ARIMA MODEL AND NEURAL NETWORK: A COMPARATIVE STUDY OF CRIME RATE MODELLING

[1]James Tolulope O. , [2]Suleiman, S., [3]Udomboso C. G and[1]Babayemi, A.W.

Department of Mathematics, Kebbi State University of Science and Technology, Aliero
Department of Mathematics, Usmanu Danfodiyo University Sokoto
Department of Statistics, University of Ibadan, Ibadan

Abstract
Crime rate is a serious issue that affects everyone in society. It affects the victims, perpetrators, their families the government and even reality of good governance. In this study forecasting of crime rate using autoregression integrated moving average (ARIMA) model was compared with feed forward neural networks. The J multi software was used for analysis of data gotten from State Police Headquarter in Kebbi State from January 2004 to December 2013 and the series was stationary at first difference and ARIMA (0, 1, 1) was obtained as the best model for the series. This was model by Neural Network using SPSS. In the training of the network, the samples were automatically partitioned in to 73.3% of training and 26.7% of testing. The computational result shows that Artificial Neural Network provides better model than ARIMA by having minimum error in the in-sample and out -of- sample in MAE, MSE, and RMSE with 3.84614, 2.00466 and 1.41586 respectively.

Keywords: Crime rate, Time Series, Neural Network, FFNN and ARIMA

1.0 Introduction

Crime is a serious issue that affects everyone in society. Crime has increased drastically within the past few years. The concept of crime is very relative, crime is a term widely used by law persons to identify such act as embezzlement, forced sex (rape), murder, robbery, insurgence, kidnapping and any other act labeled as crime in a given community. Haralambo(1980) views crime as those activities that breaks the law of the land and are subject to official punishment. There are various different types of crime, minor crime, serious crime, on serious crime Sandra, et al.,(2009) says two of its more heinous components, is, murder and kidnapping. These two types of crime have actually traumatized the society as they have increased significantly over the recent years.

1.1 Causes of Crime

Many researchers are of opinions that there are different causes of crime. These include: High rates of unemployment could be taken to influence the opportunity cost of illegal activity (Witte and Witt, 2001). Existing research suggests that higher unemployment is associated with greater occurrence of property crime, but this relationship turned to be insignificant or negative for violent crime. (Saridakis, 2004; Raphael and Winter-Ebmer, 2001; Entorf and Spengler, 2000). Lack of education can be a cause of crime and criminal behaviour (Usher, 1997 and Becker and Mull

2.0 Measures of Errors

Three measures are used to evaluate the performance of models in forecasting crime rate as follows: mean absolute error (MAE), mean squared error (MSE) and root mean square error (RMSE). These measures are defined as

MAE $$\frac{1}{N}\sum_{t=1}^{N}\left|Z_t - \hat{Z}_t\right| = \frac{1}{N}\sum_{t=1}^{N}\left|e_t\right|$$

(1.1)

MSE= $$\frac{1}{N}\sum_{t=1}^{N}\left(Z_t - \hat{Z}_t\right)^2 = \frac{1}{N}\sum_{t=1}^{N}e_t^2$$

(1.2)

RMSE= $$\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(Z_t - \hat{Z}_t\right)^2} = \sqrt{\frac{1}{N}\sum_{t}^{N}e_t^2} \quad (1.3)$$

3.0 Materials and Methodology

3.1 ARIMA Models

A time series Xt is said to be an autoregressive moving average model of order p,q if it has representation form of

$$Xt = \sum_{i=1}^{p}\phi_i x_{t-1} + \sum_{j=0}^{q}\varepsilon_t + \theta_j \varepsilon_{t_{-1}}$$

(1.4)

The lag or backshift operator is defined as $ly_t = ly_t - 1$. Polynomials of lag operator or lag polynomials are used to represent ARMA models as follows

AR(P): ϕ (L) = 0 MA(q) model:$\theta(l) = 0$ ARMA(p,q)= ϕ (L) = :$\theta(l)\varepsilon_t$ it shows that an important property of AR(p) process is invertible, i.e. an AR(p) process can always be written in terms of an MA(∞) process. Whereas for an MA(q) process to be invertible, all the roots of the equation θ (L) = 0 must lie outside the unit circle.

3.2 Neural Network Model

The statistical neural network (SNN) model structurally is composed of two parts: the predictive and the residual, as is in classical regression, given as

$$y = f(X, w) + e_i$$

(1.5)

where $f(X,w) = \alpha X + \sum_{h=1}^{H}\beta_h g(\sum_{i=0}^{I}\gamma_{hi}x_i)$.

