Economic Indicators Selection for Property Crime Rates using Grey Relational Analysis and Support Vector Regression

Razana Alwee, Siti Mariyam Hj Shamsuddin, Roselina Sallehuddin
Soft Computing Research Group, Faculty of Computing, Universiti Teknologi Malaysia
Skudai Johor, Malaysia
razana@utm.my, mariyam@utm.my, roselina@utm.my

Abstract—Features selection is very important in the multivariate models because the accuracy of forecasting results produced by the model are highly dependent on these selected features. The purpose of this study is to propose grey relational analysis and support vector regression for features selection. The features are economic indicators that are used to forecast property crime rate. Grey relational analysis selects the best data series to represent each economic indicator and rank the economic indicators according to its importance to the property crime rate. Next, the support vector regression is used to select the significant economic indicators where particle swarm optimization estimates the parameters of support vector regression. In this study, we use unemployment rate, consumer price index, gross domestic product and consumer sentiment index as the economic indicators, as well as property crime rate for the United States. From our experiments, we found that the gross domestic product, unemployment rate and consumer price index are the most influential economic indicators. The proposed method is also found to produce better forecasting accuracy as compared to multiple linear regressions.

Keywords—Grey relational analysis; support vector regression; crime forecasting; economic indicators

I. INTRODUCTION

Crime forecasting model can be developed using multivariate model that involves the use of more than one variable. For example, the crime rate can be predicted using economic indicators such as unemployment rates, gross domestic product and consumer price index [1]. By having other variables in the model, multivariate models have the ability to forecast new patterns that have never been observed in the past. Forecasting performance of the multivariate model is highly dependent on variables or features used in the model. Only significant features need to be used in the model. Hence, features selection process is very important to produce an accurate model [2].

The relationship between crime and economic condition is undeniable. Many previous studies have been conducted in order to identify the relationship between crimes and economic indicators [3,4,5]. In literatures, the most frequent economic indicators used to forecast crime rates is unemployment [1,6,7,8]. Besides unemployment, other economic indicators which also influence crimes were gross domestic product [9], consumer price index [1] and consumer sentiment index [10]. Lower circulation rates of money and property would reduce opportunities for criminals, and thus cause a decline in property crime rates. Meanwhile, prosperity reflected by rising inflation provides more opportunities for criminals, and thus would increase the property crime rate [1].

Regression analysis is a statistical method often used in correlation analysis. This method not only gives the correlation coefficient but also the level of significance. However, this method requires a sufficient amount of data and statistical assumptions such as distribution of data and a linear relationship [11]. However, most of the crimes data are not sufficient and not linearly distributed. The limitation of the statistical methods can be overcome by using grey relational analysis (GRA), which is an alternative method can be used to identify relationships between economic indicators and crimes. Based on the data series, GRA can be used to find the relationship and rank the economic indicators according to their importance.

Since GRA can only produce rank of economic indicators according to its importance or priority, another method is required to determine the significant economic indicators. Artificial neural network (ANN) is one of the methods that have been used together with GRA for features selection [12]. ANN is a popular and commonly used method in time series forecasting [13] despite of several problems such as the need for controlling numerous parameters, uncertainty in solution (network weights), and the danger of over fitting. ANN also requires sufficient data to produce a good model [14]. The drawbacks of ANN can be overcome by using support vector regression (SVR). SVR is a nonlinear model to solve regression problems. There are four factors that contributed to the success of SVR, which are good generalization, global optimal solution, the ability to handle nonlinear problems, and the sparseness of the solution. This has made SVR very robust with small training data, nonlinear, and high-dimensional problems [15]. Despite the advantages, SVR also has some limitations. For example, SVR model parameters must be set correctly as this can affect the regression accuracy.
Inappropriate parameters may lead to over-fitting or under-fitting [16]. In order to overcome this problem, researchers used particle swarm optimization (PSO) to estimate the parameters of SVR model [17,18]. PSO has the ability to escape from local optima, easy to implement and has fewer parameters to be adjusted [19].

In this study, GRA and SVR are used to select the significant economic indicators for the property crime rate. PSO is used to estimate the parameters of SVR model. GRA and SVR are suitable to be used when data is scarce. First, GRA is used to select the best data series to represent each economic indicator. After that, GRA ranks each economic indicator according to its importance or priority. On the other hand, SVR will choose the optimal features and forecast the property crime rate. The United State (US) property crime rate is used in this study. As for the economic indicators, unemployment rates (UR), gross domestic product (GDP), consumer price index (CPI) and consumer sentiment index (CSI) are selected.

