Electrical Energy Consumption Forecasting Based on Cointegration and a Support Vector Machine in China

ZHANG XING-PING, GU RUI
School of Business Administration, North China Electric Power University
No.2 Beinong Road, Zhuxinzhuang, Dewai, Beijing 102206
CHINA
Zhangxingping302@163.com

Abstract: - By undertaking a cointegration analysis with annual data over the period 1985–2005 in China, the estimation results show that there is cointegration relationship between electrical energy consumption and economic growth taking into account industry structure changes and technical efficiency. The model shows that three explanatory variables, the GDP per capita, heavy industry share and efficiency improvement are the crucial factors which influence the electric energy consumption. The three explanatory variables and the actual electrical energy consumption are input into a support vector machine (SVM), a Gaussian radial basis function is taken as the kernel function and electrical energy consumptions from 1994–2006 are forecasted. The forecast results prove that the multivariable SVM is valid in forecasting electrical energy consumption in China.

Key-Words: Cointegration analysis; Electrical energy consumption; Johansen cointegration test; Multivariate time series; Support vector machine ; Unit root test

1 Introduction

Electrical consumption forecasting is the basis for electric energy planning. Many scholars [1~5] have applied econometrics to study electricity demand and its main determining factors is usually analyzed correctly in theory, but it is greatly affected by fluctuations in the sample data. A lot of non-linear programming and combinational forecasting methods such as fuzzy logic methods are applied widely in electric load forecasting. But results produced by fuzzy logic methods are quite difficult to express and set up, and the parameters are not easy to modulate [6, 7]. A new machine learning technique called support vector machines (SVM) is not only helpful for solving problems involving small sample, devilish learning, high dimension and local minima, but also strong generalizability. So SVM was widely applied in electric load forecasting, and some research results [8~12] indicate that SVM has distinct advantages in electric load forecasting. SVM is seldom used in forecasting the electrical energy consumption, and when it is, actual electrical energy consumption is taken as the only input variable of the SVM, while the major factors which impact electrical consumption are not considered [13].

In this paper, the three variables, GDP per capita, heavy industry share and efficiency improvement, are taken as the explanatory variables, and the electrical energy consumption is taken as the explained variable. An equilibrium relationship between the explanatory and the explained variables is analyzed by the cointegration analysis. Taking these crucial factors and actual electricity consumption as the input variables of a SVM, and selecting the rational kernel function of the SVM, the electrical energy consumption is forecasted.

2 Multivariate Cointegration Analysis of Electrical energy Consumption

2.1 Cointegration Theory

Cointegration theory seeks to determine whether there is a stationary relationship among nonstationary economic variables, and whether there is a long-term equilibrium relationship among them. It avoids the disadvantages of unreliable regression results generated by spurious regression, and it can differentiate long-term stationary relationships from short-term dynamic relationships among variables. Before cointegration analysis came along, the combination of variables had to be stationary. The variable autoregression model, which includes $g$ variables and $k$ lags, is expressed as:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + \mu_t \quad (1)$$
Supposed all $y_i$ are I(1); then a suitable transformation of equation (1) is made, and the error correction model is obtained as:

$$\Delta y_t = \Pi y_{t-k} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \mu_t$$

(2)

where $\Pi = \sum_{j=1}^{g} \beta_j - I_g$, $I_g$ is the g-step unit matrix, and $\Gamma = \sum_{j=1}^{g} \beta_j - I_g$.

Matrix $\Pi$ is the coefficient matrix which reflects the long-term relationships of the variables. When the variables are in a long-term equilibrium state, the difference in the first variables of equation (2) is the zero vector, and $E(\mu_t) = 0$; so $\Pi y_{t-k} = 0$ when the variables are in a long-term equilibrium state, and this can be judged by calculating the rank and the eigenvalues of matrix $\Pi$.

When all the endogenous variables are I(1), and when all the variables of $\Pi y_{t-k}$ are I(0), the stochastic error term is a stationary process. If $0 < \text{Rank}(\Pi) = m < g$, there are matrices $\alpha$ and $\beta$, and $\Pi = a\beta^T$. So equation (2) is transformed into equation (3).

$$\Delta y_t = a\beta^T y_{t-k} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \mu_t$$

(3)

Each row of the matrix $\beta^T y_{t-k}$ is a stationary combined variable, that is, each row is a linear combined form which enables the variables $y_{1,t-1}, y_{2,t-1}, \ldots, y_{g,t-1}$ to be cointegrated.

