The Improved Adaptive LASSO Method and Applications in Autoregressive Model

Wenbin Hao, Yanling Zhu*

Abstract—**For the accuracy of the autoregressive model, variable selection and model order are two main problems. We hope to select variables and determine the order at the same time. The improved adaptive LASSO method (IALASSO) not only notices the sensitivity of the time series, but pays attention to the influence of the lag order. We show that the IALASSO estimator enjoys asymptotic properties under certain conditions. Simulation studies demonstrate that the IALASSO estimator is superior than the other method. Finally, by applying IALASSO to actual time series data, the prediction accuracy is higher.**

Index Terms—**Adaptive LASSO, variable selection, autoregressive model, lag effect**

I. Introduction

THERE are many methods of parameter estimation, including least squares estimation, maximum likelihood estimation, ridge regression estimation [1] and the other methods [2]–[4], but these methods have a common disadvantage that they cannot reduce the set of explanatory variables. Variable selection is an important method for dimensionality reduction and mining hidden structures in data, which can improve model interpretability and prediction accuracy. Classical variable selection methods include AIC [5], Mallows'C_p [6], BIC [7] and EICw [8]. These traditional methods are discrete and not stable in variable selection. In order to solve this problem, Tibshirani [9] proposed the LASSO, which makes some regression coefficients become 0 to achieve variable selection by using penalized likelihood method. At the same time, the regression coefficients of significant variables are estimated. On the basis of the LASSO, many scholars have conducted in-depth research and continuous improvement, and have proposed many model selection methods [10]–[15]. For example, Zou found that the LASSO uses the same degree of compression for all coefficients, and does not have the Oracle properties [16]. So the Adaptive LASSO method is proposed, which has the Oracle properties and uses different degrees of compression for different coefficients.

The existed reference has shown that the LASSO is an effective variable selection method, especially in the analysis of high-dimensional data models. If LASSO is applied to time series models, its good performance will further play a role in variable selection. Nardi et al. [17] showed that LASSO has good applicability in estimating regression coefficients and determining the order of AR(p) models. Che et al. [18] presented the AHO-Lasso-SVR model by applying support vector regression forecasting. However, the correlation between the time series will affect the process of parameter estimation and variable selection. Noted that the Adaptive LASSO is proposed for cross-sectional data, it does not consider the sensitivity of the time series model with the time order and ignores the impact of the lag order on the time series modeling. Therefore, these facts will cause that the Adaptive LASSO can not perform very well for the time series model, which can easily lead to inaccurate estimation and prediction.

In order to fully consider the characteristics of the time series model and retain the advantages of the Adaptive LASSO, we modify the objective function of the Adaptive LASSO and the penalty term, and propose the improved Adaptive LASSO (IALASSO) for the autoregressive model in this paper. We want to achieve the goal that the penalty function has a heavier penalty to more backward lag term and a lighter penalty to more forward lag term. Besides, we also assure that the penalty function is related with the distance to the middle term. This method gives a lighter penalty to the lag terms farther away from the middle. In this paper, we mainly focus on using the Adaptive LASSO with improved penalty term for variable selection and parameter estimation of autoregressive time series models.

The rest of this paper is organized as follows. In Section II we propose the IALASSO estimator and introduce the improved penalty term in detail. The main theoretical properties of the IALASSO estimator are given in Section III. In Section IV, simulation studies of the proposed method are presented. As an application, the stock price data is selected for fitting and prediction in Section V.

II. Adaptive LASSO with Improved Penalty Term(IALASSO)

A. Penalized likelihood method

A classical linear regression model can be set as $y_i = \sum_{j=1}^{p}\beta_{0j}x_{ij} + \varepsilon_i, i = 1,2,...,n$. There are n groups of observations, and each group of observations consists of an response variable y_i and p related predictor variables $x_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$, where $\beta_0 = (\beta_{01}, \beta_{02}, ..., \beta_{0p})^T$ are unknown parameters and ε_i is the error term. The goal of linear regression is to predict the response value from the predictor variables and to be able to find out which predictors are important. Estimating the parameter vector β_0 is the core work to achieve the above goals. The traditional

Manuscript received August 27, 2022; revised May 29, 2023. This research was supported by the National Natural Science Foundation of China (No.71803001), Excellent Young Talents Fund Program of Higher Education Institutions of Anhui Province (No. gxyqZD2019031), the Ministry of Education humanities social sciences study project of China (No.21YJAZH081), the Natural Science Foundation of Anhui Province (No.2108085MA04) and the Graduate Research and Innovation Fund project of Anhui University of Finance and Economics(ACYC2021387).

