A new robust estimation method for ARMA models

Y. Chakhchoukh[1,2], P. Panciatici[1], P. Bondon[2] and L. Mili[3]

[1] Gestionnaire du Réseau de Transport d'Electricité, France
[2] Laboratoire des Signaux et Systèmes – CNRS – Supélec – Université Paris-Sud, France
[3] Virginia Tech-NVC, USA

Our method makes use of the autocorrelations estimates based on the ratio of medians together with a robust filter cleaner able to reject a large fraction of outliers, and a Gaussian maximum likelihood estimation which handles missing values. The procedure is easy, robust and fast. Application to the forecasting of the French daily electricity consumption is considered.

Introduction

- iid data or linear regression models: many robust estimation methods have been proposed
- time series: filtered τ, filtered M, generalized M (GM) and the Residual Autocovariance (RA) estimators [1]-[2].

Median-based filtering

$\{X_t\}$ is a Gaussian stationary time series with variance σ^2 and autocorrelation function $\rho(\cdot)$. For each $k \in \mathbb{N}$, we define $\widehat{\tau}(k)$ by

$$\widehat{\tau}(k) = \widehat{\xi}_{F_{\rho(k),\sigma^2}} / \widehat{\xi}_{G_{\sigma^2}},$$

where $F_{\rho(k),\sigma^2}$, resp. G_{σ^2} are the monovariate distribution function of $\{X_t X_{t-k}\}$, resp. $\{X_t^2\}$. $\widehat{\xi}_F$ is the sample median at F. A robust estimate $\widehat{\rho}(k)$ of $\rho(k)$ is obtained by the relation

$$\widehat{\rho}(k) = r(\widehat{\tau}(k))$$

An explicit relation between $\tau(k)$ and $\rho(k)$ does not seem to exist. Fig. 1 represents ρ as a function of τ, $\rho = r(\tau)$, and is obtained numerically.

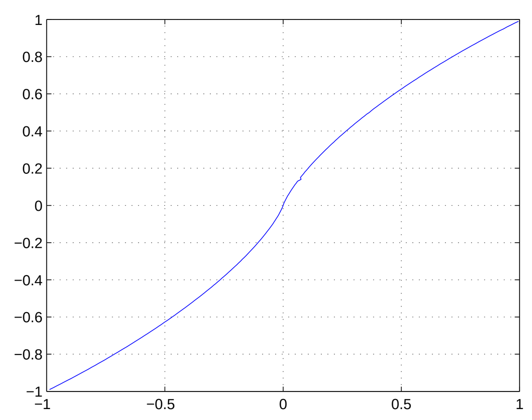

Fig. 1

Call the obtained estimator ratio-of-medians-based estimator (RME).

Robustness for an AR(1) model with $\phi = 0.5$

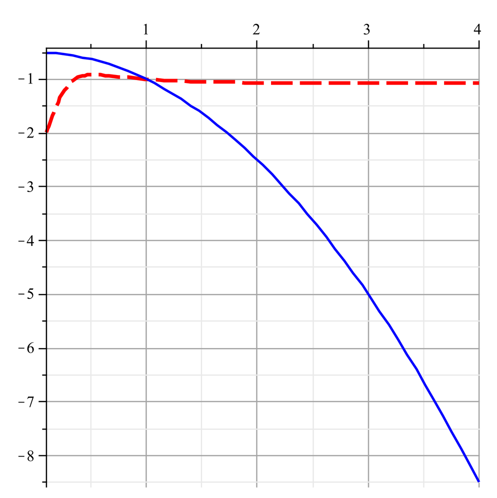

Fig. 2

The influence functional (IF) of the classical least squares estimator (solid line) is not bounded while that of our RME (dashed line) is.

