

6-17-2021

Robust Lag Weighted Lasso for Time Series Model

Tahir R. Dikheel

University of Al-Qadisiyah, Iraq, tahir.dikheel@qu.edu.iq

Alaa Q. Yaseen

University of Al-Qadisiyah, Iraq, alaakyasen@gmail.com



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Dikheel, Tahir R. and Yaseen, Alaa Q. (2021) "Robust Lag Weighted Lasso for Time Series Model," *Journal of Modern Applied Statistical Methods*: Vol. 19 : Iss. 1 , Article 14.

DOI: 10.22237/jmasm/1608553500

Robust Lag Weighted Lasso for Time Series Model

Tahir R. Dikheel
University of Al-Qadisiyah
Al Diwaniyah, Iraq

Alaa Q. Yaseen
University of Al-Qadisiyah
Al Diwaniyah, Iraq

The lag-weighted lasso was introduced to deal with lag effects when identifying the true model in time series. This method depends on weights to reflect both the coefficient size and the lag effects. However, the lag weighted lasso is not robust. To overcome this problem, we propose robust lag weighted lasso methods. Both the simulation study and the real data example show that the proposed methods outperform the other existing methods.

Keywords: Lag-weighted lasso, robust lag-weighted lasso, adaptive lasso, time series model, lag effect

Introduction

Variable selection plays a significant role in building a time series model. This technique provides shrinkage for good estimation parameters, good production, and identification of the relevant variables. The statistical procedures for selecting variables are characterized by the provision of interpretable models. Variable selection methods such as stepwise and best subset selection may suffer from a lack of stability. To overcome this problem, Tibshirani (1996) proposed the least absolute shrinkage and selection operator (lasso). This method provides shrinkage coefficients toward zero and makes some coefficients exactly zero and thus tries to keep the relevant variables with strong influences. Studies have found that the lasso estimator is sometimes inefficient, and the results of the variables selection are inconsistent (Fan & Li, 2001; Yuan & Lin, 2007; Zou, 2006).

To overcome this problem, Zou (2006) proposed the adaptive least absolute shrinkage and selection operator (alasso), which penalizes different regression coefficients by different weights. These penalties reflect the size of the coefficient to define the correct model in the regression. However, these methods cannot deal

with the lag effect, which is a base stone in the time series models. Recently, alasso has been used in time series (see, for example, [Nardi & Rinaldo, 2011](#); [Chen & Chan, 2011](#); [Liu, 2014](#); [Medeiros & Mendes, 2015](#)). However, the alasso in time series cannot reflect certain properties of a time series model, such as the lag effect. To improve the accuracy of a time series model, Park and Sakaori ([2013](#)) proposed a lag-weighted lasso (Lwlasso). The Lwlasso imposes different penalties on each coefficient based on weights that reflect not only coefficient size but also the lag effects. Under the Lwlasso, the regression coefficient vector can be estimated as follows:

$$\boldsymbol{\beta}_{\text{Lwlasso}} = \arg \min_{\boldsymbol{\beta}} \left\{ \|y_t - \mathbf{x}_t \boldsymbol{\beta}\|^2 \right\} + \lambda \sum_{j=0}^p \sum_{l=0}^{q_j} w_{j,l} |\boldsymbol{\beta}_{j,l}|, \quad (1)$$

where y_t is the time series, $\mathbf{x}_t$ is the vector of explanatory variables, $\boldsymbol{\beta}$ is the regression coefficient vector, λ is the tuning parameter, and w is a weighted function.

Equation (1) depends on the following three types of weight:

$$\hat{w}_{j,l}^1 = \frac{1}{\left[\alpha (1-\alpha)^l \right]^{\gamma}}, \quad (2)$$

$$\hat{w}_{j,l}^2 = \frac{1}{\alpha (1-\alpha)^l \left[\|\boldsymbol{\beta}_{j,l}\| \right]^{\gamma}}, \quad (3)$$

$$\hat{w}_{j,l}^3 = \frac{1}{\left[\alpha (1-\alpha)^l |\boldsymbol{\beta}_{j,l}| \right]^{\gamma}}. \quad (4)$$

Unfortunately, the Lwlasso method is sensitive to outliers because it depends on OLS which is not robust to outliers in the observations. Robust regression such as Huber's criterion ([Lambert-Lacroix & Zwald, 2011](#)), LAD ([Wang et al., 2007](#)), and quantile regression ([Wu & Liu, 2009](#)) has been used recently in variable selection. In this paper, we propose a robust lag-weighted lasso method by replacing the quadratic loss function with a Huber function.

Time Series Models

There are several models that can be used to describe time series, such as autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA). The AR model is the most common model in time series because most phenomena follow it in practice. Furthermore, the errors in MA models have non-linear functions of the parameters, so iterative estimation methods are needed to minimize the residual sum of squares (Chatfield, 2004).

