# Mathematical Model of Petroleum Consumption in Thailand

Sirisup Laohakiat

Based on historical data, the paper develops a mathematical model to describe characteristics of petroleum consumption in Thailand. The data is divided into 3 components: trend, seasonal and residue. The trend component is formulated as a transfer function model with various key economic factors as input of the system. The optimization technique based on sequential quadratic programming (SQP) is applied to identify the parameters of the model. The detrended data is characterized by its frequency spectrum then the development of the seasonal component follows. Finally, the Box-Jenkin approach is applied to the residue to develop a stochastic ARMA model.

**Key words:** Petroleum consumption model, time-series analysis and component model.

*Department of General Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.*
*Correspondence to: e-mail: Sirisup.L@chula.ac.th*

# แบบจำลองทางคณิตศาสตร์ของการบริโภคปิโตรเลียมในประเทศไทย

ศิริสรรพ เหล่าหะเกียรติ (2549)
*วารสารวิจัยวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย* **31(1)**

บทความนี้พัฒนาแบบจำลองทางคณิตศาสตร์ของการบริโภคน้ำมันปิโตรเลียมในประเทศไทย โดยแบ่งข้อมูลอนุกรมเวลาออกเป็นสามส่วนประกอบ ได้แก่ ส่วนแนวโน้ม ส่วนที่มีคาบเวลาเกิดซ้ำ และส่วนตกค้าง ข้อมูลในส่วนแนวโน้มจะจำลองด้วยแบบจำลองแบบฟังก์ชันโอนย้ายที่มีข้อมูลทางเศรษฐศาสตร์ตัวอื่นเป็นสัญญาณเข้า จากนั้นจะใช้วิธีการหาค่าเหมาะสมที่สุดมาประมาณพารามิเตอร์ของแบบจำลอง ข้อมูลส่วนที่มีคาบเวลาจะวิเคราะห์หาสเปกตรัมของข้อมูลและจำลองด้วยแบบจำลองแบบฤดูกาล ส่วนข้อมูลในส่วนตกค้างจะใช้แบบจำลองแบบ ARMA ของ Box และ Jenkin

**คำสำคัญ** แบบจำลองการบริโภคน้ำมันปิโตรเลียม การวิเคราะห์อนุกรมเวลา แบบจำลองแบบส่วนประกอบ

## INTRODUCTION

Petroleum has played a vital role in the world economy. In spite of the discovery of various alternative sources of energy, petroleum remains the main energy resource of the world. In Thailand, petroleum consumption has grown in accordance with the expansion of social and economic activities. As Thailand has to import petroleum and considering the drastic increase of petroleum prices in recent years, the amount of petroleum consumption becomes one of the major indices for making economic policy.

To understand the nature of petroleum demand, the historical data of petroleum consumption was analyzed and some deterministic characteristics were extracted. Then, a mathematical model was formulated according to the characteristics of the data. After verification of the model, it could be used for forecasting future demand which will play an important role in economic and energy policy.

### Source of data

For analyzing and testing, the model the data used were those related to historical petroleum consumption, gross national product (GNP), electric consumption and automotive sales. The data is the monthly data from 1990 to 2004.

Historical petroleum consumption, GNP and electric consumption were provided by the Energy Policy and Planning Office (EPPO) (www.eppo.go.th). Automotive sales were sourced from the Thailand automotive institute.

The data used to formulate the model was from 1990-2002, data for 2003 and 2004 will be used to calculate the forecasting error.

### The component time series model

There are many techniques proposed for modeling time series data including the component time series model,(1-5) stochastic ARMA model,(6) neural network and fuzzy model.(7)

The nature of the petroleum consumption amount would display both trend and seasonal effects which can be noted. For example, during the economic crisis of Thailand in 1999, the consumption decreased. This can be regarded as a trend of the data. Moreover, during each year economic activity and seasonal environment also affect the amount of consumption. For example, during winter there is less consumption compared with summer time use. As a result, a component time series model was chosen to describe and model the behavior of the data.