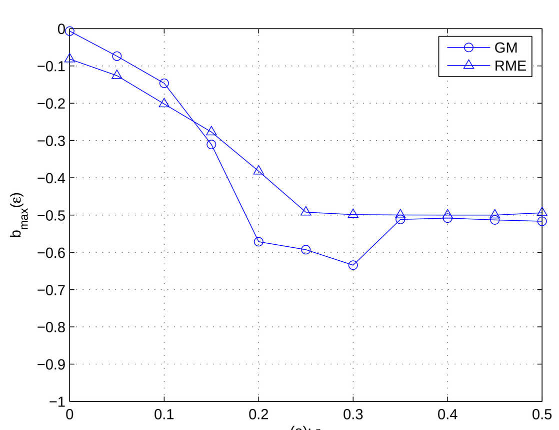

Fig. 3

The maximum bias curves of the RME shows better performance than the GM estimator when $\epsilon > 0.12$. The breakdown point is equal to 25% (BP of the RME is half that of the sample median).

Robust estimation of an ARMA using the RME

- Fit a high order AR(p^*) model using robust autocorrelations estimates (RME method), where p^* is selected by a robust order selection criterion.
- Detect the outliers by a robust filtering [3] applied to the AR(p^*) model, reject them and use the maximum likelihood estimation method of an ARMA model with missing data [4].

Application to load time series forecasting

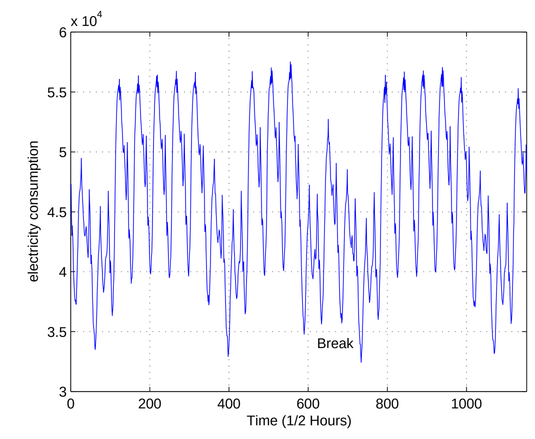

Fig. 4

There is a break on Thursday 14 and Friday 15 (July 14^{th} is a national holiday).

A SARIMA$(p,d,q) \times (p_1,d_1,q_1)_7$ model for different hours of the day using three different methods :

- the RME-based estimation method (RME)
- the filtered τ [1](τ)
- outliers are deleted ("three sigma" rejection rule) then fit an ARMA model with missing data (CM).

If $Y_1,\dots,Y_n$, $n = 500$ denotes a given series, $\widehat{Y}_{n+h}$ is the (out-of-sample) predictor of Y_{n+h} based on $Y_1,\dots,Y_n$ and an adjusted model to $Y_1,\dots,Y_n$. The mean absolute percentage error is

$$\text{MAPE} = \frac{100}{H}\sum_{h=1}^{H}\left|\frac{Y_{n+h}-\widehat{Y}_{n+h}}{Y_{n+h}}\right|, \qquad (1)$$

where H is a given final step prediction. The MAPE is

- evaluated on 200 post-sample observations.
- robust since only "normal days" are considered.

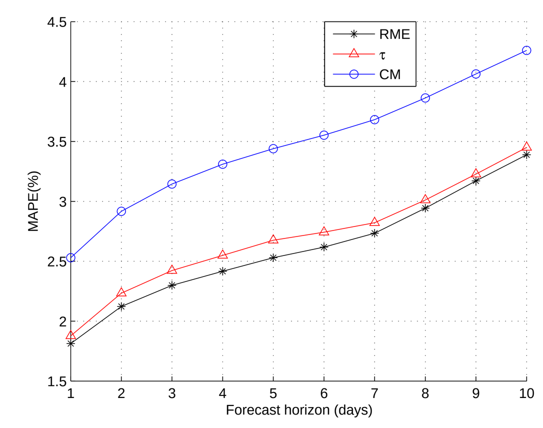

Fig. 5

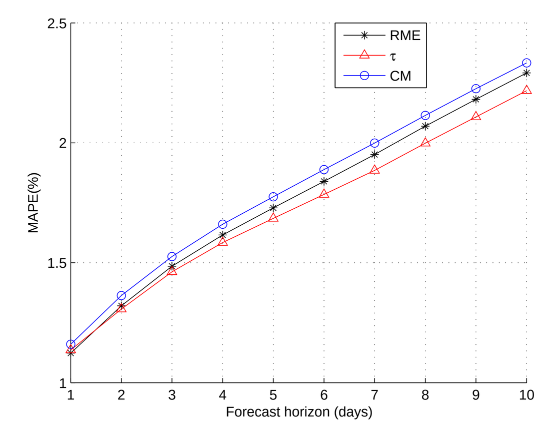

Fig. 6

- The performances vary with the series (some hours are more contaminated with outliers than others).
- The filtered-τ and RME-based are efficient since they are as good as the CM at 10:00 pm (almost clean series).

