Robust Penalized Methods for the Estimation of ARMA Models

Eleni Dretaki, supervised by Yiannis Kamarianakis

Department of Mathematics and Applied Mathematics, University of Crete

FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS

Introduction

ARMA models play a pivotal role in modern time-series analysis. A major challenge in this field is the computational complexity, especially when dealing with high true ARMA orders and periodic processes. To tackle this issue, we identify optimal subset ARMA models by employing adaptive Lasso regression [6] and variants, including adaptive Elastic Net and LAD Lasso. The above mentioned penalized estimators, are compared against stepwise model building procedures [4], in a series of simulation experiments, which mimic the scenarios presented in [2]. Finally, we conduct simulation experiments to evaluate, a Maximum Entropy Bootstrap Lasso scheme which is a new promising model selection method, [5].

ARMA Models

Assume that $\{\mathbf{Y_t} : \mathbf{t} = \mathbf{1}, \mathbf{2}, \dots, \mathbf{T}\}$ is a discrete and equally-spaced sample from a weakly stationary, homoskedastic process Y_t and the collection from a Gaussian white noise process of equal size $\{\varepsilon_\mathbf{t} : \mathbf{t} = \mathbf{1}, \mathbf{2}, \dots, \mathbf{T}\}$, then an ARMA model, is formulated as:

$$\mathbf{Y_t} = \phi_\mathbf{1}\mathbf{Y_{t-1}} + \cdots + \phi_\mathbf{p}\mathbf{Y_{t-p}} + \theta_\mathbf{1}\varepsilon_\mathbf{t-1} + \cdots + \theta_\mathbf{q}\varepsilon_\mathbf{t-q} + \varepsilon_\mathbf{t} \qquad (1)$$

Finite order parameters $\mathbf{p}, \mathbf{q} \in \mathbb{N}$ quantify the strength that past information has on prediction. Another expression of an ARMA model is through the use of the backshift operator $\mathbf{B}$, where $\mathbf{B^k Y_t} = \mathbf{Y_{t-k}}$. An ARMA$(\mathbf{p}, \mathbf{q})$ is expressed as:

$$\mathbf{\Phi_p}(\mathbf{B})\mathbf{Y_t} = \mathbf{\Theta_q}(\mathbf{B})\varepsilon_\mathbf{t}. \qquad (2)$$

where $\mathbf{\Phi_p}(\mathbf{B}) = \mathbf{1} - \phi_\mathbf{1}\mathbf{B} - \cdots - \phi_\mathbf{p}\mathbf{B^p}$ is the autoregressive operator and $\mathbf{\Theta_q}(\mathbf{B}) = \mathbf{1} + \theta_\mathbf{1}\mathbf{B} + \cdots + \theta_\mathbf{q}\mathbf{B^q}$ denotes the moving average operator.

Subset ARMA Selection

This work focuses on regularization methods for subset ARMA selection and estimation. The ARMA$(\mathbf{p}, \mathbf{q})$ model can be expressed as a conventional regression $\mathbf{y} = \mathbf{X}\beta + \varepsilon$; let $\mathbf{y} = [\mathbf{y_m}, \dots, \mathbf{y_T}]^\mathbf{T}$ denote the time series of interest and $\varepsilon = [\varepsilon_\mathbf{m}, \dots, \varepsilon_\mathbf{T}]^\mathbf{T}$ the residuals of an initial autoregressive model. The explanatory part of the model is based on the matrix $\mathbf{X}$, which is shown below.

$$\mathbf{X} = \begin{bmatrix} x'_m \\ \vdots \\ x'_T \end{bmatrix} = \begin{bmatrix} y_{m-1} & \cdots & y_{m-p} & \hat{\varepsilon_{m-1}} & \cdots & \hat{\varepsilon_{m-q}} \\ y_m & \cdots & y_{m-p+1} & \hat{\varepsilon_m} & \cdots & \hat{\varepsilon_{m-q+1}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{T-1} & \cdots & y_{T-p} & \hat{\varepsilon_{T-1}} & \cdots & \hat{\varepsilon_{T-q}} \end{bmatrix} \qquad (3)$$

The main steps of the **Hannan & Rissanen's** [3] algorithm are described as follows. First, an AR$(\mathbf{n})$ is estimated in order to obtain residuals $\hat{\varepsilon_\mathbf{t}} = \sum_{\mathbf{j}=\mathbf{0}}^{\mathbf{n}} \hat{\alpha}_\mathbf{j}\mathbf{y_{t-j}}$, with $\hat{\mathbf{a_0}} = \mathbf{1}$. The order n of the previous model can be chosen by minimizing the AIC criterion. Second, an Adaptive Lasso estimator is used to compute the unknown coefficients β.