The component time series model has been used extensively in modeling and prediction of time series data. Various types of data have been used for modeling including air pollution(3,5) and electricity load.(4)

The advantage of using this model is that the data can be separated and modeled independently. Thus, the model requires less computational effort. Moreover, the data is modeled according to a predefined structure that is easy to understand compared with the black-box model. After the model is formulated, the characteristics of the data such as trend, relationship with other factors and the periodic component can be seen clearly.

The component time series model is described as

$$Y(t) = T(t) + S(t) + \varepsilon_d(t) + \varepsilon(t)$$

where

$Y(t)$ denotes original time series data

T(t) denotes a low frequency component which represents trend of the series,

S(t) is a periodic component,

$\varepsilon_d(t)$ is random noise output generated from deterministic system,

$\varepsilon(t)$ is uncorrelated white noise.

### Trend component

Trend component is a part of the low frequency data. This part does not have a seasonal change, so it is treated constant along one year. To extract this part from the data, many techniques have been proposed.(1) However, none of those methods yields a superior result. So, the one-year average was chosen for the sake of simplicity. The trend component is shown in Figure 1.

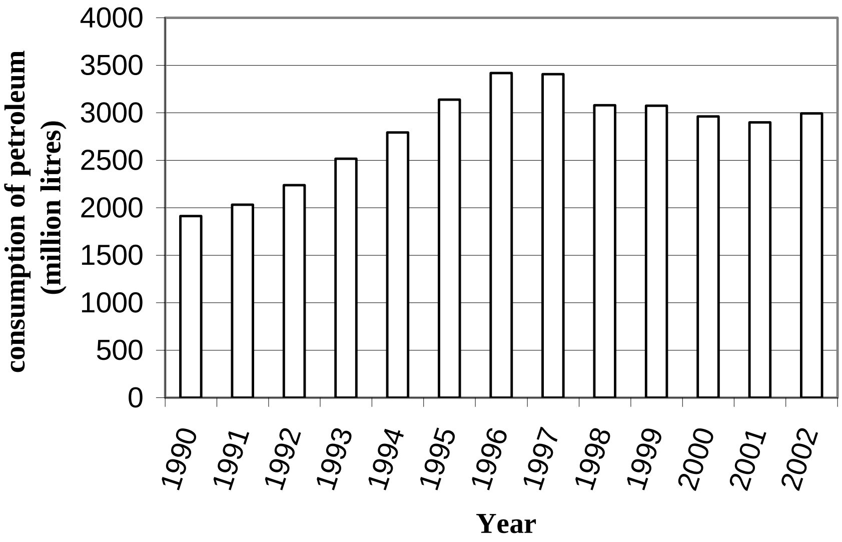


**Figure 1. The average petroleum consumption from 1990-2002.**

The trend of petroleum consumption has a close relationship with economic activity. To model the relationship between the consumption and other socio-economics indices, the correlation was tested to determine the effect of other indices on the consumption data by a covariance diagram. The data having relevancy with petroleum consumption, namely automotive sales, GNP and electricity consumption were chosen. Electricity consumption was chosen because it reflects petroleum consumption in electricity generation, one of major industrial uses of petroleum fuel.

Covariance is estimated by calculating covariance from

$$\mathrm{cov}_{xy}(k) = \sum_{i=1}^{M} (x(i) - \overline{x})(y(i-k) - \overline{y})$$

where $\mathrm{cov}_{xy}(k)$ is the covariance at lag k, M is the length of the data, x and y are data to find covariance. $\overline{x}$ and $\overline{y}$ are mean of the data. All data are taken in logarithmic function.

