Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 21, Issue 1(2019) ISSN$_{Print}$: 1816-9171

Comparison of some penalized methods in time series models

Tahir R. Dikheel Alaa Q. Yaseen
Department of Statistics,
College of Administration and Economics, University of Al-Qadisiyah, Iraq.
Corresponding Author: Tahir R. Dikheel
Alaa Q. Yaseen

Abstract **:** *In this paper , the lag weighted lasso (lwlasso) is compared with lasso and alasso to deal with effect of lag in linear time series models. The lwlasso methods are more stable than the other methods in the comparison. Consequently, lag weighted lasso methods are capable to dealing with lag effect . In particular, the lag weighted lasso methods with w^R2 and w^R3 weights gave the best results compared with the other methods.*

INTRODUCTION

Time series y_t is a set of observations arranged according to their occurrence in time such as years, seasons, months, days. ..etc. Therefore, it is a historical record that is adopted to build future expectations. There are two time series types (the discrete time series and the continuous time series). The time series is either to be non stationary in the mean and can be converted into stationary one using differences or is non_stationary in variance and can be converted to stationary using transformations. The stationary, time series can be divided into two common types, stationary of the second order and strictly stationary. Time series is said to be stationary of the second order if the first and second moment are known and the mean of the series is constant independent of time and the covariance depends only on the lag k:

$$Cov(y_t, y_{t+k}) = \gamma_k \qquad (1)$$

γ_k is the covariance function

The time series is said to be strictly stationary if for each integer $k \geq 1$, and for any partial set of time $t_1, t_2, \dots, t_k$ and the $y_{t_1}, y_{t_2}, \dots, y_{t_k}$ common distribution is constant by time difference. This means that for any positive integer k and for any integer l:

$$F(y_{t_{1+l}}, y_{t_{2+l}}, \dots, y_{t_{k+l}}) = F(y_{t_1}, y_{t_2}, \dots, y_{t_k}) \qquad (2)$$

F is the distribution function

ARMA models

Stationary time series models were introduced by Yule in (1926), he studied the autoregressive model AR (q). Wlker in (1931) introduced the general model of the autoregressive models. Stutzky (1937) studied models of moving average MA(s) and put it's general formula. Then completed his way to find the model in a mixed and complete way by Wold in (1938), where he developed these two models with a series of operations into three directions in the estimation procedure and called it the autoregressive_ moving average (ARMA) models. This model is used if data is stationary.

ARMA models are mathematical formulas that represent the continuity pattern of the phenomenon, or the type of correlation between the time series and itself. It is widely used in many sciences such as economics, geography,

aviation, agricultural sciences, physical systems and other fields. Models can help us to understand how the system works by detecting the properties associated with this system.

ARMA models can be described through a series of equations; these equations are easier if the mean time series is zero by subtracting the sample mean. Therefore, the modified series is treated with the sample mean, meaning that:

$$y_t = Y_t - \bar{Y} \qquad (3)$$

where

Y_t represents the time series.

y_t represents the modified time series.

$\bar{Y}$ represents the mean sample

Thus the ARMA model can be written as:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-q} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-s} \qquad (4)$$

where
q:represent the order of autoregressive model
s:represent the order of moving average model
β:represent the autoregressive model parameter
θ:represent the moving average model parameter
a_t:represent the random error

q and s can be found by looking at the form of the autocorrelation (ACF) function and the partial autocorrelation (PACF) function. When the autocorrelations decay exponentially to zero, this means that the model is an AR model and its order is determined by a number of PACF that
are significantly different from zero. If the PACF decay exponentially to zero, the model is an MA model whose order is determined by the number of ACF with statistical significance. If ACF and PACF decay to zero exponentially, this ARMA model is determined by AR & MA. See the ACF and PACF function, if the ACF function does not give up quickly with increasing degrees of delay, it means that the time series is non stationary, and you need to take the differences. Summed up the diagnostic process through the following :

Table(1): Determine models order according to the behavior of ACF and PACF for the stationary time series

Model	ACF	$PACF$
$AR(q)$	-Exponential decay, if $q \geq 1$	-Spike at lag 1,no correlation for other lags, if q =1; -Spikes at lags 1 to p, no correlation for other lags if $q \geq 2$.
$MA(s)$	-Spike at lag 1,no correlation for other lags, if s =1; -Spikes at lags 1 to p, no correlation for other lags if $s \geq 2$.	-Exponential decay, if $s \geq 1$
$ARMA(q,s)$	-Exponential decay, if $q \geq 1$	-Exponential decay, if $s \geq 1$

the problem here in the case, of mixed models, to determine q and s. Because the ACF and PACF functions in this case behave in a similar manner. There are several criteria have been developed to compare the models in the selection process of the model order. Selecting a order lower than the actual order of the model results in the inconsistency of the model parameters While choosing a order higher than the actual order in the model increases the variation of the model, This leads to loss of accuracy. There are several criteria use to selecting the model order such as: "Akaike information criterion" (AIC), "Bayesian Information Criterion" (BIC) and etc.