The **Adaptive Lasso** proceeds in two steps: i) coefficient specific weights $\hat{\mathbf{w}}$ are constructed are constructed by Least Squares, Ridge or Lasso regression of y on $\hat{\mathbf{X}}$, ii) weights are combined with the penalty term in the Lasso estimator and the optimal $\lambda_\mathbf{T}$ can be identified with a criterion, such as AIC and BIC.

$$\hat{\beta}_{\text{ADLASSO}}(\lambda) = \arg\min_\beta \left\{ \|\mathbf{y} - \mathbf{X}\beta\|_\mathbf{2}^\mathbf{2} + \lambda \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{p}+\mathbf{q}} \mathbf{w_i}|\beta_\mathbf{i}|_\mathbf{1} \right\} \qquad (4)$$

Subset ARMA selection can be based on alternative penalized estimators; one of those the **Adaptive LAD Lasso**, which is designed to optimize performance in terms of the mean absolute error (MAE).

$$\hat{\beta}_{\text{AD-LAD-LASSO}}(\lambda) = \arg\min_\beta \left\{ \|\mathbf{y} - \mathbf{X}\beta\|_\mathbf{1} + \lambda \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{p}+\mathbf{q}} \mathbf{w_i}|\beta_\mathbf{i}|_\mathbf{1} \right\} \qquad (5)$$

Monte Carlo Experiments

The empirical performance of Lasso-type Subset ARMA selection [2] is examined through simulations. Furthermore, the conventional model building procedures are examined. The set of experiments use series with sample sizes of 120, 240 and 360 to evaluate the performance for the aforementioned techniques within 1000 replications.

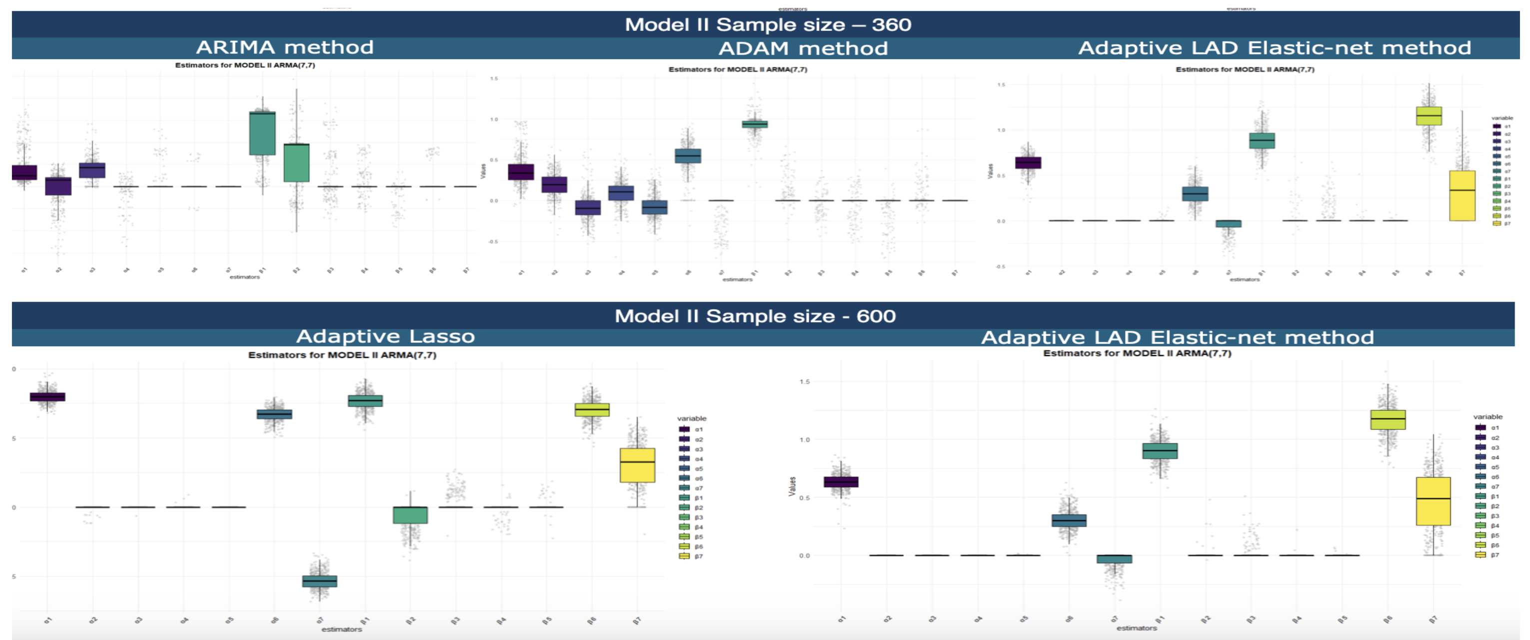


Fig. 1: This box-plots depict estimation of orders p and q of simulated Gaussian ARMA process of length $n = \{360, 600\}$ for **Model II:** $\mathbf{y_t} - \mathbf{0.8y_{t-1}} - \mathbf{0.7y_{t-6}} + \mathbf{0.56y_{t-7}} = \varepsilon_\mathbf{t} + \mathbf{0.8}\varepsilon_\mathbf{t-1} + \mathbf{0.7}\varepsilon_\mathbf{t-6} + \mathbf{0.56}\varepsilon_\mathbf{t-7}$.