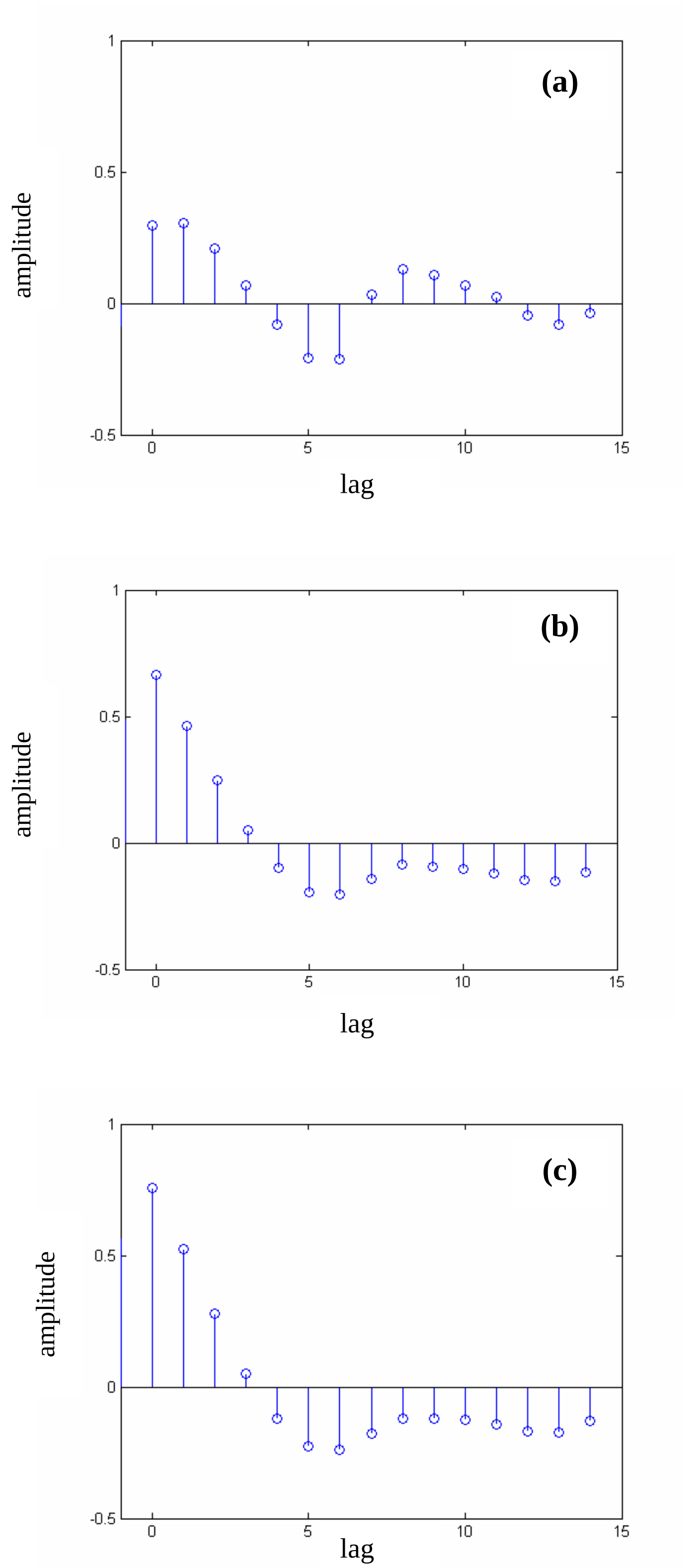


**Figure 2. Covariance diagrams between (a) automotive sales, (b) petroleum consumption and GNP and (c) petroleum consumption and electricity consumption.**

In Figure 2a we can see that the present and lag-one of automotive sales have almost the same effect on the value of petroleum consumption. Figures 2b and c show that the present data of both GNP and electricity consumption have a higher interaction with the petroleum consumption data than that of lag-one data.

