IRAQI STATISTICIANS JOURNAL
https://isj.edu.iq/index.php/isj
ISSN: 3007-1658 (Online)

Using Lasso Procedure for Variables Selection of Autoregressive Model for High Dimensional Time Series of Caenorhabditis Elegans Motion

Mohammed Khames Rasheed[1], and Osamah Basheer Shukur[2]

[1,2] Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

ARTICLE INFO

Article history:
Received 14/11/2024
Revised 16/11/2024
Accepted 15/1/2025
Available online 15/5/2025

Keywords:
High dimensional time series
Auto-regressive
Variable selection
LASSO
Caenorhabditis elegans motion

ABSTRACT

The Lasso is a common model for selection and also is a common estimation procedure for linear models. In this study, Lasso estimator will be used for obtaining more fitted autoregressive time series models. Simulation procedure has been used to generate a time series of the motion caenorhabditis elegans (CE represented by the tan-angles of wave-motion). Each observation of this time series is a recorded frame (0.5 second) of 2.5 hours video of CE motion. in this study, the real and simulated univariate time series of CE motion (tan-angles) are modelled via Lasso and autoregressive models (hybrid Lasso-AR approch) after multi-processes of variable selection. The results of simulated and real univariate time series reflect more fitted models after performing variables selection procedure. In conclusion, hybrid Lasso-AR approach can be used for best high dimensional time series modelling.

1. Introduction

Roundworms are usually used as model organisms in the study of genetics, including Caenorhabditis elegans (CE), since the movement of this worm is a useful indicator for understanding behavioral genetics. Datasets for the movement of this worm were obtained from the UEA&UCR time series classification archives that are available to the public. The data are decimal numbers that represent the movement of the worm on an agar plate containing bacterial food. This movement was determined by tracking using video clips. The length of the video is recorded as frames and time. Each movement is represented by a start frame, an end frame, and the time spent to make this movement until the next movement in the form of a set of time series values. Each time series is a video clip of the worm's movement with a length of approximately (144) minutes, where the data represent 6 dimensions for 5 strains of Caenorhabditis elegans (CE), and each worm represents a time series containing (17984) observations [1].

The lasso method can provide very good prediction accuracy, because reducing and removing coefficients can reduce variance without a significant increase in bias, and this is useful when we have a small number of observations and a large number of features and in the field of statistics they are variables and in time series they are autoregressive variables. Moreover, the lasso helps to increase the interpretability of the model by removing or eliminating irrelevant variables that are not associated with the response variable [2].

The selection of variables plays an important role in statistical modeling when

* Corresponding author.E-**mail** address: drosamahannon@uomosul.edu.iq
https://doi.org/10.62933/jxdyj297

there are a large number of variables of importance in the model among a large group of explanatory variables, where the goal of selection is to exclude variables that do not contain information related to the response variable, and thus improve the statistical models , and this can be seen not only in prediction, but also in interpreting the model and reducing computational operations [3].

Reduces the number of (predictive) explanatory variables to create the model. Variables selection algorithms search for a subset of predictive variables and determine which variables are required or excluded and the final size. The main benefits of variable selection are improved predictive, faster and more optimized predictions, and a better understanding of the data generation process. Using too many variables can lead to a deterioration in prediction performance even when all predictive variables have an effect on the response variable [4]. There are several previous studies dealing with the use of (AR and Lasso) method, where the researchers [5] studied the LASSO estimator to fit the autoregressive time series models, where they derived the conditions under which the LASSO estimators are estimated for the autoregressive coefficients. Also, the two researchers [6] developed a method to estimate the autoregressive model in the time series by reformulating the parameters of the autoregressive model by selecting the appropriate variables, where the LASSO procedure reduces the variables and estimates the model, as well as a simulation study was conducted to evaluate the performance and apply the results based on quarterly data of real gross national product of the United States for the period 1947-2009. The researchers [6, 7] also developed an inferential model for modeling high-dimensional time series using the extended LASSO method (under NEAR EPOCH Dependence NED). The two researchers [8] also studied the LASSO estimators to fit the time series and compared it with the OLS method. To this end, a large number of different time series were used.