2.2 Explained and Explanatory Variables

Lots of documents show that GDP plays the most important role in determining electricity consumption in China. Thus there is a positive correlation between electrical consumption and GDP. Taken into account the population, the GDP per capita is taken as an explanatory variable.

In China, the share of industrial electricity consumption is rising from 71.75% in 2000 to 74.89% in 2006. The most of the electrical energy is consumed by the heavy industry, in 2006 for example, electrical energy consumed by heavy industrial took up 60.26% of all electrical energy consumption, and 79.71% of all industrial electrical energy consumption. The breakdown of electrical energy consumption has been changing in China; electrical energy consumption by the light industry increased 1.87% and by heavy industry decreased 0.14% in 2006. So the heavy industry share or the ratio of heavy industry production value to gross industry production reflects changing industrial structure. So the heavy industry share is a key factor which influences the electrical energy consumption, and is taken as an explanatory variable.

As the science and technology level has steadily increased since 1997, the comprehensive social and technology level index increased by 1.5% in 2006 to 47.11%. Consequently, efficiency improvement plays an important role in electrical energy consumption; so the ratio of increase in industrial value to industrial electricity consumption is used to reflect efficiency improvement.

So electricity consumption ($Q$) is chosen as the explained variable, and GDP per capita ($PCGD$), heavy industry share ($HIS$), and efficiency improvement ($EI$) are chosen as the explanatory variables. The sample space is from 1985 to 2005. The impact of inflation is removed, and the samples are shown in table 1.

### Table 1 Sample Data from 1985 to 2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Q (10 million KWH)</th>
<th>PCGD (yuan)</th>
<th>HIS</th>
<th>EI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>4705.9</td>
<td>858</td>
<td>52.6</td>
<td>1.178</td>
</tr>
<tr>
<td>1986</td>
<td>5096.4</td>
<td>963</td>
<td>52.4</td>
<td>1.231</td>
</tr>
<tr>
<td>1987</td>
<td>5514.3</td>
<td>1112</td>
<td>51.8</td>
<td>1.311</td>
</tr>
<tr>
<td>1988</td>
<td>5956.0</td>
<td>1366</td>
<td>50.7</td>
<td>1.509</td>
</tr>
<tr>
<td>1989</td>
<td>6390.8</td>
<td>1519</td>
<td>51.1</td>
<td>1.566</td>
</tr>
<tr>
<td>1990</td>
<td>6895.7</td>
<td>1644</td>
<td>50.6</td>
<td>1.584</td>
</tr>
<tr>
<td>1991</td>
<td>7399.1</td>
<td>1893</td>
<td>51.6</td>
<td>1.750</td>
</tr>
<tr>
<td>1992</td>
<td>7991.0</td>
<td>2311</td>
<td>53.4</td>
<td>2.017</td>
</tr>
<tr>
<td>1993</td>
<td>8590.4</td>
<td>2998</td>
<td>53.5</td>
<td>2.547</td>
</tr>
<tr>
<td>1994</td>
<td>9260.4</td>
<td>4044</td>
<td>53.7</td>
<td>3.214</td>
</tr>
<tr>
<td>1995</td>
<td>10023.4</td>
<td>5046</td>
<td>52.7</td>
<td>3.744</td>
</tr>
<tr>
<td>1996</td>
<td>10764.3</td>
<td>5846</td>
<td>51.9</td>
<td>4.206</td>
</tr>
<tr>
<td>1997</td>
<td>11284.5</td>
<td>6420</td>
<td>51.0</td>
<td>4.472</td>
</tr>
<tr>
<td>1998</td>
<td>11598.5</td>
<td>6796</td>
<td>50.7</td>
<td>4.640</td>
</tr>
<tr>
<td>1999</td>
<td>12305.2</td>
<td>7159</td>
<td>50.8</td>
<td>4.646</td>
</tr>
<tr>
<td>2000</td>
<td>13471.4</td>
<td>7858</td>
<td>50.2</td>
<td>4.719</td>
</tr>
<tr>
<td>2001</td>
<td>14633.5</td>
<td>8622</td>
<td>50.6</td>
<td>4.740</td>
</tr>
<tr>
<td>2002</td>
<td>16311.5</td>
<td>9398</td>
<td>50.9</td>
<td>4.570</td>
</tr>
<tr>
<td>2003</td>
<td>19031.6</td>
<td>10542</td>
<td>64.9</td>
<td>4.492</td>
</tr>
<tr>
<td>2004</td>
<td>21971.4</td>
<td>12336</td>
<td>66.5</td>
<td>4.547</td>
</tr>
<tr>
<td>2005</td>
<td>24940.4</td>
<td>14040</td>
<td>69.0</td>
<td>4.682</td>
</tr>
</tbody>
</table>