The simulations showed that **Adaptive LASSO** and **Adaptive Elastic Net** performed similarly and could identify more frequently the true Data Generating Mechanism (DGM), for large sample sizes. Methodologies such as **Adaptive LAD Elastic Net** though, could capture more often the true **DGM**, when contaminated data were considered. The **stepwise procedures**, were not able to identify the true model.

Maximum Entropy Bootstrap Lasso

The algorithm generates samples from a uniform uninformative distribution, while trying to maximize entropy under constrains. Analogously then to [1], a **screening step** follows;

i. a Lasso, Elastic Net or other penalized estimator is fitted in each bootstrap sample,

ii. appropriate $\lambda_\mathbf{T}$ is chosen based on minimization of information criteria,

iii. the variable inclusion frequencies are recorded and

iv. variables with small inclusion frequency are excluded.

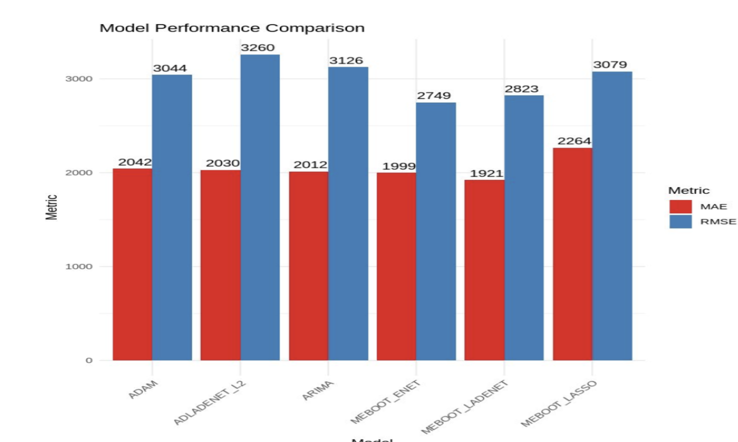
Finally, the procedure contains fitting an arima model with fixed parameters ϕ and θ, which have been obtained from the screening step. The final solution path is defined by the estimated coefficients obtained, when a information criterion used for the selection of the best model among candidates.

Shrinkage Method	Information Criterion	n	Metrics: A	esd(A)	T	+	esd(+)	-	esd(-)
LASSO	BIC	120	0.60 (0.59)	0.10 (0.10)	0.69 (0.74)	0.40 (0.41)	0.10 (0.10)	0.06 (0.05)	0.09 (0.08)
		240	0.56 (0.57)	0.06 (0.06)	0.97 (0.94)	0.44 (0.43)	0.06 (0.06)	0.01 (0.01)	0.03 (0.03)
		360	0.55 (0.55)	0.06 (0.05)	1.00 (1.00)	0.45 (0.45)	0.05 (0.05)	0.00 (0.00)	0.02 (0.00)
	AIC	120	0.53 (0.53)	0.07 (0.06)	0.91 (0.91)	0.47 (0.47)	0.07 (0.06)	0.02 (0.02)	0.06 (0.05)
		240	0.51 (0.51)	0.04 (0.00)	1.00 (0.05)	0.49 (0.98)	0.04 (0.49)	0.00 (0.05)	0.01 (0.01)
		360	0.50 (0.50)	0.04 (0.04)	1.00 (1.00)	0.50 (0.50)	0.04 (0.04)	0.00 (0.00)	0.01 (0.01)
LAD-LASSO	BIC	120	0.59 (0.58)	0.10 (0.09)	0.62 (0.70)	0.41 (0.42)	0.10 (0.09)	0.06 (0.05)	0.07 (0.08)
		240	0.57 (0.57)	0.06 (0.05)	0.92 (0.94)	0.43 (0.43)	0.06 (0.05)	0.01 (0.01)	0.04 (0.03)
		360	0.58 (0.56)	0.06 (0.05)	0.97 (1.00)	0.42 (0.44)	0.06 (0.05)	0.00 (0.00)	0.02 (0.01)
	AIC	120	0.54 (0.52)	0.07 (0.04)	0.80 (0.85)	0.46 (0.48)	0.07 (0.04)	0.04 (0.03)	0.07 (0.06)
		240	0.54 (0.54)	0.04 (0.04)	0.92 (0.96)	0.46 (0.46)	0.04 (0.04)	0.01 (0.00)	0.04 (0.02)
		360	0.55 (0.55)	0.04 (0.04)	0.96 (0.98)	0.45 (0.45)	0.04 (0.04)	0.01 (0.00)	0.03 (0.02)