In this study, the main objective is using Lasso estimator for obtaining more fitted autoregressive time series models. In addition to generate a time series of the motion caenorhabditis elegans by using simulation procedure (CE represented by the tan-angles of wave-motion).

2. Methodology

This section will deal with the general theoretical framework, autoregressive models, the LASSO procedure in selecting variables, the method for determining the tuning parameter, and the use of the Cross-Validation method to choose the best tuning parameter, as well as the standard used in measuring prediction error.

2.1 Framework of study

The framework for this study will include the following:

a. Assigning the suitable AR model.

b. Constructing three different LASSO models based on the lags AR model for 100, 500, 1000 variables.

c. Calculating the MSE measurement for AR and LASSO in of sample forecasts.

d. Comparing the accuracy results for AR and LASSO to determine which model would be provided better results and which non zero parameters will be included in each model.

2.2 Auto-regressive model

Time series are defined as a set of observations generated sequentially and in a specific chronological order, its main characteristic is that it is not independent, that is, it is linked in time and each observation depends on its predecessors, which generates an impetus to make predictions and future predictions based on the behavior of the series observations in the past [9-11].

Autoregressive can be used to express the value of the current time series using the linear regression function for the values of the previous time series. In general, the autoregressive p-order can be written as in the equations below.

$$\phi(B)x_t = a_t \qquad (1)$$

$$\Rightarrow (1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)x_t = a_t$$

$\Rightarrow x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + a_t$

where ϕ_k is the k^{th} auto-regressive parameter, x_{t-k} on x_t in the auto-regressive model, $k = 1,2,3,\ldots,p$, a_t is random error or white noise with mean zero and constant variable σ_a^2 $a_t \sim i.i.d.N(0,\sigma_a^2)$, and $\phi(B) = (1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ $W_t = (1-B)^d x_t$, W_t is after differencing to satisfy the stationarity of arithmetic mean and use W_t instead of x_t in the equations (1). With autoregressive models, the Box-Jenkins methodology is used with its four steps: identification, parameter estimation, diagnostic checking, and forecasting.

2.3 LASSO procedures

In recent years several penal methods for selecting variables and model estimation with time series data have been proposed. Accordingly, a set of penal methods have been proposed by adding a penal constraint to the Residuals sum of squares RSS. The goal of adding a penal constraint is to control the complexity of the model and provide a criterion for selecting a variable by introducing some restrictions on the parameters, and these restrictions impose some parameters that their values should be equal to zero[12]. This improves the prediction accuracy of the time series and provides an easy-to-interpret model [13]. However, the effectiveness of this system depends on the correct selection of the tuning Parameter in the penalty function. There are several ways to choose a tuning Parameter that are determined using an appropriate criterion. The desired determinant can be obtained by reducing this criterion around the tuning Parameter [14].

In 1996, the scientist (Tibshirany) proposed the LASSO penal function (Least Absolute Shrinkage and Selection Operator), as it is considered one of the most commonly used penalty methods. This method follows the estimated (L1-norm) by adding it to the residuals sum of squares, and it has become one of the basic penal methods for selecting variables due to its ability to reduce parameter values and select variables at the same time [15]. First of all, LASSO inputs are defined, which are denoted by x_i the time lags that represent the significant autoregressive variables from the AR(p) model. As for the target variable, it will represent the original series and symbolize it as y_i . Moreover, LASSO reduces the regression coefficients and makes them exactly zero. The LASSO estimator can be obtained with the minimized value of as in equation (2):

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(y_i - \beta_0 - x_i^T\beta\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right) \quad (2)$$

where β_j represents the vector of non-zero parameters and (λ) represents the tuning parameter when ($\lambda \geq 0$) as the penalty limit depends entirely on the value of (λ) which controls the shrinking (diminishing) of the parameter values, and N represents the number of observations and β_0, β the value is fixed values. When the value of ($\lambda = 0$), we get the least squares (OLS) estimates [15, 16] . On the contrary, when the value of (λ) increases, the number of variables excluded from the model increases [17]. The LASSO method can provide very good prediction accuracy in time series, because reducing and removing coefficients can reduce variance without a significant increase in bias, and this is useful when the number of features is large. Moreover, LASSO helps to increase the interpretability of the model by removing or eliminating irrelevant variables that are not associated with the response variable [18, 19].

