector and Hessian matrix of the log-likelihood. Getting accurate parameter estimates depends critically on choosing appropriate initial values or conditions. These starting values are crucial for the optimization process, significantly influencing both the final parameter estimates and the overall model fit. Appropriate initial conditions are essential for the convergence of the optimization algorithm, as they initiate the recursive calculations required for maximizing the log-likelihood function. Different assumptions about these initial values lead to variations in the calculated log-likelihood and, consequently, different parameter estimates. A common approach is to treat the initial values as fixed, drawing them from the unconditional distribution implied by the model.

An innovation transformation of $\boldsymbol{\Gamma}$ facilitates the calculation of the exact log-likelihood and its derivatives through a convenient recursive procedure. Wincek and Reinsel (1986) used the innovation transformation of $\boldsymbol{\Gamma}$ to evaluate the score vector and Hessian matrix of log-likelihood $l(\boldsymbol{\xi})$. For a covariance matrix $\boldsymbol{\Gamma}$, there exists a unique lower triangular matrix $\boldsymbol{P}$ with ones on the diagonal such that $\boldsymbol{P}^\top \boldsymbol{\Gamma} \boldsymbol{P} = \boldsymbol{D}$, where $\boldsymbol{D}$ is a diagonal matrix of positive values. The specific prediction errors, denoted by $a_{t|t-1} = w_t - w_{t|t-1}$, are called the innovations of the w_t. Hence, the vector form of innovations $\boldsymbol{a} = (a_{t|0}, a_{t|1}, \ldots, a_{t|t-1})^\top$ and $\boldsymbol{W}$ are linearly related with $\boldsymbol{a} = \boldsymbol{P}^\top \boldsymbol{W}$. Thus, the second term of Equation (5) can be written as $\boldsymbol{W}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{W} = \boldsymbol{a}^\top \boldsymbol{V}^{-1} \boldsymbol{a}/\sigma^2$, where $\boldsymbol{\Gamma} = \sigma^2 \boldsymbol{V}$, $\boldsymbol{V} = \text{diag}(\boldsymbol{v})$, $\boldsymbol{v} = (v_1, v_2, v_t, \ldots, v_n)^\top$, $v_t = \sigma_t^2/\sigma^2$, with v_t are functions of $\boldsymbol{\eta}$. The log-likelihood in Eq. (5) can be expressed in innovations form as

$$l(\boldsymbol{\xi}) = -\frac{1}{2} n \log(\sigma^2) - \frac{1}{2} \log|\boldsymbol{V}| - (2\sigma^2)^{-1} \boldsymbol{a}^\top \boldsymbol{V}^{-1} \boldsymbol{a}, \tag{6}$$

where $\boldsymbol{a} = \boldsymbol{P}^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$.

The components of score vector of Eq. (5) are $\frac{\partial l(\boldsymbol{\xi})}{\partial \boldsymbol{\beta}} = \sigma^{-2} \boldsymbol{U}^\top \boldsymbol{V}^{-1} \boldsymbol{a}$ and $\frac{\partial l(\boldsymbol{\xi})}{\partial \boldsymbol{\eta}} = \boldsymbol{F}^\top \boldsymbol{N}^{-1} \boldsymbol{g}$, where $\boldsymbol{U} = \boldsymbol{P}^\top \boldsymbol{X}$, $\boldsymbol{g} = [\boldsymbol{a}^\top (\boldsymbol{a}^\top \boldsymbol{a} - \sigma^2 \boldsymbol{v})^\top]^\top$, $\boldsymbol{F}^\top = \begin{bmatrix} -\frac{\partial}{\partial \boldsymbol{\eta}} \boldsymbol{a}^\top & \frac{\partial}{\partial \boldsymbol{\eta}} (\sigma^2 \boldsymbol{v}^\top) \end{bmatrix}$, and $\boldsymbol{N} = \text{diag}(\sigma^2 \boldsymbol{V}, 2\sigma^4 \boldsymbol{V}^\top \boldsymbol{V})$.

The Hessian matrix of from the log-likelihood in Eq. (6) is

$$\frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\top} = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\eta}^\top} \\ \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & 0 \\ 0 & \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \end{bmatrix}, \tag{7}$$

where $\frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = -\sigma^{-2} \boldsymbol{U}^\top \boldsymbol{V}^{-1} \boldsymbol{U}$, $\frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} = -\boldsymbol{F}^\top \boldsymbol{N}^{-1} \boldsymbol{F}$, $\frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\eta}^\top} = 0$, and $\frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\beta}^\top} = 0$.

2.3. *Modified Newton Raphson procedure for ML estimates*

Based on Wincek and Reinsel (1986), the approximate Hessian matrix in the Newton-Raphson procedure yields the modified Newton-Raphson equations

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + [\tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{V}}^{-1} \tilde{\boldsymbol{U}}]^{-1} \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{V}}^{-1} \tilde{\boldsymbol{a}} = [\tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{V}}^{-1} \tilde{\boldsymbol{U}}]^{-1} \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{V}}^{-1} \tilde{\boldsymbol{P}}^\top \boldsymbol{Y}, \tag{8}$$

$$\hat{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}} + [\tilde{\boldsymbol{F}}^\top \tilde{\boldsymbol{N}}^{-1} \tilde{\boldsymbol{F}}]^{-1} \tilde{\boldsymbol{F}}^\top \tilde{\boldsymbol{N}}^{-1} \tilde{g}, \tag{9}$$

where $\tilde{\boldsymbol{F}}$, $\tilde{\boldsymbol{V}}$, $\tilde{\boldsymbol{a}}$, $\tilde{\boldsymbol{U}}$, $\tilde{g}$, $\tilde{\boldsymbol{P}}$, $\tilde{\boldsymbol{N}}$ are evaluated at the $\boldsymbol{\xi} = (\tilde{\boldsymbol{\beta}}^\top, \tilde{\boldsymbol{\eta}}^\top)^\top$ of the parameter vector at the previous iteration of modified Newton-Raphson procedure. The estimator $\hat{\boldsymbol{\beta}}$ in Eq. (8) is the weighted least squares estimate of $\boldsymbol{\beta}$ of model in Eq. (6) and also called the unrestricted maximum likelihood estimator (URE). This estimator $\hat{\boldsymbol{\beta}}$ is expressed in terms of the transformed variables $\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{P}}^\top \boldsymbol{X}$ and $\tilde{\boldsymbol{P}}^\top \boldsymbol{Y}$ with $\tilde{\boldsymbol{V}}^{-1}$ as the diagonal matrix of weights. A suitable estimation procedure is to obtain $\hat{\boldsymbol{\beta}}$ in Eq. (8) and then use $\tilde{\boldsymbol{a}} = \tilde{\boldsymbol{P}}^\top (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ and $\partial \tilde{\boldsymbol{a}} / \partial \boldsymbol{\eta} = (\partial \boldsymbol{P}^\top / \partial \boldsymbol{\eta})|_{\tilde{\boldsymbol{\eta}}} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ in Eq. (9) when obtaining $\hat{\boldsymbol{\eta}}$. Similar to Eq. (8), Equation (9) for the adjustment $\hat{\boldsymbol{\eta}} - \tilde{\boldsymbol{\eta}}$ in the time series parameters also has the form of a weighted least squares estimator. Although we obtained the estimate $\hat{\boldsymbol{\eta}}$, we consider it a nuisance parameter, and primarily focus on $\hat{\boldsymbol{\beta}}$.

The above weighted least squares procedure described by Equations (8) and (9) needs to use for calculating $\boldsymbol{a} = \boldsymbol{P}^\top \boldsymbol{W}, \boldsymbol{U} = \boldsymbol{P}^\top \boldsymbol{X}$, and $\boldsymbol{P}^\top \boldsymbol{Y}$, as well as the derivatives of $\partial \boldsymbol{a} / \partial \boldsymbol{\eta} = (\partial \boldsymbol{P}^\top / \partial \boldsymbol{\eta}) \boldsymbol{W}$ and $\partial \boldsymbol{v} / \partial \boldsymbol{\eta}$. To obtain these quantities with recursive procedure, see details in Wincek and Reinsel (1986).

2.4. Restricted estimator

Sometimes, we are interested in estimating the regression parameters in Eq. (2) when some of the regression parameters may be linearly related. Instead of deleting the insignificant covariate parameters, Hossain and Lac (2018) incorporated the information from insignificant covariates as auxiliary information, in the form of linear restrictions, to obtain an improved estimator.