The remainder of this study is organized as follows. Related work on crimes and economic conditions is first discussed in Section II. In Section III, the grey relational analysis, the support vector regression, particle swarm optimization and the proposed method, GRA_PSOSVR are described. Section IV describes the data set and model evaluation. The experimental results are presented in Section V. Finally, a brief conclusion is drawn in Section VI.

II. RELATED WORK ON CRIMES AND ECONOMIC CONDITIONS

Economic conditions are often considered to be related to crimes, especially property crimes. In literatures, many studies have been done by researchers in order to relate the economic conditions with property crimes. The unemployment rate is often selected by the researchers in their studies to represent the economic conditions [10]. A study using a country level panel data set from Europe found that unemployment has a positive influence on property crimes [20]. Meanwhile, another study based on UK annual regional data has discovered that unemployment is an important explanatory variable for crimes motivated by economic gain [21]. Results produced by some other studies also found significant relationship between the unemployment and property crimes. Among the findings are motor vehicle theft is significantly associated with the unemployment rate [22], and is also co integrated with male youth unemployment [23]. Another finding shows that unemployment has a positive effect on burglary, car theft and bike theft [24].

Unemployment, especially among youth and young adults are also found to influence crimes. According to a study on the United States arrest data, unemployment has a positive relationship with theft crimes among youth and young adults (16 - 24 years) [25]. Another study investigated the relationship between crime with male adult (26 - 64 years) and youth (16 - 25 years) unemployment in Britain [26]. The results indicate that youth unemployment and adult unemployment are both significantly and positively related to burglary, theft, fraud and forgery as well as total crime rates.

In addition to unemployment, other economic indicators such as consumer price index, gross domestic product and consumer sentiment index were also studied by the researchers to examine the relationship between economic conditions with crimes. Several researchers used the consumer price index to measure the inflation [4, 27]. Inflation reduces the purchasing power and increases the cost of living. A study on the impact of inflation rate on crime in the United States using the modified Wald causality test found that the crime rate is co integrated with inflation and unemployment rates [27]. Further, another study which examined the linkages between inflation, unemployment and crime rates in Malaysia revealed that inflation and unemployment are positively related to the crime rate [4]. Meanwhile, for gross domestic product, a study to explain the relationship between national crime rates with social and economic variables has found that robbery and homicide have significant negative relationship with gross domestic product [28]. Whilst, a study on the consumer sentiment index has discovered that the consumer sentiment has significant negative effects on robbery and property crime rates [10].

In summary, economic conditions and their relationship to crimes are undeniable. However, in the literature there are several economic indicators studied by researchers. Researchers use different methods and data which vary the results of the studies. It is difficult to determine which economic indicators have the most influence on the property crimes. The purpose of this study is to propose a method for identifying important economic indicators that influence the property crime rate. Four economic indicators are considered which are unemployment rate, consumer price index, gross domestic product and consumer sentiment. The important economic indicators that have been identified can be used for forecasting purposes.

III. METHODOLOGY

In this section, grey relational analysis, support vector regression and particle swarm optimization are explained as a basis for further discussion on the proposed method, GRA_PSOSVR.

A. Grey Relational Analysis

Grey relation analysis (GRA) is an analysis method proposed in the Grey system theory and was founded by Professor Deng Julong [29,30]. This method is based on geometrical mathematics, which complies with the principles of normality, symmetry, entirety, and proximity [31,32]. GRA is a distinct similarity measurement approach that uses data series to obtain grey relational order to describe the relationship between the related series [33]. It can be used to measure the correlations between the reference series and other compared series [30]. The relative distance between a compared series and the reference series which referred as grey relational grade (GRG) represents the degree of influence between these series. A small distance indicates a significant influence [11]. GRG is a numeric value between 0 and 1. GRG value close to 1 indicates that there is a strong relationship between the two series. The following are the basic steps in the GRA [34]:

1. Normalize the reference series and compared series.
2. Calculate the absolute difference between the normalized reference series and compared series.
3. Calculate the average of the absolute difference between the normalized reference series and compared series.
4. Calculate the grey relational coefficient.
5. Calculate the grey relational grade.
6. Rank the grey relational grades to determine the importance of each economic indicator.
Step 1: Determine compare series and reference series
\[ X_0 = \{ x_0(1), x_0(2), \ldots, x_0(n) \} \quad (1) \]
\[ X_i = \{ x_i(1), x_i(2), \ldots, x_i(n) \}, i=1, 2, \ldots, m \quad (2) \]
Step 2: Transform the series to dimensionless form using (3).
\[ x_i^*(k) = (x_i(k) - \min x_i(k))/(\max x_i(k) - \min x_i(k)) \quad (3) \]
Step 3: Calculate the grey relational coefficient between \( X_0 \) and \( X_i \) using (4), which reflects the degree of influence of two compared series at one time.
\[ \xi_i = \left( \min_{\xi_0} |x_0^*(k) - x_i^*(k)| + \rho \max_{\xi_0} |x_0^*(k) - x_i^*(k)| \right) / (|x_0^*(k) - x_i^*(k)| + \rho \max_{\xi_0} |x_0^*(k) - x_i^*(k)|) \quad (4) \]
where, \( \rho \in (0, 1) \), which is called as distinguishing coefficient. Generally, \( \rho = 0.5 \). According to the mathematical proof, the value of \( \rho \) does not affect the rank of the grey relational grade [31].
Step 4: Calculate the grey relational grade (GRG) which is the average value of grey relational coefficient using (5),
\[ r_{0i} = (1/n) \sum_{k=1}^{n} \xi_i(k) \quad (5) \]
where, \( r_{0i} \) represents the level of correlation between the reference series and the compared series. The grey relational order is constructed based on the calculated value of GRG, \( r_{0i} \).
Order obtained represents the list of priorities in selecting a series that is closely related to the reference series, \( X_0 \). Generally, \( r > 0.9 \) indicates a marked influence, \( r > 0.8 \) a relatively marked influence, \( r > 0.7 \) a noticeable influence, and \( r < 0.6 \) a negligible influence [31].

B. Support Vector Regression
Support vector regression (SVR) is a nonlinear model to solve regression problems. SVR training process is equivalent to the process of solving the linearly constrained quadratic programming problems that provide a unique optimal value with no local minimum problem. The solution has sparseness, as only the essential data are used to solve the regression function. Lagrangian multipliers are introduced to solve the quadratic programming problem. The SVR model is given by (6) [36],
\[ f(x) = (z, \phi(x)) + b \quad (6) \]
where
- \( z \) = weight vector
- \( b \) = bias value
- \( \phi(x) \) = kernel function

SVR used \( \epsilon \)-insensitivity loss function which can be expressed as (7),
\[ L_\epsilon(f(x) - y) = \begin{cases} \vert f(x) - y \vert - \epsilon, & \text{if } |f(x) - y| \geq \epsilon \\ 0, & \text{otherwise} \end{cases} \quad (7) \]
where \( \epsilon \) is the region for \( \epsilon \)-insensitivity. Loss is accounted only if the predicted value falls out of the band area. The SVR model can be constructed to minimize the following quadratic programming problem (8),
\[ \min: (1/2) \sum_{i=1}^{n} (z_i^T z_i) + C \sum_{i} (\xi_i + \zeta_i^*) \quad (8) \]
subject to
\[ \begin{cases} y_i - z_i^T x_i - b \leq \epsilon + \zeta_i \\ z_i^T x_i + b - y_i \leq \epsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \]
where
- \( i = 1, 2, \ldots, n \) is the number of training data
- \( (\xi_i, \zeta_i^*) \) = the empirical risk
- \( (1/2) z^T z \) = the structure risk preventing over-learning and lack of applied universality
- \( C \) = the regularization parameter

After selecting proper regularization parameter \( (C) \), width of band area \( (\epsilon) \) and kernel function \( (K) \), the optimum value of each parameter can be resolved though Lagrange function. The commonly used kernels are linear kernel, polynomial kernel, radial basis function (RBF) or Gaussian kernel and sigmoid kernel.

The type of kernel function influences the parameters of SVR kernel. The kernel function and parameters of SVR kernel function should be set properly because it can affect the regression accuracy. Inappropriate parameters may lead to over-fitting or under-fitting [16]. This study uses the RBF kernel function because it is suitable for most forecasting problems [17]. The RBF kernel is also effective and has fast training process [37]. The RBF kernel function is given by (9) [38],
\[ K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2) \quad (9) \]

There are three important parameters to be determined in RBF kernel function [18]:

i. Regularization parameter \( C \)
\( C \) is a parameter for determining the tradeoff cost between minimizing training error and minimizing model complexity.

ii. Kernel parameter \( (\gamma) \)
\( \gamma \) represents the parameter of the RBF kernel function.

iii. The tube size of \( \epsilon \)-insensitive loss function \( (\epsilon) \)
\( \epsilon \) is the approximation accuracy placed on the training data points.