2.3 Cointegration Analysis

Because the economic variables in a time series are usually nonstationary, and there is neither randomness nor a definite tendency, the sample data should be transformed by taking the natural log so as to reduce vibration, and by taking the difference so as to eliminate instability and heteroscedasticity. Before cointegration analysis, the Augment Dickey-Fuller (ADF) test was applied to test...
whether a data series is stationary. The null hypothesis is that the data series is nonstationary. The test results are shown in Table 2. (* expresses the first order difference).

<table>
<thead>
<tr>
<th>Variables</th>
<th>ADF Test Statistic</th>
<th>5% Critical Value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNQ</td>
<td>-3.423</td>
<td>-3.710</td>
<td>non-stationary</td>
</tr>
<tr>
<td>ΔLNQ</td>
<td>-3.168*</td>
<td>-3.066</td>
<td>stationary</td>
</tr>
<tr>
<td>LNPCGDP</td>
<td>-4.107*</td>
<td>-3.733</td>
<td>stationary</td>
</tr>
<tr>
<td>ΔLNPCGDP</td>
<td>-1.093</td>
<td>-3.658</td>
<td>non-stationary</td>
</tr>
<tr>
<td>LNHIS</td>
<td>-4.413*</td>
<td>-3.674</td>
<td>stationary</td>
</tr>
<tr>
<td>LNEI</td>
<td>-1.316</td>
<td>-3.691</td>
<td>non-stationary</td>
</tr>
<tr>
<td>ΔLNEI</td>
<td>-3.802*</td>
<td>-3.733</td>
<td>stationary</td>
</tr>
</tbody>
</table>

Table 2. ADF unit root test results on variables

Note: ** expresses MacKinnon critical values for rejection of hypothesis of a unit root under the 5% significance level.

In table 2 all the original values of the variables are less in absolute value than the ADF test statistic’s critical value at the 5% significance level; so we fail to reject the null hypothesis at the 5% significance level. But all the computed ADF test statistic values of the first difference of the variables are greater in absolute value than the ADF test statistic’s critical value at the 5% significance level, and so the null hypothesis is rejected at the 5% significance level, and so all the variables are I(1), and this meets the conditions for cointegration analysis. In other words, from 1985 to 2005, there may be a cointegration relationship between electricity consumption and the explanatory variables.

The cointegration test needs to be run to find whether there is a cointegration relationship. The null hypothesis is that there is no cointegration relationship between electrical energy consumption and the explanatory variables. All the observed series contain a time trend; so the cointegration test model contains the intercept and time trend. The results of the Johansen cointegration test are shown in table 3.

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Likelihood</th>
<th>5 Percent Critical Value</th>
<th>Hypothesized No. of CE(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7561</td>
<td>60.950</td>
<td>47.856</td>
<td>None*</td>
</tr>
<tr>
<td>0.7184</td>
<td>34.141</td>
<td>29.797</td>
<td>At most 1*</td>
</tr>
<tr>
<td>0.3170</td>
<td>10.645</td>
<td>15.495</td>
<td>At most 2</td>
</tr>
<tr>
<td>0.1380</td>
<td>2.821</td>
<td>3.841</td>
<td>At most 3</td>
</tr>
</tbody>
</table>

Table 3 Results of Johansen Cointegration test

Note: ** expresses it is significant under 5% confidence level.

