A NEW ROBUST ESTIMATION METHOD FOR SHORT-TERM LOAD FORECASTING

Yacine Chakhchoukh

Laboratoire des signaux et systèmes, CNRS, Supélec, Univ. Paris-Sud XI
3, rue Joliot-Curie. 91192 Gif-sur-Yvette, France
yacine.chakhchoukh@lss.supelec.fr

ABSTRACT

This paper presents a new robust method to estimate the parameters of ARIMA models. This method makes use of the minimum Hellinger distance estimator (MHDE) 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 main advantages of the procedure are its easiness, robustness, high efficiency and practical execution. Its effectiveness is demonstrated on Monte Carlo simulations and an example of the forecasting of the French daily electricity consumptions.

Index Terms– Robustness, time series, Hellinger distance, ARIMA models, outliers, load forecasting.

1. INTRODUCTION

Accurate forecasts of electricity load in short term are necessary for electric utilities. Short-term forecasting is important to balance the electricity produced and consumed at any moment of the day. Short-term forecasting helps also to manage the production, the transmission, and the distribution of electricity in a more efficient and secure way. Errors in electricity prediction have significant cost implications for electric companies. This work is initiated by RTE, the transmission operator that manages and operates the French electric power transmission system, which is confronted to the presence of outliers in the French daily electric consumptions. In our case, the outliers, termed also breaks, are atypical days such as public holidays and exceptional events. RTE uses a SARIMA model in its daily forecasting. The load time series is first corrected from the influence of the weather by using a regression model where the exploratory variables are the temperature and the nebulosity. Then a SARIMA model is fitted to the residuals. The resulting adjusted series exhibit a trend and several major cycles (daily, weekly,seasonal, yearly, etc.). Obviously, due to the qualitative change observed in the series during the breaks, it is of paramount importance to treat them separately from the majority of the data. It is very difficult and challenging to detect these outliers by experience or eye-balling. This approach, which is widely used by electric companies, is not very efficient. This is because of the fact that an observation is judged outlying relative to some model. To improve the robustness of the parameter estimation and forecasting methods, we may resort to a robust statistical estimation.

Robust estimation is important when estimating a statistical model. When the data contains deviant observations termed outliers, the classical statistical estimators of a SARIMA model become unreliable. Thus order selection, parameter estimation, and forecasting can be affected notably. In the robust statistics literature, several methods were proposed mainly for iid data and for regression models. Some methods were proposed in the context of time series such as the filtered–τ, filtered M–, generalized M– and the so-called *Residual Autocovariance* (RA)–estimators [1, 2]. In this paper, we propose a new robust procedure based on the minimum Hellinger distance estimator. Our method compares favorably to the other methods in terms of simplicity and forecasting performance. When applied to the actual French daily electric consumptions, it exhibits the same performance as compared to the filtered-τ estimators. The filtered-τ are efficient highly robust estimators proposed in [1]. By highly robust, we intend that the estimator has a high breakdown point. We analyze the robustness of our method using a simple novel approach and the maximum bias curve, which is computed numerically. The effectiveness of our method is demonstrated on an example of the forecasting of the French daily electricity consumptions. The paper is organized as follows. In Section 2, we present the method and analyze its robustness. Section 3 presents the simulation results in the case of electricity load modeling and forecasting. Finally, Section 4 concludes the paper.

2. MHDE-BASED FILTERING

Hellinger estimator belongs to the minimum distance estimators. A *minimum distance estimator* minimizes the distance between two density functions in a functional space. An estimation process based on the Hellinger distance was first put forth by Beran [3]. In general, the Hellinger estimator offers the advantage of being as efficient as the maximum likelihood under the real or the non-contaminated model and very robust when the data observed deviates from the strict modeling assumptions [4].