These parameters determined through particle swarm optimization (PSO) must be set correctly, in order to produce accurate estimation model.

C. Particle Swarm Optimization for SVR parameter estimation (PSOSVR)
Particle swarm optimization (PSO) is one of stochastic optimization methods introduced by Kennedy and Eberhart in 1995 [39]. This method is based on the natural evolution process which uses swarming strategies in bird flocking and fish schooling. PSO is a population-based method which consists of particles. Initially, the particles are randomly generated. Each particle has a position and velocity, which
represents a potential solution to a problem in D-dimensional space. The position and velocity of \( i \)th particle are denoted by \( X_i = (x_{i1}, x_{i2}, \ldots, x_{iD}) \) and \( V_i = (v_{i1}, v_{i2}, \ldots, v_{iD}) \), respectively. While solving the search problem, each particle explores the search space by moving in the previous direction. The previous best particle (pbest) and the best solution for the entire population (gbest). The velocity and position of each particle are updated using (10) and (11), respectively [40].

\[
v_i(t+1) = w \cdot v_i(t) + c_1 \cdot \text{rand1}_i \cdot (\text{pbest}_i(t) - x_i(t)) + c_2 \cdot \text{rand2}_i \cdot (\text{gbest}_i(t) - x_i(t)) \tag{10}
\]

\[
x_i(t+1) = x_i(t) + v_i(t+1) \tag{11}
\]

where \( v_i(t) \) is the velocity of \( i \)th particle at iteration \( t \), \( x_i(t) \) is the position of \( i \)th particle at iteration \( t \), \( j=1, 2, \ldots, D \), is the dimension of the search space, and \( w \) is the inertia weight that balances the global and local search abilities of particles. Suitable selection of inertia weight can balance the global and local search and thus reduce the number of iterations in order to find the optimal solution [41]. \( c_1 \) and \( c_2 \) are two learning factors which control the influence of the social and cognitive components, rand1 and rand2 are two uniform random numbers generated independently within the range of [0, 1], \( \text{pbest}_i(t) \) is the best previous position yielding the best fitness value for \( i \)th particle at iteration \( t \), and gbest, is the global best particle by all particles at iteration \( t \). After changing the position of the particle, the particle’s fitness value is evaluated. The pbest and gbest are updated based on the current position of the particles. As this process is repeated, the whole population evolves toward the optimum solution.

In order to estimate the SVR parameters, \( C \), \( \gamma \), and \( \varepsilon \), the \( i \)th particle is represented by three-dimensional vectors, \( X_i = (x_{i1}, x_{i2}, x_{i3}) \) and \( V_i = (v_{i1}, v_{i2}, v_{i3}) \), where the first, second and third dimensions of the vectors refer to \( C \), \( \gamma \), and \( \varepsilon \), respectively. In this study, the fitness is defined by \( k \)-fold cross-validation, where \( k=5 \). In \( k \)-fold cross-validation, the training data set is divided into \( k \) subsets of equal size. One subset is used for validation. The regression function is built with a given set of parameters \((C, \gamma, \varepsilon)\) using the remaining \( k-1 \) subsets. The performance of the parameter set is measured by the root mean square error (RMSE) on the validation set. Each subset is used once for validation and the process is repeated \( k \) times. The average of RMSE on the validation set from 5 trials is used as a measure of fitness. The RMSE is defined as (13).

\[
\text{RMSE} = \sqrt{\left(\sum_{i=1}^{n} (y_i - p_i)^2\right)/n} \tag{13}
\]

where \( n \) is the number of validation data; \( y_i \) is the actual value and \( p_i \) is the predicted value. The flowchart of PSOSVR is shown in Figure 1.

