COMPARISON OF THREE FORECAST METHODS FOR POWER DEMAND IN GAZİANTEP

Mehmet Oğuz HENGİRMEN
Electrical and Electronics Engineering Department
University of Gaziantep
27310 Gaziantep/TURKEY
E-mail : oguz@gantep.edu.tr

ABSTRACT

After privatisation of Electrical Power Systems, Power Plants and Turkish National Electrical Generating and Distribution Companies, the self-power generation of the private sector plants and companies became attractive in recent years. For this type of expensive establishments, a detailed forecast should be necessary indeed. In this study, the power consumption forecast for Gaziantep is made by using Least Square, Simple Exponential and Moving Average methods in the basement of last 5 years. All methods were compared and searched for the best one.

1. INTRODUCTION

The power of forecasting comes from the necessity to plan and equivalently the need for planning comes from the necessity to work today on activities that are intended to meet future demands.

Historical data is the best media which is basing estimates of future activity on the past performance of a product is one of the most commonly used and most reliable forecasting methods available. It has the advantage of being quantifiable and objective. The best method to use for a given set of data is the one that most closely appropriates is pattern. In essence we are using estimates of past performances as practice for estimating future performance. Historical data for our study is given in Table.1, [1].

2. FORECASTING METHODS

2.1. Least Square Method

A line, which serves to describe the relationship between two variables, is referred to as regression line; process is called simple regression analysis. The method of least square yields an equation, which describe two characteristics of the line of best fit. One is the intersection point on the y-axis,

Table.1. Power consumption in Gaziantep in months (1000xkWh).

	1994	1995	1996	1997	1998
JAN	98127	100154	136082	151724	174712
FEB	91906	116611	119313	136853	163649
MRC	88660	103071	139177	156531	180402
APR	88505	101341	121146	137475	137747
MAY	82438	91653	117395	145832	153028
JUNE	90570	99187	125784	147619	159492
JULY	101013	122713	136448	164680	181937
AGST	101117	117446	139822	186384	185038
SEPT	93312	111330	127521	132831	165804
OCT	92038	109461	137942	148012	151598
NOV	116803	126129	150892	168745	169227
DEC	119383	138854	160360	185597	188654
SUM	1164000	1337000	1611000	1862000	2011000

the other is the slope of the straight line,

$$y=a+bx$$

where, y is the calculated value dependent variable, a is the y-intercept of the line of best fit, b is the slope of the line of best fit, x is the given value of the independent variable normal equations are:

$$\sum y = N.a + b.(\sum x) \tag{1}$$

$$\sum(x.y) = a(\sum x) + b.(\sum x^2) \tag{2}$$

normal equations are: where, N is the number of given paired observations, and

$$a = \left[(\sum y).(\sum x^2) - (\sum x)\sum(xy)\right] / \left[N(\sum x^2) - (\sum x)^2\right] \tag{3}$$

$$b = \left[N\sum(xy) - (\sum x).(\sum y)\right] / \left[N.(\sum x^2) - (\sum x)^2\right] \tag{4}$$

However, one of the characteristics of the least square line is the sum of the deviations (Y- Yf) will always be equal to zero. Ones the values of the deviations have been calculated, sum of the squares of these values, i.e. $\Sigma(Y-Yf)^2$ which is the second characteristics of the least squares line will yield a minimum value for this sum, [2].

Simple Average Method:
When b in the straight line equation (Y =a + b.x) is equal to zero then the line will be levelled i.e. horizontal. The forecast for the next period then becomes simple average of all Y values to date;

$$Yf = \Sigma Y / N \qquad (5)$$

Simple average method can be considered as a special case of least square method. Average-value calculations are more often associated with seasonal variations, which takes place within an overall trend. Seasonal variations are limited to fluctuations during a single year. Therefore, one must collect data for several time periods within the year to determine seasonal patterns. A pattern of seasonal variations often remains relatively constant even though the trend may rise or fall. The pattern can be easily defined by dividing the simple average of each period by the sum of the averages. The resulting figure is a percentage estimate of the amount of activity expected during each period.