In this article, we introduce a practical method to estimate a SARIMA model using the Hellinger distance. In the case of the first order autoregressive model AR(1) given by

$$Y_t = \phi Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2),$$

we show how to estimate ϕ and σ_ε using the Hellinger distance. Minimizing the Hellinger distance of the prediction residuals (d_H) with respect to both ϕ and σ gives multiple solutions and can not be used for estimating the parameters. However, ϕ can be estimated by minimizing a robust efficient MHDE estimator of scale of the prediction residuals. Thus, we define the minimum-Hellinger-based estimator in the case of an AR(1), as $\widehat{\phi}_H$ which satisfies

$$\widehat{\phi}_H = \arg\min_{\phi}\left\{\arg\min_{\sigma}(d_H(\sigma,\phi))\right\}$$

In order to improve the breakdown point of the estimator and prevent the propagation and the presence of large outliers in the explanatory variables; we replace the prediction residuals by the robust prediction residuals. For an AR(1), robust prediction residuals are defined by $r_i(\phi) = Y_i - \phi\widehat{Y}_{i-1|i-1}$. $\widehat{Y}_{i-1|i-1}$ is obtained using a robust filter cleaner [5].

The proposed estimator is computed by the following steps

- Step 1: Search for $\widehat{\sigma}_H$ that minimizes the Hellinger cost function $d_H(\sigma,\phi)$ for a certain ϕ

$$2-2\int_{-\infty}^{+\infty}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-z^2}{2\sigma^2}}\right)^{\frac{1}{2}} f_r(z)^{\frac{1}{2}}dz,$$

$f_r(z)$ is the nonparametric kernel estimate of the probability density function of the robust prediction residuals defined by

$$f_r(z)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{z-r_i(\phi)}{h}\right),$$

where n is the number of observations used, K is referred to as the kernel and h is a positive number known as the bandwidth [6]. The optimal choice of the bandwidth value and the kernel type are widely studied in the nonparametric area [6]. In this article, we propose to use a Gaussian kernel.

- Step 2: Choose the $\widehat{\phi}_H$ that minimizes the $\widehat{\sigma}_H(r_1,\dots,r_n)$ = $\underset{\sigma}{\operatorname{argmin}}(d_H(\sigma,\phi))$ estimated previously.

The algorithm proposed in this paper uses a simple grid search for ϕ, which is tractable since the interval of search is]-1,1[. More sophisticated optimization algorithms were developed for the MHDE estimator [7]. To estimate an AR(p) with $p \geq 1$, the MHDE based estimator can be combined with a Durbin-Levinson algorithm to give a robust-efficient Durbin-Levinson algorithm. This algorithm is given by

- $\widehat{\phi}_{m,i}=\widehat{\phi}_{m-1,i}-\widehat{\phi}_{m,m}\widehat{\phi}_{m-1,m-i}$, $i=1,\dots,m-1$
- $\widehat{\phi}_{m,m}=\arg\min_{\phi_{m,m}}\widehat{\sigma}_H^m(r_{m+1,m},\dots,r_{n,m})$
- $r_{i,m}=Y_i-\widehat{Y}_{i|i-1}^{(m)}$

$r_{m+1,m},\dots,r_{n,m}$ are the robust residuals of the m^{th} step and obtained by the robust filter cleaner [5]. The MHDE algorithm applied previously to estimate the ϕ in the case of an AR(1) will be used to estimate the partial autocorrelation function $\phi_{m,m}$ in each step of the Durbin-Levinson algorithm. This means that for a certain $\phi_{m,m}$ in]-1,1[, calculate $\widehat{\sigma}_H^m$ and choose the $\widehat{\phi}_{m,m}$ that minimizes $\widehat{\sigma}_H^m$. This approach allows to estimate an autoregressive model of order p, AR(p).

To estimate the parameters of an ARMA(p,q) model, we propose the following procedure

- Fit a high order AR(p^*) using the MHDE based estimator, where p^* is selected by a robust order selection criterion subject to being larger than the order of the autoregressive part p.
- Detect the outliers by filtering with the high order AR(p^*), reject them and use a classical maximum likelihood based estimation method of ARMA models with missing values [8].

The robust filtering is based on the state representation of an AR(p^*). The filter used is defined in [1] and based on the robust filter of Masreliez [5], which is termed the filter cleaner.