In this study, the population size is set to 5, maximum number of iterations is set to 50 and the value of \( c_1 \), \( c_2 \) is set to 1. The inertia weight, \( w \), initially is set to 1.4, and its value is decreased along with the iterations according to (14)[41].

\[
w = ((w-0.4) \cdot \text{Maxiter} - \text{iter}) / \text{Maxiter} + 0.4 \tag{14}
\]

D. The Proposed Method (GRA_PSOSVR)

In this study GRA is used for two purposes. First, it is used to determine the best data series to be used as experimental data. Secondly, it is used as feature selections in determining the best input features to be used in PSOSVR model, where in this case the significant economic indicators. The combination of GRA and PSOSVR is able to reduce the economic indicators that are not important by selecting the significant economic indicators only. The significant economic indicators are chosen based on their ability to enhance the forecasting accuracy and increase the speed of the learning process. There are three main steps in the proposed model. The steps are:

Step 1: Select the best data series to represent the economic indicators. There are several candidate data series that can be used to represent each economic indicator. However, only one data series is required. For each economic indicator, GRA is used to rank all the candidate data series (compared series). The rank is based on the GRG value. The data series with the greatest GRG value is selected to represent the economic indicator in step 2.

Step 2: Rank the economic indicators. In this step, the GRA is used to examine the relationship between each economic indicator and the reference series (property crime rate). The economic indicators (compared series) will be arranged according to their importance to the reference series based on the GRG value.

Step 3: Select the significant economic indicators. This step is intended to select the significant economic indicators for property crime rate using PSOSVR model. Figure 2 shows the flowchart in this step. A different set of economic indicators is used as inputs to the PSOSVR.
model. Starting with all available economic indicators, the least important economic indicators will be removed one at a time from the list of input. PSOSVR model with the current list of input is trained using the training data set. After that, the model is applied to the test data set to forecast the property crime rate. The forecasting error is calculated using four types of quantitative error measurements, namely root mean square error (RMSE), mean square error (MSE), mean absolute percentage error (MAPE), and mean absolute deviation (MAD). After all the necessary PSOSVR models have been developed, the best model is determined based on error measurements, RMSE, MSE, MAPE, and MAD. If the results given by the four error measurements are inconsistent, then MAPE is chosen as a benchmark [42, 43]. The economic indicators in the list of input for the best model are selected as the significant economic indicators.

![Flowchart to select the significant economic indicators](image)

IV. DATA SET AND MODEL EVALUATION

This section describes the data set used and the model evaluation carried out in this study.

A. Data Set

This study uses data of annual property crime rate, violent crime rate, robbery rate, murder and nonnegligent manslaughter rate, forcible rape rate and aggravated assault rate, consumer price index, gross domestic product, consumer sentiment index and unemployment rate from 1960 to 2009 in United States. The crimes data were downloaded from the Uniform Crime Reporting Statistics web site (http://www.ucrdataonline.gov), while economic indicators data were obtained by downloading the data from Economic Research Federal Reserve Bank of St. Louis web site (http://research.stlouisfed.org). There were 30 data series for the unemployment rate, 15 data series for gross domestic product, 16 data series for the consumer price index and two data series for the consumer sentiment index.

B. Model Evaluation

The forecasting performance of the proposed method is validated through comparison with a multiple linear regression model (MLR). MLR is a popular linear model in crime forecasting and is often used for multivariate analysis. Four types of quantitative error measurements are conducted, namely root mean square error (RMSE), mean square error (MSE), mean absolute percentage error (MAPE), and mean absolute deviation (MAD).

In addition, the hypothesis test is performed to prove that there is no significant means difference between the forecasting values and the actual data. Paired sample t-test is used in this study. Let, \( \mu_1 \) represents the mean of actual data, \( \mu_2 \) is the mean of forecasting values of the forecasting model and \( \mu_1, \mu_2=\mu_D \) is the difference of means. The hypothesis,

\[
H_0: \mu_D = 0 \\
H_1: \mu_D \neq 0
\]

The test statistic is shown by (19), where student \( t \) is distributed with \( n_D-1 \) degrees of freedom.

\[
t = \frac{(y_D - \mu_D)/(s_D/\sqrt{n_D})}{(y_D - \mu_D)/(s_D/\sqrt{n_D})} \tag{19}
\]

where \( y_D \) is a sample mean difference, \( s_D \) is a sample standard deviation of the difference, and \( n_D \) is a sample size. The mean of forecasting values is equal to the mean of actual values, if the hypothesis test fails to reject the null hypothesis. It indicates that the model is an appropriate forecasting model since it represents the real situation.