2.3.1 Tuning Parameter Estimation

Accurate estimation of the tuning parameter λ is important because it greatly affects the performance of penal methods because it plays an important role in choosing variables, as its value determines the number of variables chosen in the model and the amount of bias imposed on the estimated regression coefficients [20, 21] .One of the most widely used methods for tuning parameter estimation is the BIC (Bayesian Information Criterion) and the Cross-Validation method CV. The model selection criterion AIC was defined by [22] such as follows.

$$BIC = 2f - 2\ln L(RSS)$$

where RSS is the residual sum of squares, and n is the number of observations [23]. RSS can be also changed to RMSE [24] and *f* is the number of the model's parameters [25, 26].

2.3.2 Cross-Validation (CV) Method

CV cross validation is a method of model selection by dividing the data (at least once). A part of the data (the training set) is used to train the algorithm, and the remaining part (the test set) is used to estimate the error of the algorithm and choose the model corresponding to the smallest estimated error. Therefore, CV is used to evaluate the prediction performance of the statistical learning model on out-of-sample data. This method ensures that the data used to train the model is independent of the test data set in which prediction performance is evaluated. The CV K-fold process means performing the CV process by dividing the data into K groups and using one of them for testing, while the remaining groups (K-1) are used as training data. In this way we obtain several different estimates of the prediction error and choose the least of these errors for optimization.

CV is used in data analysis to validate the implemented models where the main objective is to predict and estimate the prediction performance of the statistical machine learning model. In other words, CV assesses how good a statistical machine is [27]. For the purpose of employing this method (CV) in estimating the tuning parameter (λ) the prediction error rate will be calculated for each of the imposed values. This method can be represented mathematically as in equation (4):

$$k - cv_{(\lambda)} = \frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_{i(\lambda)}^{-k(i)})^2 \quad (4)$$

Where ($\hat{y}_{i(\lambda)}^{-k(i)}$) represents the appropriate response variable when the observation (i) belongs to the investigation data as long as it is the constant value of (λ). Since there is more than one value for the tuning parameter, the best value corresponding to the smallest rate of prediction error will be chosen [21]. In the following as in equation (5):

$$\lambda_{optimal} = \underset{r=1,2,\ldots,R}{\operatorname{argmin}} \ \mathrm{k - CV}_{(\lambda_r)} \quad (5)$$

2.3.3 Mean Squared Error (MSE) measurement

MSE is imputed for the error of methods as a statistical criterion to evaluate the accuracy of these methods. The MSE is written such as follows [28]

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(e_i)^2 \quad (6)$$

where e_i is the forecasting error, n is the number of observations.

3. Results and discussion

3.1 Data used in the study

The data that was used in the research represents the movement of the (CE) warm, as the movement is according to its speed. The speed is recorded passively when a part of the body moves towards the tail (as opposed to the head). The speed is defined as the distance between the two midpoints of the start and end frame. Divided by the time between both frames, the average speed of the worm is at least equal to 5% of its length per second in each frame, where the worm must maintain this speed continuously, so the worm compensates for the delay caused by stopping by increasing its speed by increasing the sharpness of the angles of movement until reaching the required rate of speed, where the movement of the worm was recorded (the movement is forward or backward) with time as a time series of movement represented by shadow angles of the wave movement, each observation of this time series is a recorded frame (0.5) seconds from a (2.5) hour video for the CE movement, every second the worm moves at a speed of (5%) of its length, and we must maintain this speed almost continuously, with interruptions allowed at most (0.25) seconds, which may generate contradictory movements during and after stopping, such as withdrawal of the head, contractions in the body, and noise in the mouth. move movement Its different [1]. Based on the foregoing, it is possible to benefit from the motor behavior of the worm through the relationship of the angle with the speed, so when the angle is acute, the speed will be greater, and when the angle is obtuse, that is, when the angle is positive, the movement will be slower. Where the data represent 6 dimensions for 5 species of CE worm genes, and each worm represents a time series containing (17984) observations or variables. A sample of two time series (each series representing an independent worm movement) was drawn from the same strain, the reference strain (N2), which is the strain of interest in most studies.

