Enhanced Knowledge in Sciences and Technology Vol. 2 No. 1 (2022) 132-142

EKST

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/ekst
e-ISSN : 2773-6385

A Comparative Study on Criminal Cases through Economic Indicators in Malaysia using Regression Modelling

Lee Kai Seong[1], Rohayu Mohd Salleh[1]*

[1]Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia, 84600 Pagoh, Johor, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/ekst.2022.02.01.015
Received 31 December 2021; Accepted 29 January 2022; Available online 1 August 2022

Abstract: Criminal occurred when someone committed a crime and was charged under the justice system for violating the public law. Criminal analysis is commonly applied with economics in modelling of criminal cases since sluggish economy causes citizens to have no proper earning income and forces them to commit crime to get back to their regular life. This research study on the suitable regression method for prediction of the criminal case model between ordinary least squares regression, LASSO regression and Ridge regression. A total of 51 monthly observations of economic indicators were collected from the Department of Statistics Malaysia and Bank Negara Malaysia as well as criminal cases were collected from the Federal Court of Malaysia. Based on the selection features from OLS regression and LASSO regression, the study found out consumer price index and KLCI were the key variables that have the impact on criminal cases. However, there were additional two variables: unemployment rate and participation rate selected by LASSO regression do have significant impact on criminal cases. In addition, the prediction models obtained were found out to have insensible intercept due to the negative intercept observed. These insensible intercepts were caused by the fluctuation of data that occurred during COVID-19 outbreak in Malaysia started February 2020 which had largely affected the economy and social behaviours. However, the estimated coefficients of the regression equations (OLS, LASSO and Ridge) show sensible, thus allowed to make prediction for criminal cases. Lastly, the result found LASSO regression allowed to use 4 economic indicators and given a lower mean squared error with less complexity of the model, plus allowed to give a better R-squared value as compared to other regression models.

Keywords: Criminal Case, Economic Indicator, LASSO, Ordinary least Squares, Ridge, COVID-19 Pandemic.

Corresponding author: msrohayu@uthm.edu.my

1 Introduction

Criminal behaviour is the offender's reaction to committing a crime, when someone committed a crime and was charged for violating the public law, then a criminal case occurred [1]. The crime occurred due to the result of criminal opportunity that the offenders made a rational decision which selected their preferred target that has a high return with minimal commitment and risk. By referring to economic theory, a person who engages in illegal activity decreases when the economic opportunities are given to an individual to earn more income, thus helping to reduce the criminal motive of an individual [2].

Crime issues had been alerted through the news spreading all over the social media, making the society not comfortable with the news spreading and the sign of panic had been elevated in the current situation. Humans likely tend to react with negative stories in social media, and thus causing the bad news to always be chosen to dominate headlines [3]. As a result, the criminal activities were usually being reported to draw the attention of readers for the newspapers and social media. The sluggish economy is one of the reasons to increase the desire of people to get involved in illegal activities due to low purchasing power and no proper earning income. In addition, with the current scenario of COVID-19 where it becomes a world pandemic issue, the impact of COVID-19 becomes significant to be investigated since it has affected the criminal and social behaviours in the community of Malaysia.

The research of criminal cases was still rare compared to crime rate due to limited historical data, but it was discovered to have theoretical support that criminal cases had a link with crime rate. The criminal cases forecast provides additional information to the government that can be utilised to plan and develop a sustainable and healthy economy [4]. In Malaysia, there are two types of cases recorded in the Federal Court of Malaysia, namely criminal cases and civil cases. This study focused on the criminal cases and used it to develop a prediction model through econometric models: ordinary least squares (OLS) regression, LASSO regression and Ridge regression.

Apart from that, this research study had a similar concept to previous researchers with carried out OLS regression [5-7], but it was found out that OLS regression was unable to overcome the issue of over-fitting and over-selection of independent variables, therefore LASSO regression was proposed to overcome these issues [8]. Moreover, Ridge regression does have the ability to overcome the problem of multicollinearity [9] and improve the accuracy of the regression model compared to OLS regression which always gives non-zero coefficients [10]. As a result, LASSO and Ridge regression have been considered as an alternative econometric model instead of OLS regression to predict the criminal case.

Furthermore, there are several economic factors that influence criminal cases, where the unemployment rate is the major economic indicators selected by researchers [11]. In a previous study, the unemployment rate had been found to influence criminals [11-13]. In addition to unemployment, other economic indicators such as consumer price index [12], inflation rate [11-12], and tourist arrivals [11] were found to have an influence on criminal cases. Also, a recessionary economy will lead to FBM KLCI drop and vice versa. Therefore, the recession of economy generally increases the criminal case while decrease during expansions of the economy [14], then it was appropriate to study FBM KLCI could have significant impact on criminal cases. Additionally, the labour participation rate is significantly related to criminal case since high participation rate tend to have no influence on the criminal violence while compared to the countries with less participation rate [5], thus participation rate act as another economic factor that could impact on the criminal cases.

Besides regression models, this study also conducted the association between the criminal cases and economic indicators through the Pearson correlation coefficient that used to examine the degree of correlation in linear regression [15], Moreover, the variable selections by the features of OLS and LASSO regression was used to determine the significant relationship between criminal cases and

economic indicators while discarding the less important variables affect the criminal econometric models.

The remainder of this research is structured as follows: Section 2 gives a detailed description of the research methodology to determine the regression models of criminal cases. Section 3 clearly explains the results and discussion based on the methodologies utilised in this research and Section 4 concludes this whole research.

2 Methodology

This section discusses the various techniques applied for criminal cases analysis including the proposed method, namely Pearson correlation, OLS regression, LASSO regression and Ridge regression. The process started with data collection from different resources, then undergone data-pre-processing with detect outliers, collinearity and correlation. After that, the process continued to build the prediction model by using OLS, LASSO and Ridge regression. For LASSO and Ridge regression, cross-validation method had been used to obtain the tuning parameter for each regression model built.

2.1 Pre-processing Analysis

All the observations and variables need to undergo pre-processing analysis before the linear regression is conducted. Outlier is an observation that is located far away from the fitted line in the regression model [16], and Mahalanobis distance is suitable to use in this research since Mahalanobis distance could apply in multivariate statistics [17]. After the outlier's detection, multicollinearity is conducted with variance inflation factors which are used to examine the strength of correlation for a variable linear relationship with the rest variables in multicollinearity analysis [18]. Then, the analysis was conducted with Pearson correlation, it used to examine the strength and direction of the linear relationship between two variables which the coefficient is given the correlation [19].

2.2 Cross-validation

Cross-validation split the data observations into *k* number of groups in the k-fold cross-validation [20]. The measurement of k-fold cross-validation divided the dataset into k-subsets of equal size, the k-1 subsets were used to train the model by obtaining the parameter, while the one remaining subset act as a validation set to measure the performance of the trained model (k-1 subsets) [21]. In this study, the mean squared error was used to measure the performance, and the 10-fold cross-validation was used to obtain the tuning parameter, λ as the purpose to estimate the LASSO and Ridge regression coefficients.

2.3 Regression Model-Building

The process of model-building is to determine the suitable independent variable to include in the model [22]. In this process, the prediction model is built for the dependent variable with the significant independent variables and allows to study of the relationship between the variables.

2.3.1 Ordinary Least Squares Regression

Ordinary least squares (OLS) regression uses least squares method to estimate the coefficients of independent variables to create a linear regression equation by minimising the sum of squares of residuals [23]. The OLS coefficients are consistent when there is no multicollinearity problem between the independent variables, as well as the residuals are not present with heteroscedasticity and serially correlated [24]. Also, the normality of residuals is needed to ensure the analysis of the econometric model is given a correct analysis without obtaining bias estimates in the regression model [25]. The matrix form equation of the OLS linear regression model is shown in Eq. 1.

$$Y = X\beta + \varepsilon \qquad \text{Eq.1}$$

where Y is vector of dependent variable, X is the matrix constant of independent variables, β is the vector of parameters and ε is a vector error of normal independent random variables. The OLS

regression equation is estimated by using matrix approach and the formula of regression coefficients by minimise the sum of squares of residuals, $\hat{\beta}$ is shown in Eq. 2.

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \text{Eq.2}$$

where X' is the transpose of matrix X and $(X'X)^{-1}$ is the inverse matrix of $X'X$.

2.3.2 Least Absolute Shrinkage and Selection Operator (LASSO) Regression

LASSO regression has the function to shrink the less important variable toward zero and select the significant coefficients to remain in the model [26]. LASSO regression minimises the prediction error with added penalty of tuning parameter, λ or constant value toward the least squares method [27]. The optimal tuning parameter is obtained through a cross-validation approach. The formula of LASSO regression coefficients minimises the sum of squares with added tuning parameter as shown in Eq. 3.

$$L_{Lasso}(\hat{\beta}) = \arg min \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \right\} \quad \text{Eq.3}$$

where k is the number of LASSO shrinkage parameters.

2.3.3 Ridge Regression

Ridge regression is a linear regression by minimising the sum of squares of residuals through a tuning parameter, λ which controls the amount of shrinkage on the coefficients of parameters [28]. Besides that, Ridge regression shrinks the coefficient estimates to approach zero but not towards zero with the added shrinkage penalty, λ and overcomes the problem of multicollinearity issue [29]. Ridge regression also uses cross-validation method to obtain the optimal tuning parameter, λ and the formula of Ridge coefficient is shown in Eq. 4.

$$\beta_{Ridge} = (X'X + \lambda I)^{-1}(X'Y) \quad \text{Eq.4}$$

where *I* am the identity matrix.

3 Results and Discussion

Each symbol was assigned to the variables for easier understanding of analysis, namely: criminal case (y), unemployment rate (x_1), participation rate (x_2), consumer price index (x_3), inflation rate (x_4), Kuala Lumpur Composite Index (x_5) and tourist arrivals (x_6).

3.1 Data Collection

The data collected is the total criminal cases reported to the Malaysian judicial system and it was collected from January 2017 until March 2021 with a sample size of 51 monthly observations from Federal Court of Malaysia. The total number of criminal cases of Malaysia were reported by Federal Court, Court of Appeal, High Court, Sessions Court and Magistrates Court as well as these criminal cases had consisted of all the states in Malaysia including federal territories and states. Additionally, five economic indicators were collected from the Department of Statistics Malaysia, which are unemployment rate, participation rate, consumer price index, FBM KLCI index, and tourist arrivals, whereas another one economic indicator was collected from Bank Negara Malaysia namely the inflation rate. All six economic indicators were used for determining the health status of economy by using macroeconomic scale with the duration and sample size of the economic indicators corresponding to the total number of criminal cases reported.

3.2 Analysis Prior to COVID-19 Pandemic Period

Since COVID-19 happened in Malaysia it had a huge impact to the Malaysia's society and economy resulting in the citizens needed to adjust to a new normal lifestyle. This analysis was necessary to carry out as use for validating the effect of COVID-19 pandemic had affected the regression model. There are 37 monthly observations, 1 dependent variable or criminal case and 6 independent variables or

economic indicators were used to carry out analysis before the COVID-19 pandemic period which started from January 2017 until January 2020.

3.2.1 Data Pre-processing Analysis Prior to COVID-19 Pandemic Period

The results showed there were no missing values in the dataset, and there were no observations in this multivariate dataset prior to the COVID-19 pandemic period indicated as an outlier. Turning to multicollinearity analysis, participation rate was found out to exist as a multicollinearity problem, the analysis was continued by removing the participation rate from the model. It was necessary to remove the collinearity problem variable because of the variable would cause instability in the regression calculation [30].

The analysis continued with analysing the association of criminal cases with economic indicators, the result showed KLCI (-0.07), and tourist arrivals (-0.11) had a very weak correlation with criminal cases, unemployment rate (-0.33) and consumer price index (0.42) show a weak correlation with criminal cases, and inflation rate (-0.51) indicated as a moderation correlation with criminal cases which is the highest correlation among the economic indicators. This result shows inflation rate was the most important variable prior to the COVID-19 pandemic period

3.2.2 OLS Regression Model Prior to COVID-19 Pandemic Period

This model was used to validate the result of the COVID-19 period had influenced the result included the COVID-19 period become nonsensible intercept caused by the fluctuation data in the COVID-19 period. The analysis was further conducted by using backward elimination to remove those economic indicators which are less important to the OLS regression model. The final model was obtained and only the inflation rate was significant to the regression model, the prediction criminal cases model is shown in Eq. 5.

$$\hat{y} = 167156 - 4912x_4 \qquad \text{Eq. 5}$$

3.2.3 Residual Analysis of OLS Regression Model Prior to COVID-19 Pandemic Period

The results showed the residuals had the constant variance with applied of Breusch-Pagan test, the residuals were completely uncorrelated with each other via the Durbin-Watson test which concluded the residuals were independence to each other's, and Jarque-Bera test concluded the residuals were normally distributed.

3.3 Analysis Included Influence of COVID-19 Pandemic Period

Since the COVID-19 pandemic had strongly affected the community of Malaysia due to restricted movement control order had been implemented to curb the outbreak of COVID-19. The period of COVID-19 pandemic started to influence the society and economy in February 2020 and keep continued until 2021. This analysis included 51 monthly observations, 1 dependent variable or criminal cases, and 6 independent variables or economic indicators which included the pandemic period until March 2021.

3.3.1 Data Pre-processing Analysis Included COVID-19 Pandemic Period

There were no missing values observed in the dataset included COVID-19 Pandemic period and the result showed there was no outlier detected in this multivariate dataset using Mahalanobis distance. The analysis was further conducted with multicollinearity, the VIF showed there have unemployment rate (21.81) and tourist arrivals (22.23) identified as presence of multicollinearity issue, tourist arrivals was removed from the model since it contributed the higher VIF compared to unemployment rate. After removing the tourist arrivals, the model did not show any presence of multicollinearity problem.

The analysis continued with the association between economic indicators and criminal cases, the Pearson correlation coefficient was once again used to identify the association. From the overall result, it showed an improvement of correlation when compared to the association prior to COVID-19 pandemic period. Inflation rate (0.28) indicated a weak correlation with criminal cases, which is the least important variable towards criminal cases. This is surprisingly showing the inflation rate having a total change compared to the analysis before COVID-19 as indicated the highest correlation with criminal cases. This result due to the impact of COVID-19 had made a significant fluctuation on both criminal and inflation rate during the COVID-19 period. While participation rate (0.31) and KLCI (0.41) had a weak correlation with criminal cases. There is an increment of correlation for KLCI with criminal cases about 0.34 compared to prior COVID-19, this suggested the influence of COVID-19 had made a change or improved the correlation of the stock market on the criminal case. Finally, the consumer price index (0.57) and unemployment rate (-0.56) had a moderate Pearson correlation with criminal cases which the results showed increments in the correlation with criminal cases after including the COVID-19 period. This could suggest consumer price index and the unemployment rate were more important after COVID-19 occurred in Malaysia.

3.3.2 OLS Regression Model Included COVID-19 Pandemic Period

This model was used to determine the possible criminal cases model which also acts as the comparative model with LASSO and Ridge Regression models. The analysis had further continued with the student's t-test as to find out the significance of economic indicators through backward elimination which contribute to the model included influence of the COVID-19 period. After the t-test was conducted, the final OLS regression model was obtained which only had less consumer price index and KLCI. The prediction criminal case model is shown in Eq. 6.

$$\hat{y} = -1510000 + 12520x_3 + 91.47x_5 \qquad \text{Eq. 6.}$$

3.3.3 Residual Analysis of OLS Regression model include COVID-19 Pandemic Period

The results showed the residuals had the constant variance through the Breusch-Pagan test which indicated the residuals were homoscedasticity. Then, the Durbin-Watson test showed the result as the residuals were completely uncorrelated and illustrated the residuals were independent to each other. Jarque-Bera test was used and came out the result showed the residuals were normally distributed.

3.4 LASSO Regression

This LASSO regression used the dataset included COVID-19 pandemic period with a total observation of 51. The optimal lambda value, λ had been obtained through 10-fold cross-validation method with using the lambda value that had the lowest mean squared error. The 10-fold cross-validation plot of LASSO regression had been showed in Figure 1.

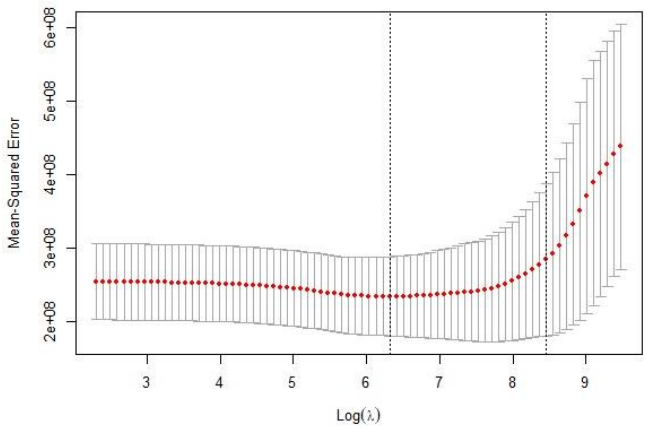

Figure 1: 10-fold cross-validation plot of LASSO regression model

The first dashed indicated the optimal value of log λ (6.12) with the minimum mean squared error, then obtained the optimal lambda, λ value of 454.0202 and further used to apply in the forming of LASSO regression model. The LASSO regression equation included COVID-19 pandemic period is shown in Eq. 7.

$$\hat{y} = -1379683 - 8826x_1 + 3863x_2 + 9920x_3 + 62.23x_5 \quad \text{Eq. 7}$$

There were two economic indicators were shrunk to zero such as inflation rate and tourist arrivals while the remaining 4 economic indicators were non-zero. The tourist arrivals shrunk to zero due to its higher collinearity value with other independent variables that lead to overfitting for the regression model. Also, LASSO regression would shrink the least important variables to zero, thus inflation rate contributes the least Pearson correlation coefficient with criminal cases, then caused it shrunk to zero and removed from the LASSO regression model.

3.5 Ridge Regression

The Ridge regression allowed to use all the predictor variables in the regression model with adding tuning parameter. This Ridge regression model included COVID-19 pandemic period, all the 51 observations and 6 independent variables or economic indicators had included in order to obtain the regression model. The 10-fold cross-validation plot was used to identify the optimal lambda, λ and the plot as shown in Figure 2.

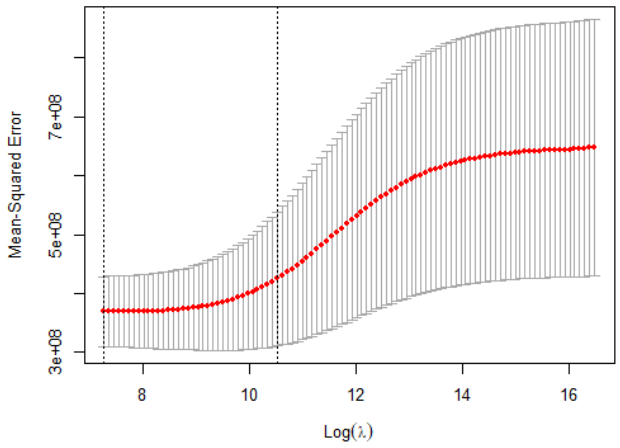

Figure 2: 10-fold cross-validation plot of Ridge regression model

The first dashed lines indicated the optimal value of log λ with 7.2 since it had the lowest mean squared error. Then, the optimal lambda, λ value (1419.138) was obtained which was allowed to obtain the Ridge regression model. The equation of Ridge regression model included COVID-19 pandemic period is shown in Eq. 8.

$$\hat{y} = -1669295 - 11589\,x_1 + 10871\,x_2 + 8374\,x_3 + 676.74\,x_4 + 69.27\,x_5 + 3038x_6 \quad \text{Eq. 8}$$

The Ridge regression model showed there were no zero coefficients was obtained from the model. All the economic indicators were included and allowed to make predictions of criminal cases in the full model.

3.6 Interpretation for Regression Models Included Influence of COVID-19 Pandemic Period

All the regression models used 51 monthly observations and there were not any observations being removed from the dataset from conducting outlier analysis through Mahalanobis distances. Additionally, all the result for intercepts obtained from OLS, LASSO and Ridge regression models were nonsensible due to the negative intercept obtained from the models. These nonsensible intercept had

been validated as compared to the OLS regression models prior to the COVID-19 period which showed the intercept is positive. This explained the fluctuation of the data that occurred during the COVID-19 pandemic period had affected the intercept of all regression models become negative. However, the estimated coefficients of the regression equations (OLS, LASSO and Ridge) show sensible have clearly show sensible and allowed the regression models to predict for criminal cases.

The performance of each regression model involved the period of COVID-19 pandemic was studied by using mean squared error (MSE), number of independent variables used, R-squared and adjusted R-squared. Table 1 shows the summaries performance of each regression model that is used to predict criminal cases using economic indicators.

Table 1: Summaries of models with influence of COVID-19 pandemic

Regression model	MSE	Number of independent variables	R-squared	Adjusted R-squared
OLS regression	312000000	2	52.5%	50.5%
LASSO regression	304019364	4	55.7%	51.8%
Ridge regression	314247886	6	56.2%	50.2%

For the OLS regression, it used only two economic indicators which were the least independent variables used but contribute the second highest MSE, the lowest R-squared value while compared to another two regression models. However, the adjusted R-squared of OLS gave the second higher result which indicated this model does not include the variables that with no or little importance toward criminal cases.

Next, the LASSO regression models were using 4 independent variables to form a criminal case regression model, as well as the total variation of criminal cases, could be explained by 55.7%. Although, the R-squared value was in between OLS and Ridge regression, but it adjusted R-squared value given the highest value which illustrated the additional input two variables had added a significant value to the model. The MSE of LASSO regression is the lowest and indicated it is the least error as compared to other regression models.

While for the last Ridge regression, it allowed to use all the 6 economic indicators to form a prediction of criminal cases model which had the highest R-squared value but given the highest MSE and the lowest adjusted R-squared value as compared to other regression models. This result show Ridge regression was the least accurate regression model.

Among the analysis had conducted for various regression models, although Ridge regression showed the R-squared value was the highest, however, ridge regression only increases by 0.5% variation by adding two more economic indicators as compared to the LASSO regression model. The additional two economic indicators had made the Ridge regression model become more complex regression and the additional variables did not contribute any significant value to the model and even increase the MSE as compared to LASSO regression. Thus, the LASSO regression model tends to do well and leave 4 economic indicators that had influence on the criminal cases. Lastly, LASSO regression was able to obtain a simpler model than Ridge regression and greater accuracy as compared to OLS and Ridge regression, thus LASSO regression was the most suitable to use as prediction of criminal case model with the influence of COVID-19 pandemic that keep continue happened in Malaysia.

4 Conclusion

In this research, it can be concluded the influence of the COVID-19 pandemic had largely affected the data and it was validated by comparing to the analysis prior to the COVID-19 pandemic period. The results for the Pearson correlation coefficient show inflation rate had a very weak correlation with

criminal cases which show the total opposite change as compared to the correlation coefficient prior to the COVID-19 pandemic period which indicated as the highest correlation coefficient with criminal cases. Furthermore, the variable selection done by LASSO regression and OLS regression showed consumer price index and KLCI were the key economic indicators to predict criminal cases after including the influence of the COVID-19 pandemic. There have two more economic indicators included in the LASSO regression such as unemployment rate and participation rate which illustrates both economic indicators were still acting as important variables to predict criminal cases after the influence of the COVID-19 pandemic. Among the performance of the regression models, the LASSO regression model could well explain the criminal cases by using 4 economic indicators in the condition of less errors and a better fit model. Although the regression models under the influence of the COVID-19 pandemic had been found for the negative intercept which had been proved by the cause of the COVID-19 pandemic that occurred in Malaysia, however the negative intercept would not affect the predictive criminal cases models since the estimated coefficients are sensible, thus allowed regression models to make prediction for criminal cases.

In future, more in-depth research on criminal cases related to economic indicators is needed. More data observations should be included in order to obtain more accurate results; thus, it could reduce the margin of error and reduce the fluctuation within the data caused by the pandemic of COVID-19.

Acknowledgement

The authors would also like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

References

[1] Hornby, A.S, Oxford Advanced Learner's Dictionary, 9th ed. Oxford: Oxford University Press, 2015.

[2] Street, B, "The Impact of Economic Opportunity on Criminal Behavior: Evidence from Fracking Boom" Logan: The center for Growth and Opportunity, Working Paper 2019 (No. 012), 2019.

[3] Tom Stafford, "Psychology: Why Bad News Dominates the headlines," BBC, July 29, 2014. [Online]. Available: https://www.bbc.com/future/article/20140728-why-is-all-the-news-bad [Accessed May 7, 2021]

[4] Yufei, Z. and J. Huifeng. 2010. Using GIS to Analyze and Forecast The Chinese Crime Rate. Information Science and Engineering (ICISE), December 4-6, 2010, Hangzhou, China. 3352-3354.

[5] Crutchfield, R. D., & Pitchford, S. R., "Work and Crime: The Effects of Labor Stratification," Social Forces, vol. 76, no. 1, pp. 93-118, 1997, doi: 10.1093/sf/76.1.93

[6] Ingilevich, V., & Ivanov, S., "Crime Rate Prediction in the Urban Environment using Social Factors," Procedia Computer Science, vol. 136, pp. 472-478, 2018.

[7] Ojog, D., "The Effect of Crime on Economic Growth," Bachelor dissertation, Eramus Univ., Rotterdam, 2014.

[8] McNeish, D. M., "Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences," Multivariate Behavioral Research, vol. 50, no. 5, pp. 471-484, 2015, doi: 10.1080/00273171.2015.1036965

[9] Hoerl, A. E., & Kennard, R. W., "Ridge Regression: Biased Estimation fro Nonorthogonal Problems," Technometrics, vol. 12, no. 1, pp. 55-67, 1970, doi: 10.1080/00401706.1970.10488634

[10] Tibshirani, R., "Regression Shrinkage and Selection Via the Lasso," Journal of the Royal Statistical Society: Series B (methodological), vol. 58, no. 1, pp. 267-288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x

[11] Chor, F. T., " An Exploration of Dynamic Relationship between Tourist Arrivals, Inflation, Unemployment and Crime Rates in Malaysia," International Journal of Social Economics, vol. 38, no. 1, pp. 50-69, 2011, doi: 10.1108/03068291111091963

[12] Rattner, A., "Social Indicators and Crime Rate Forecasting," Social Indicators Research, vol. 22, pp. 83-96, 1990, doi: 10.1007/BF00286392

[13] Wu, D. & Wu, Z., "Crime, Inequality and Unemployment in England and Wales," Applied Economics, vol. 44, no. 29, pp. 3765-3775, 2012, doi: 10.1080/00036846.2011.581217

[14] Bushway, S. D., Cook, P. J., & Phillips, M., "The Net Effect of the Business Cycle on Crime and Violence," SSRN Electronic Journal, 2010, doi: 10.2139/ssrn.1655741

[15] Profillidis, V. A., & Botzoris, G. N., "Econometric, Gravity, and the 4-Step Methods," in Modeling of Transport Demand, Amsterdam: Elsevier, 2019, pp. 271-351.

[16] Raj, S., & Kannan, S, "Detection of Outliers in Regression Model for Medical Data," International Journal of Medical Research & Health Sciences, vol. 6, no. 7, pp. 50–56, 2017.

[17] Brereton, R. G., "The mahalanobis distance and its relationship to principal component scores," Journal of Chemometrics, vol. 29, no. 3, pp. 143-145, 2015, doi: 10.1002/cem.2692

[18] Forthofer, R. N., Lee, E. S., & Hernandez, M., "Linear Regression", in Biostatistics, 2nd Ed, Cambridge: Academic Press, 2007, pp. 349-386.

[19] Hauke, J., & Kossowski, T., "Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data," Quaestiones Geographicae, vol. 30, no. 2, pp. 87-93, 2011, doi: 10.2478/v10117-011-0021-1

[20] Berrar, D., "Cross-validation," in Reference Module in Life Sciences, Tokyo: Tokyo Instititute of Technology, 2018.

[21] Alwee, R., Shamsuddin, S. M., & Sallehuddin, R., "Modeling of Violent Crime Rates with Economic Indicators using Hybridization of Grey Relational Analysis and Support Vector Regression," Jurnal Teknologi, vol. 76, no. 1, pp. 229-238, 2015, doi: 10.11113/jt.v76.3899

[22] Jenkins-Smith, H., Ripberger, J., Copeland, G., Nowlin, M., Hughes, T., Fister, A., & Wehde, W., Quantitative Research Methods for Political Science, Public Policy and Public Administration, 3rd Ed, Oklahoma: University of Oklahoma Libraries, 2017.

[23] Hyndman, R. J., & Athanasopoulos, G., Forecasting: Principles and Practice, Victoria: Otexts, 2018. [E-book] Available: Otexts e-book.

[24] Virgantari, F., Wijayanti, H., & Koeshendrajana, S., "Aitken's Generalized Least Square Method for Estimating Parameter of Demand Function of Animal Protein in Indonesia," in Journal of Physics, West Java: IOP Publishing, 2019.

[25] Thadewald, T. & Buning, H., "Jarque-Bera Test and its Competitors for Testing Normality – A Power Comparison," Journal of Applied Statistics, vol. 34, no. 1, pp. 87-105, 2007, doi: 10.1080/02664760600994539

[26] Khuneswari G. P., Ravie, T., & Padzil, S. A. M., LASSO Regression in Consumer Price Index Malaysia, AIP Conference Proceedings 2355, Maryland: AIP Publishing, 2021.

[27] Weisberg, S., Applied Linear Regression, 4th Ed. Hokoken: John Wiley & Sons, 2014.

[28] Kidwell, J. S., & Brown, L. H., "Ridge Regression as a Technique for Analyzing Models with Multicollinearity," Journal of Marriage and the Family, vol. 44, no. 2, pp. 287-299, 1982, doi: 10.2307/351539

[29] Pereira, J., Basto, M., & Silva, A., "The Logistics LASSO and Ridge Regression in Predicting Corporate Failure," Procedia Economics and Finance, vol. 39, pp. 634-641, 2016, doi: 10.1016/s2212-5671(16)30310-0

[30] Siegel, A.F., "Multiple Regression: Predicting One Variable From Several Others," in Practical Business Statistics, 7th Ed, Amsterdam: Elsevier Inc, 2016, pp 354-418.