Wenbin Hao is a postgraduate student in School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, Anhui, 233000, China(e-mail: haowenbin2021@126.com).

Yanling Zhu is a professor in School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, Anhui, 233000, China(Corresponding author, phone: +86-0552-3170748; fax: +86-0552-3170748; e-mail: zhuyanling99@126.com).

OLS approach is to minimize the following function

$$\min_{\beta}\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}\beta_j x_{ij})^2. \quad (1)$$

In general the OLS estimator of the parameters is not equal to 0, which means that the final model become difficult to interpret if p is large. When $n < p$, the OLS estimator is not unique, and there are infinite solutions. Therefore it is necessary to constrain this estimation process by a penalty function. On the basis of OLS, the constraint condition about β is added by the penalty function $p_\lambda(|\beta|)$. The penalized least squares estimation is established [19] as follows

$$\widehat{\beta}=\arg\min[\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}\beta_j x_{ij})^2+\sum_{j=1}^{p}p_\lambda(|\beta_j|)].$$

When the penalty function takes different functions, different variable selection methods can be constructed. The L_1 penalty function $p_\lambda(|\beta|)=\lambda|\beta|$ corresponds to LASSO, Ridge regression with the L_2 penalty function $p_\lambda(|\beta|)=\lambda\beta^2$, and Bridge regression with the L_q $(0<q<1)$ penalty function $p_\lambda(|\beta|)=\lambda|\beta|^q$.

B. LASSO

The LASSO method aims to preserve the important features of the independent variables by assigning some coefficients to 0. It is widely used in high dimensional data, where the number of features p is much larger than the number of observations n. The LASSO estimator $\hat{\beta}_{LASSO}$ is defined as

$$\hat{\beta}_{LASSO}=\arg\min[\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{j=1}^{p}|\beta_j|],$$

where λ is a non-negative tuning parameter and the penalty function is $\lambda\sum_{j=1}^{p}|\beta_j|$.

The complexity of LASSO is controlled by λ. The smaller the λ, the lighter the penalty for the linear model with more variables. When λ is sufficiently large, the LASSO can make some coefficients in the model strictly equal to 0, so that a model with fewer variables is finally obtained. By constructing a penalty function, it can compress the coefficients of variables and make some regression coefficients become 0, so as to achieve the purpose of variable selection.

C. Adaptive LASSO

Zou [20] proposed Adaptive LASSO, which aims to improve LASSO regression by introducing weight coefficient, and uses adaptive weight in the penalty function to penalize different parameters. It also has a sparse solution and Oracle properties, which means it has the same asymptotic distribution as the case when the real parameters are known in advance. The Adaptive LASSO estimator $\hat{\beta}_{ALASSO}$ is defined as

$$\hat{\beta}_{ALASSO}=\arg\min[\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{j=1}^{p}\omega_j|\beta_j|],$$

where the penalty function is $\lambda\sum_{j=1}^{p}\omega_j|\beta_j|$.

The adaptive weight is selected by

$$\omega_j=\frac{1}{\left|\widetilde{\beta}_j\right|^{\gamma}},(j=1,2,...,p)$$

where $\gamma>0$ and $\widetilde{\beta}$ is estimated by LASSO or the other method. The positive value γ is a power of the adaptive weight and associated with higher order. The Adaptive LASSO is obtained by selecting the tuning parameter λ and the adaptive weight order γ from two-dimensional cross-validation. The Adaptive LASSO method uses different degrees of compression for different coefficients and the superiority of this method is to effectively correct the estimation bias of the model and speed up the convergence rate.

For time series model, in view of time sensitivity, variables with higher lag order usually have poorer predictive ability for the future. Combined with time series characteristics, Wang et al. [21] proposed the MALASSO, which combined with the characteristics of the time series model on the basis of Adaptive LASSO by giving a heavier penalty to the later order. The adaptive penalty weight is as follows

$$\omega_j=\frac{j^{\gamma_2}}{\left|\widetilde{\beta}_j\right|^{\gamma_1}},$$

where γ_1 is the positive penalty parameter of the coefficient, and γ_2 is the positive penalty parameter for the lag order. The larger the j, the larger the penalty weight ω_j. By giving a heavier penalty for the lag period, we limit the size of the parameter space to reduce the complexity of the model and improve the convergence speed. Therefore the accuracy and efficiency of parameter estimation and variable selection can be improved.

D. IALASSO

For time series model, the MALASSO ignores that the penalty function should be related with the distance to the middle term, which means the penalty function should give a lighter penalty to more forward lag term. Based on this, a new penalty method can be proposed to appropriately adjust the penalty function of the MALASSO. The specific method is to divide the intermediate term of the lag order, and give a relatively heavier penalty to the coefficient whose lag order is larger than the middle term. A relatively lighter penalty for coefficients whose lag order is smaller than the middle term ensures that the leading lag terms enter the model with larger probability.

The advantage of this improvement is that the compression of different coefficients can speed up the automatic adjustment, limit the size of the parameter space to reduce the complexity of the model, and improve the convergence speed. In order to achieve the purpose of improving the accuracy and efficiency of variable selection and parameter estimation, we propose the IALASSO method with the adaptive penalty weight as follows

$$\omega_j=\frac{j^{sgn\left(j-\frac{p+1}{2}\right)\gamma_2}}{\left|\widetilde{\beta}_j\right|^{\gamma_1}}. \quad (2)$$

Therefore, for the AR(p) model the estimator of the IALASSO is defined as

$$\hat{\beta}_{IALASSO}=\arg\min[\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{j=1}^{p}\omega_j|\beta_j|],$$

where ω_j is defined by (2). When the lag order j is less than the middle term, the penalty is lighter by $\omega_j=\frac{j^{-\gamma_2}}{\left|\widetilde{\beta}_j\right|^{\gamma_1}}=$

$\frac{1}{|\tilde{\beta}_j|^{\gamma_1} j^{\gamma_2}}$. When the lag order j is larger than the middle term, the penalty will be heavier by $\omega_j = \frac{j^{\gamma_2}}{|\tilde{\beta}_j|^{\gamma_1}}$.

III. Theoretical Properties

For the convenience of statement, we introduce some notations firstly. For AR(p) model $Y = X\beta_0 + \varepsilon$, where $\varepsilon = (\varepsilon_{p+1}, \varepsilon_{p+2}, ..., \varepsilon_T)^T$, $\{\varepsilon_t\}$ are random errors with mean 0 and variance σ^2, $\{x_t\}$ is the time series, $Y = (x_{p+1}, x_{p+2}, ..., x_T)^T$ and

$$X = \begin{bmatrix} x_p & x_{p-1} & \cdots & x_1 \\ x_{p+1} & x_p & \cdots & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{T-1} & x_{T-2} & \cdots & x_{T-p} \end{bmatrix}.$$

Denote $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$, $\frac{1}{T-p}\sum_{t=1}^{T} X_t X_t^T := C_T$ and suppose $C_T \to C(T \to \infty)$, where $X_t = (x_{t-1}, ..., x_{t-p})^T$ and C is a nonnegative definite non-singular matrix.

A. Consistency of the estimator

From the previous discussion, the IALASSO estimator of $AR(p)$ model can be obtaind by

$$\begin{aligned} &\hat{\beta}_{IALASSO} \\ &= \arg\min[\frac{1}{T-p}\sum_{t=p+1}^{T}(x_t - \sum_{j=1}^{p}\beta_j x_{t-j})^2 \\ &+ \frac{\lambda_T}{T-p}\sum_{j=1}^{p}\omega_j|\beta_j|], \end{aligned} \tag{3}$$

where λ_T is the tuning parameter related to sample size T.

Theorem 1 If $\lambda_T = o(T)(T \to \infty)$, then the IALASSO estimator $\hat{\beta}_{IALASSO}$ is consistent to β_0.

Proof: Firstly denote

$$Q_T(\beta) = \frac{1}{T-p}\sum_{t=p+1}^{T}\left(Y_t - X_t^T\beta\right)^2 + \frac{\lambda_T}{T-p}\sum_{j=1}^{p}\omega_j|\beta_j|,$$

then we can easily get that

$$\begin{aligned} &Q_T(\beta) \\ &= \frac{1}{T-p}\sum_{t=p+1}^{T}\left[X_t^T(\beta_0-\beta)+\varepsilon_t\right]^2 + \frac{\lambda_T}{T-p}\sum_{j=1}^{p}\omega_j|\beta_j| \\ &= (\beta-\beta_0)^T C_T(\beta-\beta_0) + \frac{2}{T-p}(\beta_0-\beta)^T\sum_{t=p+1}^{T}X_t\varepsilon_t \\ &+ \frac{1}{T-p}\sum_{t=p+1}^{T}\varepsilon_t^2 + \frac{\lambda_T}{T-p}\sum_{j=1}^{p}\omega_j|\beta_j|, \end{aligned}$$

where $\{\varepsilon_t\}$ is not correlated with $\{x_t\}$ at the same period. Besides this, notice the fact that $Q_T(\beta)$ is convex, which yields that for any compact set S,

$$\sup_{\beta\in S}|Q_T(\beta) - Q(\beta) - \sigma^2| \xrightarrow{p} 0,$$

where

$$Q(\beta) = (\beta-\beta_0)^T C(\beta-\beta_0) + (\lim_{T\to\infty}\frac{\lambda_T}{T-p})\sum_{j=1}^{p}\omega_j|\beta_j|.$$

In view of $Q_T(\beta) \geq \frac{1}{T-p}\sum_{t=p+1}^{T}\left(Y_t - X_t^T\beta\right)^2$ and $\arg\min(\frac{1}{T-p}\sum_{t=p+1}^{T}\left(Y_t - X_t^T\beta\right)^2) = O_p(1)$, we can get that

$$\arg\min(Q_T(\beta)) = O_p(1).$$

It follows from the above formula that

$$\arg\min(Q_T(\beta)) \xrightarrow{p} \arg\min(Q(\beta)+\sigma^2),$$

which yields that the IALASSO estimator $\hat{\beta}_{IALASSO}$ is consistent to β_0 when $\lambda_T = o(T)$. ■

B. Asymptotic normality

Theorem 2 If $\lim_{T\to\infty}\frac{\lambda_T}{\sqrt{T}} = \lambda_0$, then

$$\sqrt{T-p}\left(\hat{\beta}_{IALASSO} - \beta_0\right) \xrightarrow{d} \arg\min V$$

where

$$\begin{aligned} V(u) = &-2u^T W + \lambda_0\sum_{j=1}^{p}[u_j\omega_j sgn(\beta_{0j})I(\beta_{0j}\neq 0) \\ &+ \omega_j|u_j|I(\beta_{0j}=0)] + u^T Cu, W \sim N(0,\sigma^2 C) \end{aligned}$$

and u is any p-dimensional vector. Especially when $\lambda_0 = 0$, $\sqrt{T-p}\left(\hat{\beta}_{IALASSO} - \beta_0\right) \xrightarrow{d} C^{-1}W \sim N(0, \sigma^2 C^{-1})$.

Proof: For the p-dimensional vector $\overline{\beta} = \beta_0 + u$, we can obtain that

$$\begin{aligned} V_T(u) &= Q_T(\overline{\beta}) - Q_T(\beta_0) \\ &= \sum_{t=p+1}^{T}[(\varepsilon_t - u^T X_t/\sqrt{T-p})^2 - \varepsilon_t^2] \\ &+ \lambda_T\sum_{j=1}^{p}\omega_j[|\beta_{0j} + u_j/\sqrt{T-p}| - |\beta_{0j}|] \end{aligned} \tag{4}$$

has minimum at the point $\sqrt{T-p}\left(\hat{\beta}_{IALASSO} - \beta_0\right)$. For the first part of (4), it holds that

$$\sum_{t=p+1}^{T}[(\varepsilon_t - u^T X_t/\sqrt{T-p})^2 - \varepsilon_t^2] \xrightarrow{d} -2u^T W + u^T Cu.$$

For the second part of (4), we can easily get

$$\begin{aligned} &\lambda_T\omega_j[|\beta_{0j} + u_j/\sqrt{T-p}| - |\beta_{0j}|] \\ &= \lambda_T/\sqrt{T-p}\,\omega_j|u_j| \to \lambda_0\omega_j|u_j|, \end{aligned}$$

when $\beta_{0j} = 0$ and

$$\begin{aligned} &\lambda_T\omega_j[|\beta_{0j} + u_j/\sqrt{T-p}| - |\beta_{0j}|] \\ &\geq \lambda_T/\sqrt{T-p}\,\omega_j sgn(\beta_{0j})u_j \to \lambda_0\omega_j sgn(\beta_{0j})u_j \end{aligned}$$

when $\beta_{0j} \neq 0$. Therefore for the p-dimensional vector u,

$$V_T(u) \xrightarrow{d} V(u).$$

It follows from the convexity of V_T that

$$\arg\min(V_T) = \sqrt{T-p}\left(\hat{\beta}_{IALASSO} - \beta_0\right) \xrightarrow{d} \arg\min(V).$$

When $\lambda_0 = 0$,

$$\sqrt{T-p}\left(\hat{\beta}_{IALASSO} - \beta_0\right) \xrightarrow{\text{d}} C^{-1}W \sim N(0, \sigma^2 C^{-1}),$$

which shows that the IALASSO estimator is asymptotically normal. ■

IV. Numerical Simulation

In order to evaluate the numerical performance of IALASSO estimator, we simulate data from the following four AR(p) models:

$$AR\,(1) : y_t = -0.8y_{t-1} + \varepsilon_t$$
$$AR\,(2) : y_t = 0.5y_{t-1} - 0.3y_{t-2} + \varepsilon_t$$
$$AR\,(3) : y_t = 0.8y_{t-1} - 0.7y_{t-2} + 0.3y_{t-3} + \varepsilon_t$$
$$AR\,(4) : y_t = 0.5y_{t-1} + 0.6y_{t-2} - 0.2y_{t-3} + 0.3y_{t-4} + \varepsilon_t$$

Before using ALASSO, MALASSO and IALASSO for parameter estimation, we determine the length of time series data T equals 300, 500, 700 respectively. The times of simulation N is 200. In order to comprehensively evaluate the performance of the estimation method, we select four indicators, namely EE, MSE, FP and FN. Here EE (Estimation Error) is the average value of the values obtained by $\|\widehat{\beta} - \beta_0\|_2$, which is used to evaluate the accuracy of the estimation method. MSE (Mean Square Error) equals the average value of the numerical values obtained by $\|Y - X\widehat{\beta}\|_2$, which is used to evaluate the prediction ability of the estimation method. FP (False Positive) represents the average number of variables that mistake zero variables for non-zero variables. FN (False Negative) represents the average number of variables that mistake non-zero variables for zero variables.

TABLE I
Simulation results for AR(1) model

Setting	Method	EE	MSE	FP	FN
T=300	ALASSO	0.0023	0.0578	0.2300	0
	MALASSO	0.0051	0.0579	0.3200	0
	IALASSO	0.0037	0.0581	0.2100	0
T=500	ALASSO	0.0017	0.0448	0.2100	0
	MALASSO	0.0023	0.0451	0.2400	0
	IALASSO	0.0019	0.0449	0.1900	0
T=700	ALASSO	0.0011	0.0379	0.1800	0
	MALASSO	0.0014	0.0380	0.2400	0
	IALASSO	0.0013	0.0379	0.1600	0

TABLE II
Simulation results for AR(2) model

Setting	Method	EE	MSE	FP	FN
T=300	ALASSO	0.0108	0.0578	0.2400	0
	MALASSO	0.0144	0.0578	0.4400	0
	IALASSO	0.0097	0.0577	0.2900	0
T=500	ALASSO	0.0059	0.0450	0.2400	0
	MALASSO	0.0062	0.0449	0.2700	0
	IALASSO	0.0063	0.0447	0.2200	0
T=700	ALASSO	0.0034	0.0378	0.1700	0
	MALASSO	0.0043	0.0379	0.2600	0
	IALASSO	0.0035	0.0379	0.1100	0

TABLE III
Simulation results for AR(3) model

Setting	Method	EE	MSE	FP	FN
T=300	ALASSO	0.0232	0.0582	0.2300	0.030
	MALASSO	0.0215	0.0579	0.2600	0
	IALASSO	0.0220	0.0577	0.2650	0
T=500	ALASSO	0.0127	0.0447	0.2100	0
	MALASSO	0.0119	0.0447	0.2000	0
	IALASSO	0.0114	0.0449	0.2100	0
T=700	ALASSO	0.0063	0.0378	0.1600	0
	MALASSO	0.0067	0.0378	0.2050	0
	IALASSO	0.0054	0.0377	0.1700	0

TABLE IV
Simulation results for AR(4) model

Setting	Method	EE	MSE	FP	FN
T=300	ALASSO	0.0278	0.0578	0.1450	0.1250
	MALASSO	0.0282	0.0578	0.1400	0
	IALASSO	0.0282	0.0577	0.1400	0
T=500	ALASSO	0.0161	0.0447	0.1050	0.0200
	MALASSO	0.0144	0.0446	0.1000	0
	IALASSO	0.0153	0.0447	0.0850	0
T=700	ALASSO	0.0108	0.0378	0.1100	0
	MALASSO	0.0109	0.0378	0.0900	0
	IALASSO	0.0108	0.0376	0.0700	0

The following conclusions can be obtained from TABLE I to TABLE IV.

1) For the AR model, Adaptive LASSO and its related methods can perform variable selection and parameter estimation. The simulation results of the IALASSO method are generally better than the ALASSO and MALASSO.
2) Under the same scenario, with the increase of the sample size, the performance of the IALASSO estimator is getting better and better.
3) For the IALASSO, not only the correct variables are selected, but the accuracy of parameter estimation is also higher than that of the ALASSO and MALASSO. On the whole, for AR model the IALASSO estimator can select the correct signal variables, and estimate parameters more accurately.

V. Real Data Applications

In this section, we will apply a dataset to illustrate the performance of IALASSO estimator. We select the daily opening price data of the Chinese stock of Pingan Bank from May 13th, 2019 to April 2nd, 2022, and reserve the last 25 days as test data.

A. Perliminary processing of data

Before modeling, the time series is not stationary by stationarity test, so the log-transform and first-order difference of daily opening price is taken.

By ADF test on the series, it is found that the time series is stationary, and a corresponding model can be established. By observing ACF and PACF diagrams, we try to establish an ARIMA(5,1,0) for the series. By using the maximum likelihood method to estimate the parameters, we can get the following model

$$\widehat{\nabla lnY_t} = 0.0077\nabla lnY_{t-1} + 0.0227\nabla lnY_{t-2}$$
$$+ 0.0556\nabla lnY_{t-3} - 0.0294\nabla lnY_{t-4} - 0.1019\nabla lnY_{t-5}.$$

Now the residual sequence has no autocorrelation basically, and can be regarded as a white noise series by using the Ljung-Box test. So the model of ARIMA(5,1,0) can extract most of the information of the series. Before modelling the opening price series using the penalized methods, we denote

$$Y = (\nabla lnY_{p+1}, \nabla lnY_{p+2}, ..., \nabla lnY_T)^T,$$

$$X = \begin{bmatrix} \nabla lnY_p & \nabla lnY_{p-1} & ... & \nabla lnY_1 \\ \nabla lnY_{p+1} & \nabla lnY_p & ... & \nabla lnY_2 \\ ... & ... & ... & ... \\ \nabla lnY_{T-1} & \nabla lnY_{T-2} & ... & \nabla lnY_{T-p} \end{bmatrix}.$$

ALASSO, MALASSO and IALASSO are used to model by the BIC criterion and the obtained models are as follows

$$ALASSO : \widehat{\nabla lnY_t} = 0.0002789 - 0.0502181 \nabla lnY_{t-5}$$

$$MALASSO : \widehat{\nabla lnY_t} = 0.0002755 + 0.0165881 \nabla lnY_{t-3} - 0.0553417 \nabla lnY_{t-5}$$

$$IALASSO : \widehat{\nabla lnY_t} = 0.0002751 + 0.0078352 \nabla lnY_{t-3} - 0.0466015 \nabla lnY_{t-5}$$

There is no autocorrelation and partial autocorrelation in the residual sequence, and the Ljung Box test also shows that the residual is a white noise sequence.

B. Model prediction

By using the above models we predict the daily opening price of the next 25 days. For further comparison, we calculate the mean absolute percentage error (MAPE) of the four methods respectively

$$MAPE = (\sum_{i=1}^{T-p} (|\frac{Y_i - \hat{Y}_i}{Y_i}|) * 100\%)/(T-p).$$

TABLE V
COMPARISON OF THE PREDICTIVE ABILITY OF FOUR METHODS

MAPE	ARIMA	ALASSO	MALASSO	IALASSO
In-sample	0.7125%	0.3041%	0.3429%	0.2846%
Out-of-sample	1.4615%	0.5253%	0.5262%	0.5179%

From TABLE V the values of MAPE obtained by the four methods are very small. Among them, the in-sample and out-of-sample MAPE of IALASSO are 0.2846% and 0.5179% respectively, which is the smallest among four methods. Therefore the model established by the IALASSO method is generally better than the other methods.