The relationship between petroleum consumption and other socio-economic variables is formulated in a regression model. The model adopts conventional macroeconomic analysis which relate the logarithmic function of data as

$$\log\hat{T}(y) = a_1\log T(y-1) + b_1\log a(y) + b_2\log a(y-1) + c_1\log g(y) + c_2\log g(y-1)$$
$$+d_1\log e(y) + d_2\log e(y\text{-}1)$$

where

$\hat{T}(y)$ = predicted petroleum consumption amount in year y
T(y-1) = petroleum consumption amount in year y-1
a(y) = automotive sales in year y
a(y-1) = automotive sales in year y-1
g(y) = GNP in year y
g(y-1) = GNP in year y-1
e(y) = electricity consumption in year y
e(y-1) = electricity consumption in year y-1

From the covariance diagrams in Figures 2a, b and c, it can be inferred that

1) The effect of present and lag-one data of automotive sales are equivalent, so $b_1 = b_2$
2) The effect of g(y) and e(y) are greater than that of g(y-1) and e(y-1) respectively, so we have $c_1 \geq c_2$ and $d_1 > d_2$
3) All variables have a positive effect on x(y), thus all parameters are positive

With the above constraints we can not identify the parameters by using the least square criterion. We have to adopt sequential quadratic programming (SQP)[8] to minimize the total error, $e = \sum_{i=1}^{k}[T(y) - \hat{T}(y)]^2$, where k is the number of years in the past selected for model synthesis. The SQP is performed under the following constraints

$$c_2 - c_1 \leq 0$$
$$d_2 - d_1 \leq 0$$
$$-a \leq 0,\ -b \leq 0,\ \text{-}c_1 \leq 0,\ \text{-}c_2 \leq 0,\ \text{-d}_1 \leq 0 \text{ and } \text{-d}_2 \leq 0$$

The solution for parameter [a, b, $c_1$, $c_2$, $d_1$, $d_2$] = [0.367, 0.070, 0.3477, 0, 0.042, 0]. Figure 3 shows the result of the model compared with actual value from 1990-2002.