2.1.1 Mathematical Formulas of Least Square Method

In this section, the notation and methodology of regression methods applied to estimating the parameters of time series models are discussed. Initially, leastsquare analysis of a model with linear trend is presented, and then the general procedure for the case of several independent variables is described.

Simple Linear Regression

Many time series can be adequately described by a simple linear function of time. We may write this function as where b_1 and b_2 are the intercept and slope, respectively, and E_1 is the random deviation from the mean in time period t. We shall assume that this random error component has mean 0 and variance σ^2_ϵ, and is not correlated with random deviations in other periods. Expressed mathematically, our assumptions are $E(\epsilon_t) = 0$, $Var(\epsilon_t) = \sigma_\epsilon^2$, and $E(\epsilon_t \times \epsilon_{t+j}) = 0$ for $j \neq 0$. The mean of the time series process changes linearly with time, and at time t the mean is $b_1 + b_2t$. The parameters b_1 and b_2 are unknown constants, which we shall estimate by the method of least squares.
Assume that there are T periods of data available, say $x_1, x_2, \ldots, X_T$. Denote the estimators of b_1 and b_2 by $\underline{b}_1$ and $\underline{b}_2$. The fitted model is

$$X_t = \hat{b}_1 + \hat{b}_2 t \qquad (6)$$

and the difference between the fitted model and the data is called a residual, say

$$e_t = X_t - \hat{X}_t \qquad (7)$$

To estimate b_1 and b_2 by the method of least squares, we must choose b_1 and so that the error sum of squares

$$SS_E = \sum_{t=1}^{T} (x_t - b_1 - b_2 t)^2 \qquad (8)$$

is minimised. Therefore, it is necessary that $\hat{b}_1$ and $\hat{b}_2$; satisfy

$$\left.\frac{\partial SS_E}{\partial b_1}\right|_{\hat{b}_1, \hat{b}_2} = 0 \qquad (9)$$

and

$$\left.\frac{\partial SS_E}{\partial b_2}\right|_{\hat{b}_1, \hat{b}_2} = 0 \qquad (10)$$

This results in the following equations:

$$-2\sum_{t=1}^{T}(x_t - \hat{b}_1 - \hat{b}_2) = 0 \qquad (11)$$

$$-2\sum_{t=1}^{T}(x_t - \hat{b}_1 - \hat{b}_2)t = 0 \qquad (12)$$

Equations (11) and (12) may be rewritten as

$$\hat{b}_1 \sum_{t=1}^{T}(1) + \hat{b}_2 \sum_{t=1}^{T} t = \sum_{t=1}^{T} x_t \qquad (13)$$

$$\hat{b}_1 \sum_{t=1}^{T} t + \hat{b}_2 \sum_{t=1}^{T} t^2 = \sum_{t=1}^{T} t x_t \qquad (14)$$

Equations (13) and (14) are called the least-squares normal equations. Since

$\sum_{t=1}^{T} t = T(T+1)/2$ and $\sum_{t=1}^{T} t^2$, we may easily solve the normal equations for the estimators b_1 and b_2. The solution is

$$\hat{b}_1 = \frac{2(2T+1)}{T(t-1)}\sum_{t=1}^{T} x_t - \frac{6}{T(T-1)}\sum_{t=1}^{T} t x_t \qquad (15)$$

$$\hat{b}_2 = \frac{12}{T(T^2-1)}\sum_{t=1}^{T} t x_t - \frac{6}{T(T-1)}\sum_{t=1}^{T} x_t \qquad (16)$$

The estimators $\hat{b}_1$ and $\hat{b}_2$ depend upon the point in time at which they are computed (that is, T). Therefore, it's occasionally convenient to denote them as a function of time, say $\hat{b}_1(T) \equiv b_1$ and $\hat{b}_2(T) \equiv b_2$, where T is the time at which the estimates are computed.
The forecast, made at the end of period T of an observation in some future time period, say T + τ, would be denoted by $x_{T+\tau}(T)$, and is computed from

$$\hat{x}_{T+\tau}(T) = \hat{b}_1(T) + \hat{b}_2(T)[T+\tau] \qquad (17)$$

As a new observation becomes available, new estimates of the model parameters may be computed.
It is frequently useful to estimate the variance of the random error component, $\sigma^2_\in$.
If the linear trend model adequately describes the data, then we may estimate $\sigma^2_\in$ by

$$\hat{\sigma}^2_\in = \frac{\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}{T-2} \qquad (18)$$