Recall The predicted model below

$$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$$
(1.6)

Thus equation (1.7) can be written as

$$y = \alpha X + \sum_{h=1}^{H} \beta_h g\left(\sum_{i=0}^{I} \gamma_{hi} x_i\right) + e_i$$
(1.7)

$X = (x_0, x_1, \ldots, x_I)$ is the vector of the input variable, $g(.)$ is the transfer (or activation) function and $w = (\alpha, \beta, \gamma)$ are the weights (or parameters) associated with the input vector, hidden neuron and the transfer function respectively, while e_i is the error associated with the network. We note that when there is no hidden neuron, the SNN reduces to the ordinary regression model.

3.3 Data Characteristics

This is the graphical representation of the crime rate under consideration before any transformation (at their levels). The variable considered is all cases registered at the police headquarters station in Kebbi State from January 2004 to December 2013 which is one hundred and twenty months (120 months) is depicted in the Graphs. Summary statistics of the variable, unit root analysis are shown below.

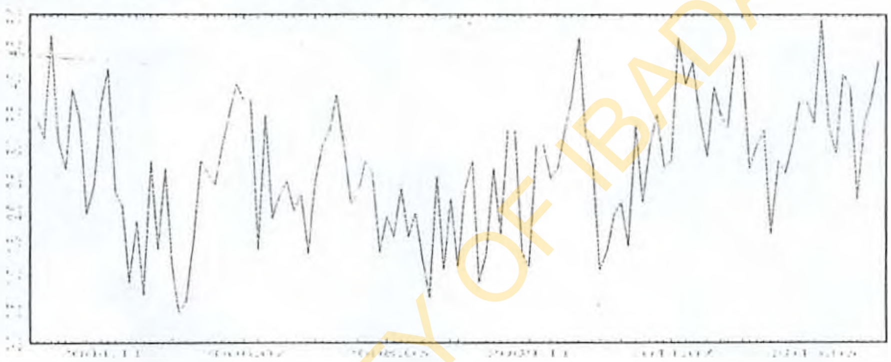

Fig. 3.0 Time plot of Number of crimes at levels

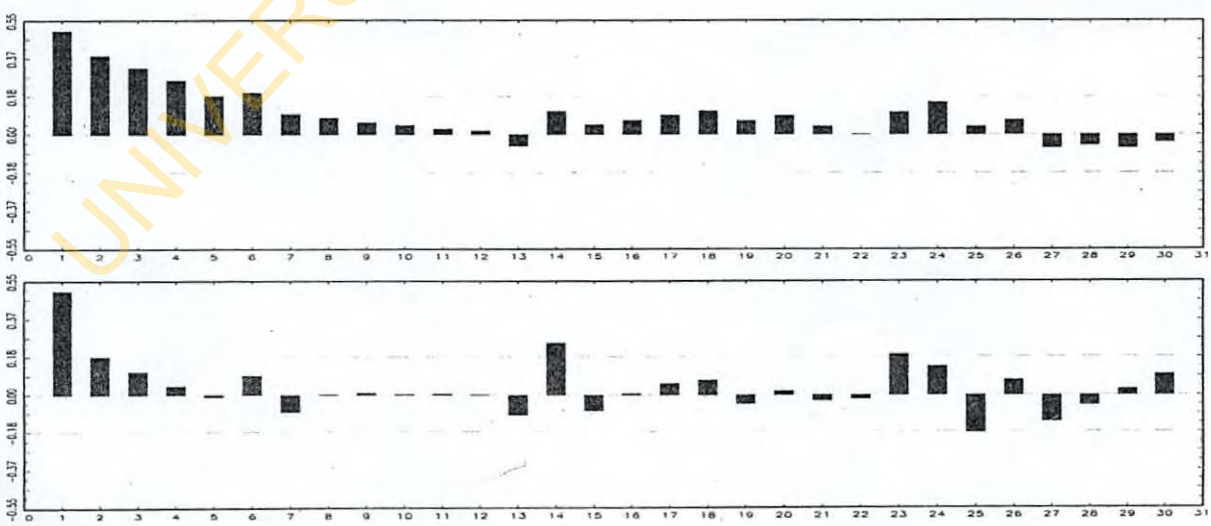

Fig3.1 Autocorrelations of crime at levels

Fig. 3.2 Time plot of Number of Crime at first difference

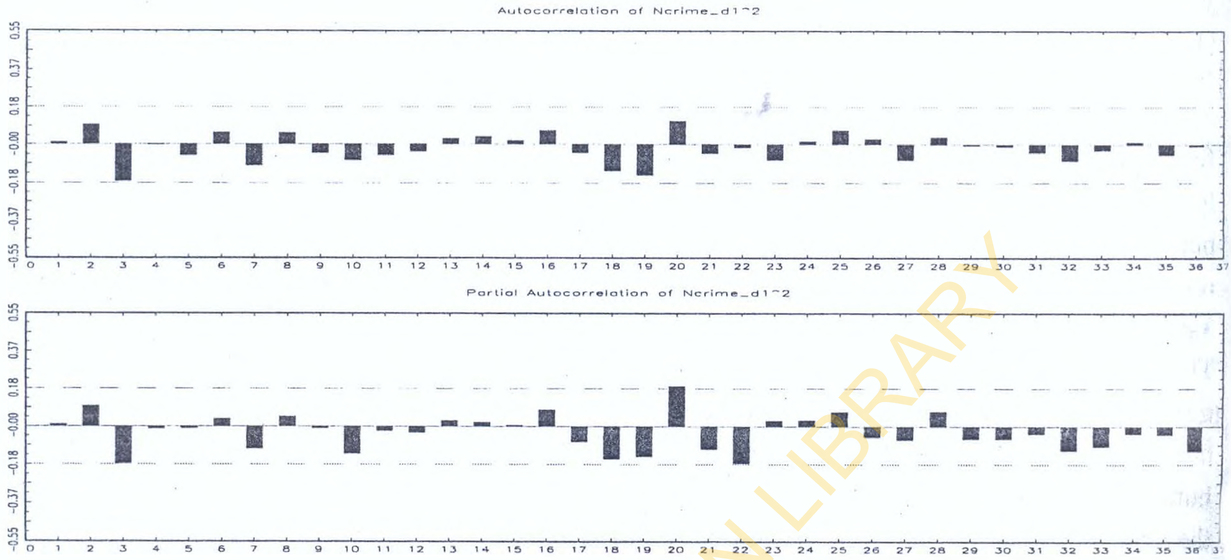

Fig. 3.3 Autocorrelations of crime at first difference

Fig. 3.0 shows that in November 2004 there was increase in crime rate and in January 2005 the least number of crimes was recorded. In March 2013 the crime rate reaches its peak in the series, more so the highest fluctuations were noticeable between the year 2009 and 2013, the reason may be as a result of increment in population between these