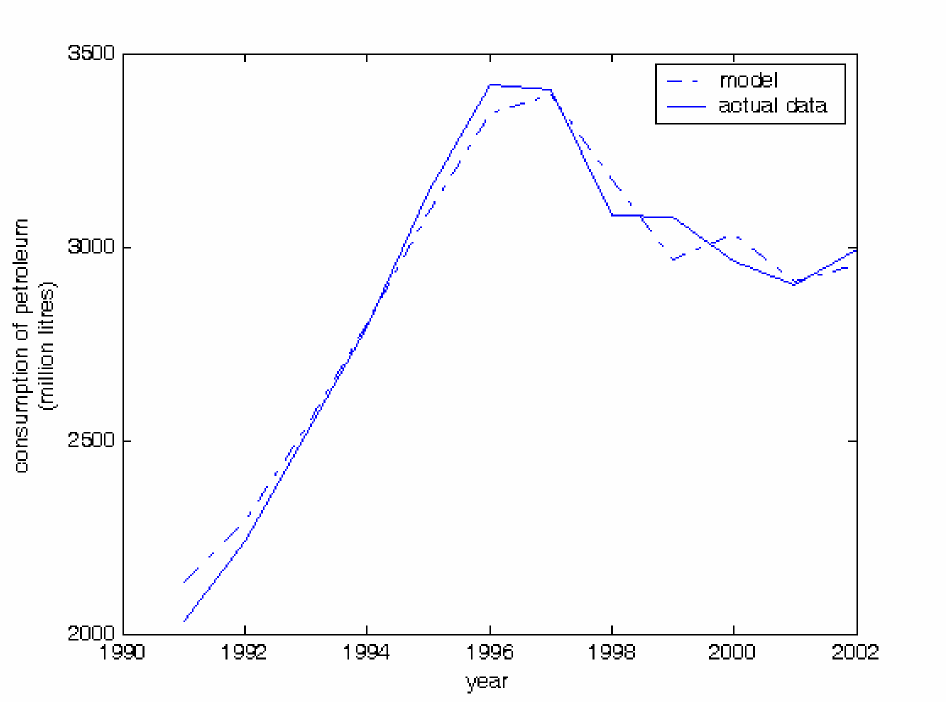


**Figure 3. The result of the model compared with actual data of petroleum consumption.**

From the trend component, the detrended data X(m) is derived by subtracting the original data from the data in the model as $X(m) = Y(m) - T(m) = S(m) + \varepsilon_d(m) + \varepsilon(m)$ by treating the trend T(m) as monthly data which is constant along each year.

**Seasonal component**

The seasonal component is the periodic part to the data. To extract the seasonal component, the frequency spectrum of the data must be estimated by periodogram. Then, the seasonal model is formulated as

$$S(m) = \sum_{k=1}^{L} s_k(m)$$

Where

L is the number of periodic components in the data, m is month, $s_k(m)$ is the k$^{th}$ periodic component at the m$^{th}$ month. Each $s_k(m)$ has periodic property, so the sum of one cycle would become zero, or it can be written as $\sum_{m=1}^{T_k} s_k(m) = 0$ where $T_k$ is the period of the k $^{th}$ cycle. From the detrended data X(m), the seasonal component is derived by the following procedure.

1) Find the number of cyclic components and their period length of X(m) from a periodogram. There would be L cycles, where $T = \{t_1, t_2, ..., t_L\}$ denotes period length of each cycle.
2) For each $t_k \in T$ the seasonal component $s_k(m)$ is formulated by partitioning X(m) into parts with length $t_k$. Then align the partitioned data into a 2-dimensional array x(i,j) such that x(i,j)= X(j.$t_k$+i) where i = 1, 2, ..,$t_k$, j = 1, 2,. ., $\lceil M / t_k \rceil$. We would have a table of data that is $t_k$ rows and $\lceil M / t_k \rceil$ columns where M denote the length of X(m).
3) Calculate mean of each row by
$$\bar{X}(i) = \frac{\sum_{j=1}^{C} x(i,j)}{C} \text{ where } C = \lfloor M / t_k \rfloor$$
4) Formulate the periodic series $s_k(m)$ from $\bar{X}(i)$ such that
$$s_k(m) = \bar{X}(i) \text{ where } i = m - t_k \left\lfloor \frac{m}{t_k} \right\rfloor = \text{mod}(m, t_k)$$
5) Subtract $s_k(m)$ from X(m), the cyclic component of the period $t_k$ would be extracted from the original series.

For the consumption data, the periodogram shows four peaks at

$$f = \frac{1}{T} = 0.42,\ 0.25,\ 0.17 \text{ and } 0.085 \text{ month}^{-1}$$

as shown in Figure 4.