2.2. EXPONENTIAL METHOD

Sometimes a smooth curve provides a better fit for data points than does a straight line. The equation for a curve may take the exponential form;

$$Y = a.b^x$$

where, Y changes at the constant rate. One can determine the values a and b by the least square method. Converting the exponential equation to its logarithmic form:

$\log Y = \log a + x . \log b$ (19)
$\Sigma \log Y = N. \log a + \Sigma(x. \log b)$ (20)
$\Sigma(x.\log Y) = \Sigma(x.\log a) + \Sigma(x^2.\log b)$ (21)
$\log a = \Sigma(\log Y) / N$ (22)
$\log b = \Sigma(x.\log Y)/\Sigma(x^2)$ (23)

and taking anti-log of a and b then finding a and b one can construct the $Y = a.b^x$ exponential equation.

2.3. MOVING AVERAGE METHOD

The moving average approach may be used to forecast future observations from a time series which has a linear trend

$x_t = b_1 + b_2 t + \in_t$ (24-a)

where b_1 and b_2 are unknown parameters and $\in_t$ is a random error component with mean 0 and variance $\sigma^2_\in$. At time T, the N-period simple moving average is

$$M_T = \frac{x_T + x_{T-1} + \ldots\ldots x_{T-N+1}}{N} \qquad (24\text{-b})$$

The expected value of M_T, assuming that observations come from the linear trend process (24-a), is

$$E(M_T) = \frac{1}{N}E(x_T + x_{T-1} + \ldots\ldots + x_{T-N+1} \qquad (25)$$

$$= \frac{1}{N}[E(x_T) + E(x_{T-1}) + \ldots\ldots + E(x_{T-N+1})]$$

$$= \frac{1}{N}[b_1 + b_2 T + b_1 + b_2(T-1) + \ldots + b_1 + b_2(T-N+1)]$$

$$= \frac{1}{N}[Nb_1 + Nb_2 T - \frac{N(N-1)}{2}b_2]$$

$$= b_1 + b_2 T - \frac{N-1}{2}b_2$$

or

$$E(M_T) = E(x_T) - \frac{N-1}{2}b_2 \qquad (26)$$

That is, the simple moving average M_T lags behind the observation at time T by an amount equal to $[(N-1)/2]b_2$. Therefore some method must be devised to correct for this lag. Consider a moving average of the moving averages, called a double moving average, say

$$M_T^{[2]} = \frac{M_T + M_{T-1} + \ldots + M_{T-N+1}}{N} \qquad (27)$$

where the bracketed superscript [2] denotes a second-order statistic, not a squared quantity. An alternate way to compute $M_T^{[2]}$ is

$$M_T^{[2]} = M_{T-1}^{[2]} + \frac{M_T - M_{T-N}}{N} \qquad (28)$$

which is analogous to Eqn. (24-b) for the simple moving average. It can be shown that

$$E(M_T^{[2]}) = E(x_T) - (N-1)b_2 \qquad (29)$$

Solving Eqns. (26) and (29) for b_1 and b_2, we obtain

$$b_1 = 2E(M_T) - E(M_T^{[2]}) - b_2 T \qquad (30)$$

$$b_2 = \frac{2}{N-1}[E(M_T) - E(M_T^{[2]})] \qquad (31)$$

Therefore, it seems logical to estimate b_1 and b_2 by

$$\hat{b}_1 = 2M_T - M_T^{[2]} - \hat{b}_2 T \qquad (32)$$

$$\hat{b}_2 = \frac{2}{N-1}(M_T - M_T^{[2]}) \quad (33)$$

The estimate of the observation in period T would be

$$x_T = \hat{b}_1 + \hat{b}_2 T = 2M_T - M_T^{[2]} \quad (34)$$

3. RESULTS AND CONCLUSION

By using consumed energy data since 1994 to 1997, we have calculated the forecast results for 1998 by three methods and compared with the original data of 1998 as seen in Fig.1. The individual results of Least Square Method, Exponential Method and Moving Average Methods are given in Table.2, 3 and 4 respectively. While comparing the results of the graphics in Fig.1, it is seen that the most similarity becomes with the original data and the moving average method results.