Consider the series y_t ; the model of AR(q) can be written with response lag variables and a disturbance error term as:

$$y_t = a + \sum_{l=0}^q \beta_l L^l y_t + e_t, \quad (5)$$

where a is the intercept, β_l is the l^{th} regression coefficient for $l = 0, 1, 2, \dots, q$, e_t refers to a white noise term with zero mean, constant variance σ^2 , and $\text{cov}(e_t, y_{t-1}) = 0$ for all $l \neq 0$, and L represents the lag operator (i.e., $L^0 y_t = y_t$, $L^1 y_t = y_{t-1}$). Sometimes the past series y_{t-1} cannot describe the present value y_t . To overcome this problem and to improve the forecast accuracy, more explanatory variables $\mathbf{x}_{j,t-1}$ ($j = 1, \dots, p$ and $l = 0, \dots, q_j$) are added (Pesaran & Shin, 1997). Specifically, Pesaran and Shin (1997) proposed an autoregressive distributed lag (ADL) model with response lag variables, current and lagged explanatory variables. The ADL($q_0, q_1, q_2, \dots, q_p$) model can be formulated as

$$y_t = a + \sum_{l=0}^{q_0} \beta_{0,l} L^l y_t + \sum_{l=0}^{q_1} \beta_{1,l} L^l \mathbf{x}_{1,t} + \dots + \sum_{l=0}^{q_p} \beta_{p,l} L^l \mathbf{x}_{p,t} e_t. \quad (6)$$

Equivalently,

$$y_t = a + \sum_{j=0}^p \sum_{l=0}^{q_j} \beta_{j,l} L^l \mathbf{x}_{j,t} + e_t \quad (7)$$

In the above model, $\beta_{0,0} = 0$ and $\mathbf{x}_{0,t} = y_t$ are assumed. In addition, Equation (3) satisfies the following assumptions (Park & Sakaori, 2013):

1. $E(e_t | y_{t-1}, y_{t-2}, \dots, \mathbf{x}_{1,t}, \mathbf{x}_{1,t-1}, \dots, \mathbf{x}_{p,t-1}, \mathbf{x}_{p,t-2}, \dots) = 0$.
2. $(y_t, \mathbf{x}_{1,t}, \dots, \mathbf{x}_{p,t})$ are stationary.

3. The correlation coefficients between $(y_t, \mathbf{x}_{1,t}, \dots, \mathbf{x}_{p,t})$ and $(y_{t-l}, \mathbf{x}_{1,t-l}, \dots, \mathbf{x}_{p,t-l})$ decline as l increases.

Penalized Robust Methods for Rime Series

Outliers in time series are more complex than in the regression due to the presence of time effects in the data. There are different types of outliers in time series, such as isolated outliers, patchy outliers, and level shifts in mean (Maronna et al., 2006). Another classification was made for outliers, divided into outliers and structure changes: additive outlier (AO) and innovational outlier (IO). The presence of these extraordinary values affects the time series analysis and their results.

The main objective of penalized robust methods is to provide accurate results rather than penalized least squares methods in the presence of outliers. To reach this goal, deleting the outliers, limiting the influence of outliers by the robust method of reducing the weight of outliers, changing the value of outliers, and penalized robust estimation techniques are used. The general formulation of the penalized robust methods in the time series is as follows (Fan & Li, 2001):

$$\sum_{t=1}^n \rho \left(y_t - \sum_{j=0}^p \sum_{l=0}^{q_j} \beta_{j,l} L^l \mathbf{x}_{j,t} \right) + p_{\lambda}^R \left(\left| \beta_{j,l} \right| \right), \quad (8)$$

where $\rho(\cdot)$ is the Huber loss function (Huber, 1981) defined by

$$\rho(u) = \begin{cases} u^2 & \text{if } u \leq k \\ 2k|u| - k^2 & \text{if } u > k \end{cases} \quad (9)$$

with tuning constant k , and $p_{\lambda}^R \left(\left| \beta_{j,l} \right| \right)$ is the penalty function. Here, R refers to the robust proposed penalty.

The efficiency properties, as well as the breakdown point (BP), are used as a measure to determine the effectiveness of penalized robust methods. The BP is a measure of the resistance of an estimator when the data have a large ratio of contamination. The least squares estimator has BP as low as $1/n$, meaning that even a single outlying observation can turn out an estimator of OLS to be useless. In contrast, there are some estimators that have a high BP of approximately 50%.