The results in table 3 show that the Likelihood ratio of the first two eigenvalues is greater than the critical value at the 5% significance level; therefore there is a long-term equilibrium relationship between electricity consumption and the three explanatory variables. The normalized cointegration coefficients are shown in table 4.

<table>
<thead>
<tr>
<th>Variables</th>
<th>LNHIS</th>
<th>LNEI</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNQ</td>
<td>1.0000</td>
<td>0.13</td>
<td>-0.86</td>
</tr>
<tr>
<td>ΔLNPCGDP</td>
<td>1.01</td>
<td>0.065</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Table 4 Normalized Cointegration Coefficients

Note: the number in parenthesis in the table is the asymptotic standard error.

So the cointegration function is stated as:

\[
LNQ = 1.27 + 1.01LNPCGDP + 0.13LNHS - 0.86LNEI
\] (4)

If the residual series of equation (4) is stationary, there is a cointegration relationship between electrical energy consumption and the three explanatory variables; otherwise, there is no cointegration relationship. So the Johansen cointegration test is run to test whether the residual series is stationary, and the test results are shown in table 5.

<table>
<thead>
<tr>
<th>ADF Test Statistic</th>
<th>1 Percent</th>
<th>5 Percent</th>
<th>10 Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Critical Value</td>
<td>Critical Value</td>
<td>Critical Value</td>
</tr>
<tr>
<td></td>
<td>-4.01</td>
<td>-4.62</td>
<td>-3.71</td>
</tr>
</tbody>
</table>

Table 5 ADF Unit Root Test Results on Residual Series

The 5% critical value of the ADF test statistic is -3.71; the computed ADF test statistic value of -4.01 indicates that there are no unit roots in the residual series; that is, the residual series is stationary. So there is a cointegratin relationship between electricity consumption and the explanatory variables.

In equation (4) the coefficients of the explanatory variables are the elasticity of $Q$ with respect to the three explanatory variables. That is, a 1% increase in $PCGDP$ leads to, on average, a 1.01% increase in $Q$, a 1% increase in $HIS$ increases $Q$ by 0.13%, and a 1% improvement in $EI$ decreases $Q$ by 0.86% on average.

## 3 Multivariate SVM Model

### 3.1 Regression Arithmetic of SVM

Suppose

\[
T = \{(x_1, y_1), \ldots, (x_p, y_p)\},
\]
\( x_i \in R^m \) is the input variable, \( y_i \in R \) is the corresponding output value and \( p \) is the total number of the data points. Then the SVM regression function is:

\[
f(x) = (\omega \cdot \Phi(x)) + b \tag{5}
\]

where \( \Phi(\cdot) \) is a non-linear mapping function, \( \omega \) is a weight vector, and \( b \) is the error term. \( \omega \) and \( b \) are estimated by:

\[
\text{min} \| \omega \|^2 + C \sum_{i=1}^{p} L_i (y_i, f(x_i)) \tag{6}
\]

where \( C \) is the punishment parameter, which is considered to specify the trade-off between empirical risk and the model’s flatness. \( \frac{1}{2} \| \omega \|^2 \) is the normalization term. \( L_i (y_i, f(x_i)) \) is called the \( \epsilon \)-insensitive loss function, which is defined as:

\[
L_i (y_i, f(x_i)) = \max\{0, |y_i - f(x_i)| - \epsilon \} \tag{7}
\]

In equation (7) the loss equals zero if the forecasting error is less than \( \epsilon \); otherwise the loss not less than \( \epsilon \). In order to represent the distance from actual values to the corresponding boundary values of the \( \epsilon \)-band, two positive slack variables \( \xi_i \) and \( \xi_i^* \) are introduced. Then, equation (6) is transformed into the following form:

\[
J = \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{p} (\xi_i^* + \xi_i) \tag{8}
\]

\[
\begin{aligned}
&\text{s.t.} \quad y_i - [\omega, \Phi(x)] - b \leq \epsilon + \xi_i^* \quad \xi_i^* \geq 0 \\
&\quad [\omega, \Phi(x)] + b - y_i \leq \epsilon + \xi_i \quad \xi_i \geq 0
\end{aligned}
\]