3.2 Auto-Regressive Model

After selecting the two series randomly from the data, we drew these two series using the (Minitab) program. The functions (ACF) and (PACF) were drawn for the two series to determine the significant parameters and entered them into the AR model to obtain the values of the significant parameters, which represent a number of values that we will use in building a model the (AR). The figure below shows the functions ACF and PACF for the first sample of the original series.

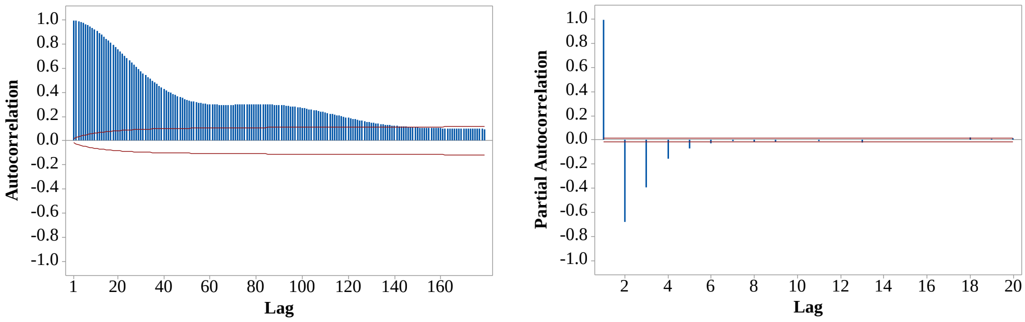

Figure 1. ACF and PACF for the time series of sample 1.

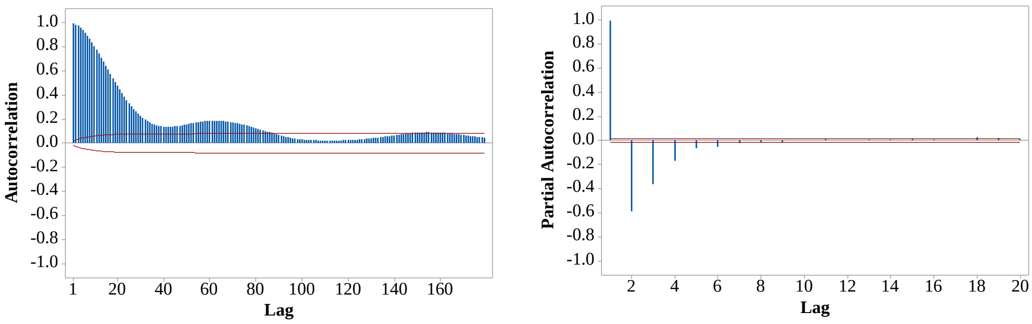

Figure 2. ACF and PACF for the time series of sample 2.

We note from the previous figures, Figure 1 and Figure 2, that the best models, regardless of the stability of the time series, are AR (6) for both samples, after noting the significant of the first six time lags of the ACF function in both figures. Information criteria also have been used to satisfying AR(6) is the most fitted model for studied datasets.

As for the estimates of the AR (6) model for both samples, it is not possible to obtain them with high accuracy through traditional statistical programs, so it was resorted to matlab and y is the target variable that represents the original series. When applying several instructions, the MSE of the model was calculated and the X and Y were obtained, and then we obtained the LASSO estimator, and the cross validation method was used to determine the values of the tuning parameter.

After applying the LASSO directive to the data and to all the different (AR) values, which contain several variables (p = 100, p = 500, p = 1000), the results were obtained. The following table, Table 1, shows the values of the significant parameters for the first and second samples when the regression values are Self AR(6).