This filter adapts the outliers with their expected values from the other observations and the structure of the model. While at this stage, we can apply a maximum-likelihood estimator on the 'cleaned' series, we prefer to delete the outliers and apply a classical estimator with missing values [8]. An AR(p^*) of high order is described by the following state space representation

$$\begin{cases}X_t=\Phi X_{t-1}+D\varepsilon_t\\ y_t=GX_t\end{cases},\ \Phi=\begin{pmatrix}\phi_1 & I_{p^*-1}\\ \vdots & \\ \phi_{p^*} & 0'_{p^*-1}\end{pmatrix} \quad (1)$$

Here, Φ is the transition matrix, $D=(1,0,\dots,0)'$, $G=(1,0,\dots,0)$, I_k is the $k\times k$ identity matrix and 0_k the zero vector in $\mathbb{R}^k$; dim(Φ)=$k\times k$.

2.1 Robustness analysis of the MHDE based estimator

We propose in this section a simple novel approach to understand the robustness and the efficiency of the MHDE in general and in the case of an AR(1). Beran[3] put forth an informal proof of the efficiency and robustness of the MHDE in the location case. The theoretical robustness and efficiency was studied by Lindsay[4]. Tamura and Boos[9] studied the case of multivariate location and covariance estimates. In this section, we compare the MHDE solution to an equivalent M-estimator solution. The Hellinger in this approach is found to be equivalent to several redescending M-estimators that tends to the maximum likelihood estimator when the fraction of contamination ε tends to 0.

2.1.1 M-estimator equivalent to minimum Hellinger distance estimator in the location case

$Z_1,\dots,Z_n$ are independent contaminated Gaussian observations. We consider an infinitesimal contamination. The contaminated model is $g=(1-\varepsilon)N(\mu,1)+\varepsilon\delta_{\mu+m}$; $\delta_{\mu+m}$ is the point-mass at $\mu+m$. The contaminated distribution function is $G(z)=(1-\varepsilon)\Phi_\mu(z)+\varepsilon\Delta_{\mu+m}(z)$, where Φ_μ is the distribution function of the normal distribution with mean μ and variance 1 denoted by $N(\mu,1)$. An M-estimator $\widehat{\mu}_M$ is solution of the equation $\sum_{i=1}^{n}\psi(z_i-\widehat{\mu}_M)=0$. Asymptotically, $\widehat{\mu}_M \xrightarrow{as} \mu_M$, with $\mathrm{E}_g(Z-\mu_M)=\int_{-\infty}^{\infty}\psi(z-\mu_M)g(z)dz=0$. On the other hand, the asymptotic Hellinger estimator satisfies

$$\int_{-\infty}^{\infty}\frac{(z-\mu_H)}{(2\pi)^{\frac{1}{4}}}e^{-\frac{(z-\mu_H)^2}{4}}\sqrt{g(z)}dz=0 \quad (2)$$

Equation (2) is obtained by deriving d_H with respect to μ_H; d_H is the Hellinger distance between g and $N(\mu_H,1)$. The M-estimator which is equivalent to the previous Hellinger estimator is given by

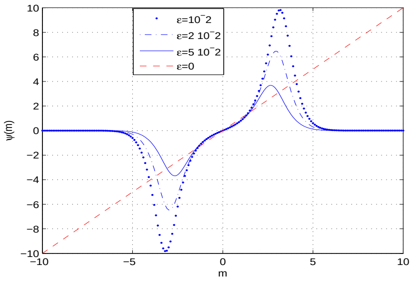
$$\psi(z-\mu)g(z)=\frac{(z-\mu)}{(2\pi)^{\frac{1}{4}}}\sqrt{g(z)}e^{-\frac{(z-\mu)^2}{4}} \quad (3)$$

The MHDE is a very robust estimator. In the presence of a very small fraction of contamination ε and $d(z)$ given previously, we obtain:

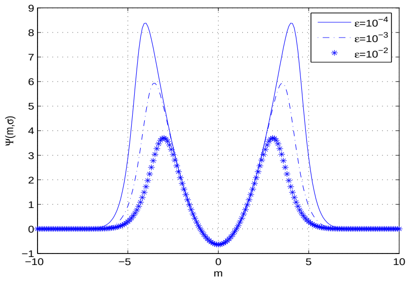
$$\psi(z-\mu)=\frac{(z-\mu)}{(2\pi)^{\frac{1}{4}}}\frac{e^{-\frac{(z-\mu)^2}{4}}}{\sqrt{(1-\varepsilon)e^{-\frac{(z-\mu)^2}{2}}+\varepsilon\delta(z-\mu-m)}} \quad (4)$$

If we consider r as a residual then

$$\psi(r)=(2\pi)^{-\frac{1}{4}}\left(r\frac{e^{-\frac{r^2}{4}}}{\sqrt{(1-\varepsilon)e^{-\frac{r^2}{2}}+\varepsilon\delta(r-m)}}\right) \quad (5)$$



(a) The function $\psi(m)$, $\varepsilon = 0, 10^{-2}, 2\times 10^{-2}, 5\times 10^{-2}$



(b) The function $\Psi(m,1)$, $\varepsilon = 10^{-4}, 10^{-3}, 10^{-2}$

Figure 1:

We want to analyze the impact of the outlier $r = m$, the weight $\psi(m)$ is given by:

$$\psi(m) = (2\pi)^{-\frac{1}{4}} \left(\frac{m e^{-\frac{m^2}{4}}}{\sqrt{(1-\varepsilon)e^{-\frac{m^2}{2}} + \varepsilon}} \right) \quad (6)$$

From the Fig. 1(a), we remark that large additive outliers receive a 0 weight and thus do not influence the estimation which is then robust. When $\varepsilon = 0$, the ψ-function is the linear curve $\psi(m) = m$ which is equivalent to the sample mean or the maximum likelihood estimator. For $\varepsilon > 0$, $\psi(m)$ is a redescending function. The MHDE adapts the ψ to the data.

2.1.2 M-estimator equivalent to minimum Hellinger distance estimator in the dispersion case

$Z_1, \ldots, Z_n$ are independent contaminated Gaussian observations. The contaminated model $g = (1-\varepsilon)N(0,\sigma) + \varepsilon\delta_m$. We want to estimate the parameter σ. The Hellinger distance between the contaminated distribution function g and the Gaussian distribution function of $N(0,\beta)$ is given by

$$HD(g, N(0,\beta)) = 2 - \frac{2}{(2\pi)^{\frac{1}{4}}} \int_{-\infty}^{\infty} \sqrt{g(z)}\,\beta^{-\frac{1}{2}} e^{-\frac{z^2}{4\beta^2}} dz$$

Deriving the Hellinger distance with respect to β and replacing by $\beta = \sigma_H$, we obtain

$$\frac{1}{(2\pi)^{\frac{1}{4}}} \int_{-\infty}^{\infty} \sigma_H^{-\frac{3}{2}} \sqrt{g(z)}\, e^{-\frac{z^2}{4\sigma_H^2}} \left(-1 + \frac{z^2}{\sigma_H^2} \right) dz = 0.$$

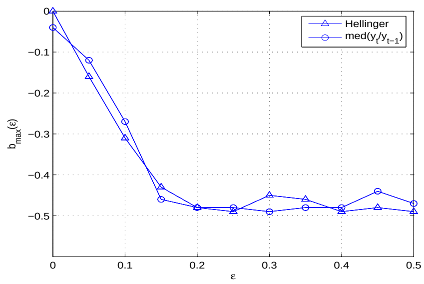
An M-estimate of scale satisfies asymptotically [1]

$$\begin{aligned} \mathrm{E}_g \left[\rho\left(\frac{z}{\sigma_M} \right) - \delta \right] &= \int_{-\infty}^{\infty} \left[\rho\left(\frac{z}{\sigma_M} \right) - \delta \right] g(z) dz \\ &= 0 \end{aligned}$$

