Forecasting Final Energy Consumption using the Centered Moving Average Method and Time Series Analysis

Janina POPEANGĂ, Ion LUNGU
University of Economic Studies, Bucharest, Romania
janina.popeanga@yahoo.com, ion.lungu@ie.ase.ro

The forecasting of energy consumption has become one of the major fields of research in recent years. Accurate energy demand forecasting is essential in energy system operations and planning.
In this paper, we will describe a method to determine the information that is useful for a good forecasting. Further, we adopt the time series modeling approach to model final energy consumption in Romania using previous data of 2010 to 2013. This method is implemented using stored procedures, developed in Oracle PL/SQL programming language.
Finally, the developed model is compared for goodness of fit to the historical data and forecasting accuracy, and results are encouraging, showing that the forecast model is in control and is working correctly.
Keywords: *Forecasting, Energy, Centered Moving Average Method, Time Series, Accuracy*

1 Introduction

The analysis of temporal data and the prediction of future values of time series are among the most important problems that data analysts face in many fields, ranging from finance and economics, to production operations management or telecommunications. [1]
According to NIST/ITL (1997), time series is generally an ordered sequence of values of a variable at equally spaced time intervals.
Energy consumption recorded over a period of time at fixed interval is a classic time series modeling problem, which is generally used for forecasting.
The forecast for the energy consumption and power also is the scientific activity with the main purpose: the forecast for the energy consumption and power based on calculations analysis and based on the interpretation of different dates, so we will obtain a more precise concordance between the estimated consumptions and the one effectively realized. [2]
A forecast is a prediction of some future event(s).
Accurate final energy consumption forecasting is essential in energy system operations such as during startup and shut-down schedules of generating units as well as for fix planning and spot market energy pricing.
Energy consumption forecasting can be divided into three categories [3]:

- *short term forecasting* - predicts the load demand from one day to several weeks. It helps to estimate and allocate resources in a power grid to supply the demand continuously, to prevent overloading and so lead to more economic and secure energy system.
- *medium term forecasting* – provides information for power system planning and operations, predicting the consumption from a month to several years.
- *long term load forecasting* - predicts the final energy consumption from a year up to twenty years and it is mainly for system planning, allowing decision makers of a power supplying company to decide when to build new power plants, transmission and distribution networks.

2. Methodological framework

The methodology of elaboration of a forecast study for the energy consumption has few main steps [4]:

- collecting, selection and analyze the initial dates;
- establishing the mathematical model for the consumption;
- the analyze for the variance which has been obtained for the forecast and establishing the final decision.

This paper investigates the effectiveness of a model developed for time series forecasting. Here, we utilize the time series modeling approach to model final energy consumption in Romania using previous data of 2010 to 2013 shown in Table 1.

Table 1. Quarterly Data for Final energy consumption

Year	Quarter	Final Consumption
2010	1	13268.50
	2	11832.00
	3	12401.60
	4	13006.70
2011	1	13650.10
	2	12618.60
	3	12853.00
	4	13416.70
2012	1	13735.70
	2	13106.50
	3	12695.20
	4	13419.00
2013	1	13108.80
	2	11773.00
	3	11975.80
	4	12932.00

Any time series can contain some or all of the following components:
1. *Trend (T)* - Is the long term pattern of a time series;
2. *Cyclical (C)* – up and down movement repeating over long time frame;
3. *Seasonal (S)* - Seasonality occurs when the time series exhibits regular fluctuations during the same month (or months) every year, or during the same quarter every year;
4. *Irregular (I)* - In prediction, the objective is to “model" all the components to the point that the only component that remains unexplained is the random component.

LINEAR TREND EQUATION:

$$Y' = a + b*t$$

where:
Y' is the projected value of the Y variable for a selected value of t.
a is the Y-intercept. It is the estimated value of Y when t=0, so a is the estimated value of Y where the line crosses the Y-axis when t is zero.
b is the slope of the line, or the average change in Y' for each change of one unit in t.
t is any value of time that is selected.

