International Journal of Advanced Smart Convergence Vol.10 No.4 59-65 (2021)
http://dx.doi.org/10.7236/IJASC.2021.10.4.59

IJASC 21-4-8

A Multiple Variable Regression-based Approaches to Long-term Electricity Demand Forecasting

Lan Dong Thi Ngoc[1], Khai Phan Van[2], Ngo-Thi-Thu-Trang[3], Gyoo Seok Choi[4], Ha-Nam Nguyen[5]

[1]PHD course Student, Academy of Finance, Ha Noi, Viet Nam
[2]Master course Student, Electric Power University, Ha Noi, Viet Nam
[3] Senior Researcher(PHD), Posts and Telecommunications Institute of Technology, Hanoi city, Vietnam
[4]Professor, Department of Computer Science, Chungwoon University, Incheon, Korea
[5]Professor, Vietnam Institute for Advanced study in mathematics, Ha Noi, Viet Nam
E-mail : landn0101@gmail.com, khaipv.epu@gmail.com,
lionel@chungwoon.ac.kr,nhnam@viasm.edu.vn,

Abstract

Electricity contributes to the development of the economy. Therefore, forecasting electricity demand plays an important role in the development of the electricity industry in particular and the economy in general. This study aims to provide a precise model for long-term electricity demand forecast in the residential sector by using three independent variables include: Population, Electricity price, Average annual income per capita; and the dependent variable is yearly electricity consumption. Based on the support of Multiple variable regression, the proposed method established a model with variables that relate to the forecast by ignoring variables that do not affect lead to forecasting errors. The proposed forecasting model was validated using historical data from Vietnam in the period 2013 and 2020. To illustrate the application of the proposed methodology, we presents a five-year demand forecast for the residential sector in Vietnam. When demand forecasts are performed using the predicted variables, the R square value measures model fit is up to 99.6% and overall accuracy (MAPE) of around 0.92% is obtained over the period 2018-2020. The proposed model indicates the population's impact on total national electricity demand.

Keywords: *demand forecasting, regression, electricity consumption.*

1. Introduction

In daily life as well as in production, electricity plays a very important role. To develop the economy, the electricity industry must be promoted. Thus, electricity is considered as a driving force to promote the development of economic sectors. Electricity is both a production industry and an infrastructure industry for the entire socioeconomic, and it is also an important criterion to evaluate the development of a country. Therefore, the more accurate the forecast of electricity demand is, the more effective it will be for the electricity industry in particular and the whole economy in general. Apart from this, short-term electricity demand forecasting in the residential sector is a complex problem because its rise and fluctuation are caused by differences in demand [1]. Understanding the importance of forecasting, we have chosen to study "Forecast of electricity demand for Vietnam's residential electricity industry in the period of 2021-2025". Based on the results of this forecast, it is possible to develop a plan for allocating electricity for the domestic electricity industry and give directions to adjust the distribution of power sources reasonably, which provides a plan for electricity use economically and efficiently for the residential sector. As a result, there is no shortage of

Manuscript Received: October. 5, 2021 / Revised: October. 8, 2021 / Accepted: October. 12, 2021
Corresponding Author: lionel@chungwoon.ac.kr, hnam@viasm.edu.vn
Tel:*** - **** - **** +84-90-467-0471, Fax: +82-32-770-8165
Professor, Department of Computer Science, Chungwoon University, Incheon, Korea
Professor, Vietnam Institute for Advanced study in mathematics, Ha Noi, Viet Nam

electricity during peak hours.

There are many approaches to calculate electricity demand forecast such as the Decomposition method, Long-range Energy Alternative Planning System, Regression analysis, etc. Each method has its advantages and disadvantages different. It is very important to apply any method to suit the reality in our country. From the characteristics of the above methods, we choose the multi-regression method to forecast the civil electricity demand in Vietnam.

In this study, the forecast analysis data is collected within 8 years from 2013-2020. Secondly, using Multi Regression analysis to estimate and forecast electricity demand based on the relationships between average income, average electricity price, population, and electricity consumption of the respective sector will provide accurate results through the tests and model fit. Besides, the performance was evaluated using MAPE values, which are scale-independent. This value is based on relative errors, hence, it is significant [2].