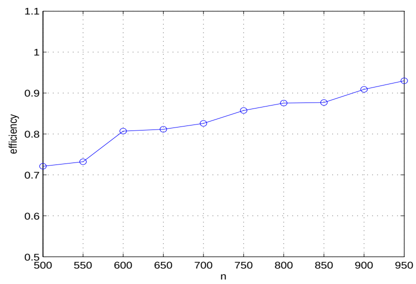
The ψ-function of an M-estimate of scale is given by $\Psi(z,\sigma) = \rho(z/\sigma) - \delta$. By comparing the M-estimator equation to the Hellinger equation, we obtain for $z = m$

$$\Psi(m,\sigma) = \frac{\sigma^{-\frac{3}{2}} e^{-\frac{m^2}{4\sigma^2}} \left(-1 + \frac{m^2}{\sigma^2} \right)}{(2\pi)^{\frac{1}{4}} \sqrt{(1-\varepsilon)e^{-\frac{m^2}{2\sigma^2}} + \varepsilon}}$$

Fig. 1(b) shows that the MHDE based estimator is equivalent to a robust scale M-estimator. If ε tends to 0 then the equivalent M-estimator tends to $\psi(z) = z^2$. The MHDE estimator is equivalent to the classical standard deviation for $\varepsilon = 0$. This shows that the Hellinger is equivalent to several redescending M-estimators where their weight function depends on ε. When $\varepsilon = 0$, the Hellinger estimator is equivalent to the maximum likelihood estimator. This computation shows the robustness and the high efficiency of this estimator in presence and absence of contamination respectively. Since the estimated scale of the residuals is robust and efficient, we conclude that the proposed filtered-MHDE-based estimator defined previously is robust. This estimator can be considered as a filtered S-estimator as defined by [1].



(a) Maximum bias curves of two robust estimators of an AR(1) , ϕ=0.5



(b) Efficiency of the filtered-MHDE-based estimator with increasing sample size under standard Gaussian distribution for $n = 500$ to $n = 950$, ϕ=0.5

Figure 2:

2.1.3 Maximum bias curves in the case of AR(1)

The maximum bias curves of the filtered-MHDE-based estimator are calculated following the Monte Carlo procedure described in [1, page 305]. For AR(1), Fig. 2(a) depicts the maximum bias curve of our MHDE together with that of another robust estimator, namely the median of slopes estimator, which has bias-optimality properties [1, Chapter 5]. The filtered-MHDE exhibits almost similar maximum bias behavior as the median of slopes. It is observed from these plots that the filtered-MHDE-based estimator is robust and has a breakdown point superior to 25%. The median of slopes has a breakdown point of 25 %, that is, it can handle up to 25% of outliers among the data samples. Simulation results seem to exhibit a constant breakdown point regardless of the order of the AR model for the filtered-MHDE. This result is interesting since the percentage of outliers in load time series is around 10 to 20 %. The efficiency of the proposed estimator can be verified empirically for a certain AR(1). We do this by calculating the variance of our estimator for increasing sample size n. The efficiency is calculated using Monte Carlo replications of the sample. Fig. 2(b) shows that for $\phi = 0.5$, the efficiency of the filtered-MHDE-based estimator tends asymptotically toward unity with increasing n.

3. APPLICATION TO LOAD TIME SERIES FORECASTING

SARIMA models are widely used to forecast electricity consumption time series [10, 11]. Fig. 3 illustrates the load demand from Saturday July 2^{nd}, 2005 to Saturday July 23^{rd}, 2005. We notice that there is a break appearing on July 14^{th} and lasts until July 17^{th}, 2005 (approximately from observation 600 to 800 on figure (3). July 14^{th} is a public holiday in France. These breaks or outliers give rise to problems with online forecasting systems.