M-Lag Weighted Lasso (Mlwlasso) Method

The least squares method with penalty function in Lwlasso is unprotected because it is affected by the abnormal values. Therefore, in this paper, we replace the loss function in Lwlasso with the M robust loss function (which has a BP of about 0.5), then we replace the penalty function in Lwlasso with the M robust penalty function, to obtain the Mlwlasso method, which is defined as (Park & Sakaori, 2013; Maronna et al., 2006):

$$\hat{\boldsymbol{\beta}}_{\text{Mlwlasso}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho\left(\frac{e_t}{s}\right) + p_{\lambda}^{\text{R}}\left(\left|\boldsymbol{\beta}_{j,l}\right|\right), \quad (10)$$

or, equivalently,

$$\hat{\boldsymbol{\beta}}_{\text{Mlwlasso}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho\left(\frac{y_t - \sum_{j=0}^p \sum_{l=0}^{q_j} \boldsymbol{\beta}_{j,l} L^l \mathbf{x}_{j,t}}{s}\right) + p_{\lambda}^{\text{R}}\left(\left|\boldsymbol{\beta}_{j,l}\right|\right), \quad (11)$$

where s is robustly determined according to the following formula:

$$s = \frac{MAD}{0.6745} = \frac{\text{median}\left|e_t - \text{median}(e_t)\right|}{0.6745} \quad (12)$$

and $p_{\lambda}^{\text{R}}\left(\left|\boldsymbol{\beta}_{j,l}\right|\right)$ is the Mlwlasso penalty function with the following form:

$$p_{\lambda}^{\text{R}}\left(\left|\boldsymbol{\beta}_{j,l}\right|\right) = \lambda \sum_{j=0}^p \sum_{l=0}^{q_j} \hat{w}_{j,l}^{\text{R}} \left|\boldsymbol{\beta}_{j,l}\right|. \quad (13)$$

As the three weight types, (2), (3), and (4), reported in Park and Sakaori (2013) are not robust, we propose the following robust weights:

$$\hat{w}_{j,l}^{\text{R1}} = \frac{1}{\left[\alpha(1-\alpha)^l\right]^{\gamma}}, \quad (14)$$

$$\hat{w}_{j,l}^{R2} = \frac{1}{\alpha(1-\alpha)^l \left[\left| \hat{\beta}_{j,l}^R \right| \right]^\gamma}, \quad (15)$$

$$\hat{w}_{j,l}^{R3} = \frac{1}{\left[\alpha(1-\alpha)^l \left| \hat{\beta}_{j,l}^R \right| \right]^\gamma}. \quad (16)$$

where $\hat{\beta}_{j,l}^R$ is the $(j, l)^{\text{th}}$ element in $\hat{\beta}^R$, which can be estimated as follows:

$$\hat{\beta}^R = \arg \min_{\beta} \left\{ \sum_{i=1}^n \rho \left(y_t - \sum_{j=0}^p \sum_{l=0}^{q_j} \beta_{j,l} L^l \mathbf{x}_{j,t} \right) \right\}. \quad (17)$$

To summarize, the proposed Mlwllasso estimators and associated robust weights are listed below:

$$\begin{aligned} w^{R1}, \quad \hat{\beta}_{\text{Mlwllasso}} &= \arg \min_{\beta} \sum_{i=1}^n \rho \left(y_t - \sum_{j=0}^p \sum_{l=0}^{q_j} \beta_{j,l} L^l \mathbf{x}_{j,t} \right) + \lambda \sum_{j=0}^p \sum_{l=0}^{q_j} \hat{w}_{j,l}^{R1} |\beta_{j,l}| \\ w^{R2}, \quad \hat{\beta}_{\text{Mlwllasso}} &= \arg \min_{\beta} \sum_{i=1}^n \rho \left(y_t - \sum_{j=0}^p \sum_{l=0}^{q_j} \beta_{j,l} L^l \mathbf{x}_{j,t} \right) + \lambda \sum_{j=0}^p \sum_{l=0}^{q_j} \hat{w}_{j,l}^{R2} |\beta_{j,l}| \\ w^{R3}, \quad \hat{\beta}_{\text{Mlwllasso}} &= \arg \min_{\beta} \sum_{i=1}^n \rho \left(y_t - \sum_{j=0}^p \sum_{l=0}^{q_j} \beta_{j,l} L^l \mathbf{x}_{j,t} \right) + \lambda \sum_{j=0}^p \sum_{l=0}^{q_j} \hat{w}_{j,l}^{R3} |\beta_{j,l}| \end{aligned}$$

In practice, the choice of tuning parameters is important. k -fold cross validation is used widely in selecting the tuning parameters (Bengio & Grandvalet, 2004; Rodriguez et al., 2010). Here, we find optimal tuning parameters (a, γ, λ) by 10-fold cross validation. Then, we compare the forecast accuracy of each method based on the weight's relative prediction error (RPE^w):

$$RPE^w = \frac{E \left[\left(\hat{y}_t - y_t^w \right)^2 \right]}{\left(\hat{\sigma}^w \right)^2} \quad (18)$$

where

$$\sigma^w = \frac{\sum_{t=1}^n (y_t^w - \bar{y}_t)^2}{n-1}, y_t^w = y_t * h \left(y_t - \sum_{j=0}^p \sum_{l=0}^{q_j} \beta_{j,l} L^l \mathbf{x}_{j,t} \right),$$

and $h(\cdot)$ is the Huber function defined by

$$h(u) = \begin{cases} 1 & \text{if } |u| \leq k \\ \frac{k}{|u|} & \text{if } |u| > k \end{cases}$$