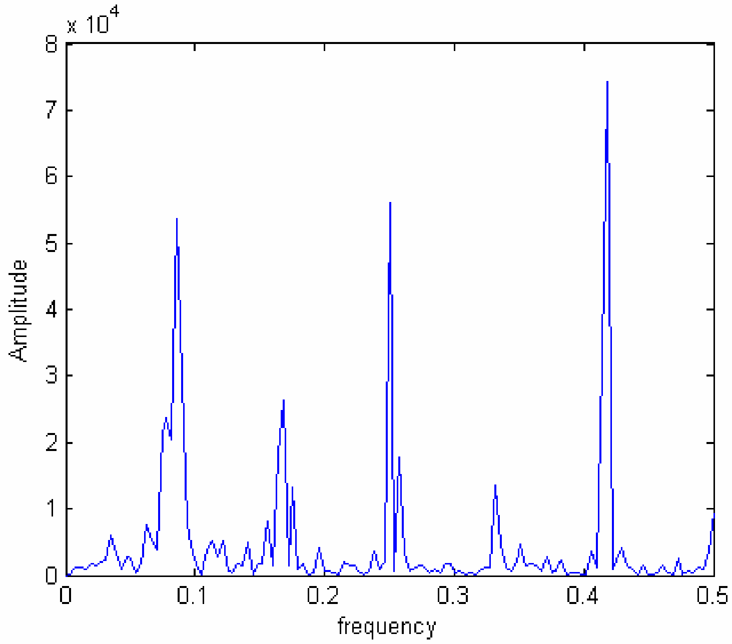


**Figure 4. The periodogram of detrended data.**

So there are 4 cycles with the period of 3, 4, 6 and 12 months respectively. So we have L = 4 and T = {3, 4, 6, 12}. After formulating the seasonal component, it appears that a peak at the period around 12 months still remains in the periodogram, so two more periodic series with t=11 and 13 months are formulated and subtracted from the detrended data. Figure 5 shows the frequency spectrum of the data after extracting all seasonal components; it can be seen that the periodogram does not significantly exhibit a peak so this data can be treated as a residue component that combines deterministic and random noise. Figure 6 shows the residue component.

