Forecasting of short term electricity load demand in Cameroon using semi parametric model

NGAMOU TCHUANA William*,
(*) Department of Analysis and Economic Policies
Ministry of Economy, Planning and Regional Development
Yaoundé, Cameroon Po. Box 660 DAPE/DGEPIP/MINEPAT
E.mail: wngamou@yahoo.fr

Abstract

The rationalization of electricity supply is a key activity of power system planning and development in all countries. This involves forecasting load demand with a higher accuracy. If the demand is under estimated, this will affect economic activities, because of power cut; if then the demand is over estimated, it will imply financial penalty due to waste of resources. Therefore, the minimization of forecasting errors is a strategic bet for the electricity supply industry, and also for the entire economy. Recent studies in Cameroon experimented a parametric model and non parametric model to forecast the electricity load demand, but they gave mitigate results, since no model surpassed the other one during some period of forecasting. We implemented in this study a semi parametric model, which combines parametrical and non parametrical methods, and therefore has the advantage of dealing with some drawbacks of both methods. The application of semi parametric model on hours load demand during three period of the year 2009 has globally given interesting results, comparing to individuals models (parametric and non parametric) results.

Key Words: dimension reduction model, electricity load curves, kernel function

1. Introduction

Forecasting electricity load demand is one of the most important activities while planning and managing the electricity production system of a country.

A parametric model and non parametric model were recently experimented in Cameroon to forecast the electricity load demand in Cameroon[1]. The parametric model was first used, and it was a seasonal autoregressive integrated Mobil Average (SARIMA), and the methodology of Box and Jenkins has been used to estimate the parameters of the model. Secondly, a non parametric model (a kernel predictor) has been used to correct the rigidity of the link function facing by the SARIMA models.

But, this non parametric model has a major inconvenience which is the "curse of the dimension ". Indeed, when the number of exogenous variables of the model is very high, the speed of convergence of the non parametric estimator is less fast. This phenomenon is susceptible to blemish the quality of the forecasting produced by the model.

Facing this problem, recent development in statistics experienced the semi parametric model, which combine the parametrical and the non parametrical approach, and thus have the particularity to deal with both drawbacks presented by the two individual's models (LEFIEUX V.2007). In this paper, the semi parametric model we experiment is a dimension reduction model, which deals with the problem of "curse of dimension" face by non parametric models. We then confront the results of forecasts given by individuals models (parametric, non parametric and semi parametric) in the essence of improving the forecasting of electricity load demand in Cameroon.

2. The semi parametric model

Let assume $Y \in \mathbb{R}$ an endogenous variable, and $X = (X_1, X_2, \dots X_p)$ a set of p exogenous variables of Y. The link equation between Y and X is:

$Y = g(X_1, X_2, \dots X_p) + \epsilon \qquad (1.1)$

[1] See NYAWA S.(2009)

Where:

- g : $\mathbb{R}^p \to \mathbb{R}$ is an unknown function
- ε is the error term.

The main idea of the semi parametric model is to estimate a non parametric tie function on a set of parametric variables. It proceeds by constructing a set of new exogenous variables $U_1, U_2, \dots, U_D$ that are linear combinations of the previous exogenous variables $X_1, X_2, \dots, X_p$ and then estimating on these new exogenous variables $U_1, U_2, \dots, U_D$ a non parametric function.

In the vast range of the semi parametric models, we will choose here the model of dimension reduction. The particularity of this model is that the number of the new set of exogenous variables $U_1, U_2, \dots, U_D$ is less than previous one $X_1, X_2, \dots, X_p$ (D<p), and this helps dealing with the problem of the ‘‘curse of dimension”.