The algorithm of proposed methods is summarized by the following steps:

- Step 1: Compute $\hat{\beta}^R$ as in Equation (17).
- Step 2: Depending on the $\hat{\beta}^R$ we find the weights $\hat{w}_{j,l}^{R1}$, $\hat{w}_{j,l}^{R2}$, and $\hat{w}_{j,l}^{R3}$.
- Step 3: Define $\mathbf{x}_{j,t-1}^* = \mathbf{x}_{j,t-1} / \hat{w}_{j,l}^R$.
- Step 4: Solve the problem for all λ as in Equation (11).
- Step 5: Output $\beta_{j,t-1} = \beta_{j,t-1}^* / \hat{w}_{j,l}^R$.
- Step 6: Compute RPE^w as in Equation (18)

Simulation

To compare the Mlwllasso, Madlasso, and Mlasso methods with Lwllasso, lasso, and alasso, consider the following ADL(5, 3, 3) model:

$$y_t = a + \sum_{l=0}^5 \beta_{0,l} L^l y_t + \sum_{l=0}^3 \beta_{1,l} L^l \mathbf{x}_{1,t} + \sum_{l=0}^3 \beta_{2,l} L^l \mathbf{x}_{2,t} + e_t \quad (19)$$

Generate a time series with 100 observations when $\sigma = 1, 3, 5$ in each situation. Shown in Table 1 are the true model settings for the simulation study.

Case 1

Table 1. True parameters for the simulation study Case 1

y_{t-l}					$\mathbf{x}_{1,t-l}$				$\mathbf{x}_{2,t-l}$			
$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{0,3}$	$\beta_{0,4}$	$\beta_{0,5}$	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$	$\beta_{2,3}$
0.3	-0.2	0.1	0.0	0.0	0.9	0.7	0.5	0.0	1.0	-0.7	0.5	0.0

The optimal set of tuning parameters (a, γ, λ) was selected by 10-fold cross validation. Compute the median RPE^w for each method with 1000 replications.

Table 2. RPE^w for methods when $\sigma = 1$

Outlier rate	Dist.	lwlasso			lasso	alasso	Mlwlasso			Mlasso	Malasso
		$\hat{w}^1$	$\hat{w}^2$	$\hat{w}^3$			$\hat{w}^{R1}$	$\hat{w}^{R2}$	$\hat{w}^{R3}$		
0%		3.981	3.433	3.549	4.004	4.005	3.207	3.109	3.262	3.507	3.979
5%	N(0,6)	3.959	3.858	3.920	3.991	3.948	3.194	3.131	3.137	3.332	3.256
	N(0,10)	4.014	3.873	3.948	4.043	3.971	3.242	3.134	3.276	3.266	3.237
	t(1)	3.998	3.916	3.980	4.024	3.998	3.215	3.180	3.315	3.230	3.254
	t(5)	3.982	3.937	4.014	4.011	4.029	3.218	3.149	3.284	3.243	3.231
10%	N(0,6)	4.501	4.380	4.473	4.535	4.974	3.349	3.278	3.339	3.379	3.353
	N(0,10)	4.549	4.514	4.593	4.581	4.619	3.279	3.208	3.350	3.305	3.275
	t(1)	4.084	4.010	4.082	4.109	4.102	3.136	3.053	3.248	3.159	3.127
	t(5)	3.930	3.851	3.920	3.962	3.937	3.137	3.086	3.220	3.162	3.147
15%	N(0,6)	4.987	4.918	4.993	5.322	5.308	3.303	3.209	3.301	3.323	3.296
	N(0,10)	5.090	5.050	5.134	5.113	5.154	3.309	3.240	3.470	3.324	3.324
	t(1)	4.393	4.353	4.421	4.422	4.449	3.153	3.091	3.325	3.183	3.169
	t(5)	3.899	3.808	3.876	3.929	3.900	3.117	3.035	3.197	3.142	3.127
20%	N(0,6)	5.166	5.107	5.187	5.184	5.204	3.665	3.582	3.647	3.945	3.736
	N(0,10)	5.639	5.577	5.671	5.670	5.700	3.358	3.352	3.541	3.373	3.425
	t(1)	4.612	4.580	4.647	4.635	4.675	3.122	3.045	3.323	3.155	3.125
	t(5)	3.837	3.724	3.807	3.873	3.833	3.050	2.957	3.100	3.075	3.053