Table 1: Significant parameters of AR(6) model for sample 1 and sample 2

p.	AR(6) sample 1 Parameters	AR(6) sample 2 Parameters
1	- 1.338	-1.294
2	0.0297	0.02602
3	0.1884	0.1439
4	0.0666	0.08593
5	0.0283	0.0036
6	0.0281	0.05235

3.3 LASSO Model.

The general framework of the LASSO model implementation algorithm includes the implementation of several sequential steps as follows.

1- Using autoregressive variables with p number represented by time lags based on Table 1 to determine the input variables for the LASSO model.
2- Inputting the target variable, which is the original time series.
3- Building the best LASSO model using time series data for the input and target variables, as well as selecting the cross validation (CV) method with the default K-Fold when K = 10. Using the directive (lasso(x,y,'CV',10)) in Matlab program when the number of variables is 100, 500, and 1000, respectively.
4- Use the model in the previous step to forecasting of the data after setting the default values for the parameters.
5- Calculate the accuracy of the LASSO method for forecasting by means of the MSE scale.

Non-zero parameter values and forecasting accuracy were measured by MSE. Table 2 shows the variables of the first sample when (AR(6)) which were selected using the LASSO method for different (p) values. When using the LASSO method for the first sample when the value of (p = 100) we notice that the variables that were chosen are (x_1,x_5,x_6,x_7,x_8,x_9,x_{17}) which are the important variables according to this method and the remaining variables were excluded. It can be compared with the significant AR (6) variables as in Table (1), namely (x_1,x_2,x_3,x_4,x_5,x_6). By comparison, it is clear that the LASSO method excluded the significant variables (x_2,x_3 x_4) in Table (1), and also chose the variables (x_7,x_8,x_9,x_{17}) as important variables, while they were not significant in AR (6) in Table (1).

When using the LASSO method for the first sample when the value is (p = 500), we notice that the variables that were chosen are (x_1,x_5,x_6,x_7,x_8,x_9,$x_{17,\dots,}$,x_{422}) which are the important variables according to this method as The remaining variables were excluded. It can be compared with the significant AR (6) variables as in Table (1), namely (x_1,x_2,x_3,x_4,x_5,x_6). By comparison, it is clear that the LASSO method excluded the significant variables (x_2,x_3 x_4) in Table (1), and also chose the variables (x_7,x_8,x_9,$x_{17,\dots,}$,x_{422}) as important variables, while they were not significant in AR (6) in Table (1).

Table 2: The parameters of LASSO model for sample 1

I	P=100 Parameters	i	P=500 Parameters	i	P=1000 Parameters
x_1	1.12842	x_1	1.12872	x_1	1.12144
x_5	0.02876-	x_5	-0.02864	x_5	-0.01739
x_6	-0.04892	x_6	-0.05111	x_6	-0.04745
x_7	-0.02175	x_7	-0.02206	x_7	-0.02336
x_8	-0.01740	x_8	-0.01672	x_8	-0.02006
x_9	-0.01174	x_9	-0.01023	x_9	-0.01339
x_{17}	-0.00724	x_{17}	-0.00754	x_{16}	-0.00165
⋮	⋮	⋮	⋮	⋮	⋮
x53	0.00078	*X422*	0.00005	*X756*	-0.00046
MSE	0.09680	*MSE*	0.09740	*MSE*	0.09830

When using the LASSO method for the first sample when the value is (p = 1000), we note that the values that were chosen are (x_1,x_5,x_6,x_7,x_8,x_9,$x_{16,\dots,}$,x_{756}) which are the important variables according to this method as The remaining variables were excluded. It can be compared with the significant AR (6) variables as in Table (1), namely (x_1,x_2,x_3,x_4,x_5,x_6). By comparison, it is clear that the LASSO method excluded the significant variables (x_2,x_3 x_4) in Table (1), and also chose the variables (x_7,x_8,x_9,$x_{16,\dots}$,x_{756}) as important variables, while they were not significant in AR (6) in Table (1). Table 1 shows the variables of the second sample when (AR(6)) which were selected using the LASSO method for different (p) values

Table 3: The parameters of LASSO model for sample 2

i	P=100 Parameters	i	P=500 Parameters	i	P=1000 Parameters
x_1	1.201547303	x_1	1.174394173	x_1	1.168376691
x_3	-0.034020129	x_3	-0.00000998	x_4	-0.081672128
x_4	-0.099200217	x_4	-0.092005648	x_5	-0.028521723
x_5	-0.022422905	x_5	-0.027381626	x_6	-0.049359947
x_6	-0.039569996	x_7	-0.045734031	x_7	-0.004227875
x_7	-0.001807311	x_9	-0.011552096	x_9	-0.010158665
x_9	-0.009778673	x_{10}	-0.00685326	x_{10}	-0.00571434
⋮	⋮	⋮	⋮	⋮	⋮
x_{84}	-9.82E-05	x_{484}	-0.000106263	x_{905}	-0.00036975
MSE	0.14871	MSE	0.15003	MSE	0.14905

when using the LASSO method for the second sample when the value is (p = 100), we note that the values that were chosen are ($x_1,x_3,x_4,x_5,x_6,x_7,,x_{9,\dots},x_{84}$) which are the important variables according to this method and the remaining variables were excluded. It can be compared with the significant AR (6) variables as in Table (1), namely (x_1,x_2,x_3,x_4,x_5,x_6). By comparison, it is clear that the LASSO method excluded the significant variable (x_2) in Table (1), and also chose the variables ($x_7,x_{9,\dots},x_{84}$) as an important variables, while they were not significant in AR (6) in Table (1).

when using the LASSO method for the second sample when the value was (p = 500), we note that the values that were chosen are($x_1,x_3,x_4,x_5,x_7,,x_{9,\dots},x_{484}$) which are the important variables according to this method and the remaining variables were excluded. It can be compared with the significant AR (6) variables as in Table (1), namely (x_1,x_2,x_3,x_4,x_5,x_6). By comparison, it is clear that the LASSO method excluded the significant variables (x2, x6) in Table (1) and also chose the variables ($x_7,x_{9,\dots},x_{484}$) as an important variables, while they were not significant in AR (6) in Table (1).

when using the LASSO method for the second sample when the value is (p = 1000), we note that the values that were chosen are($x_1,x_4,x_5,x_6,x_7,,x_{9,\dots},x_{905}$) which are the important variables according to this method and the remaining variables were excluded. It can be compared with the significant AR (6) variables as in Table (1), namely (x_1,x_2,x_3,x_4,x_5,x_6). By comparison, it is clear that the LASSO method excluded the significant variables (x_2,x_3) in Table (1) and also chose the variables ($x_7,x_{9,\dots},x_{905}$) as an important variables, while they were not significant in AR (6) in Table (1).

4. Conclusions

In this study, the LASSO method was used as a proposed method to improve the selection of optimal non-zero parameters among a large number of parameters, which may be difficult to accommodate a traditional method such as AR when the data are for the time series of CE with very close and very accurate forecasting results for both LASSO and AR methods. Two samples of data were used and the results showed superiority with high agreement on important parameters of the LASSO method over AR as a traditional method to modeling the time series and forecasting it. The MSE criterion was used to indicate the quality of the forecasting, which reflected a great convergence in the accuracy of the forecasting, despite the different parameters chosen by both the proposed and traditional methods. The LASSO method reflects absolute and true preference with different numbers of autoregressive variables, as it chooses the best variables that actually reflect the data of the study. It is possible to conclude the possibility of using LASSO as an optimal method with high-dimensional time series data for one type of roundworms , which carries as a very large number of observations represented as autoregressive variables.

References

[1]Yemini, E., Jucikas, T., Grundy, L.J., Brown, A.E.X., and Schafer, W.R. (2013). *A database of*

Caenorhabditis elegans behavioral phenotypes. Nature Methods, **10**(9): p. 877-879.

[2] Adamek, R., Smeekes, S., and Wilms, I. (2022). *Lasso inference for high-dimensional time series.* Journal of Econometrics.

[3] Zhang, Y., Li, R., and Tsai, C.-L. (2010). *Regularization parameter selections via generalized information criterion.* Journal of the American Statistical Association, **105**(489): p. 312-323.

[4] Guyon, I. and Elisseeff, A. (2003). *An introduction to variable and feature selection.* Journal of machine learning research, **3**(Mar): p. 1157-1182.

[5] Nardi, Y. and Rinaldo, A. (2011). *Autoregressive process modeling via the lasso procedure.* Journal of Multivariate Analysis, **102**(3): p. 528-549.

[6] Chan, N.H., Yau, C.Y., and Zhang, R.-M. (2015). *LASSO estimation of threshold autoregressive models.* Journal of Econometrics, **189**(2): p. 285-296.

[7] Adamek, R., Smeekes, S., and Wilms, I. (2023). *Lasso inference for high-dimensional time series.* Journal of Econometrics, **235**(2): p. 1114-1143.

[8] Iturbide, E., Cerda, J., and Graff, M. (2013). *A comparison between LARS and LASSO for initialising the time-series forecasting auto-regressive equations.* Procedia Technology, **7**: p. 282-288.

[9] Brockwell, P.J. and Davis, R.A. (2009). *Time series: theory and methods*. Springer science & business media.

[10] Liu, L.-M. (2006). *Time Series Analysis and Forecasting*. 2nd ed. Illinois, USA: Scientific Computing Associates Corp.

[11] Shukur, O.B. (2015). *Artifical Neural Network and Kalman Filter Approaches Based on ARIMA for Daily Wind Speed Forecasting*, Universiti Teknologi Malaysia.

[12] Muhammadullah, S., Urooj, A., Khan, F., Alshahrani, M.N., Alqawba, M., and Al-Marzouki, S. (2022). *Comparison of weighted lag adaptive LASSO with Autometrics for covariate selection and forecasting using time-series data.* Complexity, **2022**: p. 1-10.

[13] Müller, I.M. (2021). *Feature selection for energy system modeling: Identification of relevant time series information.* Energy and AI, **4**: p. 100057.

[14] Fahrmeir, L., Kneib, T., Lang, S., Marx, B., Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Generalized linear models.* Regression: Models, methods and applications: p. 269-324.

[15] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1): p. 267-288.

[16] Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1): p. 1.

[17] Wood, S.N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. Biometrics, 62(4): p. 1025-1036.

[18] Freijeiro-González, L., Febrero-Bande, M., and González-Manteiga, W. (2022). A critical review of LASSO and its derivatives for variable selection under dependence among covariates. International Statistical Review, 90(1): p. 118-145.

[19] Wang, T. (2021). A combined model for short-term wind speed forecasting based on empirical mode decomposition, feature selection, support vector regression and cross-validated lasso. PeerJ Computer Science, 7: p. e732.

[20] Androulakis, E., Koukouvinos, C., and Mylona, K. (2011). Tuning parameter estimation in penalized least squares methodology. Communications in Statistics-Simulation and Computation, 40(9): p. 1444-1457.

[21] Li, Y., Dicker, L., and Zhao, S.D. (2014). The Dantzig selector for censored linear regression models. Statistica Sinica, 24(1): p. 251.

[22] Hirotugu, A. (1973). Information theory and an extension of the maximum likelihood principle, in Selected Papers of Hirotugu Akaike, Springer. p. 199-213.

[23] Burnham, K.P. and Anderson, D.R. (2002). Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.

[24] Dawson, C. and Wilby, R. (2001). Hydrological modelling using artificial neural networks. Progress in physical Geography, 25(1): p. 80-108.

[25] Billah, B., Hyndman, R.J., and Koehler, A.B. (2005). Empirical information criteria for time series forecasting model selection. Journal of Statistical Computation and Simulation, 75(10): p. 831-840.

[26] Bozdogan, H. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity. Journal of Mathematical Psychology, 44(1): p. 62-91.

[27] Zhang, Y. and Yang, Y. (2015). Cross-validation for selecting a model selection procedure. Journal of Econometrics, 187(1): p. 95-112.

[28] Hyndman, R.J. and Koehler, A.B. (2006). Another look at measures of forecast accuracy. International journal of forecasting, 22(4): p. 679-688.