years, thereby increasing the number of crime committed by the populace. However, because of this difference in fluctuations we may conclude that the mean and variance are not equal. In fig. 3.2 the autocorrelations shows that there is level of correlation between the series hence stationary and in fig. 3.1 the series plot on number of crime at first difference shows the fluctuations are about the mean of zero (0) and the number of upward movement and their levels are equal to the number of downward movement.

4.0 Result and Discussion

4.1 Testing For Stationarity

Testing for stationarity of a series involves the use of two tests;

1 AUGUMENTED DICKEY FULLER TEST (ADF TEST).

H_0 = It is stationary

H_1 = The number of crime at level contains unit root

If the value of the test statistics is less than the critical value, we can reject the null hypothesis and conclude that the series is stationary at level

2 KWIATKOWSKI PHILIPS SCHMIDT SHIN TEST (KPSS TEST).

H_0 = It is not stationary

H_1 = The number of crime at level is stationary

If the value of the test statistics is greater than the critical value, we can reject the null hypothesis and conclude that the series is not stationary

Table: 4.1 Stationary test of NCrime at a level

Test	Significance test 1% 5% 10%	Test statistics	Decision
ADF	-3.96 -3.41 -3.13	-2.7703	Reject H_o
KPSS	0.347 0.463 0.739	1.0187	Accept H_o

Table: 4.2 Stationary test of NCrime at a level

Test	Significance test 1% 5% 10%	Test statistics	Decision
ADF	-3.96 -3.41 -3.13	-11.2899	Accept H_o
KPSS	0.347 0.463 0.739	0.0324	Reject H_o

The results in table 1 and 2 indicate that the series were not stationary at a level using both ADF and KPSS test statistics but were made stationary after one difference is taken.