2. Problem Definition

A. Choose a forecasting method

Usually, there are many aspects when experts want to determine forecast methodology, which could be the most suitable for the residential sector. A review and analysis of publications that have used LEAP (Long-range Energy Alternative Planning System) methodology to forecast electrical power demand for the residential sector of Pakistan are presented by Mariam Gul and Waqar A. Qureshi [3]. In this study, authors have been collected and structured for the essential factors including the growth of end-use appliances, domestic consumption trends, and the changes in urban-rural ratio. However, one disadvantage of using this method is that statistical data must be large enough, highly accurate, and slight change from year to year. José Francisco Moreira Pessanha used Long-term demand forecast methodology, which has three models: demographic, macroeconomic, and microeconomic for the residential sector in Brazil [4]. Nevertheless, this methodology is based on the decomposition of total electricity residential consumption. This forecasting method is complex and detailed, requires data to be grouped, suitable for each country. For instance, to forecast the total occupied households taking into account the effects from demographic transition authors applied the household headship rate method, classifying households according to the age and sex of the head of household.

Sunil Malla and Govinda R. Timilsina presented two types of techniques that are used for energy demand forecasting: an econometric approach, an end-use accounting approach for Romania [5]. For the residential sector, authors have to choose a base year to forecast long-term electricity according to each energy-using device in the family. Nonetheless, this form of statistical data is only suitable for each country. Furthermore, this forecasting will encounter large errors if the data have large differences in the future years. Hyojoo Son and Changwan Kimal proposed a method to provide a precise model for the one-month-ahead forecast of electricity demand in the residential sector using weather and social variables [1]. Francesco Apadula built a model based on relationships between meteorological variables and monthly electricity demand in Italy [6]. Using weather variables is only suitable for short-term forecasting. Especially, this method is difficult to apply to countries in temperate regions, where is a difference in weather between seasons and regions. A few research studies have been directed to explore the complex problem of monthly electricity demand forecasting by multivariate time series analysis[7] [8] [9][10].

The energy demand forecasting model by Multivariable regression is widely applied in research and practice. The choice of variables to be included in the model depends on the research objectives as well as the availability of data [11]. As consequence, we choose the Multivariable-regression method to forecast the civil electricity demand in Vietnam, which is explained by the following reasons. Firstly, the method Multivariable Regression method requires a large enough sample of research data that these data can be collected at the Institute of Energy Science, General Statistics Office, and reports. Secondly, using regression analysis to estimate and forecast electricity demand based on the relationships between average income, average electricity price, population, and electricity consumption of the respective sector will provide accurate results through the tests and model fit.

B. Select variables

We mentioned above that the regression method is particularly interested in estimating or predicting the mean value of the dependent variable based on knowing the values of the independent variable. To estimate the sample regression function, we use the Ordinary Least Square method (OLS).

Suppose we consider the relationship between two variables X and Y, where X is considered as the independent variable (the explanatory variable), and Y is considered the dependent variable. Because there can be simultaneous effects of n factors, determining the relationship form of each factor and the whole equation is very complicated. Therefore, it is necessary to analyze the theory carefully in tandem with research experience to choose the appropriate form. For example, when forecasting a country's electricity demand, electricity demand is not generally related to one or more factors such as national income, population, gross domestic product, and consumer price index. Therefore, in this paper, we only choose the independent variables: population, electricity price, and average income, to explain the dependent variable (electricity demand for the residential sector).

3. Building Model

Based on the table of the collected sample values of input variables such as electricity consumption of the industry and factors affecting electricity consumption demand such as average income, price, and population, we establish the forecast function A = f(P, DS, TNBQ...) according to the common statistical method for the residential sector in Vietnam.

Many types of functions are used to forecast power demand, be it normal linear function or more complex functional forms such as squared trend function or exponential function. The general trend is to linearize these complex functions and solve them using least squares estimation techniques.