In contrast to the least squares method, which expresses the trend in terms of a mathematical equation ($Y' = a + b*t$), the moving-average method “smooths” the fluctuations in the time series, to see its trend.
When calculating a moving average, placing the average in the middle time period makes sense. But, giving the fact that our data are quarterly, and since there are four quarters in a year, technically, the MA(4) would fall at $t = 2{,}5; 3{,}5; ....$
The first value that can be calculated for this series by a 4-period MA process would use observations X_1, X_2, X_3, and X_4. So, first 4-period average has a center between quarter 2 and quarter 3.
Thus we have:

$$X_{2,5} = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

The second MA values would use observations X_2, X_3, X_4 and X_5. So, second 4-period average has a center between quarter 3 and quarter 4:

$$X_{3,5} = \frac{X_2 + X_3 + X_4 + X_5}{4}$$

For the time series, the general formula is:

$$X_{t,5} = \frac{X_{t-1} + X_t + X_{t+1} + X_{t+2}}{4} \quad (1)$$

To avoid this problem, when we average an even number of values, we need to smooth the smoothed values. This method it is called *Double Moving Average for a Linear Trend Process*.

To get a 4-period double moving average that is centered at quarter 3 we take the average of $X_{2,5}$ and $X_{3,5}$:

$$X^{CMA}_3 = \frac{X_{2,5} + X_{3,5}}{2}$$

The general formula is:

$$X^{CMA}_t = \frac{X_{(t-1),5} + X_{(t+1),5}}{2} \quad (2)$$

From (1) and (2), results that:

$$X^{CMA}_t = \frac{1}{2} * \left(\frac{X_{t-2} + X_{t-1} + X_t + X_{t+1}}{4} + \frac{X_{t-1} + X_t + X_{t+1} + X_{t+2}}{4} \right)$$

=>

$$X^{CMA}_t = \frac{X_{t-2} + 2 * X_{t-1} + 2 * X_t + 2 * X_{t+1} + X_{t+2}}{8}$$

Table 2. Calculating *MA*(4) and *CMA*(4)

Year	Quarter	Final Consumption	MA(4)	CMA(4)
2010	1	13268.50		
	2	11832.00		
			12627.20	
	3	12401.60		12674.90
			12722.60	
	4	13006.70		12820.93

The specific seasonal for each quarter is calculated by dividing final consumption in column 3 by the centered moving average in column 5. The specific seasonal reports the ratio of the original time series value to the moving average. Algebraically, we compute final consumption/*CMA*(4) = *SI* and this result is the seasonal component.

Next we calculate a typical seasonal index for the corresponding quarters:

$$S_{t(Q1)} = AVG\ (SI_{(2011,Q1)}, SI_{(2012,Q1)}, SI_{(2013,Q1)})$$

$$S_{t(Q2)} = AVG\ (SI_{(2011,Q2)}, SI_{(2012,Q2)}, SI_{(2013,Q2)})$$

$$S_{t(Q3)} = AVG\ (SI_{(2010,Q3)}, SI_{(2011,Q3)}, SI_{(2012,Q3)})$$

$$S_{t(Q4)} = AVG\ (SI_{(2010,Q4)}, SI_{(2011,Q4)}, SI_{(2012,Q4)})$$

The set of typical indexes is very useful in adjusting the time series, for example, for seasonal fluctuations. Deseasonalizing the final consumption series is to remove the seasonal fluctuations so that the trend can be studied.

To remove the effect of seasonal variation, the final energy consumption for each quarter is divided by the seasonal index for that quarter:

$$Deseasonalize = final\ consumption\ /\ S_t$$

Next we determine the regression equation of the trend data and use it to forecast future energy consumptions.

$$Regression\ (Deseasonalize,\ t)$$

The trend equation is:

$$Trend_t = Intercept + Slope * t$$

$$Forecast_t = S_t * T_t$$

3. Implementing the method

In order to implement the method described, we create the following tables:

- **tb_FCons**

```
([t] [int] IDENTITY(1,1) NOT
NULL,
[Year] [int] NOT NULL,
[Quarter] [int] NOT NULL,
[Final_cons] [decimal](10, 2) NOT
NULL,
[CMA] [decimal](10, 2),
[SIt] [decimal](10, 2),
[St] [decimal](10, 2),
[Deseasonalize] [decimal](10, 2),
[Tt] [decimal](10, 2),
[Forecast] [decimal](10, 2))
```