Table 3. RPE^w for methods when $\sigma = 3$

Outlier rate	Dist.	lwlasso			lasso	alasso	Mlwlasso			Mlasso	Malasso
		$\hat{w}^1$	$\hat{w}^2$	$\hat{w}^3$			$\hat{w}^{R1}$	$\hat{w}^{R2}$	$\hat{w}^{R3}$		
0%		3.923	3.870	3.882	4.033	4.975	3.236	3.129	3.230	3.628	3.658
5%	N(0,6)	3.942	3.884	3.969	3.983	3.979	3.161	3.109	3.167	3.474	3.970
	N(0,10)	3.961	3.886	3.959	3.998	3.969	3.163	3.118	3.264	3.187	3.189
	t(1)	3.989	3.871	3.965	4.016	3.997	3.216	3.125	3.280	3.248	3.217
	t(5)	3.969	3.853	3.921	4.000	3.944	3.180	3.096	3.240	3.201	3.193
10%	N(0,6)	5.779	5.765	5.822	5.811	5.856	3.350	3.230	3.237	3.393	3.380
	N(0,10)	6.416	6.408	6.476	6.443	6.502	3.224	3.147	3.435	3.244	3.219
	t(1)	8.888	8.905	8.982	8.925	8.994	3.206	3.146	3.537	3.230	3.220
	t(5)	4.047	3.952	4.030	4.078	4.047	3.180	3.160	3.292	3.209	3.219
15%	N(0,6)	10.694	10.765	10.811	10.740	10.837	3.291	3.232	3.276	3.946	3.939
	N(0,10)	14.274	14.389	14.372	14.279	14.375	3.259	3.184	3.783	3.276	3.265
	t(1)	5.568	5.610	5.694	5.617	5.747	3.189	3.118	3.487	3.215	3.208
	t(5)	4.085	3.983	4.069	4.117	4.089	3.257	3.182	3.290	3.279	3.256
20%	N(0,6)	17.134	17.217	17.290	17.163	17.288	3.397	3.306	3.321	3.422	3.463
	N(0,10)	20.008	20.022	20.042	20.026	20.055	3.246	3.187	3.853	3.270	3.260
	t(1)	8.067	8.172	8.206	8.088	8.216	3.242	3.166	3.704	3.266	3.254
	t(5)	4.126	3.996	4.077	4.156	4.103	3.207	3.131	3.301	3.231	3.210

ROBUST LAG WEIGHTED LASSO FOR TIME SERIES MODEL

Table 4. RPE^w for methods when $\sigma = 5$

Outlier rate	Dist.	lwllasso			lasso	alasso	Mlwlasso			Mlasso	Malasso
		$\hat{w}^1$	$\hat{w}^2$	$\hat{w}^3$			$\hat{w}^{R1}$	$\hat{w}^{R2}$	$\hat{w}^{R3}$		
0%		3.910	3.362	3.400	4.032	3.977	3.194	3.105	3.152	3.953	3.859
5%	N(0,6)	3.983	3.874	3.952	4.006	4.974	3.214	3.139	3.211	3.238	3.209
	N(0,10)	4.023	3.940	4.013	4.046	4.039	3.236	3.184	3.350	3.262	3.257
	t(1)	3.949	3.847	3.910	3.981	3.934	3.218	3.111	3.229	3.240	3.201
	t(5)	3.995	3.921	3.973	4.019	3.985	3.227	3.143	3.279	3.252	3.231
10%	N(0,6)	4.002	3.904	3.969	4.038	4.993	3.213	3.107	3.211	3.238	3.299
	N(0,10)	4.199	4.107	4.181	4.229	4.209	3.250	3.209	3.364	3.277	3.296
	t(1)	4.091	3.986	4.059	4.114	4.079	3.156	3.131	3.284	3.179	3.194
	t(5)	3.948	3.812	3.876	3.972	3.903	3.157	3.051	3.195	3.181	3.139
15%	N(0,6)	4.062	3.985	4.060	4.660	4.849	3.272	3.194	3.237	3.296	3.276
	N(0,10)	4.323	4.235	4.301	4.348	4.323	3.283	3.205	3.366	3.308	3.287
	t(1)	4.173	4.098	4.164	4.197	4.181	3.098	3.030	3.211	3.118	3.102
	t(5)	3.893	3.808	3.861	3.923	3.882	3.080	2.981	3.126	3.096	3.056
20%	N(0,6)	4.139	4.058	4.137	4.171	4.158	3.305	3.228	3.300	3.321	3.322
	N(0,10)	4.507	4.457	4.533	4.527	4.551	3.392	3.341	3.472	3.412	3.406
	t(1)	4.186	4.118	4.181	4.212	4.200	3.062	2.996	3.179	3.082	3.069
	t(5)	3.794	3.715	3.777	3.831	3.804	2.988	2.901	3.046	3.008	2.977

Shown in Tables 2, 3, and 4 are results of the ordinary methods are very close to our proposed robust methods when $\sigma = 1, 3, 5$. However, when contaminating the data at different rates and distributions (normal and t), our proposed methods are more stable than the ordinary methods based on the values of RPE^w across deferent error distributions.

Case 2

Select the optimal set of tuning parameters (a, γ, λ) by 10-fold cross validation. Compute the median RPE^w for each method with 1000 replications.

Table 5. True parameters for the simulation study Case 2

y_{t-l}					$x_{1,t-l}$				$x_{2,t-l}$			
$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{0,3}$	$\beta_{0,4}$	$\beta_{0,5}$	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$	$\beta_{2,3}$
0.3	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	1.0	0.0	0.0	0.0

Table 6. RPE^w for methods when $\sigma = 1$

Outlier rate	Dist.	lwlasso			lasso	alasso	Mlwlasso			Mlasso	Malasso
		$\hat{w}^1$	$\hat{w}^2$	$\hat{w}^3$			$\hat{w}^{R1}$	$\hat{w}^{R2}$	$\hat{w}^{R3}$		
0%		0.483	0.625	0.357	0.365	0.979	0.304	0.264	0.269	0.349	0.406
5%	N(0,6)	1.322	1.324	1.325	1.360	1.980	0.267	0.265	0.266	1.048	0.972
	N(0,10)	3.478	3.621	3.354	3.364	3.983	0.300	0.263	0.266	0.343	0.407
	t(1)	1.486	1.617	1.356	1.366	1.979	0.299	0.263	0.267	0.343	0.407
	t(5)	0.854	0.621	0.553	0.633	0.978	0.299	0.264	0.265	0.344	0.409
10%	N(0,6)	1.718	1.772	1.688	1.611	1.985	0.298	0.201	0.269	1.047	0.977
	N(0,10)	4.869	3.881	4.757	4.764	4.000	0.336	0.297	0.301	0.383	0.439
	t(1)	2.587	2.699	2.448	2.484	2.987	0.314	0.276	0.279	0.360	0.417
	t(5)	0.914	0.628	0.643	0.733	0.984	0.305	0.266	0.268	0.349	0.407
15%	N(0,6)	2.827	2.849	2.779	2.722	2.992	0.332	0.323	0.329	1.048	0.973
	N(0,10)	4.952	4.951	4.860	4.860	5.011	0.375	0.333	0.338	0.422	0.469
	t(1)	3.707	3.774	3.579	3.636	4.004	0.330	0.289	0.291	0.377	0.427
	t(5)	1.000	0.932	0.968	0.980	0.983	0.308	0.269	0.271	0.350	0.411
20%	N(0,6)	3.887	3.899	3.835	3.784	4.000	0.365	0.360	0.361	1.047	0.968
	N(0,10)	6.983	6.979	6.902	7.899	7.021	0.427	0.374	0.379	0.475	0.516
	t(1)	5.825	5.849	5.698	5.753	7.004	0.341	0.300	0.303	0.387	0.438
	t(5)	0.504	0.632	0.678	0.689	0.982	0.313	0.274	0.275	0.355	0.414

Table 7. RPE^w for methods when $\sigma = 3$

Outlier rate	Dist.	lwlasso			lasso	alasso	Mlwlasso			Mlasso	Malasso
		$\hat{w}^1$	$\hat{w}^2$	$\hat{w}^3$			$\hat{w}^{R1}$	$\hat{w}^{R2}$	$\hat{w}^{R3}$		
0%		0.483	0.318	0.356	0.365	0.976	0.302	0.264	0.266	0.344	0.407
5%	N(0,6)	0.646	0.679	0.600	0.852	0.979	0.308	0.305	0.307	1.048	0.967
	N(0,10)	1.584	1.718	1.754	1.764	1.978	0.298	0.262	0.265	0.344	0.405
	t(1)	0.682	0.620	0.653	0.636	0.979	0.300	0.265	0.268	0.344	0.408
	t(5)	0.480	0.617	0.540	0.520	0.981	0.303	0.266	0.269	0.344	0.406
10%	N(0,6)	1.514	1.648	1.515	1.396	1.983	0.277	0.204	0.275	1.047	0.974
	N(0,10)	1.701	1.789	1.868	1.961	1.983	0.327	0.286	0.291	0.371	0.423
	t(1)	0.502	0.639	0.370	0.388	0.983	0.297	0.261	0.265	0.340	0.405
	t(5)	0.473	0.618	0.468	0.956	0.977	0.292	0.252	0.255	0.337	0.399
15%	N(0,6)	1.947	1.693	1.741	1.943	2.978	0.297	0.293	0.295	1.047	0.970
	N(0,10)	2.644	2.735	2.953	2.543	2.984	0.354	0.312	0.316	0.401	0.461
	t(1)	0.526	0.655	0.934	0.916	0.984	0.292	0.255	0.258	0.334	0.400
	t(5)	0.465	0.610	0.522	0.643	0.981	0.276	0.243	0.246	0.319	0.384
20%	N(0,6)	2.656	2.742	2.633	2.548	2.982	0.373	0.353	0.369	1.049	0.971
	N(0,10)	2.899	2.889	2.550	3.475	2.998	0.519	0.229	0.242	0.744	0.701
	t(1)	1.900	1.895	1.526	1.526	0.997	0.475	0.183	0.198	0.728	0.681
	t(5)	1.877	1.884	1.481	1.387	1.992	0.470	0.167	0.182	0.729	0.692