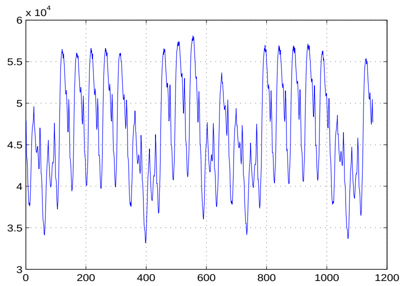


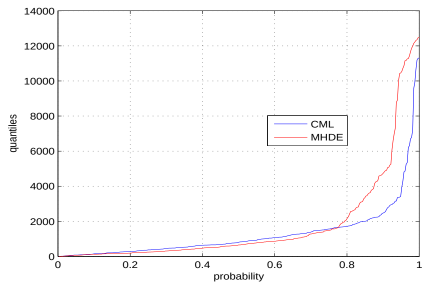
Figure 3: Half-hourly electricity consumption on July 2-23,2005, France.

3.1 Estimation and post-sample forecasting results

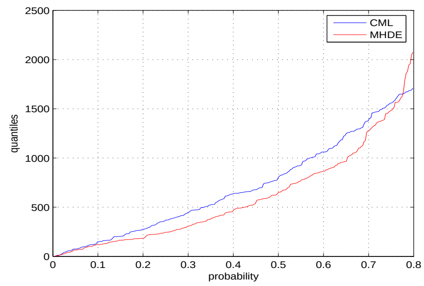
A seasonal ARIMA model, SARIMA$(p,d,q)\times(p_1,d_1,q_1)_{s_1}$ follow the equation

$$\phi_p(B)\Phi_{p_1}(B^{s_1})\nabla^d\nabla_{s_1}^{d_1}Y_t = \theta_q(B)\Theta_{q_1}(B^{s_1})\varepsilon_t,$$

where Y_t is the electricity demand at time t, s_1 is the number of periods in the different seasonal cycles. B is the lag operator. ∇ is the difference operator, ∇_{s_1} is the seasonal difference operator $(B^l Y_t = Y_{t-l}, \nabla = 1-B, \nabla_{s_1} = 1-B^{s_1})$. ϕ_p, Φ_{p_1}, θ_q, Θ_{q_1} are polynomials of order p, p_1, q, q_1. ε_t is a Gaussian white noise from $N(0,\sigma_\varepsilon^2)$. On the daily series of a certain time (12:00), s_1 is equal to 7 to model the within-week seasonal cycle. In Fig. 4, we show the quantiles of the absolute values of the residuals of two estimates. Namely, the proposed filtered-MHDE-based estimate and the classical approach based on maximum likelihood estimation applied after smoothing the unusual observations by expertise, denoted by CML, in the French daily load forecasting. It is seen that the MHDE based estimate yields the smallest quantiles, and hence gives the best fit to the bulk of data (80 %). In Fig. 5, we show the evolution of their mean absolute percentage error MAPE=$\frac{100}{h}\sum_{t=1}^{h}\left|\frac{Y_t-\widehat{Y}_t}{Y_t}\right|$, where h is



(a) 12:00



(b) 12:00

Figure 4: Quantiles of absolute residuals of estimates for load series at 12:00

the length of prediction. The estimation is done on a six-months period and the forecasting is computed on a period of one hundred (100) days. Only "normal" days are taken in account for the MAPE evaluation, which is natural since the goal of the method is to forecast "normal" days with better precision. The filtered-MHDE-based estimator has smaller MAPE than the CML for all leading times as illustrated in Fig. 5 . This means that our method improves the forecasting quality for the majority of data. In Fig. 6, the MAPE is calculated for the SARIMA estimation method based on the filtered-τ estimates and the previously defined approach. The filtered-τ estimates are highly efficient with high breakdown point of 50 %. The mean absolute percentage error of our robust approach and the filtered-τ are denoted by MAPEH and MAPEτ respectively. At 12:00, we remark that for all leading times MAPEH $<$ MAPEτ. This means that the forecasting quality is improved with the new approach for this hour of the day. At 20:00 MAPE$\tau <$ MAPEH for small forecasting times and they are almost the same for large leading times. In this example, the MHDE is slightly superior to the τ-estimates for the majority of day hours. The proposed method is much simpler in execution and algorithm than the filtered τ-estimates.