- **tb_Regr**

```
([n] [int] NOT NULL,
[slope] [decimal](10, 2),
[intercept] [decimal](10, 2))
```

tb_FCons contains quarterly information on final energy consumption in Romania using previous data of 2010 to 2013.
tb_Regr table stores data about the indicators calculated by using the Linear Regression technique (intercept and slope), for n number of observations.
The main procedures that are used in the application are:

- *QUARTERLY_ANALYSIS* - for calculating the regression indicators and forecast actual values of the time series, in order to measure the accuracy of the model.
- *FORECASTING* – for forecasting final energy consumption for the following 'nQ' quarters (nQ – number of quarters, given as parameter).

All of them are shown below.

```
CREATE PROCEDURE
QUARTERLY_ANALYSIS
AS

DECLARE
      @nr int,
      @Xtim2 decimal(10,2),
      @Xtim1 decimal(10,2),
      @Xti decimal(10,2),
      @Xtip1 decimal(10,2),
      @Xtip2 decimal(10,2),
      @cma decimal(10,2),
      @SIt decimal(10,2),
      @StQ decimal(10,2),
      @Slope decimal(10,2),
      @Intercept decimal(10,2),
      @ti int =3,
      @i int =1

SELECT @nr=count(*) FROM tb_FCons;

WHILE @ti<@nr-1
    BEGIN
        SELECT @Xtim2=Final_cons
FROM tb_FCons WHERE t=@ti-2;
        SELECT @Xtim1=Final_cons
FROM tb_FCons WHERE t=@ti-1;
        SELECT @Xti=Final_cons FROM
tb_FCons WHERE t=@ti;
        SELECT @Xtip1=Final_cons
FROM tb_FCons WHERE t=@ti+1;
        SELECT @Xtip2=Final_cons
FROM tb_FCons WHERE t=@ti+2;

        SET @cma = (@Xtim2 +
2*@Xtim1 + 2*@Xti + 2*@Xtip1 +
@Xtip2)/8;
        SET @SIt = @Xti/@cma;

        UPDATE tb_FCons
        SET CMA=@cma, SIt=@SIt
        WHERE t=@ti;

        SET @ti=@ti+1;

      END

WHILE @i<=4
      BEGIN
         SELECT @StQ = avg(SIt)
         FROM tb_FCons
         WHERE Quarter=@i
         GROUP BY Quarter;

         update tb_FCons
         SET St=@StQ
         WHERE Quarter=@i;

         SET @i=@i+1;
      END

UPDATE tb_FCons
SET Deseasonalize=Final_cons/St;

SELECT
      @Slope = ((@nr *
sum(t*Deseasonalize) -
(sum(t)*sum(Deseasonalize)))/ ((@nr
* sum(Power(t,2)))-Power(Sum(t),2)),
      @Intercept = avg(Final_cons) -
((@nr * sum(t*Deseasonalize) -
```

```
(sum(t)*sum(Deseasonalize)))/((@n
r * sum(Power(t,2)))-
Power(Sum(t),2)) * avg(t)
FROM tb_FCons

UPDATE tb_FCons
SET Tt = @Intercept + @Slope*t,
    Forecast = St*Tt;

INSERT INTO tb_Regr VALUES
(@nr,@Slope, @Intercept);
```

The second algorithm was implemented by creating the temporary table *#Temp_FCons* that will capture the forecasting final energy consumption values for the period desired.

```
CREATE PROCEDURE FORECASTING
          @nQ int
AS

DECLARE
    @t int,
      @nr int,
      @y int,
      @q int,
      @StQ decimal(10,2),
      @Slope decimal(10,2),
      @Intercept decimal(10,2);

IF
object_id('tempdb..#Temp_FCons')
is not null DROP TABLE
#Temp_FCons

CREATE TABLE #Temp_FCons(
      [t] [int] NOT NULL,
      [Year] [int] NOT NULL,
      [Quarter] [int] NOT NULL,
      [St] [decimal](10, 2),
      [Tt] [decimal](10, 2),
      [Forecast] [decimal](10, 2))

SELECT @t=count(*)+1 FROM tb_FCons;
SELECT @y=max(YEAR) FROM tb_FCons;
SET @nr=@t+@nQ;

WHILE @t<=@nr-1

BEGIN

IF @t% 4=1 SELECT @y=@y+1, @q=1
ELSE IF  @t% 4=0 SELECT @y=@y, @q=4
      ELSE SELECT @y=@y, @q=@t % 4

SELECT @StQ = avg(SIt)
FROM tb_FCons
WHERE Quarter=@q
GROUP BY Quarter;

SELECT @Slope= slope,
       @intercept = intercept
FROM tb_Regr;

INSERT INTO #Temp_FCons (t, Year,
Quarter,St,Tt) VALUES
(@t,@y,@q,@StQ,@Intercept +
@Slope*@t);

SET @t=@t+1;
  END

  UPDATE #Temp_FCons
  SET Forecast = St*Tt;

  SELECT * FROM #Temp_FCons
```

4. Results and analysis

The execution of *QUARTERLY_ANALYSIS* procedure will produce the result set shown in *Table 2*.

Table 2. Forecasting final energy consumption using the QUARTERLY_ANALYSIS stored procedure

	t	Year	Quarter	Final_cons	CMA	SIt	St	Deseasonalize	Tt	Forecast
1	1	2010	1	13268.50	NULL	NULL	1.04	12758.17	12931.97	13449.25
2	2	2010	2	11832.00	NULL	NULL	0.96	12325.00	12921.99	12405.11
3	3	2010	3	12401.60	12674.90	0.98	0.97	12785.15	12912.01	12524.65
4	4	2010	4	13006.70	12820.93	1.01	1.02	12751.67	12902.03	13160.07
5	5	2011	1	13650.10	12975.68	1.05	1.04	13125.10	12892.05	13407.73
6	6	2011	2	12618.60	13083.35	0.96	0.96	13144.38	12882.07	12366.79
7	7	2011	3	12853.00	13145.30	0.98	0.97	13250.52	12872.09	12485.93
8	8	2011	4	13416.70	13216.99	1.02	1.02	13153.63	12862.11	13119.35
9	9	2012	1	13735.70	13258.25	1.04	1.04	13207.40	12852.13	13366.22
10	10	2012	2	13106.50	13238.81	0.99	0.96	13652.60	12842.15	12328.46
11	11	2012	3	12695.20	13160.74	0.96	0.97	13087.84	12832.17	12447.20
12	12	2012	4	13419.00	12915.69	1.04	1.02	13155.88	12822.19	13078.63
13	13	2013	1	13108.80	12659.08	1.04	1.04	12604.62	12812.21	13324.70
14	14	2013	2	11773.00	12508.28	0.94	0.96	12263.54	12802.23	12290.14
15	15	2013	3	11975.80	NULL	NULL	0.97	12346.19	12792.25	12408.48
16	16	2013	4	12932.00	NULL	NULL	1.02	12678.43	12782.27	13037.92

Fig.1 compares actual final energy consumption histories to forecasts for the entire time period analyzed.

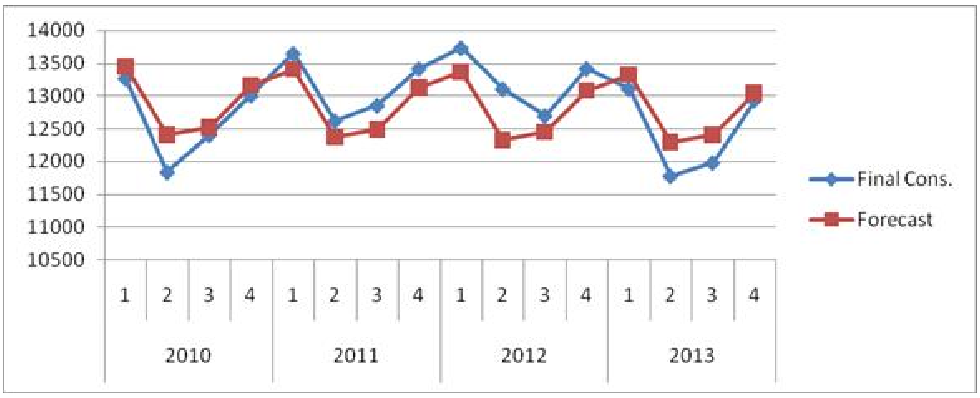

Fig. 1 – Actual and Forecast Final energy consumption (mill. kWh)

Further, we can call this store procedure using EXEC, and specifically specifying the parameter of the procedure:

```
EXEC FORECASTING @nQ=4
```

will display the results for the future 4 quarters (*Table 3*).

Table 3. Forecasting results

	t	Year	Quarter	St	Tt	Forecast
1	17	2014	1	1.04	12772.29	13283.18
2	18	2014	2	0.96	12762.31	12251.82
3	19	2014	3	0.97	12752.33	12369.76
4	20	2014	4	1.02	12742.35	12997.20

According to our analysis, the final energy consumption is expected to increase around 3% in the first quarter and then decrease by 7,77% in quarter Q2. Total energy consumption will increase from 12251,82 million kWh in Q2 to

12369,76 million kWh in Q3 and 12997,20 million kWh in Q4. (*Fig. 2*)

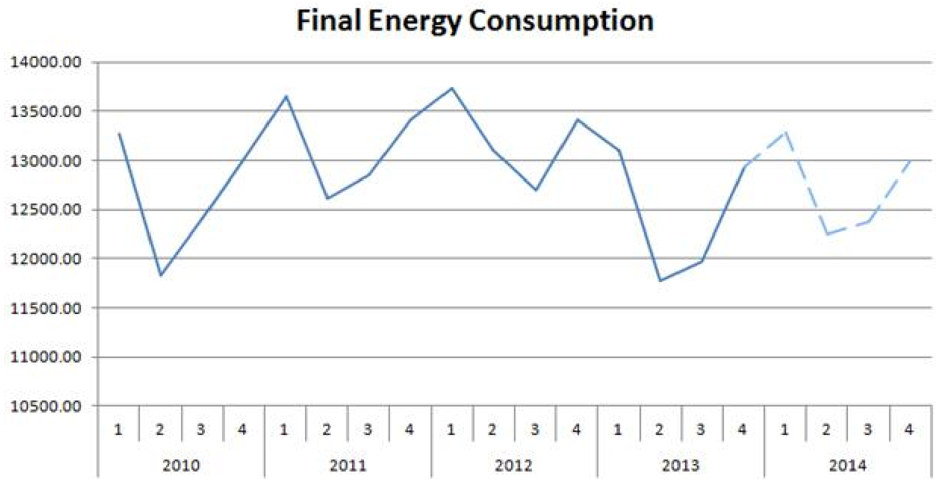

Fig. 2 – Final energy consumption (mill. kWh)

5. Evaluation of Forecasting Model

Forecast error = Difference between actual and forecasted value (also known as residual).

$$\textit{Forecast error} = \textit{Actual} - \textit{Forecast}$$

The Mean Absolute Deviation (MAD) is calculated by adding the absolute value of forecast errors in each period, and taking the average of this total.

$$MAD = \frac{\sum_{t=1}^{n} |A_t - F_t|}{n}$$

The Mean Squared Error (MSE) measures the average of the squares of the errors, that is, the difference between the actual energy consumption and what is estimated.

$$MSE = \frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}$$

The Tracking Signal indicates if the forecast is consistently biased high or low and is calculated as the ratio of cumulative error and MAD.

$$\textit{Tracking Signal} = \frac{\sum_{t=1}^{n} A_t - F_t}{MAD}$$

Tracking signal values are compared to predetermined limits (+4,-4) based on judgment and experience.

If TS > 4 or < -4 => investigate!

If TS ≥ 0 => most of the time actual values are above the forecasted values.

If TS < 0 => most of the time actual values are below the forecasted values.