VI. CONCLUDING REMARKS

Since the Adaptive LASSO model ignores the influence of the lag order for time series model, it is impossible to predict the time series data well. Therefore, we propose the IALASSO method for the AR model in this paper. The IALASSO estimator is to divide the middle term of lag order, which can put relatively heavier penalty to the coefficients whose lag order are larger than the middle term and put relatively lighter penalty to the coefficients whose lag order are less than the middle term. The IALASSO can ensure the front lag term enter the model with a high probability. The advantage of this improvement is that the compression of different coefficients can speed up the automatic adjustment, limit the size of the parameter space, and improve the convergence speed. So the IALASSO method can improve the accuracy and efficiency of variable selection and parameter estimation. In the parts of numerical simulation, we give several simulation studies to examine the asymptotic results, which show that the IALASSO method is feasible and efficient in variable selection and parameter estimation.

REFERENCES

[1] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, vol. 12, no. 1, pp55-67, 1970.
[2] I. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools (with discussion)", *Technometrics*, vol. 35, no. 12, pp109-148, 1993.
[3] J. Lv and Y. Fan, "A unified approach to model selection and sparse recovery using regularized least squares", *The Annals of Statistics*, vol. 37, no. 6A, pp3498-3528, 2009.
[4] C. Zhang, "Nearly unbiased variable selection under minimax concave penalty", *The Annals of Statistics*, vol. 38, no. 2, pp894-942, 2010.
[5] H. Akaike, "Maximum likelihood identification of Gaussian autoregressive moving average models", *Biometrika*, vol.60, no. 2, pp255-265, 1973.
[6] C. L. Mallows and N. J. A. Sloane, "An upper bound for self-dual codes", *Information and Control*, vol. 22, no. 2, pp188-200, 1973.
[7] G. Schwarz, "Estimating the dimension of a model", *The Annals of Statistics*, vol. 6, no. 2, pp461-464, 1978.
[8] P. Hingley and G. Dikta, "Use of information criteria for finding Box-Jenkins time series models for patent filings counts forecasts", *IAENG International Journal of Applied Mathematics*, vol. 52, no. 1, pp7-17, 2022.
[9] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp267-288, 1996.
[10] L. Breiman, "Better subset regression using the nonnegative garrote", *Technometrics*, vol. 37, no. 4, pp373-384, 1995.
[11] H. Wang, B. Li, and C. Leng, "Shrinkage tuning parameter selection with a diverging number of parameters", *Journal of the Royal Statistical Society: Series B*, vol. 71, no. 3, pp671-683, 2009.
[12] Y. Ming and Y. Lin, "Model selection and estimation in regression with grouped variables", *Journal of the Royal Statistical Society: Series B*, vol. 68, no. 1, pp49-67, 2006.
[13] C. Emmanuel and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n", *The Annals of Statistics*, vol. 35, no. 6, pp2313-2351, 2007.
[14] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty", *The Annals of statistics*, vol. 38, no. 2, pp894-942, 2010.
[15] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society: Series B*, vol. 67, no. 2, pp301-320, 2005.
[16] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties", *Journal of the American Statistical Association*, vol. 96, no. 456, pp1348-1360, 2001.
[17] Y. Nardi and A. Rinaldo, "Autoregressive process modeling via the lasso procedure", *Journal of Multivariate Analysis*, vol. 102, no. 3, pp528-549, 2011.
[18] J. X. Che, H. F. Xian, and Y. H. Zhang, "Adaptive hybrid optimized support vector regression with Lasso feature selection for short-term Load forecasting", *IAENG International Journal of Computer Science*, vol. 48, no. 4, pp1095-1107, 2021.
[19] M. Kohler and A. Krzyzak, "Nonparametric regression estimation using penalized least squares", *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp3054-3058, 2001.
[20] H. Zou, "The adaptive lasso and its oracle properties", *Journal of the American Statistical Association*, vol. 101, no. 476, pp1418-1429, 2006.
[21] G. C. Wang, B. T. Liang, and J. Z. Wang, "The modified adaptive Lasso method and its application in stock market", *Mathematical Statistics and Management*, vol. 38, no. 4, pp750-760, 2019.(in Chinese)

Wenbin Hao is a postgraduate student in Anhui University of Finance and Economics, China. His main research interests are in statistical modelling.

Yanling Zhu is a professor in Anhui University of Finance and Economics, China. Her main research interests are in mathematical statistics, econometrical theory and applications.