We use two types of functions: linear function and linear logarithmic function. Depending on the specific conditions of the collected data set, the forecasting models can be full of independent variables or not. Electricity demand for the residential sector:

$$ASH = f(DS, TNBQ, PSH) \tag{1}$$

Where:

- ASH: Electricity demand for the residential sector
- TNBQ: Average income per capita
- DS: Population
- PSH: Electricity price

A - Design, estimate, and test linear model

$$ASH = \beta_0 + \beta_1 * TNBQ + \beta_2 * PSH + \beta_3 * DS + \varepsilon_i \tag{2}$$

We use collected data as input values for building a forecasting model and from the output data, we can build a forecasting model:

$$ASH = (-3.65E+11) - 37074750 * TNBQ + 4781296 * DS - (1.36E+10) * PSH$$

With the significant level: $\alpha = 0.05$, we see that:

$$Prob(tTNBQ) = 0.5981 > \alpha$$

$$Prob(tPSH) = 0.1196 > \alpha$$

$$Prob(tDS) = 0.0002 < \alpha$$

Thus, the variables TNBQ and PSH do not affect the consumption demand of the industry. We remove the variable TNBQ and variable PSH from the model.

From the new output data:

We can build a forecasting model:

$$ASH= (-3.25E+11) + 4082565*DS$$

- T-test:

Null hypothesis: $H0 : \beta j = 0$

$H1 : \beta j \neq 0$

With the significant level $\alpha = 0.05$ we see that Prob(tDS) = 0.000000 < α

Reject H_0, accept H_1. Thus, there is at least one variable that affects the power consumption demand of the sector. The model satisfies the T-test.

- F-test:

Null hypothesis: $H_0 : \beta_1 = \ldots = \beta_k = 0$

H_1 : There is at least one $Bj \neq 0$

With the significant level $\alpha = 0.05$ we see that Prob(F-statistic) = 0.000000 < α

Reject H_0, accept H_1. Thus, there is at least one variable that affects the power consumption demand of the sector. The model satisfies the F-test.

- d_Durbin-Watson test:

Based on the results of Eviews, we have d = 2.147645, with the number of observed samples n = 8 and the number of independent variables k' = 1, look up the table of dL, dU values of Durbin-Watson statistics with significance level 5%, we have:

dL = 0.763

dU = 1.332

dU < d < 4 - dU

So the above model does not occur autocorrelation.

Performing the T-test, F-test, and d_Durbin-Watson test, we get the accepted forecasting model as:

$$ASH= (-3.25E+11) + 4082565*DS$$

B - Design, estimate and test the linear logarithmic model:

The forecast model has the form:

$$Ln(ASH) = \beta_0 + \beta_1*\ln(TNBQ) + \beta_2*\ln(PSH) + \beta_3*\ln(DS) + \varepsilon_i \quad (3)$$

We use the data collected and analyzed in the previous section as input data for building forecasting models and from the output data, we can build a forecasting model:

$$Ln(ASH) = -59.18039 + 0.125658*\ln(TNBQ) - 0.439914*\ln(PSH) + 7.311265*\ln(DS)$$

With the above prediction model, we see that the coefficients of the variable TNBQ and the coefficient of the variable DS have positive signs. That is, when TNBQ increases by 1%, the power demand also increases by 0.125658%, when the DS increases by 1%, the power demand increases to 7.311265%. This is consistent with economic theory. Besides, we see that the coefficient of the variable PSH has a negative sign. This means that when PSH is reduced by 1%, the power demand increases by 0.439914%. All of this is consistent with economic theory.

In addition, with the significance level $\alpha = 0.05$, we see:

Prob(tlog(TNBQ)) = 0.4332 > α

Prob(tlog(PSH)) = 0.4327 > α

Prob(tlog(DS)) = 0.0065 < α

Therefore, the variables PSH and TNBQ do not affect the power consumption demand of the industry. We proceed to remove the PSH and TNBQ variables from the model and rebuild the predictive model:

ASH= (-3.25E+11) + 4082565*DS

Performing the T-test, F-test, and d_Durbin-Watson test, we get the accepted forecasting model as:

ASH= (-3.25E+11) + 4082565*DS

4. Experiments and Results

We have models to forecast electricity consumption for the residential sector in Vietnam as follows:

Table 1. Electricity consumption forecasting models for the residential sector

ID	Type	Forecasting model	R^2
1	Linear	ASH= (-3.25E+11) + 4082565*DS	0.9968
2	Logarithmic linear	ln(ASH) = -24.33274+ 3.265663*ln(DS) + 0.474747*ln(ASH(-1))	0.9932

Based on the models built above, we get the results of electricity demand for Vietnam's residential sectors in the period of 2018-2020 To evaluate the error of the forecast, we can use the MAPE criterion. The model with the smallest of these criteria, choosing that model as the forecasting model will give better results.