We investigate our forecasting model by querying the tb_FCons table:

```
SELECT (Final_cons - Forecast) as
Error, abs(Final_cons - Forecast) as
AbsValue, power((Final_cons -
Forecast),2) as SError
FROM tb_FCons
```

Table 4. Forecasting accuracy

Error	AbsValue	SError
-180.75	180.75	32670.56
-573.11	573.11	328455.07
-123.05	123.05	15141.30
-153.37	153.37	23522.36
242.37	242.37	58743.22
251.81	251.81	63408.28
367.07	367.07	134740.38
297.35	297.35	88417.02
369.48	369.48	136515.47
778.04	778.04	605346.24
248.00	248.00	61504.00
340.37	340.37	115851.74
-215.90	215.90	46612.81
-517.14	517.14	267433.78
-432.68	432.68	187211.98
-105.92	105.92	11219.05

Further, we calculate the following indicators using the accuracy results obtained at the previous step:

$$MAD = \text{AVG}\ (AbsValue)$$

$$MSE = \text{AVG}(SError)$$

$$TS = \text{SUM}(Error)\ /\ MAD$$

Sum Error	MAD	MSE	TS
592.57	324.775625	136049.5788	0.004355545

Model tends to over-forecast, with an average absolute error of 324. 77 units.

TS control limits of ±2 to ±4 are used most frequently. Values outside this rage indicate that the model should be investigated and reevaluated.

Therefore, our example shows that this forecast is in control and this model is working correctly.

5. Conclusions

Forecasting the energy consumption is the scientific activity with the main objective: to obtain a more precise concordance between the estimated consumptions and the one effectively realized, based on calculations analysis and based on the interpretation of different dates.

In the first part of our investigation, we have collected, selected and analyzed the initial dates. The role of the Double Moving Average for a Linear Trend Process is explained and detailed.

Second, we have established the mathematical method for calculating the expected energy consumption values;

This method is implemented through stored procedures that calculate the regression indicators, forecast actual values of the time series, in order to measure the accuracy of the model and forecast final energy consumption for the following quarters.

Evaluation of forecasting model shows that the model developed works correctly.

References

[1] Martinez Alvarez, F. Troncoso, A., Riquelme, J.C.; Aguilar Ruiz, J.S., *Energy Time Series Forecasting Based on Pattern Sequence Similarity*, Knowledge and Data Engineering, IEEE Transactions on (Volume:23 , Issue: 8), 2011.

[2] Hamilton, J. "Time series analysis", Princeton University Press, 1994.

[3] Intan Azmira binti Wan Abdul Razak, Shah bin Majid, Mohd Shahrieel bin Mohd. Aras and Arfah binti Ahmad, "Advances in Data Mining Knowledge Discovery and Applications", Chapter 11 - Electricity Load Forecasting Using Data Mining Technique, ISBN 978-953-51-0748-4, Published: September 12, 2012

[4] P. Mihai, M. Popescu, "Mathematical Models used in Quality Management of the Electrical Energy"

Janina POPEANGĂ graduated in 2010 the Faculty of Cybernetics, Statistics and Economic Informatics, Economic Informatics specialization. The title of her Bachelor's thesis is "*Distributed Databases*". In 2012, she graduated the Databases for Business Support master program with the thesis "*Monitoring and management of electric power consumption using sensorial data*". Janina's interests are broadly in the fields of databases and distributed systems. Since 2012 she is a Ph.D. Student in the Doctoral School of Bucharest Academy of Economic Studies. Her research focuses on real-time database systems, business intelligence analytics, sensor data management, smart grid and renewable energy.

Ion LUNGU graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 1974. He got the title of doctor in economy in the specialty economic informatics in 1983. He has been directing graduates who study towards getting a doctor's degree since 1999. At present he is a professor in the department of the faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies of Bucharest. He had documentary activity and specialization with the Eindhoven Technical University of Holland, the Economic University of Athens and Economic University of Milan. His domains of work are: informatics systems and databases. Among his books are: "Databases, organization, design and implementation", (1995), "Information Systems for Management" (1994), "SGBD Oracle Applications" (1998); "Let's learn Oracle in 28 lessons" (2003), "Database systems" (2003), "Information Systems – Analysis, Design and Implementation" (2003).