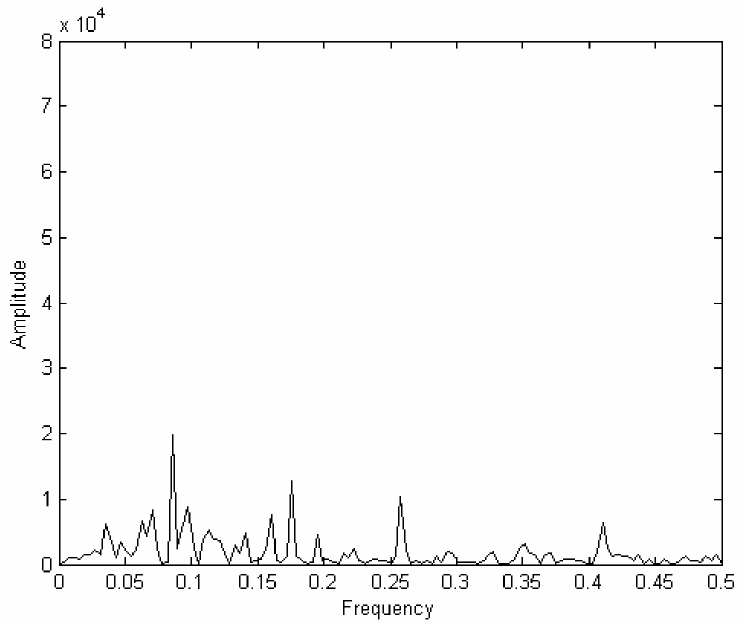


**Figure 5. The result of seasonal component after combining all periodic component series.**

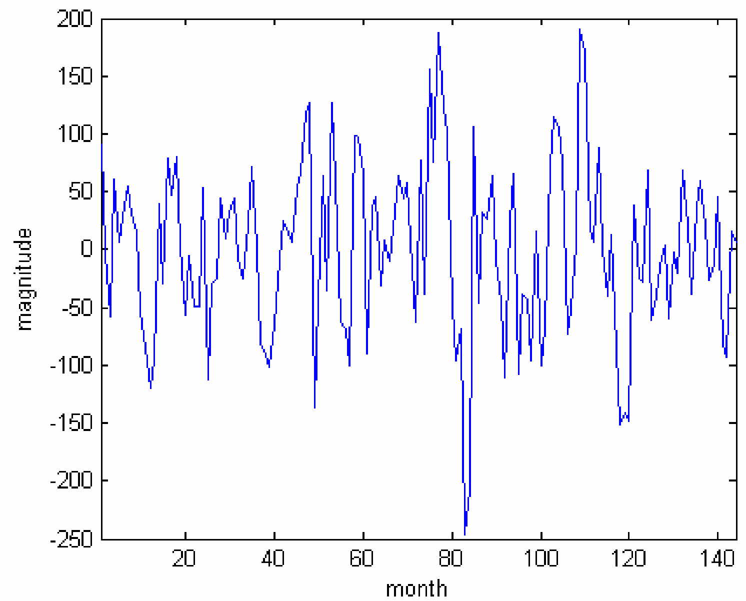


**Figure 6. The residue component.**

**Residue component**

The residue component was modeled further by the Box-Jenkin method in order to extract the deterministic component $\varepsilon_d(t)$ from the random noise. The correlation diagram is plotted as shown in Figure 7. The correlation diagram yields high magnitude from lag-one to lag-four. This data can not be treated as random noise because there exists dependency among the data.

The ARMA(4,4) model is selected for modeling the data. The parameters of the model are identified as

$$y(t) - 0.1837y(t-1) - 0.57y(t-2) - 0.22y(t-3) + 0.22y(t-4)$$
$$= e(t) + 0.126e(t-1) - 0.58e(t-2) - 0.456e(t-3) - 0.11e(t-4)$$

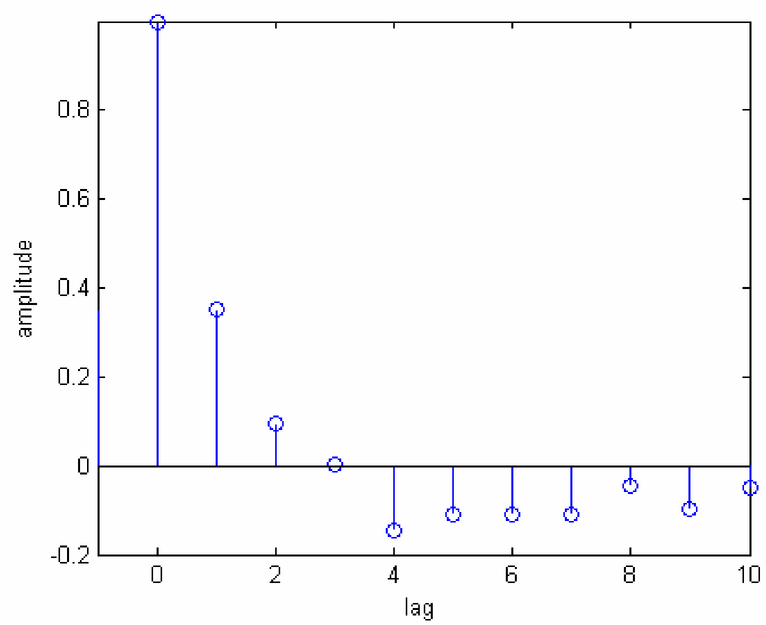


**Figure 7. Correlation diagram of residue data.**

**Model evaluation**

The model was validated by using data from January 2003 to December 2004. The comparison between actual data and predicted data is shown in Figure 8.