We compute the 200 out-of-sample first-step forecast errors obtained with the different methods. For each case, we calculate

- the median $\widehat{\xi}(\epsilon_t)$
- the MADN $\widehat{\sigma}(\epsilon_t) = 1.48 \times \widehat{\xi}\left(|\epsilon_t - \widehat{\xi}(\epsilon_t)|\right)$
- the Robust Mean Squared Error $\left(\text{RMSE}(\epsilon_t) = \widehat{\xi}(\epsilon_t)^2 + \widehat{\sigma}(\epsilon_t)^2\right)$

Method	$\widehat{\xi}(\epsilon_t)$	$\widehat{\sigma}(\epsilon_t)$	RMSE(ϵ_t) $(\times 10^6)$
RME	006.25	0834.31	0.6961
τ	-30.26	0842.02	0.7081
CM	177.39	1265.15	1.6321

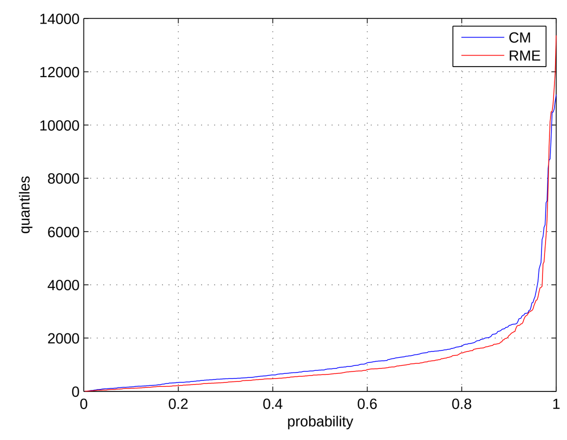

Fig. 7

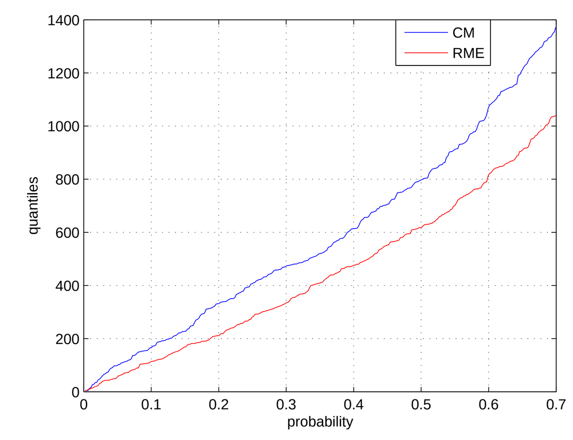

Fig. 8

The quantiles of the absolute values of the residuals in the RME and CM methods for the series at 10:00 am show that RME-estimates give the best fit to the bulk of data.

Conclusions

- RME is simpler to implement than the GM and the filtered-τ methods.
- Robust methods are useful tools for automatic on-line estimation and forecasting load series.
- Good tradeoff between robustness and efficiency.

References

[1] Ricardo A. Maronna, R. Douglas Martin, and Victor J. Yohai, *Robust statistics*, Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2006, Theory and methods.

[2] Oscar H. Bustos and Víctor J. Yohai, "Robust estimates for ARMA models," *J. Amer. Statist. Assoc.*, vol. 81, no. 393, pp. 155–168, 1986.

[3] C. Johan Masreliez and R. Douglas Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *IEEE Trans. Automatic Control*, vol. AC-22, no. 3, pp. 361–371, 1977.

[4] Richard H. Jones, "Maximum likelihood fitting of ARMA models to time series with missing observations," *Technometrics*, vol. 22, no. 3, pp. 389–395, 1980.