V. RESULT AND DISCUSSION

In this study, four economic indicators have been used to develop forecasting model for property crime rates. In the first stage, GRA functionality is to find the best data series to be employed. Table I and II show top ranked three unemployment and gross domestic product data series relationship with property crime rate, respectively. The unemployment rate for 20 to 24 years data series is the chosen data series to represent unemployment rate. This result is consistent with the findings by [25] in such that young adult unemployment has a relationship with theft (property crime). As for the gross domestic product, grey relational grade of these data series are not much different, so any of these data series can be selected to represent the gross domestic product. Real gross domestic product is used in previous study [9]. However, this study selects the data series with the largest GRG value. Gross domestic product in United States natural log of billions of chained 2005 US dollars has the highest GRG value and is chosen to represent gross domestic product.

<table>
<thead>
<tr>
<th>Data Series</th>
<th>GRG</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment Rate - 20 to 24 years</td>
<td>0.634</td>
<td>1</td>
</tr>
<tr>
<td>Unemployment Rate - 20 to 24 years, Men</td>
<td>0.625</td>
<td>2</td>
</tr>
<tr>
<td>Unemployment Rate - 20 to 24 years, Women</td>
<td>0.621</td>
<td>3</td>
</tr>
</tbody>
</table>
Table III show the top three consumer price index data series relationship with property crime rate. The CPI for all urban consumer apparel data series is chosen to represent consumer price index for property crime rate. Previous studies used the percentage of change in the annual consumer price index [1] and consumer price index in the United State [27]. Consumer price index in the United State is ranked sixth with the GRG value of 0.616. This study found that the CPI for all urban consumers-apparel is more influential to the property crime as compared with other data series for the consumer price index. Youths and young adults (16 - 24 years) are those who often keep up to date with the latest fashions and apparels. These groups of people are often labeled as to be likely to commit crimes [22] due to unemployment but possessed greater desires and needs for the latest apparel, which may lead them to commit crimes, especially property crime.

Table IV shows the rank of consumer sentiment index data series relationship with property crime rate. There is only a little difference in the GRG values between the data series. University of Michigan Consumer Sentiment- Natural Log of Index 1st Quarter 1966=100 data series is selected to represent consumer sentiment. Previous researcher used the annualized index of consumer sentiment (1966 = 100) values as their measure of consumer sentiment [10].

In the second stage, GRA is used as features selection to remove the irrelevant economic indicators. Table V shows the rank of economic indicators based on GRG. The GRG for gross domestic product and unemployment rate are similar to three decimal places which is 0.675. This indicates that both indicators have the same level of priority. Since they have the highest GRG values, it implies that both of them are most significant factors. Whereas the GRG obtained for the consumer sentiment index (CS) is 0.623 which is relatively low as compared to other economic indicators. Therefore, it can be concluded that, by comparing to others economic indicators, the consumer sentiment index is the least influencing economic indicators for the property crime rates.

In order to select the significant economic indicators, crime forecasting models have been developed using PSOSVR model. In addition to economic indicators, two crime indicators are also used in the forecasting models. One year lagged of property crime rate and one year lagged of selected violent crime are employed. The selection of one year lagged of violent crime is based on GRA as shown in Table VI. Table VI shows that the robbery rate has the highest GRG value, 0.743. Therefore, the robbery rate was selected to represent the one-year lagged of violent crime. One year lagged is to allow crime rates to change according to the patterns of criminal behavior. The one year lagged crime rate has a positive effect on the change in current crime trend [1].

There are five models with a different set of economic indicators developed for property crime rates. Table VII shows the RMSE, MSE, MAPE and MAD for property crime rate models. The minimum errors are obtained from the model with three economic indicators, namely consumer price index, gross domestic product, and unemployment rate. The SVR parameters (C, σ, ε)=(60.1924, 0.0625, 0.01). Therefore, consumer price index, gross domestic product, and unemployment rate are selected as significant economic indicators for property crime rate.
The proposed method, GRA_PSOSVR, is compared with MLR to evaluate its forecasting performance. Table VIII shows the difference of significant economic indicators selected by the MLR model and GRA_PSOSVR. For the MLR model, the significant economic indicators are consumer sentiment index, unemployment rate and consumer price index. The economic indicators selected by the MLR and the GRA_PSOSVR are different because both models are different in term of relationship existing among the data. For example MLR has linear relationship while GRA_PSOSVR has nonlinear relationships.