Pe Penalized least square

(PLS) methods are a convenient and common method to deal with high-dimensional data, especially when the number of explanatory variables are greater than the sample size. it is not possible to use the ordinary least square method.

PLS methods are used to overcome computational problems in high-dimensional data as well as improve prediction accuracy (Darwish & Buyuklu 2015).nalized Least Square Method(PLS)

PLS methods are based on the principle of minimizing the SSE with some limitations on parameters. The estimates of the least square are obtained by reducing the objective function, which consists of two parts, the loss function and the penalty function. Which are in accordance with the following formula: (Flexeder 2010)

$$p_{LS}(\lambda,\beta) = (y - X\beta\grave{)}(y - X\beta) + p(\lambda,\beta) \qquad (5)$$

where

$p(\cdot)$: penalty function.

λ : penalty parameter.

Accordingly, the penalized estimator is obtained according to the following formula:

$$\hat{\beta} = \text{argmin}_{\beta}\{p_{LS}(\lambda,\beta)\} \qquad (6)$$

PLS methods are the process of estimation and variables selection in the same time (Mylona & Goos 2010).

The good penalty function gives an estimator of three characteristics, including (Fan& Li 2001).

1-Unbiasedness: PLS estimator should be unbiased or almost unbiased when the real anonymous parameter is large.

2-Sparsity: PLS estimator should be the threshold rule, which sets estimates with small coefficients to zero.

3-Continuity: PLS estimator should be a continuous function in the data, meaning that little change in the data does not lead to a significant change in the estimates.

Fan & Li also stated that ideally estimator has the characteristics of Oracle Properties, meaning:

1- Probability Consistency The real model is one when ($T \to \infty$). Where T is the sample size. This property is called Sparsity, i.e.:

$$\lim_{T\to\infty} P_r\left(\hat{\theta}_{An} = \theta\right) = 1 \qquad (7)$$

Where (*An*) refer to the estimator has an asymptotical normal distribution

2-The estimator has an asymptotical normal distribution as in the case of the Oracle estimator, i.e.:

$$\sqrt{T}\left(\hat{\theta}_{An} - \theta\right) \sim N(O, \sum) \qquad (8)$$

1- Least absolute shrinkage and selection operator (lasso)

In (1996), Tibshirani suggested a lasso function, an abbreviation for "least absolute shrinkage and selection operator", to estimate linear model parameters and variable selection together. The main idea of the lasso method is to

minimized sum square residuals, plus a constraint representing the absolute sum of the coefficients. For the linear model, the lasso estimator of the ARMA model is obtained according to the following formula: (Tibshirani 1996)

$$\hat{\beta}_{lasso} = argmin_{\beta}\left\{\|y_t - X\beta\|^2 + \lambda\sum_{l=0}^{q}|\beta_l|\right\} \qquad (9)$$

Where

λ: Penalty Parameter or Regularization Parameter

$\lambda\sum_{l=0}^{q}|\beta_l|$: penalty function or regularization function.

lasso is often used in practice because the L_1 penalty limit allows the coefficients to be reduced to exactly zero **(**Konzen & A. Ziegelmann 2013**,** Nardi & Rinaldo,2011**).**

2- adaptive lasso (alasso)

The lasso estimator as in previous studies may be inefficient and that the results of choosing the real model may be inconsistent (Fan & Li, 2001; Yuan & Lin, 2007; Zou, 2006). To deal with these problems, Zou (2006) suggested alasso. The alasso method has assigns different penalty limits for each coefficient based on weights. These penalty limits have reflected the size of the parameter for each variable, and alasso is able to determine the correct consistent model and efficient coefficients.

$$\beta_{\mathbf{alasso}} = \arg\min_{\beta}\left\{\|y_t - X_t\beta\|^2 + \lambda\sum_{l=0}^{q} w_l|\beta_l|\right\} \qquad (10)$$

Where, $\widehat{w}_l = \frac{1}{|\hat{\beta}_l|^{\gamma}}$, γ > 0 and $\hat{\beta}$ refers to OLS estimators. (Zou 2006, Audrino & Camponovo 2013)

3- The lag weighted lasso (lwlasso)

The alasso method can determine the correct form in regression models. However, it can not calculate the lag effect period, which is necessary for a time series model. Thus, it can not reflect the properties of the time series model. To improve the prediction accuracy of the time series model, Park and Sakaori (2013) suggested the (lwlasso) method. The lwlasso has imposed

different penalty limits for each coefficient on the basis of weights that reflect coefficients size and also the lag effect period, the coefficient vector can be measured as follows:

$$\beta_{lwlasso} = arg\ min_{\beta}\{\|y_t - x_t\beta\|^2 + \lambda\sum_{l=0}^{q} w_l|\beta_l| \qquad (11)$$

where y_t is the time series, x_t is the matrix of explanatory variables,β is the regression coefficient vector, λ is the tuning parameter and w is a weighted function.