MAPE (Mean Absolute Percent Error) indicator:

Table 2. Mean absolute percentage error (MAPE) results

Sector	MAPE	
	Linear MAPE	Logarithmic MAPE
Residential sector	0.92%	0.50%

At the same time, we compare the R^2 coefficients of the models to choose the most suitable forecasting model for the residential sector in Vietnam. The coefficient R^2 evaluates the fit of the model.

We have a table of data on the population of Vietnam in the period 2013-2020 shown in the table below:

Table 3. The population of Vietnam in the period 2013-2020

Year	2013	2014	2015	2016	2017	2018	2019	2020
Population *(Thousands of people)*	89.760	90.729	91.713	92.695	93.672	94.666	96.484	97.338

Based on the above data, we get the graph along with the trend function (Trendline)

We see that, with the coefficient $R^2 = 0.9918$, it shows that the appropriateness of the trend function is 99.18%. From there, we get the results of Vietnam's population forecast for the period 2021-2025 as follows:

Table 4. The results of Vietnam's population forecast for the period 2021-2025

Year	2021	2022	2023	2024	2025
Population *(Thousands of people)*	98.293	99.384	100.475	101.566	102.658

Substituting the forecasted data of the variables into the models above, we have the following results for forecasting electricity demand for the civil industry in Vietnam:

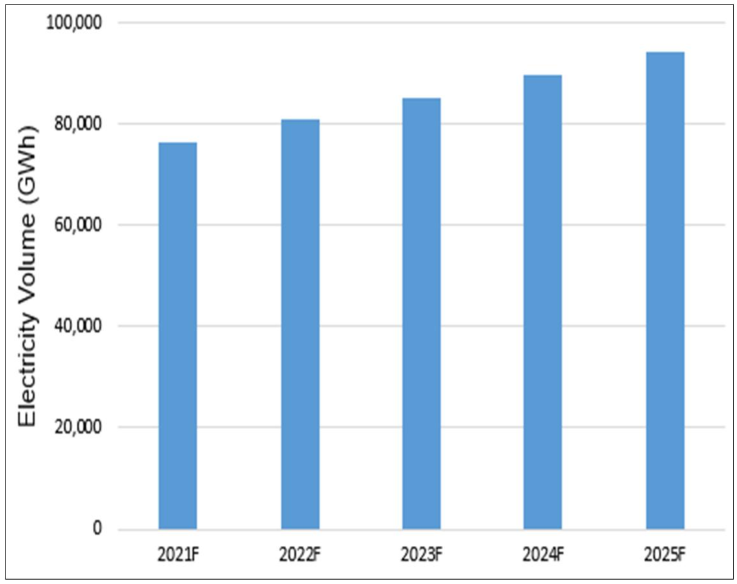

Figure 1. Forecasting results of electricity demand in the residential sector in the period of 2021-2025

The obtained results also are shown in Fig. 2.

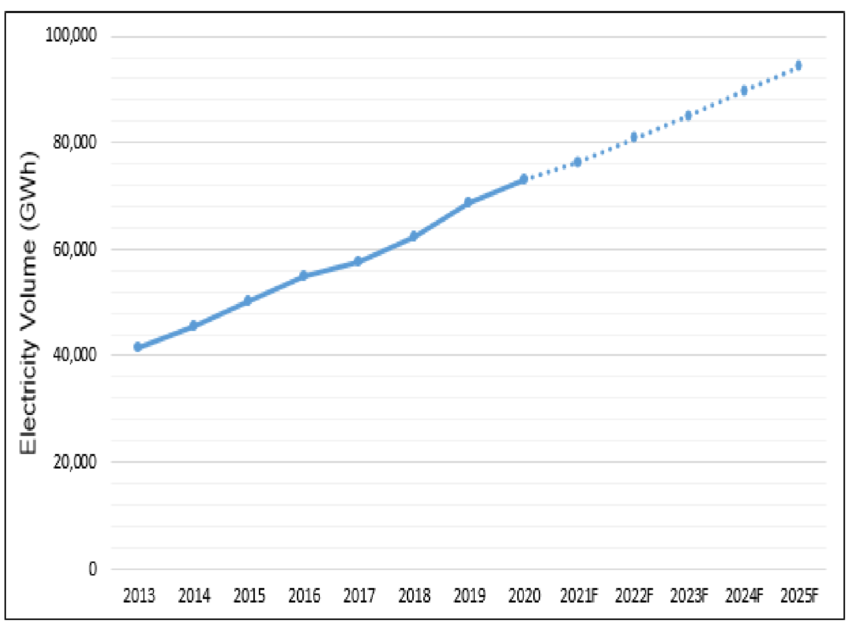

Figure 2. The electricity demand in the residential sector in the period of 2013-2025