4.2 MODEL IDENTIFICATION

Table 4.3 Possible ARIMA models for the series

Possible models	AIC values
ARIMA (0,1,1)	1232.969336
ARIMA (1,0,0)	1218.549035
ARIMA (0,1,1)	1206.421380
ARIMA (1,1,1)	1218.205197
ARIMA (2,1,1)	1213.701165
ARIMA (2,1,2)	1221.823987
ARIMA (2,2,2)	1276.309463
ARIMA (2,2,3)	1235.307230
ARIMA (3,2,3)	1257.500679
ARIMA (3,1,3)	1226.222273
ARIMA (3,0,3)	1207.086099

In this table, ARIMA type models with various combinations of (p, d,q) parameters ranging from (0,1, 1) to (3,0, 3). Therefore the best model for this series is ARIMA (0,1,1). Since it have the least number of AIC values.

4.3 Model Estimation

The estimated model ARIMA (0, 1, 1) can be represented as follows:

$Xt = e_t + 0.34638076e_{t-1}$

Adequacy of the model is tested using Portmanteau test

DIAGNOSTICS CHECKING MODEL

PORTMANTEAU TEST with 1 lags

Portmanteau: 0.8478 p-Value (Chi^2): -1.0000

Ljung & Box: 0.8693 p-Value (Chi^2): -1.0000

H_0 : all the lags correlation are zero

H_1 : atleast one lag correlation is not equal to zero

Since p. value < Portmanteau and Ljung & Box, we reject H_0 and conclude that atleast one lag correlation is not zero.

JARQUE-BERA TEST:

test statistic:	1.5207	p-Value(Chi^2):	0.4675
skewness:	-0.0944		
kurtosis:	2.4794		

H_0 : the series is symmetric hence normal

H_1 : the series is non symmetric hence non normal.

P. value > skewness and p. value < kurtosis we reject the null hypothesis and conclude that the series is non symmetric hence non normal.

ii) The feed forward neural network model

In the training of the network, the samples were automatically partitioned in to 73.3% of training, 26.7% of testing. In the architecture, the number of hidden neurons was also automatically selected in order the select the best one in terms of the predicted values. However, the number of hidden layer is fixed to be one throughout the period of the training. The hyperbolic tangent was used as the activation in the hidden layer because the output of the hidden neuron is continuous. Whatever is the output of the hidden neurons is similarly transformed by identity activation function of the output neurons by comparing it with the target variables of the output neurons. During the training, the network parameters were adjusted based on the set values as; initial learning rate is η_0=0.4, lower boundary of learning rate is 0.001, learning reduction, in epochs of 10, momentum rate is $\alpha = 0.9$, interval center 0 and interval offset ±0.5. The algorithm used for the training is called resilient back propagation/ online training algorithms. The network is trained until the number of epochs is equivalent to 10, SPSS software is used to train the network. With the above specification the following synaptic weights are obtained.

Table 4.4 Synoptic weight of FFNN Model

Predictor		Predicted			
		Hidden Layer 1			Output Layer
		H(1:1)	H(1:2)	H(1:3)	Crime
Input Layer	(Bias)	.275	.234	-.197	
	Jan	.419	-.369	-.300	
	Feb	-.368	-.027	-.185	
	March	.515	-.257	.000	
	Apr	-.372	-.608	-.018	
	May	.364	.274	.012	
	Jun	.002	-.181	-.467	
	Jul	-.413	.370	-.070	
	Aug	.368	.168	-.759	
	Sep	-.251	-.141	.420	
	Oct	-.474	-.510	.559	
	Nov	-.033	-.103	.132	
	Dec	.163	-.117	-.172	
Hidden Layer 1	(Bias)				-.094
	H(1:1)				-.003
	H(1:2)				.139
	H(1:3)				-.328

FFNN forecasting model can be constructed using above synoptic weights as follows

$$\hat{Z}_i = \mu + \delta \hat{Z},$$

$\hat{Z}_i = \mu + \delta \hat{Z}$, where μ and δ are the mean and standard deviation of the in-sample data set and

$\hat{Z}_s$ = -0.093 -0.003H (1:1) + 0.139(1:2)-0.328(1:3)

Table 4.5 Model Summary

Training	Sum of Squares Error	41.474
	Relative Error	.953
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.06
Testing	Sum of Squares Error	12.159
	Relative Error	.897

Dependent Variable: Crime

a. Error computations are based on the testing sample.