### Table VIII. Selected Significant Economic Indicators by MLR and GRA_PSOSVR

<table>
<thead>
<tr>
<th>Method</th>
<th>Property Crime</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td>Consumer sentiment index, Unemployment rate, Consumer price index</td>
</tr>
<tr>
<td>GRA_PSOSVR</td>
<td>Consumer price index, Gross domestic product, Unemployment rate</td>
</tr>
</tbody>
</table>

Figure 3 shows the comparative values between the actual values and forecast values of property crime rates. The predicted values obtained from GRA_PSOSVR are closer to the actual value and have a similar pattern with the actual data compared to MLR.

![Fig. 3. Forecasting of Test Data Set](image)

Table IX shows the results of paired samples t-test. Paired sample t-test is used to compare the actual data of property crime rates with the forecast values of the proposed model, GRA_PSOSVR, and MLR. The test is conducted to ensure that there is no statistically significant difference of means between the actual data and the forecast values from the forecasting models. The results show that between the actual data and the forecasted values from the forecasting models, the P-value > 0.05 (0.463 and 0.190). The difference between the upper and lower value for 95% interval is ranging between negatives and positives values. This result indicates that the hypothesis test fails to reject the null hypothesis which implies that there is no statistically significant difference of means between the actual data of property crime rates and the forecasted values. However, GRA_PSOSVR shows smaller mean, standard deviation and standard error mean as compared to MLR. In addition, RMSE, MSE, MAPE and MAD by GRA_PSOSVR for property crime rates model are found to be smaller than MLR as shown in Table X. Therefore, the assumption that can be made is that GRA_PSOSVR is able to give better forecasting performance in comparison to MLR.

### Table IX. Paired Samples Test

<table>
<thead>
<tr>
<th>Pair</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower Difference</td>
<td>1</td>
<td>4.63</td>
<td>12.8</td>
<td>-11.5</td>
<td>0.463</td>
</tr>
<tr>
<td>Upper Difference</td>
<td>2</td>
<td>-52.4</td>
<td>74.2</td>
<td>-144.5</td>
<td>0.190</td>
</tr>
</tbody>
</table>

(legend: Pair 1: Actual data and GRA_PSOSVR, Pair 2: Actual data and MLR)

**Table X. Comparison of Errors**

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MSE</th>
<th>MAPE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Linear Regression</td>
<td>84.55</td>
<td>7148.96</td>
<td>1.83</td>
<td>59.22</td>
</tr>
<tr>
<td>GRA_PSOSVR</td>
<td>12.33</td>
<td>152.08</td>
<td>0.31</td>
<td>9.96</td>
</tr>
</tbody>
</table>

Based on the model evaluation that has been carried out, we can conclude that the proposed method, GRA_PSOSVR is suitable to be employed in crime rates forecasting with economic indicators because it has better forecasting performance as compared to MLR. The optimal parameters are very important for the accuracy of SVR models. The use of PSO has facilitated the searching process for the optimal parameters of SVR model. MLR only identifies features based on a linear relationship, but the GRA_PSOSVR is able to identify important features based on non-linear relationships in the data. The features selection based on non-linear relationship has been able to produce accurate forecast model.

## I. Conclusion

This paper has presented the use of GRA and SVR in the selection of economic indicators to forecast the property crime rates. PSO is used to estimated the SVR parameters. The proposed method, GRA_PSOSVR is used to identify and obtain the significant economic indicators for the property crime rates. The rank of each economic indicator on the crime rates can be obtained using GRA. PSO is used to select the significant economic indicators and to forecast the crime rates. Gross domestic product is the economic indicator that has the strongest relationship with property crime rate, followed by unemployment rate, consumer price index and consumer sentiment index. The significant economic indicators for property crime rate are gross domestic product, unemployment rate, and consumer price index. GRA_PSOSVR also shows a better forecasting performance than MLR. Even though this study only focuses on the economic indicators, the GRA_PSOSVR can also be used to study other indicators that may have relationships with crime, such as demographic and social indicators. In conclusion, GRA_PSOSVR can be used as alternative method that is suitable to be applied in the crime forecasting.

**Acknowledgment**
This work is supported by Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (Vote No. 4F086). Authors would like to thank Research Management Centre (RMC) Universiti Teknologi Malaysia, for the research activities and Soft Computing Research Group (SCRG) for the support and motivation in making this study a success.

REFERENCES