ROBUST LAG WEIGHTED LASSO FOR TIME SERIES MODEL

Table 8. RPE^w for methods when $\sigma = 5$

Outlier rate	Dist.	lwlasso			lasso	alasso	Mlwlasso			Mlasso	Malasso
		$\hat{w}^1$	$\hat{w}^2$	$\hat{w}^3$			$\hat{w}^{R1}$	$\hat{w}^{R2}$	$\hat{w}^{R3}$		
0%		0.903	0.888	0.494	0.988	1.028	0.500	0.188	0.203	0.752	0.679
5%	N(0,6)	0.790	0.618	0.784	1.362	0.979	0.263	0.259	0.263	1.047	0.972
	N(0,10)	0.901	0.876	0.951	1.425	1.029	0.494	0.188	0.205	0.742	0.661
	t(1)	0.902	0.882	0.909	1.429	1.034	0.488	0.193	0.207	0.731	0.656
	t(5)	0.906	0.880	0.892	1.420	1.032	0.480	0.188	0.201	0.746	0.663
10%	N(0,6)	0.875	0.733	0.977	1.368	0.993	0.471	0.200	0.206	0.702	0.697
	N(0,10)	0.991	0.896	0.571	1.425	1.029	0.494	0.188	0.205	0.742	0.661
	t(1)	0.992	0.892	0.585	1.429	1.034	0.488	0.193	0.207	0.731	0.656
	t(5)	0.996	0.900	0.592	1.420	1.032	0.480	0.188	0.201	0.746	0.663
15%	N(0,6)	0.864	0.762	0.984	1.377	0.990	0.275	0.175	0.272	1.047	0.967
	N(0,10)	0.900	0.889	1.093	1.449	1.025	0.505	0.202	0.218	0.746	0.672
	t(1)	0.912	0.890	1.022	1.452	1.038	0.482	0.184	0.198	0.752	0.666
	t(5)	0.894	0.873	0.485	1.409	1.021	0.475	0.180	0.196	0.727	0.654
20%	N(0,6)	0.851	0.764	0.852	1.403	0.984	0.217	0.215	0.216	0.747	0.673
	N(0,10)	0.904	0.886	0.953	1.467	1.027	0.501	0.214	0.229	0.751	0.670
	t(1)	0.902	0.879	0.951	1.459	1.030	0.475	0.180	0.196	0.722	0.654
	t(5)	0.896	0.882	0.923	1.414	1.029	0.475	0.172	0.189	0.725	0.664

Shown in Tables 6, 7, and 8 are the results of the ordinary methods are very close to our proposed robust methods when $\sigma = 1, 3, 5$. However, when contaminating the data at different rates and distributions (normal and t), our proposed methods are more stable than the ordinary methods based on the values of RPE^w across deferent error distributions. When polluting the data with different pollutant rates, the RPE values for the ordinary methods increase the percentage of pollution and are greatly affected by the increased variance of polluted data when polluting by natural distribution or the higher the degree of freedom when polluting the by t distribution. The penalized robust methods remain resistant to the outlier, the greater the percentage of pollution. The Mlwlasso method with w^2 is the best in most experiments, followed by the preference of the Mlwlasso with w^3 and Mlwlasso with w^1 according to the lowest value of RPE .

Lung Cancer Data

The performance methods are illustrated using lung cancer data. These data were collected by the authors from an Iraqi medical center in Diwaniyah City, Iraq. They represent monthly numbers of people with lung cancer in Diwaniyah City from April 2004 to September 2015. These data consist of one response variable (lung

cancer) and 6 chemical water pollutants (Temp, Turb, PH, EC, Alk, TH), as shown in Table 9.

Table 9. Chemical water pollutants

Variables	
x_1	Temperature (Temp)
x_2	Turbidity (Turb)
x_3	Power of hydrogen (PH)
x_4	Electric conductivity (EC)
x_5	Alkalinity (Alk)
x_6	Total hardness (TH)

Table 10. RPE^w for the methods

lwlasso			lasso	alasso	Mlwlasso			Mlasso	Malasso
$\hat{w}^1$	$\hat{w}^2$	$\hat{w}^3$			$\hat{w}^{R1}$	$\hat{w}^{R2}$	$\hat{w}^{R3}$		
1.055	0.998	0.872	1.184	1.135	0.673	0.565	0.407	0.843	0.727