Table.2. Forecast results for next five years by using Least Square Method (1000xkWh)

	1999	2000	2001	2002	2003
Jan	193581	214055	234529	255003	275477
Feb	174784	191157	207530	223903	240276
Mch	204651	228345	252040	275734	299428
Apr	157628	171090	184551	198013	211475
May	176676	196212	215748	235284	254820
June	180413	199040	217668	236296	254923
July	202502	222884	243265	263647	284028
Ags	216971	240637	264303	287969	311635
Sep	176105	192753	209402	226050	242699
Oct	175111	190878	206645	222412	238179
Nov	190598	205344	220091	234837	249583
Dec	214155	232683	251212	269740	288269
Sum	2263000	2485000	2706000	2930000	3150000

Table.3. Forecast results for next five years by using Exponential Method (1000xkWh)

	1999	2000	2001	2002	2003
Jan	206235	241273	282263	330216	386317
Feb	183057	208762	238077	271508	309634
Mch	224182	269432	323815	389175	467728
Apr	165072	185926	209414	235870	265668
May	190989	226415	268414	377227	318203
June	192418	224216	261267	304442	354750
July	215085	249169	288654	334395	387386
Aug	233943	276442	326660	386001	456121
Sep	184414	210563	240420	274510	313444
Oct	185409	211145	240453	273829	311837
Nov	197270	218729	242522	268904	298155
Dec	224287	253018	285428	321990	363236
Sum	2402361	2775090	3207387	3673721	4232479

Table.4. Forecast results for next five years by using Moving Average Method (1000xkWh)

	1999	2000	2001	2002	2003
Jan	175922	186354	195889	204703	212924
Feb	161503	169713	177197	184054	190413
Mch	188335	201934	214506	226244	237280
Apr	148460	155503	161873	167708	173106
May	162000	172691	182520	191653	200210
June	165849	175633	184574	192837	200542
July	187057	197781	207565	216596	225008
Aug	199550	212557	224508	235605	245997
Sep	162343	170660	178209	185143	191575
Oct	165945	174726	182706	190047	196864
Nov	179336	186585	193102	199038	204502
Dec	201134	210746	219443	227411	234784
Sum	2097434	2214883	2322092	2421039	2513205

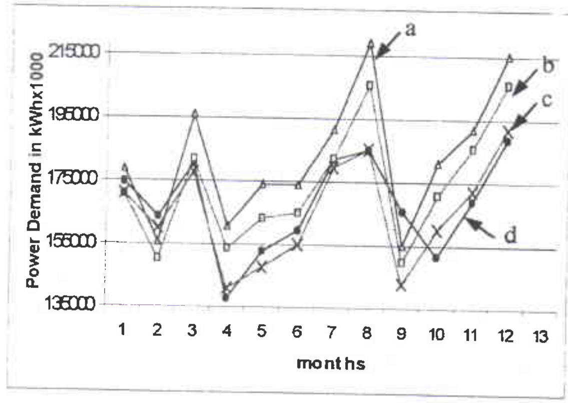

Fig.1. Comparison of three methods with real data of 1998.
a) Least Square Method
b) Exponential Method
c) Moving Average Method
d) Original Data

Consequently we can say that, the moving average method can be used for future power demands under the ubsence of unexpected conditions. That means, this is a forecast, it is not a guaranteed and exact result.

4. REFERENCES

1- Hengirmen, M.O., Kabak, S., " Gaziantep ve Yöresinde 5 Yıllık Elektrik Enerjisi İhtiyaç Tahminleri", EMO Conference in Gaziantep-TURKEY, 6-12 Sept. 1999.
2- Montgomery, D. C., Johnson, L. A., "Forecasting and Time Series Analysis", Book, McGraw-Hill Co. 1976.