Through forecast data, we can see that the electricity consumption demand of Vietnam's residential sector in the period of 2020-2025 tends to increase year by year. It is forecasted that by 2025, the demand for domestic electricity consumption will reach 94106 GWh, an increase of 23.36% compared to 2021. The demand will increase with an average growth rate of 4.7%/year.

5. Conclusion

With the current economic development, the demand for electricity consumption is increasing, especially the increase in the demand for household electricity. Therefore, the state needs to have reasonable policies to encourage people to use electricity economically and efficiently, reduce the burden on the national electricity system, and especially expand the model of using renewable energy sources when the current situation of hydroelectric and thermal power plants are running at full capacity, but they still cannot meet the national electricity demand.

Faced with that situation, we have studied and applied the multi-regression method to forecast the power consumption demand for the domestic electricity industry of Vietnam in the period 2021-2025. The model we choose has been calculated with an error of 0.92% and the fitness level is up to 99.6%. To ensure a sufficient supply of electricity for daily life, the power sector must always be one step ahead in source development as well as grid system planning.

Acknowledgments

This research is sponsored and funded by Ho Chi Minh City's University of Food Industry under Contract No. 149/HĐ-DCT.

References

[1] Hyojoo Son, Changwan Kim, "Short-term forecasting of electricity demand for the residential sector using weather and social variables", Resources, Conservation and Recycling, Vol 123, pp 200-207, Aug 2017.
[2] SpyrosMakridakis, "Accuracy measures: theoretical and practical concerns", International Journal of Forecasting, Volume 9, Issue 4, pp 527-529, Dec 1993
[3] Mariam Gul, Waqar A. Qureshi, "Long term electricity demand forecasting in the residential sector of Pakistan", 2012 IEEE Power and Energy Society General Meeting, 22-26 July 2012
[4] José Francisco Moreira Pessanhaa, Nelson Leon, "Forecasting long-term electricity demand in the residential sector", Procedia Computer Science, Vol 55, pp 529-538, 2015.
[5] Sunil Malla, Govinda R. Timilsina, "Long-Term Energy Demand Forecasting in Romania: An End-Use Demand", Policy Research Working Paper, 7697, Jun 2016.
[6] Francesco Apadula, Alessandra Bassini, Alberto Elli, Simone Scapin, "Relationships between meteorological variables and monthly electricity demand", Applied Energy, Vol 98, pp 346-356, Oct 2012.
[7] Chen, Z., Yang, S., Wang, X, "PLS-SVR optimized by PSO algorithm for electricity consumption forecasting", Applied Mathematics & Information Sciences, Vol 7, No 1L, 331–338, 2013.
[8] Bundanoon, P., Chalermyanont, K., Limsakul, C, "Multi-substation control central load area forecasting by using HP-filter and double neural networks (HP-DNNs)'', International Journal of Electrical Power & Energy System, Vol 44, Issue 1, pp 561-570, Jan 2013.
[9] Chang, P.C., Fan, C.Y., Lin, J.J., 2011, "Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach", International Journal of Electrical Power & Energy Systems Vol 33, Issue 1, pp 17–27, Jan 2011
[10] Adam, N.R.B., Elahee, M.K., Dauhoo, M.Z., 2011. Forecasting of peak electricity demand in Mauritius using the non-homogeneous Gompertz diffusion process, Energy 3, Issue 12, pp 6763–6769, Dec 2011.
[11] Phan Diệu Hương, *Phân tích và dự báo nhu cầu năng lượng*, NXB Bách Khoa Hà Nội, pp 298, 2015{12}
[12] Yong-Jin Lee, "Machine Learning Based Neighbor Path Selection Model in a Communication Network" Vol.10 No.1 55-61, Mar 2021
[13] Vietnam Electricity. htpp://*www.evn.com.vn,* May 2021
[14] General Statistics Office of Vietnam. htpp://*gso.gov.vn,* May 2021