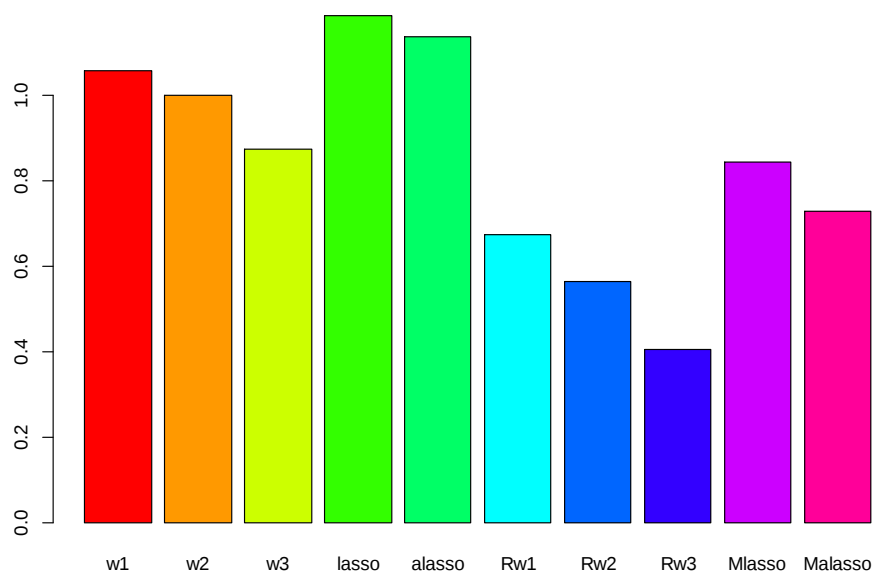


Figure 1. RPE^w values for the real data

Table 10 shows the relative prediction error (RPE^w) for the proposed methods and the other methods in comparison. From Table 10, note the ordinary methods are significantly affected, while the results of the proposed robust methods are stable. The proposed robust methods with $\hat{w}^{R3}$ give the best results because they have the least RPE . Figure 1 graphically shows the values of RPE^w for the methods. The proposed methods have the smallest values of RPE^w compared with the other methods.

Conclusion

An M-lag weighted lasso was proposed to deal with contamination data. The proposed methods were illustrated using a simulation study and a real data example. The results show the proposed methods are more stable than the other methods in the comparison. Consequently, these proposed methods are capable to dealing with outlier. In particular, the method Mlwllasso with $\hat{w}^{R2}$ and $\hat{w}^{R3}$ weights gave the best results compared with the others. The proposed methods can be extended to other methods such as, robust lag-weighted elastic net, robust lag-weighted group lasso, robust lag-weighted fused lasso, robust lag-weighted graphical lasso, and so on. Furthermore, the proposed methods can be used for a fixed lag effect, and with a varying lag effects across time.

References

- Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k -fold cross-validation. *Journal of Machine Learning Research*, 5(Sep), 1089-1105. Retrieved from <https://www.jmlr.org/papers/v5/grandvalet04a.html>
- Chatfield, C. (2004). *The analysis of time series: An introduction* (6th edition). Boca Raton, FL: CRC Press.
- Chen, K., & Chan, A. H. S. (2011). A review of technology acceptance by older adults. *Gerontechnology*, 10(1), 1-12. doi: 10.4017/gt.2011.10.01.006.00
- 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. doi: 10.1198/016214501753382273
- Huber, P. J. (1981). *Robust statistics*. New York: John Wiley & Sons.
- Lambert-Lacroix, S., & Zwald, L. (2011). Robust regression through the Huber's criterion and adaptive lasso penalty. *Electronic Journal of Statistics*, 5, 1015-1053. doi: 10.1214/11-EJS635

- Liu, Z. Z. (2014). *The doubly adaptive LASSO methods for time series analysis* [Unpublished doctoral thesis]. University of Western Ontario.
- Maronna, R. A., Martin, R. D., & Yohai, V. (2006). *Robust statistics: Theory and methods*. Chichester, UK: John Wiley & Sons.
- Medeiros, M. C., & Mendes, E. F. (2015). *ℓ_1 -regularization of high-dimensional time-series models with flexible innovations* [Working paper]. Pontifical Catholic University of Rio de Janeiro.
- Nardi, Y., & Rinaldo, A. (2011). Autoregressive process modeling via the lasso procedure. *Journal of Multivariate Analysis*, 102(3), 528-549. doi: [10.1016/j.jmva.2010.10.012](https://doi.org/10.1016/j.jmva.2010.10.012)
- Park, H., & Sakaori, F. (2013). Lag weighted lasso for time series model. *Computational Statistics*, 28(2), 493-504. doi: [10.1007/s00180-012-0313-5](https://doi.org/10.1007/s00180-012-0313-5)
- Pesaran, M. H., & Shin, Y. (1999). An autoregressive distributed-lag modelling approach to cointegration analysis. In S. Strøm (Ed.), *Econometrics and economic theory in the 20th century: The Ragnar Frisch centennial symposium* (pp. 371-413). Cambridge, UK: Cambridge University Press. doi: [10.1017/CCOL521633230.011](https://doi.org/10.1017/CCOL521633230.011)
- Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569-575. doi: [10.1109/TPAMI.2009.187](https://doi.org/10.1109/TPAMI.2009.187)
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288. doi: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x)
- Wang, H., Li, G., & Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics*, 25(3), 347-355. doi: [10.1198/073500106000000251](https://doi.org/10.1198/073500106000000251)
- Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19(2), 801-817.
- Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1), 19-35. doi: [10.1093/biomet/asm018](https://doi.org/10.1093/biomet/asm018)
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418-1429. doi: [10.1198/016214506000000735](https://doi.org/10.1198/016214506000000735)