Fig. 2: This table captures the empirical performance of LASSO and LAD LASSO for **Model II:** $\mathbf{y_t} - \mathbf{0.8y_{t-1}} - \mathbf{0.7y_{t-6}} + \mathbf{0.56y_{t-7}} = \varepsilon_\mathbf{t} + \mathbf{0.8}\varepsilon_\mathbf{t-1} + \mathbf{0.7}\varepsilon_\mathbf{t-6} + \mathbf{0.56}\varepsilon_\mathbf{t-7}$. Metric **A** denotes the probability of picking correctly all significant variables; **T** presents the relative frequencies of picking the correct model; **"-"** reports the false negative rates (FNR) and **"+"** the false positive rates (FPR). The values in blue color denote the performance when handling contaminated data.

Application

ARMA models are powerful tools for modeling and forecasting time series data in a wide range of applications. The figures below depict selected results, regarding one-step ahead forecasts and predictive intervals for daily measurements of **solar irradiance**. The performance of the best six methodologies are presented.



In terms of short-term forecasting, Adaptive LAD Elastic Net was the variant of Subset ARMA that standed out. It managed to compete the stepwise models. ME Bootstrap penalization techniques performed much better than the majority of Subset ARMA variants. The best one was the LAD Elastic Net.

Fig. 4: The plots depict the forecasts and predictive intervals of the stepwise models; ARIMA and ADAM respectively.

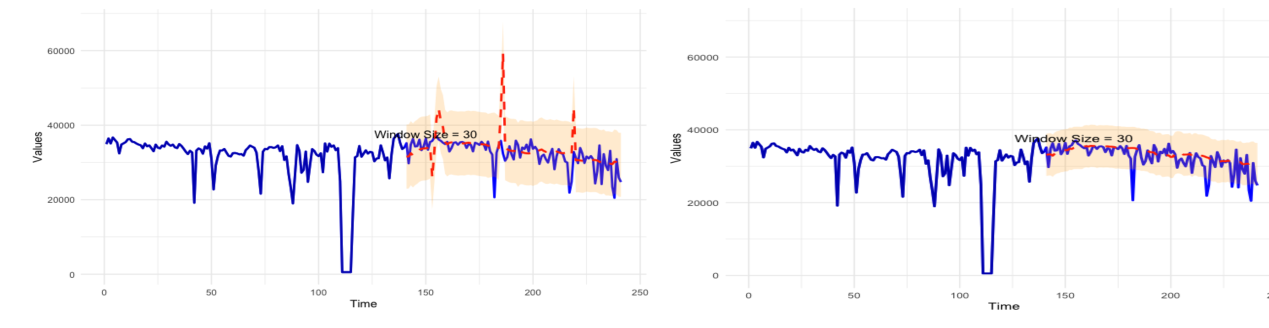


Fig. 5: The plots present the forecasts and predictive intervals of the worst and best methodology; ME Bootstrap LASSO and ME Bootstrap LAD Elastic Net, respectively.

References

[1] Anika Buchholz, Norbert Holländer, and Willi Sauerbrei. "On properties of predictors derived with a two-step bootstrap model averaging approach—A simulation study in the linear regression model". In: *Computational Statistics Data Analysis* 52 (Jan. 2008), pp. 2778–2793. DOI: 10.1016/j.csda.2007.10.007.

[2] Kun Chen and Kung Sik Chan. "Subset ARMA selection via the adaptive Lasso". eng. In: *Statistics and its Interface* 4.2 (2011), pp. 197–206. ISSN: 1938-7997. DOI: 10.4310/SII.2011.v4.n2.a14.

[3] E. J. Hannan and L. Kavalieris. "A Method for Autoregressive-Moving Average Estimation". In: *Biometrika* 71.2 (1984), pp. 273–280. ISSN: 00063444. URL: http://www.jstor.org/stable/2336243 (visited on 11/14/2024).

[4] I. Svetunkov. *Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM)*. 1st. 2023. DOI: https://doi.org/10.1201/9781003452652.

[5] Hrishikesh D. Vinod and Javier Lopez-de-Lacalle. "Maximum Entropy Bootstrap for Time Series: The meboot R Package". In: *Journal of Statistical Software* 29.5 (2009), pp. 1–19. DOI: 10.18637/jss.v029.i05. URL: https://www.jstatsoft.org/index.php/jss/article/view/v029i05.

[6] Hui Zou. "The Adaptive Lasso and Its Oracle Properties". In: *Journal of the American Statistical Association* 101 (2006), pp. 1418–1429. URL: https://EconPapers.repec.org/RePEc:bes:jnlasa:v:101:y:2006:p:1418-1429.