This constrained optimization problem is solved by using the following Lagrangian form:

\[
\text{max} \mathcal{H}(\xi_i, \xi_i^*) = -\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\xi_i^* - \xi_j^*) (\xi_i - \xi_j) K(x_i, x_j) + \sum_{i=1}^{p} \xi_i (y_i - \epsilon) - \sum_{i=1}^{p} \xi_i (y_i + \epsilon) \tag{9}
\]

\[
\begin{aligned}
&\text{s.t.} \quad \sum_{i=1}^{p} (\xi_i^* - \xi_i^*) = 0 \\
&\quad \xi_i, \xi_i^* \in [0, C]
\end{aligned}
\]

where \( \xi_i, \xi_i^* \) are Lagrangian multipliers, and \( \xi_i - \xi_i^* \neq 0 \) i.e. corresponding data points are a support vector. \( \xi_i \) and \( \xi_i^* \) calculated by the Lagrange multipliers, an optimal desired weight vector of the regression hyperplane is obtained:

\[
\omega^* = \sum_{i=1}^{p} (\xi_i - \xi_i^*) K(x_i, x) \tag{10}
\]

Hence, the regression function is:

\[
f(x) = \sum_{i=1}^{p} (\xi_i - \xi_i^*) K(x_i, x) + b \tag{11}
\]

where \( K(x_i, x) \) is called the kernel function. The value of the kernel function equals the inner product of \( \Phi(x_i) \) and \( \Phi(x) \), which are produced by mapping \( x_i \) and \( x \) into a higher dimensional feature space; that is:

\[
K(x_i, x) = \Phi(x_i, x) \tag{12}
\]

### 3.2 Multivariate SVM Model

For a univariate time series \( \{x_1, x_2, \ldots, x_n\} \), training sample sets, \( \{x_1, x_2, \ldots, x_m\} \rightarrow \{x_{m+1}\} \), \( \{x_2, x_3, \ldots, x_m\} \rightarrow \{x_{m+2}\}, \ldots \), are established. Suppose

\[
X = \begin{bmatrix}
x_1 & x_2 & \cdots & x_m \\
x_2 & x_3 & \cdots & x_{m+1} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n-m} & x_{n-m+1} & \cdots & x_{n-1}
\end{bmatrix}
\]

and

\[
Y = \begin{bmatrix}
x_{m+1} & x_{m+2} & \cdots & x_n
\end{bmatrix}^T
\]

where \( \{x_i, x_{i+1}, \ldots, x_{i+m-1}\} \) is the input vector, \( \{x_{i+m}\} \) is the output value and \( m \) is the embedded dimension.

Supposed that we have observed an \( l \)-dimensional multivariate time series:

\[
\{X_n\}_{n=1}^{N} = \{(x_{1,n}, x_{2,n}, \ldots, x_{l,n})\}
\]

As in the case of a univariate time series, we make a state space reconstruction:

\[
V_n = [x_{1,n}, x_{1,n-1}, \ldots, x_{1,n-m+1}, \ldots; x_{l,n}, x_{l,n-1}, \ldots, x_{l,n-m+1}]^T
\]

\( m_i \) is the embedded dimension of \( i \)-th variable, \( i = 1, 2, \ldots, l \). The node quantity is the sum of the embedded dimensions in the multivariate time series:

\[
m = m_1 + m_2 + \cdots + m_l
\]

The structure of the multivariate SVM is shown in figure 1.
4 Case Study

4.1 Kernel Function and Parameters
Comparing the results calculated by 4 kinds of kernel function, the Gaussian radial basis function, \( K(x, x) = \exp(-|x_i - x|/2 \sigma^2) \), was applied in the SVM.

In the model of SVM, the forecast results are sensitive to the parameter of \( \sigma \), and insensitive to the other parameters. According to the simulation results and the principle of minimum error [14], let \( \varepsilon = 0.0008 \), and \( C=10000 \). When \( \varepsilon \) and \( C \) are selected, the simulation results between \( \sigma \) and the mean of absolute proportional error (MAPE) of the forecast results is simulated and shown in figure 2. According to the figure 2, when the \( \sigma = 3.5 \), the MAPE is minimal, so \( \sigma \) is set as 3.5.