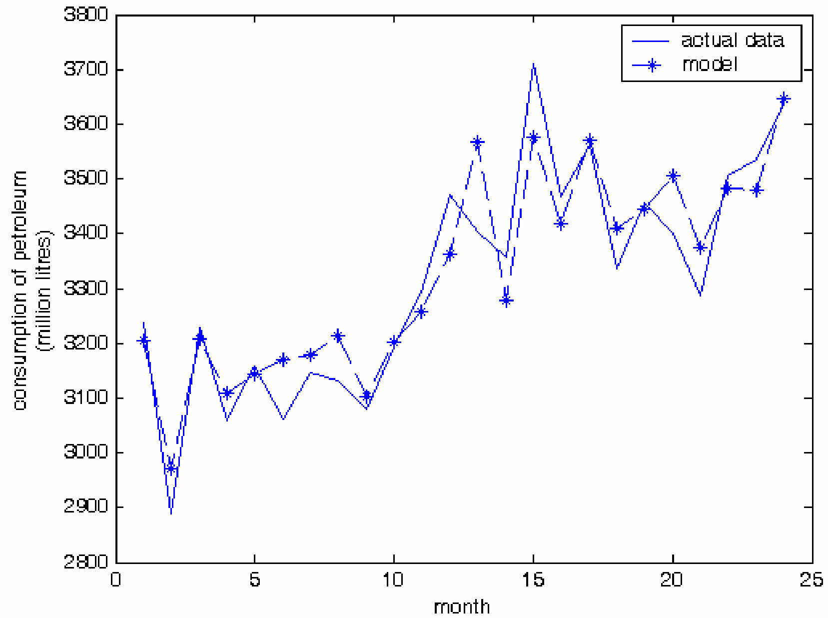


**Figure 8. Petroleum consumption from January 2003 to December 2004: comparison between actual and predicted data in 1-month ahead prediction. The monthly data is shown.**

Two indices are used for evaluating the model. Correlation coefficient between actual data and the result from the model is 0.92. Mean square error (MSE) which is computed from

$$MSE = \frac{1}{N}\sum_{m=1}^{N}\left(\frac{(Y(m)-\hat{Y}(m))}{Y(m)}\right)^2 \times 100\%$$

where $\hat{Y}(m)$ denotes the predicted value, $Y(m)$ denotes actual data, N denotes the number of data. From the model, MSE was calculated as 0.00044%. The MSE result was quite low because the data is separated into component models and the magnitude of data in the trend component model is much bigger than those of other models, so if the trend component can be predicted quite accurately the other component would yield less significance in calculating MSE.

**CONCLUSIONS**

In this paper a mathematical model for petroleum consumption in Thailand is formulated. The trend component model reveals the relationship between petroleum consumption and other socioeconomic data. The seasonal model shows that the consumption amount has periodicity around 3-4 months, 6 months and 12 months according to the quarter of economic activity. The residue data is best fitted with the ARMA(4,4) model.

The model developed for the component model shows the relationship and some characteristics of the behavior of petroleum consumption in Thailand. From the trend component we can see the impact of each socioeconomic index on the data. GNP is the most significant factor on the amount of consumption while other indices yield less impact. From the seasonal model, the cyclic pattern of the data is revealed. In each year the data repeats in the period of 3, 4, 6 months as well as for the whole year. So our assumption about the periodicity of the data is confirmed.

The model can be used for forecasting as well. Each year, there are forecasts for the value of GNP and other socioeconomic indices from various government organization. Those data can be used to predict the amount of petroleum consumption for that year through this model.

**ACKNOWLEDGMENT**

The author would like to express his gratitude to the reviewers for a number of helpful comments and suggestions which improved the quality to this paper.

**REFERENCES**

1. Kendall, M. G. (1990), "Time series", Sevenoaks, Kent, England.
2. Fu, C. W and Nguyen, T. T. (2003) "Models for long-term energy forecasting" Power engineering society meeting, IEEE, 13-17.
3. Brockwell, P. J. and Davis R. A. (1987) "Times series: theory and method" Springer Verlag, NY.
4. Ng, C. N. and Yan, T. L., (2004) "Recursive estimation of model parameters with sharp discontinuity in non-stationary air quality data", *Environmental Modelling & Software* **19**, 19-25.
5. Park, J. H., Park, Y. M. and Lee, K. Y., (1991) "Composite modeling for adaptive short-term load forecasting" *IEEE Transaction on Power systems*, **6(2)**, 450-457.
6. Karatzas, K. (2002) "Statistical analysis of environmental data as the basis of forecasting" The science of the total environment, **288,** 227-237.
7. Jung, L. (1999) "System identification-theory for user" Prentice Hall PTR.
8. Arora, J. (1989) "Introduction to optimum design" McGraw-Hill, Singapore.

*Received: May 31, 2005*
*Accepted: September 26, 2005*