To have the new set of D exogenous variables $U_1, U_2, \dots, U_D$, we project the p initial variables $X_1, X_2, \dots, X_p$ in a space H of D dimensions (D<p), which is direct by the set of vector $(\beta_1, \beta_2, \dots, \beta_D)$. The new variables $\beta_1^T X, \beta_2^T X, \dots, \beta_D^T X$ are linear combinations of $X_1, X_2, \dots, X_p$, and we have the following system :

$$\begin{cases} U_1 = \beta_1 X = \beta_1^1 X_1 + \beta_1^2 X_2 + \cdots + \beta_1^p X_p \\ U_2 = \beta_2 X = \beta_2^1 X_1 + \beta_2^2 X_2 + \cdots + \beta_2^p X_p \\ \quad \dots \\ U_D = \beta_D X = \beta_D^1 X_1 + \beta_D^2 X_2 + \cdots + \beta_D^p X_p \end{cases} \qquad (1.2)$$

In this new set of variables, we have D<p, which means that we have reduce the number of exogenous variables.

Thus, the model (1.1) is now:

$$Y = g(X_1, X_2, \dots X_p) + \epsilon = g_0(\beta_1^T X, \beta_2^T X, \dots, \beta_D^T X) + \varepsilon = g_0(U_1, U_2, \dots, U_D) + \varepsilon \qquad (1.3)$$

où :

- g_0: $\mathbb{R}^D \to \mathbb{R}$ is the new non parametric function to be estimated;
- βj∈ $\mathbb{R}^p$ with j∈{1,2,…,D}, a set of parameters to be estimated;
- ε is the error term, with $E(\varepsilon / U_1, U_2, \dots, U_D)=0$ almost sure.

The matrical form of the system (1.2) is:

$$\begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_D \end{pmatrix} = \begin{pmatrix} \beta_1 X \\ \beta_2 X \\ \dots \\ \beta_D X \end{pmatrix} = \begin{pmatrix} \beta_1^1 & \beta_1^2 \dots \dots & \beta_1^p \\ \beta_2^1 & \beta_2^2 \dots \dots & \beta_2^p \\ & \dots & \\ \beta_D^1 & \beta_D^2 \dots \dots & \beta_D^p \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{pmatrix} = B^T X \qquad (1.4)$$

With $B = \begin{pmatrix} \beta_1^1 & \beta_2^1 \dots \dots & \beta_D^1 \\ \beta_1^2 & \beta_2^2 \dots \dots & \beta_D^2 \\ \beta_1^p & \beta_2^p \dots \dots & \beta_D^p \end{pmatrix} = (\beta_1, \beta_2, \dots \beta_D)$

B is the matrix of projection of the exogenous variables $X_1, X_2, \dots, X_p$ into the space H of new variable $U_1, U_2, \dots, U_D$. It is the dimension reduction matrix.

The semi parametrical model (1.2) is now:

$$Y = g_0(B^T X) + \varepsilon \quad (1.5)$$

The core hypothesis of this semi parametric model is that the dimension reduction matrix B contains all the information of the set of initials variables $(X_1, X_2, \dots X_p)$.

The estimator of the model (1.5) is given by:

$$\hat{Y} = \widehat{g_0}(\hat{B}^T X) \quad (1.6)$$

So, the estimation of our semi parametric model is done into two consecutive steps:

(i) Estimation of the dimension reduction matrix $\hat{B}$ by the determination of D vectors of projection $\widehat{\beta_1}, \widehat{\beta_2}, \dots \widehat{\beta_D}$, with D<p

(ii) Estimation of the non parametric function $\widehat{g_0}$ on the set of D variables.

2.1 Estimation of the dimension reduction matrix $\hat{B}$

The algorithm of Hristache and al. (2001) has been used to estimate the matrix $\hat{B}$ of dimension reduction.

This algorithm also estimates the percentage of informations I_D captured by the D first vectors of the matrix B. So the determination of the value of D is base on this criteria of information. The formula of calculation of this percentage is given by:

$$I_D = \frac{\sum_{j=1}^{D} \lambda_j}{\sum_{j=1}^{p} \lambda_j} * 100$$

Where λ_j represent the eigen values of the eigen vectors $(\beta_1, \beta_2, \dots \beta_D)$ and are classified in crescents order.

2.2 Estimation of the non parametric function $\widehat{g_0}$

The estimation of the non parametric function is made using the kernel predictor.

The principle of the kernel predictor is to use a block of exogenous variables called block witness (for example the last daily load curve) and to compare it to similar backward blocks size. The model then defines an indicator of similarity named kernel function. More a past block will look like the block witness, more it will have a big weight in the forecasting of the set. The forecasting is therefore in the same way a past value average date, weighted by the kernel function define previously.

In conclusion, we can see that the semi parametrical model deal with some drawbacks of the parametric model by estimating a non linear function of link $\widehat{g_0}$. The semi parametrical model also deals with some drawbacks of the non parametrical model, by reducing the number of exogenous variables $(\hat{B})$, thus deal with the "curse of the dimension".

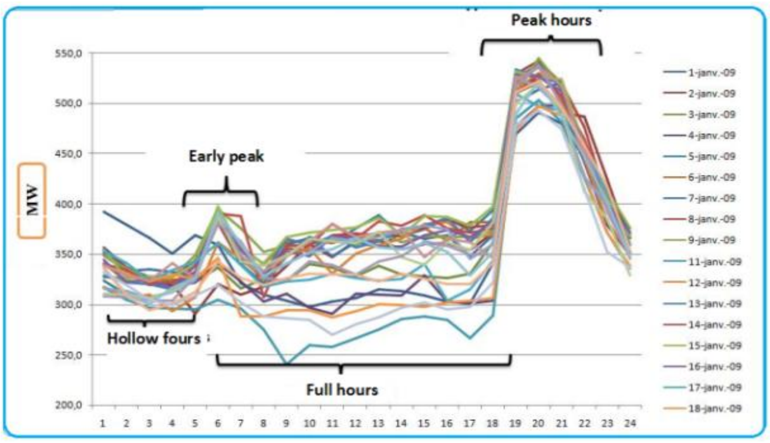
Indeed, theorically, the semi parametrical model gives more accurate forecasting than individuals forecasting issue from parametric and non parametric model.

3. Presentation of data

The load curves are constituted of the electricity powers called by the customers subscribed on the network. These powers characterize the evolution of the electricity demand coming from these customers. Thus, for one precise period, the load curves of a residential, of an enterprise, and even of a region can be represented. The data are collected from the national electricity Company (AES Sonel).

While superposing several daily load curves in Cameroon as indicated on the figure 1, we can note four big classes which can be grouped into hollow hours, early pick, full hours and peak hours.

Figure 1: *Profile of daily load curves of the month of January 2009*



Source *: AES-SONEL*

From this graph, denote a seasonal pattern in the structure of load curves in Cameroon, which repeat after every 24 hours.

The following table summarizes the information characterizing every group of hours.

***Table 1:** characteristics of daily profile (Mega Watts)*

Group	Mean	Standard deviation	Minimum	Maximum
Hollow hours (12:00-5 :00 a.m)	**319,55**	**27,74**	**222,00**	**450,20**
Early peak (6 :00 a.m- 7:00 a.m)	**337,72**	**30,65**	**251,00**	**419,00**
Full hours (8:00 a.m -6:00 p.m)	**345,81**	**34,99**	**208,30**	**520,50**
Peak hours (7:00 p.m-10:00 p.m)	**480,81**	**39,98**	**308,20**	**584,80**

***Source** : AES Sonel*

From this table, we can note that the big deal of the company is to satisfy the electricity demand during peak hours, because it is the period of the day where the power called is at his maximum.

For the simulation of data, we have chosen three period of the year 2009:

Trois périodes de simulations ont été choisis pour les différents modèles. Ces périodes sont :

- ✓ Period T1 : 1st January 2009 to 13 April 2009 : it is the period of low water;
- ✓ Period T2 : 1st March 2009 to 14 July 2009 : it is the period of flood ;
- ✓ Period T3: 1sr October 2009 au 6 March 2010: as the first period, it is into the period of low water.

4. Empirical results

4.1 Calibration of the semi parametrical model

From the previous analysis of load curves, we noted a daily seasonal character in the data. Thus, the forecasting will be made using the set of 24 exogenous variables $(X_1, X_2, \dots X_{24})$.

The principle of semi parametrical model is to reduce the dimension of this set of 24 exogenous variables, by determining the matrix B which has a lower dimension D<24, and which will contains all the information of the initials variables $(X_1, X_2, \dots X_{24})$.

While using the algorithm of Hristache and al. (2001) contained in the EDR package of the software R, we can calculation the percentage of information of the first D vectors of matrix B of the reduction of the dimension.

***Table 2:** Evolution of information percentage contains is D first vectors of dimension reduction matrix B (period T1)*

D	1	2	3
I_D	96,86%	98,56%	99,11%

***Source** : Our simulations in R software*

For the period T1, the first three vectors $\beta_1, \beta_2, \beta_3$ of the matrix B captures 99,11% of total variability of the g function, therefore nearly the entirely information contained in the set of exogenous variables $(X_1, X_2, \dots X_{24})$. On the basis of this criteria, we will choose D=3.

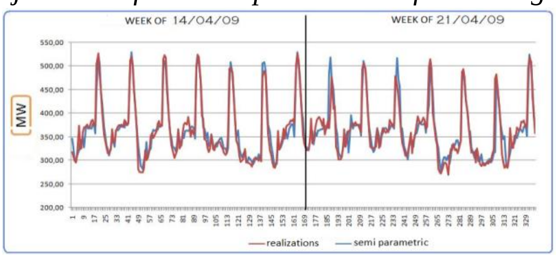
The same procedure is applied for the two other periods T2 and T3.

Then, we just have to estimate the non parametrical function $\widehat{g_0}$ on the set of the new exogenous vectors $\widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}$, using the kernel predictor.

4.2 Quality Adjustment of the semi parametrical model

Figure 2 illustrates the adjustment quality of load curves forecasting using semi parametrical model in period T1. We can see that the forecasting issue from the semi parametric model fit well the realisations.

***Figure 2**: Adjustment of the semi parametrical forecasting in period T1*



***Source**: AES Sonel*

To calculate the model quality adjustment, we will use the Mean relative error (MRE) and the Mean Absolute error (MAE).
The table 3 shows the quality adjustment of the semi parametric model. Globally, the results are interesting, the mean relative error (MRE) is less than 5%, even though it is a little beat higher during peak hours.

***Table 3**: Forecasting adjustment of the semi parametric model during period T1*

Indicators	globally	Peak hours (7:00 p.m-10:00 p.m)	Full hours (8:00 a.m -6:00 p.m)	Hollow hours (12:00-5 :00 a.m)
MRE (%)	3,95	5,17	3,14	4,71
EAM (Megawatts)	14,28	23,47	11,00	14,67

***Source :** AES-SONEL, our computation in Excel*

For the following tables:

- Semi para stands for semi parametric model
- Para stands for parametric model
- Non para stands for non parametric model
- MW stands for Megawatts

The table 4 gives the global qualities of adjustments of the semi parametric model, of the parametric model and the non parametric model. We note that on the three test periods of the year, the semi parametric model fits globally well, and present of the best performances that those of the parametric model and the non parametric model.

***Tableau 4:** Global adjustment quality of different models*

	Indicators	Semi para	Para	Non para
Period T1	MRE (%)	3,95	4,13	5,01
	AME (MW)	14,28	15,07	17,54
Period T2	MRE (%)	3,1	3,49	4,41
	AME (MW)	10,49	11,94	14,82
Period T3	MRE (%)	3,44	3,52	4,46
	AME (MW)	13,58	13,86	17,79

***Source :** AES-SONEL, our computation in Excel MW=Megawatts*

This result is very interesting. In fact, while combining the parametric and non parametric methods, the semi parametric model captures more information necessary to the forecasting, compared to the individual models (parametric and non parametric).
A more detailed look permits to examine the different model qualities at the different periods of the day.
While compiling all the three periods T1, T2 and T3, the mean perfomances of differents models show a globaly interesting result of the semi parametric model during all the period of the day, as illustrate in table 5. In fact, the semi parametric model presents better performances that the parametric model and the non parametric model. The revolution of this model is on the improvement of the full hours, because it permits to achieve a gain of about 1,24MW per hour of forecasting in relation to the parametric model, and of about 8,02 MW in relation to the non parametric model.

***Tableau 5**: mean adjustments quality of different models in all the period test.*

	Indicators	Semi para	Para	Non para
Global	MRE (%)	3,5	3,71	4,62
	AME (MW)	12,78	13,62	16,72
Peak hours	MRE (%)	3,85	4,22	3,92
	AME (MW)	18,02	19,93	18,29
Full hours	MRE (%)	3,34	3,71	5,61
	AME (MW)	11,73	12,97	19,71
Hollow hours	MRE (%)	3,65	3,94	3,97
	AME (MW)	11,64	12,67	12,67

Source : AES-SONEL, our computation in Excel

From the confrontation of the different models, it appears that the semi parametric model fits globally well on the periods test that the parametric and non parametric model. The full hours that constituted the real problem of the parametric model and the non parametric model have been improven by the model semi parametric, as well as the hollow hours.

These results join those of Vincent Lefieux (2007) on the forecasting of the load curves in France. In his study, he improves the results provided by a parametric model and a non parametric model by the use of the semi parametric approach.

5. Conclusion

The implementation of the semi parametric model has permit to improve the forecasting of electricity load demand in Cameroon.

However, the forecasting of the peak hours stand a real difficulty for the semi parametric model, because even though it is on average the best model on this period of the day, it was less effective than the two other models on the periods test T2 and T3.

Even though the semi parametric approach has several advantages, it also presents some drawbacks. Indeed:

- It is a method that is numerically very expensive. It requires several hours of simulations and optimization.
- The semi parametric model didn't take into consideration external variables, and focus only on the part autoregressive part of the sequence;

Facing these inconveniences, we intend in following studies to experiment other combining method like the aggregation model in order to perform more accurate forecast of electricity load demand in Cameroon.

References:

CHENLEI L., XIA Y., and JINFENG X.(2008), « An Adaptive Estimation Method for Semi parametric Model and Dimension Reduction », *University of Singapore*

HRISTACHE M., JUDITSKY A., POLZEHL J. and SPOKOINY V.(2001), « Structure adaptive approach for dimension reduction », *The Anals of Statistics*, Vol 29, N° 6 , ENSAI

KIM E. et DELECROIX M.(2009), « Modèles de régression à directions révélatrices : Application en assurance non-vie »

LEFIEUX V.(2007), « Modèle semi-paramétriques appliqués à la prévision des séries temporelles : Cas de la consommation d'électricité », Thèse de Doctorat, Université Rennes 2- Haute Bretagne

MIGUEL A. and CARREIRA P.(1997), « A Review of Dimension Reduction Techniques », University of Sheffield

MING-CHUNG L. and CHAN K.(2003) « Semi parametric Reduced-rank Regression»

NYAWA S.(2009), « Prévision à très court terme des courbes de charge dans le réseau interconnecté sud », *University Press,* Subregional Institute of Statistics and Applied Economics, Cameroon

PAVEL C. and HARDLE W.(2005), « Robust estimation of dimension reduction space », *TILBURG UNIVERSITY*

XIA Y., HOWELL T. and LI W.(2002), « An adaptive estimation of dimension reduction space», J. *R. Statist. Soc*, Part 3, PP. 363-410