[Diagram: Multivariate Time Series SVM]

**4.1 Forecast Results**
The electrical consumption is the output variable. The model (4) indicates that the \( PCGDP, HIS \) and \( EI \) are the determining factors of \( Q \), and there is a long term equilibrium relationship among them. So \( PCGDP, HIS, EI \) and actual \( Q \) are the input variables of the SVM, which is called multivariable SVM. In order to eliminate dimensional diversity in the variations in each time series, data is normalized into the interval \([0, 1]\). The forecasted results of electrical consumption by multivariable SVM are shown in table 6.

To compare the forecast results of the multivariable SVM with that of univariable SVM, the forecast results by univariable SVM are shown in table 6. There is only one input variable, the actual electrical consumption, in the model of univariable SVM. That is, the future electrical consumptions are obtained by inputting the previous electrical consumptions into the model of SVM.

<table>
<thead>
<tr>
<th>Year</th>
<th>Multivariable SVM</th>
<th>Univariable SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast value</td>
<td>Forecast value</td>
</tr>
<tr>
<td>1994</td>
<td>9141.1</td>
<td>9221.3</td>
</tr>
<tr>
<td>1995</td>
<td>9833.3</td>
<td>9946.6</td>
</tr>
<tr>
<td>1996</td>
<td>10292.2</td>
<td>10867.2</td>
</tr>
<tr>
<td>1997</td>
<td>11263.3</td>
<td>11605.2</td>
</tr>
<tr>
<td>1998</td>
<td>11652.4</td>
<td>11591.9</td>
</tr>
<tr>
<td>1999</td>
<td>12029.5</td>
<td>11789.4</td>
</tr>
<tr>
<td>2000</td>
<td>13059.2</td>
<td>13817.2</td>
</tr>
<tr>
<td>2001</td>
<td>14404.7</td>
<td>14169.4</td>
</tr>
<tr>
<td>2002</td>
<td>15890.8</td>
<td>15902.1</td>
</tr>
<tr>
<td>2003</td>
<td>13823.2</td>
<td>18510.3</td>
</tr>
<tr>
<td>2004</td>
<td>22794.3</td>
<td>22738.3</td>
</tr>
<tr>
<td>2005</td>
<td>24762.4</td>
<td>25604.5</td>
</tr>
<tr>
<td>2006</td>
<td>28727.5</td>
<td>28758.6</td>
</tr>
</tbody>
</table>

The absolute proportional error (APE) between the actual value and the forecasted value is calculated by: \( APE = \left| \frac{x_i - x_i^*}{x_i} \right| \times 100\% \),

where \( x_i^* \) is the forecast value, and \( x_i \) is the actual value.

According to the table 6, the average values and the variances of APE forecasted by multivariable SVM and univariable SVM are shown in table 7.

<table>
<thead>
<tr>
<th>Year</th>
<th>Average</th>
<th>Variance</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>1.9359</td>
<td>1.2499</td>
<td>2.2261</td>
<td>1.4166</td>
</tr>
</tbody>
</table>

From table 6, the APE forecasted by multivariable SVM is in the open interval of [0.1613,
3.7444], and the APE forecasted by univariable SVM is in the open interval of [0.4222, 4.1917]. From table 7, the average value and variance of APE forecasted by multivariable are less than that forecasted by univariable SVM. It proves that the forecast effect by multivariate SVM is better than that by univariable SVM in forecasting electrical consumption in China.

5 Conclusion
Two conclusions are obtained:

(1) By undertaking a cointegration analysis with annual data over the period 1985~2005 in China, the estimation results show that there is cointegration relationship between electrical energy consumption and the three explanatory variables; the cointegration model (4) explains how the three explanatory variables influence the electrical energy consumption.

(2) Input PCGDP, HIS, EI and actual electrical energy consumption into the model of SVM, the electrical energy consumption are forecasted. Comparing the forecast results by multivariable SVM with that by univariable, we conclude that the forecast effect by multivariate SVM is better than that by univariable, because the crucial factors which influence the electrical energy consumption are considered in the multivariable SVM.

References: