

FaR 2021

International Forum “Freedom and responsibility in pivotal times”

**THE USE OF TIME-SERIES METHODS FOR BOTH CRIME
MODELING AND PREDICTION**

Andrey M. Terehov (a), Sergey I. Kuvychkov (b), Oxana G. Kolarkova (c),
Denis V. Nametkin (d)*
*Corresponding author

- (a) Privolzhsky branch of Russian State University of Justice, 17A, Gagarin Ave., Nizhny Novgorod, 603022, Russia, terehoff.t@yandex.ru
(b) Privolzhsky branch of Russian State University of Justice, 17A, Gagarin Ave., Nizhny Novgorod, 603022, Russia, redsrxjd@mail.ru
(c) Privolzhsky branch of Russian State University of Justice, 17A, Gagarin Ave., Nizhny Novgorod, 603022, Russia, kog0208@inbox.ru
(d) Privolzhsky branch of Russian State University of Justice, 17A, Gagarin Ave., Nizhny Novgorod, 603022, Russia, nametkindv@mail.ru

Abstract

The goal of the research is to analyze contemporary time-series methods which are used to identify socio-economic reasons for generating crime in the society, to predict crime for the medium term. The literary analysis on the matters of both crime prediction and modeling is given in the article. The multiple regression model has been implemented which is based on the factors affecting the law enforcement system. The predictive model of registered crimes has been constructed. The source data of the research are official statistical figures taken from Russian and foreign statistics for the period of 1992–2020. The peculiarities of using time-series models are revealed in theoretical research. Various groups of factors are highlighted affecting crime. Both a multifactorial regression model and a predictive ARMA-model with three prediction variants have been constructed. The quality of the models and crime predictions depends on the choice of the optimal method and the prediction period, on the complete information database including social, economic, legal and other characteristics of social phenomena and processes that can affect criminal situations. Econometric analysis methods of predicting crimes can allow one to quantify complex interactions between variables characterizing crime. Practical realization of modeling tools will definitely predict crime for the future and develop some actions to fight crime.

2357-1330 © 2022 Published by European Publisher.

Keywords: Crime, prediction, crime model, regression, statistical analysis, statistical data



This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Both crime modelling and prediction are an integral element of the government administration system that controls social processes in the country. Methods of predicting crime are widely used in criminology and based on existing scientific crime concepts, on using statistical, experimental methods, mathematical modelling. The article considers both theoretical and practical approaches towards the use of time-series methods for modelling and predicting the state of crime.

The objectives of the research work aim at analysing current experience of using time-series methods in different countries to predict the state of crime and reveal the most significant factors that affect crime; at developing and evaluating both multifactorial regression and predictive models by the example of the Russian Federation.

2. Problem Statement

The relevance of the study deals with analysis of contemporary methods such as crime modeling and prediction that allows one to determine more suitable and effective models which can be used by public authorities to fight crime with identification of more significant factors affecting the state of crime

3. Research Questions

Analysis and practical implementation of time-series methods to reveal socio-economic reasons for generating crime in the society to predict the state of crime for the medium term are the research questions.

4. Purpose of the Study

The goal of the research work is to analyze contemporary time-series methods which are used to identify socio-economic reasons for generating crime in the society, to predict the state of crime for the medium term.

5. Research Methods

The following scientific methods are used in the research work: analysis, synthesis, logical method, modeling, generalization. The literary analysis shows time-series methods as the most common ones to model and predict crime. Statistical figures, determined due to generalization method, were involved into the multifactorial regression model when constructing. Modeling the method was used to develop both a predictive model of registered crimes in the Russian Federation and a multifactorial regression model. The developed models have passed basic tests for statistical significance that confirm their effectiveness. The information database of the research work is Rosstat data, Eurostat data, Statistical department of UN, MIA of RF for the period of 1992–2020, legal acts and reference documents.

6. Findings

While reviewing contemporary methods for modelling and predicting the state of crime, it should be particularly focused on time-series methods that help to evaluate future results on the basis of historical data. These methods include Moving average, Weighted moving average, Kalman filtering, Exponential smoothing, ARMA, ARIMA, Extrapolation, Linear prediction, Trend estimation, Growth curve (statistics) and others. Let us consider briefly some authors' research results devoted to this question.

Shoesmith (2013) in his work deals with specific features of developing the spatial-temporal autoregressive model (ST-AR) for predicting the rate of violent and property crimes in the USA. The author explains the model of violent crimes in the state of Florida and the model of property crimes which is dedicated to develop experimental predictions. The modelling results are compared with the results of AR(p), VAR and BVAR models that predict the future crime rate. The author points out that modelling due to technology ST-AR is superior to other reviewed approaches to predicting crimes.

Some scientists present spatial-temporal model for predicting crime that contains a number of regions with high crime rate and its crime predictors, each of which is given as a predictive model to evaluate a number of crimes that is likely to happen in the region concerned. This experimental evaluation has been carried out on the basis of two real datasets taken in the cities of Chicago and New York. It shows good results of the chosen approach that provides acceptable accuracy of spatial and spatial-temporal crime prediction on sliding time horizons (Catlett et al., 2019).

Ingilevich and Ivanov (2018) represent three tested types of prediction models: linear regression, logistic regression and CatBoost. Prediction factors in these models were chosen with the use of feature selection methods. Such strategy allowed one to improve prediction accuracy and at the same time to avoid model retooling. The models were tested on the basis of St. Peterburg's criminal data. It was determined during the research that the model CatBoost is considered to be more appropriate to predict the crime rate in a certain urban area.

The research work of Zaman et al. (2019) and other scientists deals with the role of quality education in ensuring access to justice which is presented on the basis of data analysis in 21 countries for the period of 1990-2015. The model of panelised fixed-effect is used in the research. VDA-analysis shows the following results: there is a positive connection between crime rate and existence of articles in scientific journals. Increase in both per capita income and in applications for trademarks and technical cooperation grants lead to significant reduction of the crime rate in different countries.

The study of cause and effect relationships of crimes in Brazilian cities is represented in the work of Alves et al. (2018). The authors use several linear models for predicting crimes on the basis of existing correlations between crimes and social figures in the cities. The method of random forest regression has shown the prediction accuracy up to 97 %. The authors come to conclusion that unemployment and illiteracy are the most important elements that influence a number of murders.

Xu et al. (2018) concern with the fact how presence or absence of street light can affect crimes by the example of the USA. The authors suggest a generalized model of crime least squares in 879 groups of census areas. The results show the inverse relationship between the density of street lighting and crime rate,

in this case the influence of 24-hour period becomes various in accordance with the nature of crimes committed.

Such scientists as Vomfell et al. (2018) present the influence of crime prediction on optimal decisions in criminal justice to prevent crime. The authors evaluate the patterns of human activity taken from social websites: Taxi trip, Twitter and Foursquare. The analysis of crime data in New York for 6 months period confirms the value of the information used in terms of improving prediction accuracy of property crimes (increased by 19 %).

Osipov (2019) describes approaches to predicting extremist crimes. The following prediction methods are used: extrapolation, modelling, expert evaluation methods. The scientific work deals with specific features of implementing those methods and evaluation of prospects for using the targeting mode. Based on the figures that characterize migration processes, the author predicts the development of extremist crimes in the country for the short and medium term. The research work considers the matters of factors' influence on crime rates among which special attention is paid to the legal basis.

Pilugina et al. (2015) touch upon the problems of mathematical crime modelling together with various interrelated types of socio-economic factors. The authors take quantitative data analysis using both systemic and structure-function strategies to describe current social processes. They identify the tendency towards crime decrease/increase in accordance with social transformations in the legislative sphere. The scientists emphasize the significance of mathematical modelling of crime time-series with regard to prediction of the enforcement system.

Markina (2019) describes crime interrelation in the country with its level of economic development; implements vertical and horizontal analysis, factor analysis, data comparing and contrasting, prediction of economic crimes figures. According to research results the author develops some proposals to optimize enforcement activity on the basis of evaluation of contemporary state and trends in economic crime development in the Russian Federation.

Specific features of implementing criminological prediction of recidivism crime are regarded in the scientific work of Artemiev and Pancratov (2016). It is established that prediction of recidivism crime allows one to minimize uncertainty of main regularities and trends of its development, to identify quality status of factors that most intensively affects it. The practical use of this method contributes to improving the quality of management decisions made by authorities of penal correction policy Bogdanova et al. (2019) describe the methodology of crime prediction in RF for the period of 2010–2019, based on time-series specificity that takes into account the following important elements: possibility to consider temporary information value, model self-control. The prediction has been performed with the use of exponential smoothing.

Pilugina and Natura (2017) deal with analysis of mathematical methods for modelling crime and socio-legal factors affecting its rate. The authors prove high efficiency of the simple linear extrapolation method, seasonal component allocation, quantitative analysis with the use of structure-function and systemic approaches to predicting crimes.

Thus, a wide range of methods can be used for modelling and predicting crimes. There are the following methods with the positive record: VAR, BVAR, Extrapolation, Logistical Regression, Gradient Boosting, VDA-analysis and some others. The prediction periods range from a few weeks to a year

(including time of the day). There are various groups of factors that affect crime. The following methods deserve special attention: linear regression that helps to find out crime predictors; autoregressive models are sure to be effective ones for prediction. These methods are involved further in the practical part of our research work.

We have carried out the linear multiple regression model to find out socio-economic factors that have an interrelated influence on the amount of registered crime. For our research we selected statistical data for 20 indicators characterizing the socio-economic sphere, the time period of 1992–2019. The indicator “number of registered crime (in thousands)” is taken as a dependent variable (Y) that characterizes the state of crime. A preliminary correlation analysis showed the closest relation between Y and some factors that had been involved into the basis for constructing the model.

In the Gretl software environment we have constructed the linear multiple regression model using the least squares method and this model has got the following form:

$$Y \approx -4.543.93 + 2.22 * X1 - 0,00254 * X2 + 21.9 * X3 + 1.23 * X4 + 0.513 * X5 + 0.00226 * X6,$$

where $X1$ – number of prisoners, thousands; $X2$ – number of divorces, units; $X3$ – population with money incomes below subsistence minimum, millions; $X4$ – number of students who take further vocational education in colleges, thousands; $X5$ – number of students who take Bachelor’s, Specialist’s and Master’s courses, thousands; $X6$ – those people who arrived in RF from CIS and more distant countries, thousands.

Multiple determination coefficient (R-square) $D = 0.9182$, i.e. 91.82 % of the variation in the number of registered crimes is due to variation of the factors studied. Significance testing of the multiple correlation coefficient carried out through comparison of actual and tabular values of the F-criterion (Fisher’s test), showed the model adequacy ($F_{\text{fact}} = 37,39 > F_{\text{tabl}} = 2,60$, with the probability of error of $\alpha = 0.05$). The reliability model level is 95 %. The model meets the main qualitative characteristics and this fact is confirmed by other tests for statistical significance: tests for non-linearity, for normal error distribution, for ARCH-processes, White and Breusch-Pagan tests for heteroscedasticity and for multicollinearity, LM-test for autocorrelation. The resulting model shows the interrelated influence of the factors included into the model on the state of crimes.

On the basis of the developed model it is not an easy matter to predict the state of crime because it is impossible to determine the behaviour of all affecting factors included into the model for the future. In order to predict in the environment Gretl, we offer our ARMA-model (2, 0) with the following characteristics: reliability level – 95 %, $R^2 = 0.79$. The values of information criteria are as follows: $AIC = 404.42$, $HQC = 406.13$, $BIC = 409.89$. The precise method of maximum likelihood is used. Time-series (1992–2020) was preliminarily tested for stationarity (ADF-test), the test showed the stationarity of time-series (with constant and trend) in its levels. The model successfully passed all main tests for statistical significance. The model equation is the following:

$$y_t = 1.18243 * y_{(t-1)} - 0.371075 * y_{(t-2)} + 2626.42$$

where y_t – a number of registered crimes, thousands.

Within the framework of this model we have calculated prediction variants for the medium term (2021–2023), thousands: calculated prediction (the period of 2021 – 2161.418; 2022 – 2292.636; 2023 – 2404.295), optimistic prediction (2021 – 1732.938; 2022 – 1629.096; 2023 – 1608.084), pessimistic prediction (2021 – 2589.899; 2022 – 2956.177; 2023 – 3200.506). Thus, the model prediction shows a slight increase in registered crimes for the medium term, while remaining variants show, accordingly, a decrease and a significant increase in the indicator. At the same time the implementation of a specific scenario will depend on the effectiveness of the state's socio-economic policy. It should be noticed that this model could not predict real crime rates as its significant part is latent and is not registered. At the same time the development of methods for predicting the total amount of crime is relevant and requires new approaches to quantitative evaluation of indicators that are not taken into account (due to their latency).

7. Conclusion

The article represents the analysis of both Russian and foreign scientists' research results that highlight effective predicting models and methods, and the groups of factors affecting greatly the development of criminal situation. Practical implementation of those methods shows different confidence (quality) of predictive values. The most common methods are extrapolation and modelling. The method of random forest regression (prediction accuracy – up to 97 %), the method of Holt exponential smoothing with monthly seasonality, logistic regression, simple linear extrapolation proved to be the best. The promising direction is the development of methods for modelling and predicting latent crime.

The research work implements the linear model of multiple regression and determines statistically significant factors. These are a number of prisoners and divorces; population with money incomes below subsistence minimum; a number of students who take further vocational education in colleges; a number of students who take Bachelor's course; those who arrived in RF from CIS and more distant countries.

Such model is constructed to reveal socio-economic factors that affect crime in an interrelated way.

The developed predictive shows ARMA-model shows three variants of predicting registered crimes. The implementation of this or that variant will depend on effectiveness of socio-economic policy realized by the state.

The use of a predictive approach in the practice of the public administration apparatus contributes to improving the quality of making management decisions. Further research aimed at improving models and modelling tools is sure to contribute to formulating more precise conclusions concerning relations between various parameters of crime.

References

- Alves, L. G., Ribeiro, H. V., & Rodrigues, F. A. (2018). Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and its Applications*, 505, 435-443.
- Artemiev, N. S., & Pancratov, V. A. (2016). General theoretical aspect of predicting recidivism (penitentiary) crime. *Human: crime and punishment*, 2(93), 6–10.
- Bogdanova, M. V., Parshintseva, L. S., & Kvachko, V. Y. (2019). Methodology of crime modeling and predicting in the Russian Federation. *Legal information technology*, 4, 50–62.
- Catlett, C., Cesario, E., Talia, D., & Vinci, A. (2019). Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing*, 53, 62–74.

- Ingilevich, V., & Ivanov, S. (2018). Crime rate prediction in the urban environment using social factors. *Procedia Computer Science*, 136, 472–478.
- Markina, S. A. (2019). Evaluation and prediction of economic crime in the system of ensuring economic security. *Bulletin of South-West State University. Series: Economics. Sociology. Management*, 9(6(35)), 147–162.
- Osipov, V. A. (2019). Some aspects of predicting extremist crime development in the Russian Federation. *Bulletin of Belgorod Legal Institute of MIA of Russia after I.D. Putilin*, 4, 17–22.
- Pilugina, T. V., & Natura, D. A. (2017). The use of mathematical modeling method for predicting regional crime in matters of its prevention. *All-Russian criminological journal*, 11(1), 61–70.
- Pilugina, T. V., Pichkurenko, E. A., & Vendina, A. A. (2015). Mathematical methods of crime prediction. *Aspectus*, 1, 98–103.
- Shoesmith, G. L. (2013). Space–time autoregressive models and forecasting national, regional and state crime rates. *International Journal of Forecasting*, 29, 191–201.
- Vomfell, L., Härdle, W. K., & Lessmann, S. (2018). Improving crime count forecasts using Twitter and taxi data. *Decision Support Systems*, 113, 73–85.
- Xu, Y., Fu, C., Kennedy, E., Jiang, S., & Owusu-Agyemang, S. (2018). The impact of street lights on spatial-temporal patterns of crime in Detroit, Michigan. *Cities*, 79, 45–52.
- Zaman, K., Usman, B., Sheikh, S. M., Khan, A., Kosnin, A. B. M., Rosman, A. S. B., Ismail, S., Ali, D. F., & Hishan, S. S. (2019). Managing crime through quality education: a model of justice. *Science & Justice*, 59(6), 597-605.