We can define the hypothesis as

$$H_0 : \boldsymbol{\xi}_2 = \mathbf{0} \quad \text{vs.} \quad H_a : \boldsymbol{\xi_2} \neq \mathbf{0}. \tag{10}$$

Researchers typically test the above hypothesis in Eq. (10), as they are believed to be a reasonable reduction of the unrestricted model. Using these restrictions, we can construct a modified log-likelihood that is, maximize $l(\boldsymbol{\xi})$ subject to $\boldsymbol{R\beta} = \boldsymbol{h}$ is equivalent to finding

$$\hat{\boldsymbol{\xi}}_r = (\hat{\boldsymbol{\beta}}_r^\top, \hat{\boldsymbol{\eta}}_r^\top)^\top = \underset{\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2}{\text{argmax}} \{l(\boldsymbol{\xi}) : \boldsymbol{\xi}_2 = \mathbf{0}\}, \tag{11}$$

where $\hat{\boldsymbol{\xi}}_r$ is the restricted maximum likelihood estimator(RE). The objective function in Eq. (11) can be maximized by using the Newton Raphson method discussed in previous Section.

For testing the particular hypothesis $H_0 : \boldsymbol{\xi}_2 = \mathbf{0}$, we need the following partition of the expected information matrix

$$\boldsymbol{I}(\boldsymbol{\xi}) = E \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_1 \partial \boldsymbol{\xi}_1^\top} & \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_1 \partial \boldsymbol{\xi}_2} \\ \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_2 \partial \boldsymbol{\xi}_1} & \frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_2 \partial \boldsymbol{\xi}_2^\top} \end{bmatrix}. \tag{12}$$

We may consider testing the restriction by testing the null hypothesis $H_0 : \boldsymbol{\xi}_2 = \mathbf{0}$ using the likelihood ratio statistic

$$\hat{\Lambda} = 2l(\hat{\boldsymbol{\xi}}) - 2l(\hat{\boldsymbol{\xi}}_r), \tag{13}$$

where $l(\cdot)$ is the logarithm of the maximum likelihood function, and $\hat{\boldsymbol{\xi}}_r$ and $\hat{\boldsymbol{\xi}}$ are the corresponding maximum likelihood estimates under the null and alternative hypotheses, respectively. Under H_0 and for large n, the statistic $\hat{\Lambda}$ approximately follows the $\chi^2_{k_2}$ distribution with k_2 degrees of freedom.

2.5. Shrinkage and positive shrinkage estimators

The shrinkage estimator (SE) of $\boldsymbol{\xi}$ is defined as

$$\hat{\boldsymbol{\xi}}_S = \hat{\boldsymbol{\xi}}_r + (1 - (k_2 - 2)\hat{\Lambda}^{-1})(\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}}_r), \quad k_2 \geq 3.$$

This comes down to a convex combination function with the form $\hat{\boldsymbol{\xi}}_S = \lambda\hat{\boldsymbol{\xi}} + (1-\lambda)\hat{\boldsymbol{\xi}}_r$, where $\lambda \in [0,1]$. When $\lambda = 1$, no shrinkage occurs, and the estimates are the same as the URE. If $\lambda = 0$, the RE is chosen. The drawback of this estimator is that the factor $(1-(k_2-2)\hat{\Lambda}^{-1})$ can be negative. This happens for small values of $\hat{\Lambda}$. This phenomenon is known as over-shrinkage. This can be relieved by taking its positive part which makes it not only a shrinkage estimator but also a thresholding estimator. The positive part shrinkage estimator (PSE) is defined as

$$\hat{\boldsymbol{\xi}}_{S+} = \hat{\boldsymbol{\xi}}_r + (1-(k_2-2)\hat{\Lambda}^{-1})^+(\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}}_r), \quad k_2 \geq 3,$$

where $z^+ = \max(0, z)$.

3. Asymptotic results

3.1. *Asymptotic distributional results: bias*

This Section deals with the asymptotic features of the proposed estimators of $\boldsymbol{\xi}$ for the RegARIMA model. First, the asymptotic distributional bias (ADB) of the estimators of $\boldsymbol{\xi}$ will be discussed in detail, and afterward, the asymptotic distributional risk (ADR).

Under nonlocal (fixed) alternatives $H_a : \boldsymbol{\xi}_2 \neq \mathbf{0}$, $\hat{\boldsymbol{\xi}}_S$ and $\hat{\boldsymbol{\xi}}_{S+}$ are asymptotically converges to $\hat{\boldsymbol{\xi}}$, while $\hat{\boldsymbol{\xi}}_r$ holds unbounded risk. To capture the purposeful comparisons of the proposed estimators in terms of their biases and risks, we consider the sequence of local alternatives,

$$K_{(n)} : \boldsymbol{\xi}_2 = \frac{\boldsymbol{h}}{\sqrt{n}}. \tag{14}$$

Note that under H_0, $\boldsymbol{h} = \mathbf{0}$ implies that $\boldsymbol{\xi}_2 = \mathbf{0}$, which is a special case of Eq. (14). We require the following Assumptions to derive the asymptotic distributions of the estimators, as well as their ADBs and ADRs.

Assumption 1. *All the zeros of $\phi(B) = 0$ and $\theta(B) = 0$ are out-side the unit circle, with no single root common to the polynomials $\phi(B)$ and $\theta(B)$.*

Assumption 2. *The regressors $\boldsymbol{x}_t$ are weakly exogenous i.e $E(\omega_t|\boldsymbol{x}_t) = 0$, stating that ω_t and $\boldsymbol{x}_t$ are uncorrelated.*

Assumption 3. *The regressors and the error term must have finite fourth moments, i.e $E(||x_t||^4) < \infty$ and $E[\eta_t^4] < \infty$.*

Assumption 4. *The log-likelihood function $l(\xi)$ of the model must be sufficiently smooth, which typically involves assuming that the third derivatives of the log-likelihood function are bounded in probability.*

ASSUMPTION 5. *The information Matrix* $E(-\frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\top})$ *OR Hessian Matrix* $\frac{\partial^2 l(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\top}$ *is finite and positive definite.*

Assumption 1 ensures that the true model is stationary and ergodic, and that p and q are the true model orders. Assumption 2 is a standard set of assumptions on the error terms in regression models. The moment condition is mild since $E(||x_t||^4) < \infty$ is required for the existence of $\boldsymbol{I}(\boldsymbol{\xi})$ (Assumption 3). This condition, while stricter than some other assumptions, is used to guarantee the robustness and validity of asymptotic properties.

Under local alternative $K_{(n)}$, the following Theorem facilitates the derivation and numerical computation of the ADBs and the ADRs of the estimators outlined below.

THEOREM 6. *Under the local alternatives* $K_{(n)}$ *in Eq. (14) and the usual Assumptions above*

1. $\sqrt{n}\hat{\boldsymbol{\xi}}_2 \xrightarrow{\mathscr{L}} N(\mathbf{0}, \boldsymbol{I}_{22.1})$ *as* $n \to \infty$ *where* $\boldsymbol{I}_{22.1} = \boldsymbol{I}_{22} - \boldsymbol{I}_{21}\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}$ *is a positive definite matrix and the information matrix for* $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)^\top$

$$\boldsymbol{I}_{\boldsymbol{\xi}_1 \boldsymbol{\xi}_2} = \begin{bmatrix} \boldsymbol{I}_{11} & \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} \end{bmatrix}$$

2. *As* $n \to \infty$, *the distribution of* $\hat{\Lambda}$, *converge to a non-central chi-squared distribution* $\Psi_{k_2}(x; \Delta)$ *with* k_2 *degrees of freedom and the non-centrality parameter,* $\Delta = \boldsymbol{h}^\top \boldsymbol{I}_{22.1}\boldsymbol{h}$, *and* $\Psi_{k_2}(x;\Delta) = P(\chi^2_{k_2}(\Delta) \le x)$, $x \epsilon \Re^+$.

PROOF. The outline of a similar proof for linear models can be found in Sen and Ehsanes Saleh (1987) and Thomson *et al.* (2016). The information matrices for linear models and time series regression model with ARIMA error are different. □

Usually, the shrinkage estimators are biased, however, bias is accompanied by a reduction in variance. We define ADB of an estimator $\boldsymbol{\xi}^*$ as

$$\text{ADB}(\boldsymbol{\xi}^*) = E\left[\lim_{n\to\infty} \sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi})\right].$$

where $\boldsymbol{\xi}^*$ be a generic notation for any of $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\xi}}_r$, $\hat{\boldsymbol{\xi}}_S$, and $\hat{\boldsymbol{\xi}}_{S+}$.

THEOREM 7. *Using the above definition of ADB and Theorem 6, under the local alternatives* $K_{(n)}$ *in Eq. (14), as* $n \to \infty$,

$$ADB(\hat{\boldsymbol{\xi}}_{S+}) \quad = \quad ADB(\hat{\boldsymbol{\xi}}_S) - \boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{h}\left[\Psi_{\nu+4}(\nu, \Delta) - \nu E\left(Z_1^{-1} I(Z_1 < \nu)\right)\right],$$

where $\nu = k_2 - 2$, $Z_1 = \chi^{-2}_{k_2+2}(\Delta)$, $ADB(\hat{\boldsymbol{\xi}}_S) = -\nu \boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{h}E(Z_1)$, $ADB(\hat{\boldsymbol{\xi}}_r) = -\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{h} = -\gamma$ *(say), and* $\Psi_g(x, \Delta)$ *is the distribution function of a non-central chi-square with g degrees of freedom and non-centrality parameter* Δ*, and*

$$E\left(\chi_g^{-2j}(\Delta)\right) = \int_0^\infty x^{-2j}\, d\Psi_g(x, \Delta).$$

PROOF. The outline of the proof of the Theorem is given in the Appendix. □

The constant term $\boldsymbol{h}$ is common to the ADBs of $\hat{\boldsymbol{\xi}}_r$, $\hat{\boldsymbol{\xi}}_S$, and $\hat{\boldsymbol{\xi}}_{S+}$ and the ADBs differ only by a constant factor Δ. Therefore, it is sufficient to compare only Δ. It is clear that the ADB of the $\hat{\boldsymbol{\xi}}_r$ is an unbounded function of Δ. On the other hand, the ADBs of both $\hat{\boldsymbol{\xi}}_S$, and $\hat{\boldsymbol{\xi}}_{S+}$ are bounded in Δ. Since $E\left(\chi^{-2}_{k_2+2}(\Delta)\right)$ is a decreasing function of Δ, the ADB of $\hat{\boldsymbol{\xi}}_S$ starts from the origin, increases to a maximum, and then decreases towards 0 as $\Delta > 0$. The characteristics of $\hat{\boldsymbol{\xi}}_{S+}$ are similar to those of $\hat{\boldsymbol{\xi}}_S$.

3.2. *Asymptotic distributional results: risk*

We use the following quadratic loss function to obtain the ADRs of the proposed estimators:

$$\mathscr{L}(\boldsymbol{\xi}^*; \boldsymbol{A}) = \left[\sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi})\right]^\top \boldsymbol{A}\left[\sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi})\right],$$

where $\boldsymbol{A}$ is a positive semidefinite weight matrix. In the literature, the researchers considered $\boldsymbol{A} = \boldsymbol{I}$ as an identity matrix in the simulation study. For instance, Gupta *et al.* (1989) suggested that weight matrix with different arbitrary diagonal elements can not ensured the outperform of shrinkage estimators to the full model estimators.

Therefore, the expected loss function is defined as

$$E[\mathscr{L}(\boldsymbol{\xi}^*; \boldsymbol{A})] \equiv \mathrm{ADR}(\boldsymbol{\xi}^*, \boldsymbol{\xi}; \boldsymbol{A}) \equiv \mathrm{ADR}(\boldsymbol{\xi}^*, \boldsymbol{\xi}),$$

which is called the risk function. Under the sequence of local alternatives, we define the asymptotic distribution function of an estimator $\boldsymbol{\xi}^*$ as

$$G(\boldsymbol{y}) = \lim_{n\to\infty} P\left[\sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi}) \le \boldsymbol{y} | K_{(n)}\right],$$

where $G(\boldsymbol{y})$ is nondegenerate distribution function for the estimators. We define the asymptotic distributional risk (ADR) by

$$\begin{aligned} \mathrm{ADR}(\boldsymbol{\xi}^*; \boldsymbol{A}) &= \int \cdots \int \boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{y}\, dG(\boldsymbol{y}) \\ &= \mathrm{trace}(\boldsymbol{A}\boldsymbol{\Gamma}), \end{aligned} \tag{15}$$

where $\boldsymbol{\Gamma} = \int \ldots \int \boldsymbol{y}\boldsymbol{y}^\top dG(\boldsymbol{y})$ is the dispersion matrix for the distribution $G(\boldsymbol{y})$ and $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of $\boldsymbol{\xi}^*$.

An estimator $\boldsymbol{\xi}^*$ is then said to dominate an estimator $\boldsymbol{\xi}^0$ asymptotically if $\mathrm{ADR}(\boldsymbol{\xi}^*;\boldsymbol{\xi}) \leq \mathrm{ADR}(\boldsymbol{\xi}^0;\boldsymbol{\xi})$. If, in addition, $\mathrm{ADR}(\boldsymbol{\xi}^*;\boldsymbol{\xi}) < \mathrm{ADR}(\boldsymbol{\xi}^0;\boldsymbol{\xi})$ for at least some $(\boldsymbol{\xi}, \boldsymbol{A})$, then $\boldsymbol{\xi}^*$ strictly dominate $\boldsymbol{\xi}^0$.

THEOREM 8. *Under the local alternatives $K_{(n)}$ in Eq. (14) and the Theorem 6, as $n \to \infty$, we obtain the ADR functions of the proposed estimators:*

$$\begin{aligned}
ADR(\hat{\boldsymbol{\xi}}_{S+};\boldsymbol{A}) &= ADR(\hat{\boldsymbol{\xi}}_{S};\boldsymbol{A}) - E\Big(\big(1-\nu Z_1^{-1}\big)^2 I(Z_1<\nu)\Big)\, trace(\boldsymbol{A}\boldsymbol{C}) \\
&+ \big(2\Gamma_{\nu+4}(\nu,\Delta) - 2\nu E\big(Z_1^{-1} I(Z_1<\nu)\big)\big) \\
&- E\Big(\big(1-\nu Z_2^{-1}\big)^2 I(Z_2<\nu)\Big)\, trace(\gamma^\top \boldsymbol{A}\gamma), \ \ where \\
ADR(\hat{\boldsymbol{\xi}}_{S};\boldsymbol{A}) &= \boldsymbol{I}_{11.2}^{-1} + \big(\nu^2 E\big(Z_1^2\big) - 2\nu E(Z_1)\big)\, trace(\boldsymbol{A}\boldsymbol{C}) \\
&+ \big(\nu^2 E(Z_2^2) + 2\nu E(Z_1) - 2\nu E(Z_2)\big)\, trace(\gamma^\top \boldsymbol{A}\gamma)
\end{aligned}$$

$Z_2 = \chi_{k_2+4}^{-2}(\Delta)$, $ADR(\hat{\boldsymbol{\xi}}_r;\boldsymbol{A}) = ADR(\hat{\boldsymbol{\xi}};\boldsymbol{A}) - trace(\boldsymbol{A}\boldsymbol{\Phi}) + \gamma^\top \boldsymbol{A}\gamma$, $ADR(\hat{\boldsymbol{\xi}};\boldsymbol{A}) = trace[\boldsymbol{A}\boldsymbol{I}_{11.2}^{-1}]$, $\gamma = \boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{h}$, $\boldsymbol{\Phi} = \boldsymbol{I}_{11.2}^{-1} - \boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{I}_{22.1}^{-1}\boldsymbol{I}_{21}\boldsymbol{I}_{11}^{-1}$, $\boldsymbol{I}_{11.2} = \boldsymbol{I}_{11} - \boldsymbol{I}_{12}\boldsymbol{I}_{22}^{-1}\boldsymbol{I}_{21}$, and $\boldsymbol{C} = \boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{I}_{22.1}^{-1}\boldsymbol{I}_{21}\boldsymbol{I}_{11}^{-1}$.

PROOF. The outline of the proof of the Theorem is given in the Appendix. □

4. SIMULATION RESULTS

This Section presents a simulation study comparing the performance of the estimators $\hat{\boldsymbol{\xi}}_r$, $\hat{\boldsymbol{\xi}}_S$, and $\hat{\boldsymbol{\xi}}_{S+}$ relative to $\hat{\boldsymbol{\xi}}$. Data are generated under three scenarios, described below:

In the simulation model, we consider the following model with ARIMA$(1,1,1)$ error for generating responses

$$y_t{}^* = \boldsymbol{x}_t^{*\top}\boldsymbol{\beta} + w_t, \tag{16}$$

where $\omega_t = \phi_1\omega_{t-1} + a_t - \theta_1 a_{t-1}$ with $\omega_t = \nabla\eta_t$, Also $y_t{}^* = \nabla y_t$, $\boldsymbol{x}_t^{*\top} = \nabla\boldsymbol{x}_t^\top$ with $t = 1,2,\ldots,n$. We take $\phi_1 = -0.49$, $d = 1$, $\theta_1 = -0.79$, $\sigma = 1$ and three covariates with regression coefficients -0.95, 1.05, and 1.9. Therefore, $\boldsymbol{\beta}_{\text{sim}}$ is a vector of regression coefficients, and all the model parameters together are $\boldsymbol{\xi}_{\text{sim}} = (-0.49, -0.79, \boldsymbol{\beta}_{\text{sim}}^\top)^\top$ $=(-0.49,-0.79,-0.95,1.05,1.9)^\top$, that is $k_1 = 5$. Finally, the response is generated from simulation model.

In the unrestricted model specified in Eq. (2), let $\hat{\boldsymbol{\xi}} = (\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)^\top$, where $\boldsymbol{\xi}_1 = \boldsymbol{\xi}_{\text{sim}}$ and $\boldsymbol{\xi}_2$ is the vector of regression coefficients for k_2 insignificant covariates. In this setting, the response is generated from the simulation model in which we consider the true $\boldsymbol{\xi}_2$ is a zero vector. Therefore, the true values of the parameters are $\boldsymbol{\xi}_1 = (-0.49,-0.79,-0.95,1.05,1.9)^\top$ and $\boldsymbol{\xi}_2 = \boldsymbol{\beta}_2 = (0,0,0,0,...,0)^\top$. We consider six values: $k_2 = 4, 8, 12, 15, 18$ and 21.

The restricted model is established by considering a constraint $\xi_2 = 0$ in the unrestricted model. The restricted model is not significantly different from the unrestricted model. In this situation, $\Delta = ||\hat{\xi} - \hat{\xi}_r||^2 = 0$, where $||\cdot||$ is the Euclidian norm. To explore the behavior of shrinkage estimators when the restricted model is substantially different from the unrestricted model, we consider $\xi_2 = (\sqrt{c}, 0, .., 0)^\top$ so that $\Delta = ||\hat{\xi} - \hat{\xi}_r||^2 = c$, where c is a positive constant. Note that Δ is the difference between the restricted and the unrestricted model in the spirit of the local alternative in Eq. (14), and the performance of the shrinkage estimators is assessed under both $H_0 : \Delta = 0$ and $H_1 : \Delta = c$, for $0 < c < 1.5$

All the covariates are generated from standard normal distributions. We use $n = 500$, 600, 700 and 800 to investigate the impact of n, k_2 and Δ on shrinkage estimators. Here, we consider $\Delta = (0, 0.03, 0.07, 0.10, 0.15, 0.3, 0.55, 1, 1.5)$.

The estimates $\hat{\xi}$, $\hat{\xi}_r$, $\hat{\xi}_S$, and $\hat{\xi}_{S+}$ are obtained from each of the 1000 simulated datasets with different combinations of n, k_1 and k_2. The mean squared error (MSE) is used to evaluate the performance of the estimators. We have calculated MSE for any estimator, $\hat{\xi}^*$ based on the decomposition $\text{MSE}(\hat{\xi}^*) = \text{tr}[\text{var}(\hat{\xi}^*)] + ||\text{bias}(\hat{\xi}^*)||^2$; we took the trace of the covariance matrix of $\hat{\xi}^*$ and the average of $||\text{bias}(\hat{\xi}^*)||^2$ across the 1000 simulated datasets to compute $\text{MSE}(\hat{\xi}^*)$. The RMSE of ξ^* to $\hat{\xi}$ is defined as $\text{RMSE}(\hat{\xi} : \xi^*) = \text{MSE}(\hat{\xi})/\text{MSE}(\xi^*)$. Thus, a RMSE value exceeding one means the estimators have lower risk than the URE, and a RMSE less than one has higher risk.

TABLE 1
RMSEs of $\hat{\xi}_r$, $\hat{\xi}_S$, and $\hat{\xi}_{S+}$ with respect to $\hat{\xi}$ when the restricted parameter space is correct ($\Delta = 0$).

	n=500	n=600	n=700	n=800
	$k_1 = 5, k_2 = 4$	$k_1 = 5, k_2 = 4$	$k_1 = 5, k_2 = 4$	$k_1 = 5, k_2 = 4$
RE	1.88	1.85	1.81	1.83
SE	1.31	1.31	1.29	1.31
PSE	1.41	1.42	1.42	1.41
	$k_1 = 5, k_2 = 21$	$k_1 = 5, k_2 = 21$	$k_1 = 5, k_2 = 21$	$k_1 = 5, k_2 = 21$
RE	5.63	5.95	5.76	5.69
SE	3.99	4.06	3.86	3.88
PSE	4.40	4.54	4.39	4.44

The findings demonstrated in Table 1 for $\Delta = 0$ and in the Figures 1 and 2 for $\Delta \geq 0$. The RMSEs of all estimators relative to URE are initially highest at $\Delta = 0$, although subject to random fluctuation. Table 1 and Figures 1-2 show that the RE outperforms other estimators near the null hypothesis due to its unbiasedness; the RMSEs of SE and PSE relative to the RE asymptotically converge to 1. On the other hand, as Δ

increases, the risk of the RE increases and becomes unbounded, that is, the RMSE of the RE decreases and becomes zero, as reported in Figures 1-2. However, as Δ increases, the RE's risk increases without bound, resulting in a decreasing RMSE that approaches zero, as illustrated in Figures 1-2.

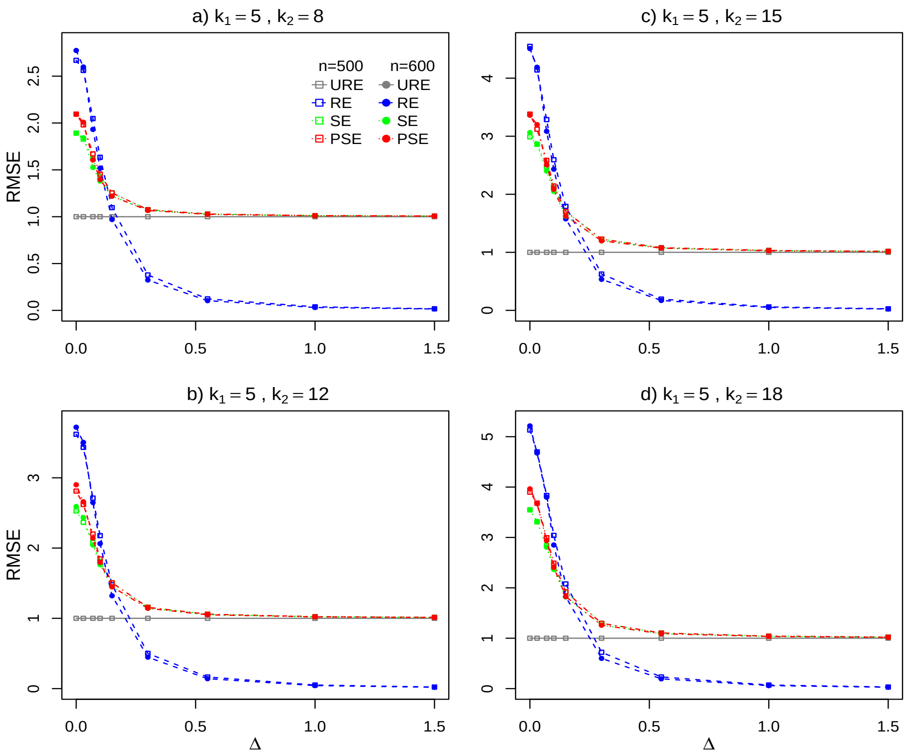

Figure 1 – RMSEs of RE ($\hat{\xi}_r$), SE($\hat{\xi}_S$), and PSE($\hat{\xi}_{S+}$) with respect to URE($\hat{\xi}$) when the subspace misspecifies $\Delta \geq 0$, $n = 500, 600$, and $k_1 = 5, k_2 = 8, 12, 15, 18$.

Shrinkage estimators offer substantial risk reduction compared to the URE, regardless of the auxiliary information used in the chosen restricted model. This aligns with Theorem 8. Since Δ represents the deviation from the null hypothesis, shrinkage estimators remain advantageous even when $\Delta > 0$. As shown in the Tables and Figures, they are highly efficient (low risk) relative to the URE when $\Delta = 0$. Furthermore, performance degrades gradually if the restricted model is misspecified. Therefore, shrinkage estimators are particularly useful in real-world applications where, as is often the case, a perfectly specified restricted model is unattainable. These results are supported by the findings in many studies in the literature.

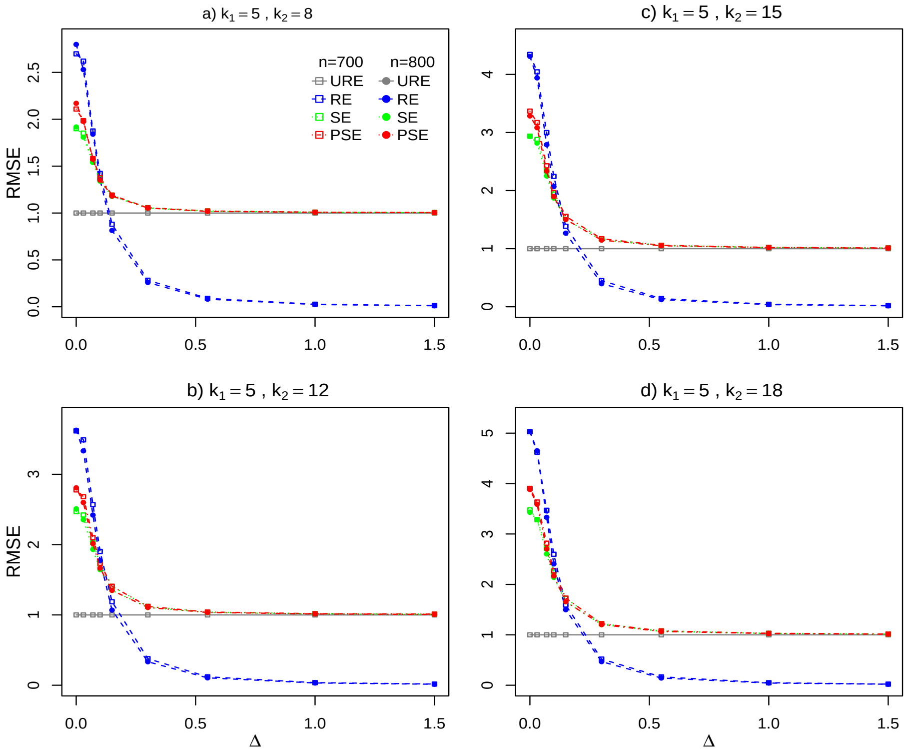

Figure 2 – RMSEs of RE ($\hat{\boldsymbol{\xi}}_r$), SE ($\hat{\boldsymbol{\xi}}_S$), and PSE ($\hat{\boldsymbol{\xi}}_{S+}$) with respect to URE($\hat{\boldsymbol{\xi}}$) when the subspace misspecifies $\Delta \geq 0$, $n = 700, 800$, and $k_1 = 5, k_2 = 8, 12, 15, 18$.

The PSE outperforms the SE as Δ approaches zero. Above $\Delta = 0.3$, the RMSEs of both SE and PSE are equal and approach one as Δ increases. The estimation accuracy of both SE and PSE is higher for $n = 500$ and $n = 600$ than for $n = 700$ and $n = 800$, because their RMSEs increase with decreasing sample size. For example, Table 1 shows that when $k_1 = 5$, $k_2 = 21$, and $n = 600$, the RMSEs of SE and PSE are 4.06 and 4.54, respectively; while for $n = 700$, $k_1 = 5$, and $k_2 = 21$, the RMSEs are 3.86 and 4.39. Increasing the sample size accelerates the convergence of the RE's RMSE to zero. However, the RMSEs of SE and PSE remain constant with increasing sample size when $\Delta > 0$.

Figure 3 shows MSE curves illustrating the impact of sample size (n) and the number of insignificant covariates (k_2) on the accuracy and uncertainty of shrinkage estimators. As expected, for a fixed sample size, accuracy decreases with increasing k_2. Conversely, for a fixed k_2, accuracy improves with increasing sample size (compare Figures 3, a-d).

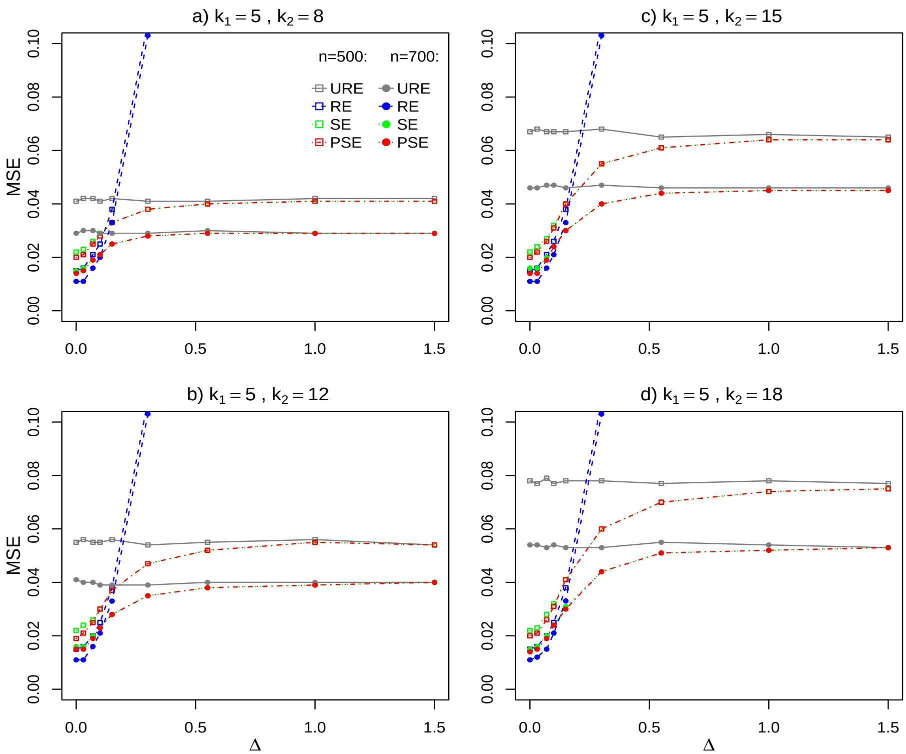

Figure 3 – MSEs of URE ($\hat{\boldsymbol{\xi}}$), RE ($\hat{\boldsymbol{\xi}}_r$), SE ($\hat{\boldsymbol{\xi}}_S$), and PSE ($\hat{\boldsymbol{\xi}}_{S+}$) when the subspace misspecifies $\Delta \geq 0$, $n = 500, 700$, and $k_1 = 5, k_2 = 8, 12, 15, 18$.

5. A REAL DATA EXAMPLE

We apply the shrinkage estimation approach to the Canadian crime rates data set (Fox and Weisberg, 2019). This data set contains 7 macroeconomic variables: female indictable-offense conviction (response) and six predictor variables: total fertility rate (x_1), women's labor force participation rate (x_2), women's post secondary degree rate (x_3), female theft conviction rate (x_4), male indictable offense conviction rate (x_5), and male theft conviction rate (x_6). Figure 4 shows a yearly time series plot of the female indictable offense conviction rate per 100,000 Canadian women aged 15 years and older. This is for the period 1931 to 1968. The conviction rate rose from the mid-1930s until 1940, then declined until the mid-1950s, and rose again. We are interested in relating variations in women's crime rates to changes in their position within Canadian society using the six covariates.

The results from the multiple linear regression of women's conviction rate on these

Figure 4 – Time series plot of the female indictable offense conviction rate per 100,000 Canadian women aged 15 years and older.

six covariates show that women's labor force participation rate and male indictable offense conviction rate are significant covariates. The coefficients are not estimated very precisely after all and the data set is quite small. A useful next step is to plot the residuals against time. Plots of the residual autocorrelations and partial autocorrelations for the multiple linear regression using this data are shown in Figure 5. As a rough guide to the statistical significance of residual autocorrelations and partial autocorrelations, reference lines have been placed in Figure 5. The pattern is clearly indicative of an AR(2) or ARIMA(1,0,1) process, with a positive autoregressive coefficient at lag 1 and a negative coefficient at lag 2. As a result of using the `auto.arima` function under the `forecast` `R-package`, it appears that ARIMA(1,0,1) is the most appropriate model to fit the data set (Rob and George, 2021). The Ljung-Box test for white noise, using 8 lags, resulted in a p-value of 0.94, confirming the presence of white noise.

In the first step of the shrinkage approach, we apply backward elimination to select

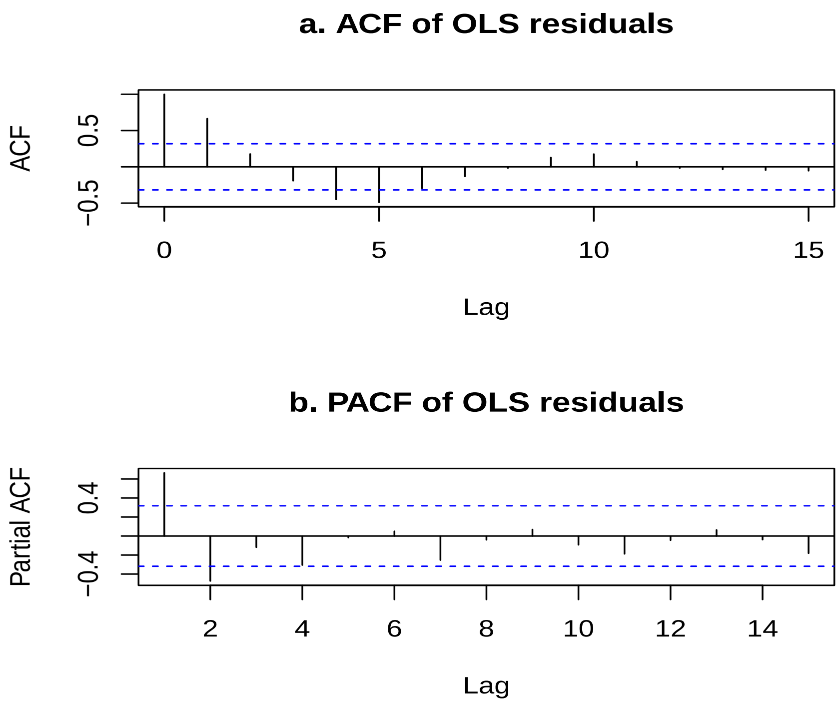

Figure 5 – ACF and PACF plots of residuals for Canadian crime rates data.

the significant covariates. It shows that women's labor force participation rate and male indictable offense conviction rate are significantly related to female indictable-offense conviction, while the remaining four covariates are insignificant. Then we define a restricted subspace using insignificant covariates. Therefore, the restricted subspace is $\boldsymbol{\xi_2} = (\xi_1, \xi_3, \xi_4, \xi_6) = (0, 0, 0, 0)$ together with $p = 1, q = 1$, $k_1 = 4$, and $k_2 = 4$.

We used non-overlapping block bootstrap (Härdle *et al.*, 2003) to compute point estimates, standard errors, and the RMSEs of the proposed estimators. Only the significant coefficients are reported in Table 2. We apply the following bootstrap Algorithm.

ALGORITHM 9.

1. *Generate the B replicates for bootstrap of the series from the crime data set:*

   i) *Divide the series into b blocks $y_1, y_2, y_3, \ldots, y_b$ with length g, where $b \times g \approx n$.*

   ii) *Randomly select b blocks with replacement from these blocks, to get a new bootstrap sample, $y_1^*, y_2^*, y_3^*, \ldots, y_b^*$.*

2. *The ARIMA model was fitted to each of the bootstrap samples by using the maximum likelihood estimation method and obtain the estimate* $\hat{\boldsymbol{\xi}}$.

3. *Repeat steps 1-2,* 1000 *times.*

Based on the above Algorithm, we calculated the estimates, standard errors, and RMSEs. The findings presented in Table 2 show that the RE, SE, and PSE estimators are superior to the URE, which is corroborated by theoretical and simulation results that identify shrinkage estimate as an improvement over the URE.

TABLE 2
Estimates (first row) and standard errors (second row) for AR parameter ϕ_1*, MA parameter* θ_1*, and significant covariates,* x_2 *and* x_5*.*

Estimators	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$	RMSE
URE	0.46	0.11	0.11	0.06	1.00
	(0.43)	(0.34)	(0.18)	(0.05)	
RE	0.55	0.02	0.12	0.03	1.88
	(0.41)	(0.18)	(0.13)	(0.03)	
SE	0.49	0.08	0.12	0.05	1.34
	(0.42)	(0.31)	(0.17)	(0.05)	
PSE	0.48	0.08	0.12	0.05	1.39
	(0.41)	(0.30)	(0.17)	(0.05)	

6. Discussion

Due to its favorable efficiency characteristics, the unrestricted maximum likelihood estimator (URE) is one of the most commonly used and widely accepted methods for estimating parameters. However, significant research has focused on improving ML estimators, particularly for time series regression model parameters. For instance, shrinkage estimators have been shown to outperform URE under certain conditions. In this study, we introduce shrinkage estimators for multiple linear regression models with ARIMA errors when some regression parameters lie in a subspace. The shrinkage method proceeds in four steps: First, a full model is fit using all available covariates. Second, covariates that do not improve the model's maximum likelihood fit are identified using a variable selection procedure with AIC criterion. Third, a restricted model is fit, assuming the coefficients of these redundant covariates are zero. Finally, the full model's estimates are optimally shrunk toward the restricted model's estimates, along the direction defined by the zero-coefficient restriction.

This paper examines the asymptotic bias and risk properties of shrinkage estimators, supported by extensive simulation studies. These simulations demonstrate that the restricted estimator outperforms the unrestricted estimator when the true parameter is near the imposed restriction, but this advantage diminishes as the true parameter moves further away. The shrinkage estimators are most effective when (1) the unrestricted and restricted models are similar, indicating that assumptions about insignificant covariates are valid; and (2) the number of insignificant covariates is large. The positive shrinkage estimator outperforms the shrinkage estimator at and near the null hypothesis. A real-world application to Canadian crime rates further demonstrates the superiority of the shrinkage estimators over the URE. Future research will explore shrinkage estimation methods for nonlinear time series regression models, and compare their performance against machine learning algorithms applied to time series models with ARIMA errors.

The paper would significantly benefit from a more thorough exploration of the practical limitations inherent in employing shrinkage methods within the context of ARIMA models. The effectiveness of these methods critically on the validity of ARIMA model assumptions, including stationarity of the time series and the normality of error terms. However, deviations from these assumptions—such as non-homogeneous variance, the presence of structural breaks, non-normal error distributions (e.g., heavy-tailed distributions or outliers), or model misspecification—can severely compromise the accuracy and reliability of the shrinkage estimators. Specifically, violations of stationarity will affect the covariance structure upon which the shrinkage is based, potentially leading to biased and inaccurate forecasts. Similarly, outliers and heavy-tailed errors will distort the estimated covariance structure, negatively impacting the shrinkage performance. A detailed discussion addressing these limitations, including strategies for robust model specification and outlier handling, is necessary to fully assess the applicability and reliability of the proposed methodology in real-world scenarios where ARIMA assumptions might be violated.

Acknowledgements

We express our sincere thanks to the editor, associate editor and referees for their constructive and valuable suggestions, which led to improvements in this paper. The research of Shakhawat Hossain and Andrei Volodin was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Appendix

A. Proofs

PROOF. Theorem 7

The bias expressions of the proposed estimators is here derived. It is obvious that

ADB($\hat{\xi}$) = 0. The ADB of ($\hat{\xi}_r$), ($\hat{\xi}_S$), and ($\hat{\xi}_{S+}$) estimators are as follows:

$$
\begin{aligned}
\mathrm{ADB}(\hat{\xi}_r) =& \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}_r-\xi)\right) = -\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{h} = -\gamma,\\
\mathrm{ADB}(\hat{\xi}_S) =& \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}_S-\xi)\right) = -\mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}((k_2-2)\hat{\Lambda}^{-1}(\hat{\xi}-\hat{\xi}_r))\right) = \nu\gamma\mathrm{E}(\mathrm{Z}_1),\\
\mathrm{ADB}(\hat{\xi}_{S+}) =& \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}_{S+}-\xi)\right)\\
=& \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}_S-\xi) - \sqrt{n}\left(1-(k_2-2)\hat{\Lambda}^{-1}\right)I\left(\hat{\Lambda}<(k_2-2)\right)(\hat{\xi}-\hat{\xi}_r)\right)\\
=& \mathrm{ADB}(\hat{\xi}_S) - \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}-\hat{\xi}_r)\left(1-(k_2-2)\hat{\Lambda}^{-1}\right)I\left(\hat{\Lambda}<(k_2-2)\right)\right)\\
=& \mathrm{ADB}(\hat{\xi}_S) + \gamma\mathrm{E}\left(I\left(\mathrm{Z}_1<(k_2-2)\right)\right)\\
& -\gamma(k_2-2)\mathrm{E}\left(\sqrt{n}(\hat{\xi}-\hat{\xi}_r)\hat{\Lambda}^{-1}I\left(\hat{\Lambda}<(k_2-2)\right)\right)\\
=& \mathrm{ADB}(\hat{\xi}_S) + \gamma\Psi_{\nu+4}(\nu,\Delta) - \gamma\nu\mathrm{E}\left(\mathrm{Z}_1^{-1}I\left(\mathrm{Z}_1<\nu\right)\right).
\end{aligned}
$$

□

PROOF. Theorem 8

To assess the proposed estimators, we derive their asymptotic covariance matrices. The covariance matrix of any estimator $\hat{\xi}^*$ is defined as:

$$\mathrm{Cov}(\hat{\xi}^*) = \mathrm{E}\left(\lim_{n\to\infty} n(\hat{\xi}^*-\xi)(\hat{\xi}^*-\xi)^\top\right).$$

First, we will start deriving the covariance matrices of the URE and RE:

$$
\begin{aligned}
\mathrm{Cov}(\hat{\xi}) &= \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}-\xi)\sqrt{n}(\hat{\xi}-\xi)^\top\right) = \boldsymbol{I}_{11.2}^{-1}\\
\mathrm{Cov}(\hat{\xi}_r) &= \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}_r-\xi)\sqrt{n}(\hat{\xi}_r-\xi)^\top\right) = \boldsymbol{\Phi} + \gamma\gamma^\top,
\end{aligned}
$$

where $\boldsymbol{\Phi} = \boldsymbol{I}_{11.2}^{-1} - \boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}\boldsymbol{I}_{22.1}^{-1}\boldsymbol{I}_{21}\boldsymbol{I}_{11}^{-1}$. Secondly, we derive the covariance matrices of the shrinkage and positive shrinkage estimators:

$$
\begin{aligned}
\mathrm{Cov}(\hat{\xi}_S) &= \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}(\hat{\xi}_S-\xi)\sqrt{n}(\hat{\xi}_S-\xi)^\top\right)\\
&= \mathrm{E}\left(\lim_{n\to\infty}\sqrt{n}\left(\hat{\xi}-\xi-\nu\hat{\Lambda}^{-1}(\hat{\xi}-\hat{\xi}_r)\right)\sqrt{n}\left(\hat{\xi}-\xi-\nu\hat{\Lambda}^{-1}(\hat{\xi}-\hat{\xi}_r)\right)^\top\right)\\
&= \mathrm{E}(\sqrt{n}(\hat{\xi}-\xi)\sqrt{n}(\hat{\xi}-\xi)^\top) + \nu^2\mathrm{E}(\sqrt{n}(\hat{\xi}-\hat{\xi}_r)\sqrt{n}(\hat{\xi}-\hat{\xi}_r)^\top\lim_{n\to\infty}\hat{\Lambda}^{-2})\\
&\quad -2\nu\mathrm{E}(\sqrt{n}(\hat{\xi}-\hat{\xi}_r)\sqrt{n}(\hat{\xi}-\xi)^\top\lim_{n\to\infty}\hat{\Lambda}^{-1})\\
&= \boldsymbol{I}_{11.2}^{-1} + \nu^2\boldsymbol{C}\mathrm{E}\left(\mathrm{Z}_1^2\right) + \nu^2\gamma\gamma^\top\mathrm{E}\left(\mathrm{Z}_2^2\right) - 2\nu\mathrm{E}(\sqrt{n}(\hat{\xi}-\hat{\xi}_r)\sqrt{n}(\hat{\xi}-\xi)^\top\lim_{n\to\infty}\hat{\Lambda}^{-1}).
\end{aligned}
$$

Consider the last term:

$$
\begin{aligned}
& \mathrm{E}\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\sqrt{n}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^{\top}\lim_{n\to\infty}\hat{\Lambda}^{-1}\Big)\\
= \;& \mathrm{E}\Big(\mathrm{E}(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\sqrt{n}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^{\top}\lim_{n\to\infty}\hat{\Lambda}^{-1}|\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}}))\Big)\\
= \;& \mathrm{E}\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\mathrm{E}(\sqrt{n}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^{\top}|\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}}))\lim_{n\to\infty}\hat{\Lambda}^{-1}\Big)\\
& +\mathrm{E}\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})^{\top}-\mathrm{E}(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}}))\Big)^{\top})\lim_{n\to\infty}\hat{\Lambda}^{-1}\Big)\\
= \;& \mathrm{E}\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})^{\top}\lim_{n\to\infty}\hat{\Lambda}^{-1}\Big)-\mathrm{E}\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\lim_{n\to\infty}\hat{\Lambda}^{-1}\Big)\mathrm{E}\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\Big)^{\top}\\
= \;& \boldsymbol{C}\mathrm{E}\Big(\chi^{-2}_{k_2+2}(\Delta)\Big)+\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}\mathrm{E}\Big(\chi^{-2}_{k_2+4}(\Delta)\Big)-\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}\mathrm{E}\Big(\chi^{-2}_{k_2+2}(\Delta)\Big)\\
= \;& \boldsymbol{C}\mathrm{E}(\mathrm{Z}_1)+\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}\mathrm{E}(\mathrm{Z}_2)-\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}\mathrm{E}(\mathrm{Z}_1).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{\xi}}_S) &= \boldsymbol{I}^{-1}_{11.2}+\nu^2\Big(\boldsymbol{C}\mathrm{E}(\mathrm{Z}_1^2)+\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}\mathrm{E}(\mathrm{Z}_2^2)\Big)\\
&\quad -2\nu\Big(\boldsymbol{C}\mathrm{E}(\mathrm{Z}_1)+\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}\mathrm{E}(\mathrm{Z}_2)-\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}\mathrm{E}(\mathrm{Z}_1)\Big)\\
&= \boldsymbol{I}^{-1}_{11.2}+\Big(\nu^2\mathrm{E}\big(\mathrm{Z}_1^2\big)-2\nu\mathrm{E}(\mathrm{Z}_1)\Big)\boldsymbol{C}\\
&\quad +\Big(\nu^2\mathrm{E}(\mathrm{Z}_2^2)+2\nu\mathrm{E}(\mathrm{Z}_1)-2\nu\mathrm{E}(\mathrm{Z}_2)\Big)\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top}.
\end{aligned}
$$

Let $F_m(\Delta)=\Big(1-\nu\hat{\Lambda}^{-1}\Big)^m I\Big(\hat{\Lambda}<\nu\Big)$, where $m=1,2$.

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{\xi}}_{S+}) &= \mathrm{E}\Big(\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\xi}}_{S+}-\boldsymbol{\xi})\sqrt{n}(\hat{\boldsymbol{\xi}}_{S+}-\boldsymbol{\xi})^{\top}\Big)\\
&= \mathrm{E}\Big(\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\xi}}_{S}-\boldsymbol{\xi})\sqrt{n}(\hat{\boldsymbol{\xi}}_{S}-\boldsymbol{\xi})^{\top}\Big)\\
&\quad +\mathrm{E}\Big(\lim_{n\to\infty}F_2(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})^{\top}\Big)\\
&\quad -2\mathrm{E}\Big(\lim_{n\to\infty}F_1(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\sqrt{n}(\hat{\boldsymbol{\xi}}_{S}-\boldsymbol{\xi})^{\top}\Big)\\
&= \mathrm{Cov}(\hat{\boldsymbol{\xi}}_S)+\mathrm{E}\Big(\lim_{n\to\infty}F_2(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\sqrt{n}(\hat{\boldsymbol{\xi}}_{\boldsymbol{r}}-\boldsymbol{\xi})^{\top}\Big)\\
&\quad -2\mathrm{E}\Big(\lim_{n\to\infty}F_1(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\Big(\sqrt{n}(\hat{\boldsymbol{\xi}}_{\boldsymbol{r}}-\boldsymbol{\xi})^{\top}+\Big(1-\nu\hat{\Lambda}^{-1}\Big)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})^{\top}\Big)\Big)\\
&= \mathrm{Cov}(\hat{\boldsymbol{\xi}}_S)+\mathrm{E}\Big(\lim_{n\to\infty}F_2(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})^{\top}\Big)\\
&\quad -2\mathrm{E}\Big(\lim_{n\to\infty}F_1(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_{\boldsymbol{r}})(\sqrt{n}(\hat{\boldsymbol{\xi}}_{\boldsymbol{r}}-\boldsymbol{\xi})^{\top}\Big).
\end{aligned}
$$

Consider the second term:

$$\begin{aligned}
& -\mathrm{E}\left(\lim_{n\to\infty} F_2(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)^\top\right) \\
= & -\mathrm{E}\left(\lim_{n\to\infty} \left(1-\nu\hat{\Lambda}^{-1}\right)^2 I\left(\hat{\Lambda}<\nu\right)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)^\top\right) \\
= & -\boldsymbol{C}\mathrm{E}\left(I(Z_1<\nu)\left(1-\nu Z_1^{-1}\right)^2\right)-\gamma\gamma^\top\mathrm{E}\left(I(Z_2<\nu)\left(1-\nu Z_2^{-1}\right)^2\right).
\end{aligned}$$

Consider the third term:

$$\begin{aligned}
& -2\mathrm{E}\left(\lim_{n\to\infty} F_1(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)(\sqrt{n}(\hat{\boldsymbol{\xi}}_r-\boldsymbol{\xi})^\top\right) \\
= & -2\mathrm{E}\left(\lim_{n\to\infty} \sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)\mathrm{E}\left(F_1(\Delta)\sqrt{n}(\hat{\boldsymbol{\xi}}_r-\boldsymbol{\xi})^\top|\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)\right)\right) \\
= & -2\mathrm{E}\left(\lim_{n\to\infty} \sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)\mathrm{E}\left(\sqrt{n}(\hat{\boldsymbol{\xi}}_r-\boldsymbol{\xi})^\top\right)F_1(\Delta)+0\right) \\
= & -2\mathrm{E}\left(\lim_{n\to\infty} \sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)I\left(\hat{\Lambda}<\nu\right)-\nu\hat{\Lambda}^{-1}\sqrt{n}(\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\xi}}_r)I\left(\hat{\Lambda}<\nu\right)\right) \\
& \times\mathrm{E}\left(\sqrt{n}(\hat{\boldsymbol{\xi}}_r-\boldsymbol{\xi})^\top\right) \\
= & \; 2\Gamma_{\nu+4}(\nu,\Delta)\gamma\gamma^\top-2\nu\mathrm{E}\left(Z_1^{-1}I(Z_1<\nu)\right)\gamma\gamma^\top \\
= & \left(2\Gamma_{\nu+4}(\nu,\Delta)-2\nu\mathrm{E}\left(Z_1^{-1}I(Z_1<\nu)\right)\right)\gamma\gamma^\top.
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathrm{Cov}\left(\hat{\boldsymbol{\xi}}_{S+}\right) = & \; \mathrm{Cov}\left(\hat{\boldsymbol{\xi}}_S\right)-\mathrm{E}\left(\left(1-\nu Z_1^{-1}\right)^2 I(Z_1<\nu)\right)\boldsymbol{C} \\
& +\left(2\Gamma_{\nu+4}(\nu,\Delta)-2\nu\mathrm{E}\left(Z_1^{-1}I(Z_1<\nu)\right)\right. \\
& -\left.\mathrm{E}\left(\left(1-\nu Z_2^{-1}\right)^2 I(Z_2<\nu)\right)\right)\gamma\gamma^\top.
\end{aligned}$$

The proof of Theorem 8 now follows using Eq. (3) and the above covariance matrices.
□

REFERENCES

A. M. BIANCO, G. C. BEN, M. G. AND. MARTINEZ, V. YOHAI (2001). *Outlier detection in regression models with ARIMA errors using robust estimates*. Journal of Forecasting, 20, pp. 565–579.

K.-S. CHAN, K. CHEN (2011). *Subset ARMA selection via the adaptive LASSO*. Statistics and its Interface, 4, no. 2, pp. 197–205.

S. CHEN, X. WANG, J. ZHAO, Y. ZHANG, X. KAN (2022). *Application of the ARIMA model in forecasting the incidence of tuberculosis in anhui during COVID-19 pandemic from 2021 to 2022*. Infection and Drug Resistance, 15, pp. 3503–3512.

R. A. DAVIS, W. T. M. DUNSMUIR (1997). *Least absolute deviation estimation for regression with ARMA errors*. Journal of Theoretical Probability, 10, no. 2, pp. 481–497.

J. FOX, S. WEISBERG (2019). *An R Companion to Applied Regression*. Sage publications, Thousand Oaks CA.

M. FURNO (1996). *The information matrix test in the linear regression with ARMA errors*. Journal of the Italian Statistical Society, 5, pp. 369–385.

E. GEORGE, G. JENKINS, G. REINSEL (2008). *Time Series Analysis: Forecasting and Control*. John Wiley and Sons, New Jersey.

A. K. GUPTA, A. M. E. SALEH, P. SEN (1989). *Improved estimation in a contingency table: independence structure*. Journal of the American Statistical Association, 84, no. 406, pp. 525–532.

W. HÄRDLE, J. HOROWITZ, J. P. KREISS (2003). *Bootstrap methods for time series*. International Statistical Review, 71, no. 2, pp. 435–459.

S. HOSSAIN, M. GHAHRAMANI (2016). *Shrinkage estimation of linear regression models with GARCH errors*. Journal of Statistical Theory and Applications, 15, no. 4, pp. 405–423.

S. HOSSAIN, L. A. LAC (2018). *Non-penalty shrinkage estimation of random effect models for longitudinal data with AR(1) errors*. Journal of Statistical Computation and Simulation, 88, no. 16, pp. 3230–3247.

S. JOHANSEN, M. RIANI, A. ATKINSON (2012). *The selection of ARIMA models with or without regressors*. University of Copenhagen, Department of Economics Discussion Paper, 12-17, pp. 1–32.

M. C. OTTO, W. R. BELL, J. P. BURMAN (1987). *An Iterative GLS Approach to Maximum Likelihood Estimation of Regression Models with ARIMA Errors*. Research Report No. 87/34, Statistical Research Division, US Bureau of the Census, Washington, D.C.

A. PANKRATZ (1991). *Forecasting With Dynamic Regression Models*. John Wiley and Sons, New York.

M. PAOLELLA (2019). *Linear Models and Time Series Analysis: Regression, ANOVA, ARMA and GARCH*. John Wiley and Sons, New York.

J. H. ROB, A. GEORGE (2021). *Forecasting: Principles and Practice*. OTexts, Melbourne, Australia.

P. K. SEN, A. K. M. EHSANES SALEH (1987). *On preliminary test and shrinkage M-estimation in linear models*. The Annals of Statistics, 15, no. 4, pp. 1580–1592.

T. THOMSON, S. HOSSAIN, M. GHAHRAMANI (2015). *Application of shrinkage estimation in linear regression models with autoregressive errors*. Statistical Computation and Simulation, 85, no. 16, pp. 3335–3351.

T. THOMSON, S. HOSSAIN, M. GHAHRAMANI (2016). *Efficient estimation for time series following generalized linear models*. Australian & New Zealand Journal of Statistics, 58, no. 4, pp. 493–513.

H. WANG, G. LI, C. TSAI (2007). *Regression coefficient and autoregressive order shrinkage and selection via the LASSO*. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 69, no. 1, pp. 63–78.

R. B. WILLIAM (2011). *REGCMPNT–A fortran program for regression models with ARIMA component errors*. Journal of Statistical Software, 41, no. 7, pp. 1–23.

M. A. WINCEK, G. C. REINSEL (1986). *An exact maximum likelihood estimation procedure for regression ARMA time series models with possibly nonconsecutive data*. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 48, no. 3, pp. 303–313.

R. WU, Q. WANG (2012). *Shrinkage estimation for linear regression with ARMA errors*. Journal of Statistical Planning and Inference, 142, no. 7, pp. 2136–2148.

M. XU, Z. QIN (2021). *A novel hybrid ARIMA and regression tree model for the interval-valued time series*. Journal of Statistical Computation and Simulation, 91, no. 5, pp. 1000–1015.