Equation (9) depends on the following three types of weight: (Park and Sakaori, 2013)

$$\widehat{w}_l^{(1)} = \frac{1}{[\alpha(1-\alpha)^l]^{\gamma}} \qquad (12)$$

$$\widehat{w}_l^{(2)} = \frac{1}{\alpha(1-\alpha)^l\left[|\beta_l|\right]^{\gamma}} \qquad (13)$$

$$\widehat{w}_l^{(3)} = \frac{1}{\left[\alpha(1-\alpha)^l|\beta_l|\right]^\gamma} \qquad (14)$$

The relative prediction error (RPE) is used to compare the methods forecast accuracy:

$$RPE = E[(\hat{y}_t - y_t)^2]/(\hat{\sigma})^2\,, \qquad (15)$$

Real world data: temperature

A real data is used to compare the lwlasso with the alasso and lasso. The data is monthly mortality of temperature from April 2004 to September 2015. This data was collected from Iraqi medical center in Diwaniyah city, Iraq. the LARS algorithm is used (Efron et al. 2004), and the OLS estimators β, to analyze this data and compare lwlasso with the alasso and lasso methods. As follow:

Table (2):RPE values for each method

RPE	lwlasso with w^1	lwlasso with w^2	lwlasso with w^3	Lasso	Adlasso
	1.143095	1.088992	1.080726	1.144328	1.143055

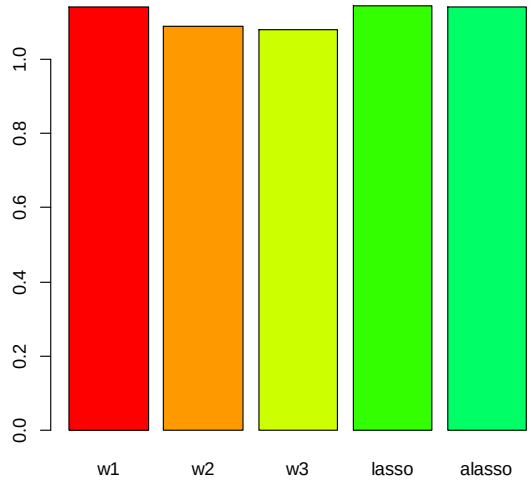

Fig (1): RPE for each method

It was noted in Table (2) and Fig (1) that the efficiency for alasso and lasso methods decreased. The good results of the lwlasso methods were observed. The lwlasso method with w^R3 is the best with ⟦RPE⟧^w value of 1.080726 followed by the lwlasso method with w^R2 at 1.088992 and then the lwlasso method with w^R1 at 1.143095.

Conclusions

The results in the section of real data showed that the lag weight lasso with w3 outperforms than both the lag weight lasso with w1 and lag weight lasso with w2 in real data. For these studies, the lwlasso with w3 enables improving the accuracy of forecast.

References

Audrino, F., & Camponovo, L. (2013). Oracle properties and finite sample inference of the adaptive lasso for time series regression models. arXiv preprint arXiv:1312.1473.

Chang, I.H., Tiao, G.C. and C. Chen (1988). Estimation of Time Series Parameters in the Presence of Outliers. Technometrics, 30, 193-204.

Darwish, K., & Buyuklu, A. H. (2015). Robust Linear Regression Using L1-Penalized MM-Estimation for High Dimensional Data. American Journal of Theoretical and Applied Statistics, 4(3), 78-84.

Fan, J., & Li, R. (2001). Variable selection via non concave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.

Flexeder, C. (2010). Generalized lasso regularization for regression models (Doctoral dissertation, Institut für Statistik).

Liu, Z. Z. (2014). The doubly adaptive LASSO methods for time series analysis.

Medeiros, M. C., & Mendes, E. F. (2015). l1-estimation of high-dimensional time-series models with flexible innovations. Working Paper.

Mylona, K., & Goos, P. (2010). Penalized Generalized Least Squares for Model Selection under Restricted Randomization

Nardi, Y., &Rinaldo, A. (2011).Autoregressive process modeling via the lasso procedure. Journal of Multivariate Analysis, 102(3), 528-549.

Park, H., Sakaori, F. (2013). Lag weighted lasso for time series model. Comp Stat, 28, 493-504.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.