4. CONCLUSION

In this paper, a new robust filtered-MHDE estimation method for SARIMA models is proposed. Its maximum bias curve is derived and its robustness discussed. We compare the performance of our MHDE method to that of the classical approach based on maximum likelihood estimation after smoothing the outliers by experts in the French daily load forecasting series. It is found that our MHDE method outperforms the current methods. We compare also the proposed method with the highly robust efficient estimators known by filtered-τ estimates [1]. Our method shows equivalent forecasting performance while being simpler and less complicated. Sophisticated robust methods are useful tools for automatic online

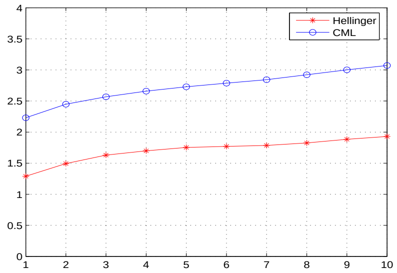
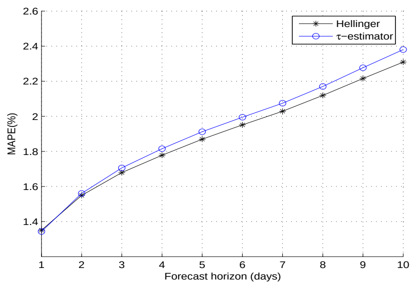
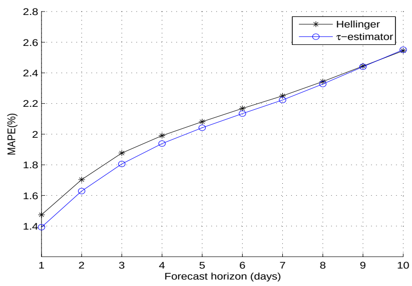


Figure 5: MAPE forecast accuracy versus lead time for the series at 10:00.



(a) 12:00



(b) 20:00

Figure 6: MAPE forecast accuracy versus lead time for the series at 12:00, 20:00

estimation and forecasting load series. They constitute also better alternatives to the intervention analysis based on experience used by the electric companies. These methods can offer a good tradeoff between robustness and efficiency. Ongoing effort has been concentrating on improving the execution algorithm to reduce the online calculating time. Hence, hybrid algorithms which combine natural gradient descent with Newton's method can be investigated [7]. Another research work will be to derive a robust order selection criterion in the same spirit as the Akaike's information criterion (AIC) proposed by Akaike [12] but using the Hellinger disrepancy instead of the Kullback discrepancy[13].

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, 1986.

[3] Rudolf Beran, "Minimum Hellinger distance estimates for parametric models," *Ann. Statist.*, vol. 5, no. 3, 1977.

[4] Bruce G. Lindsay, "Efficiency versus robustness: the case for minimum Hellinger distance and related methods," *Ann. Statist.*, vol. 22, no. 2, pp. 1081–1114, 1994.

[5] 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, 1977.

[6] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., 1992.

[7] Philip A. D'Ambriosio, "A differential geometry-based algorithm for solving the minimum hellinger distance estimator," M.S. thesis, VirginiaTech, 2008.

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

[9] Roy N. Tamura and Dennis D. Boos, "Minimum Hellinger distance estimation for multivariate location and covariance," *J. Amer. Statist. Assoc.*, vol. 81, no. 393, 1986.

[10] Marcelo C. Medeiros and Lacir J. Soares, "Robust statistical methods for electricity load forecasting," in *RTE-VT workshop*, PARIS, May 2006.

[11] L. Mili, P. Panciatici, and Y. Chakhchoukh, "Robust short-term load forecasting using projection statistics," in *RTE-VT workshop*, PARIS, May 2006.

[12] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pp. 267–281. Akadémiai Kiadó, Budapest, 1973.

[13] H. Linhart and W. Zucchini, *Model selection*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1986.