Table 4.3 above shows information on the result of training and applying the network to the *testing and hold out samples. Sums of squares error is displayed because* the output layer has scale-dependent variables. This is the error function that the network tries to minimize during training. It should be noted that the sums of squares and all errors values are computed for the rescaled values of the dependent variable as shown in the table below. The relative error for the scale-dependent variable is the ratio of the sum of squares error for dependent variable to the sum of squares error for the "null" model in which the mean value of the dependent variable is used as the predicted value for each case. In this case the relative error for the dependent variable is 0.953. The average overall relative error and relative errors are fairly constant for training and testing, which gives us some confidence that the model is not over trained and that the error in future cases scored by the network will be close to the error reported in the table

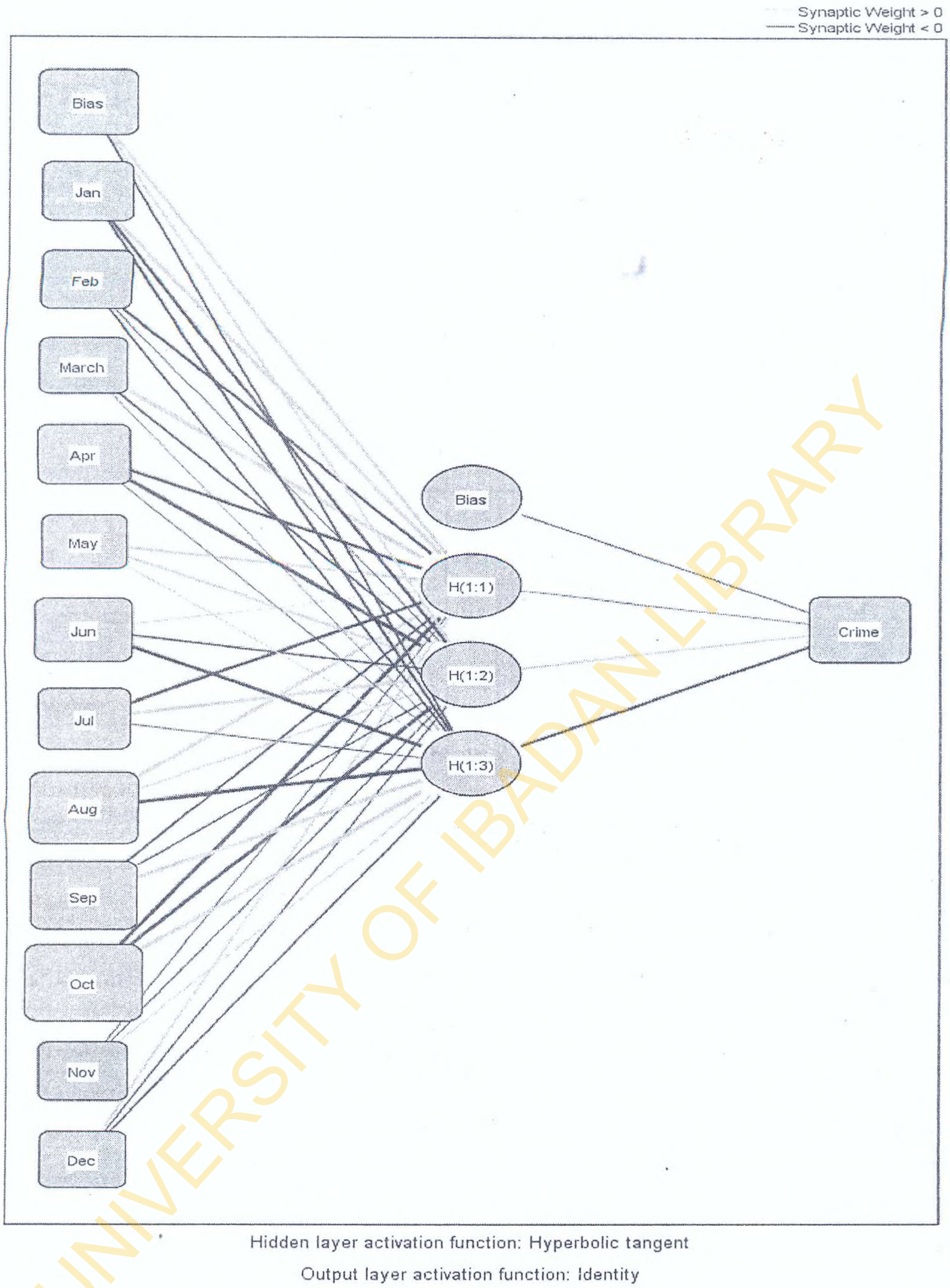

Figure 3.6 Neural network Architecture for Crime Rate

Table 4.4. Network Information

Input Layer	Covariates	1	Jan
		2	Feb
		3	March
		4	Apr
		5	May
		6	Jun
		7	Jul
		8	Aug
		9	Sep
		10	Oct
		11	Nov
		12	Dec
	Number of Units[a]		12
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		3
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Crime
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

The network information table displays information about neural networks and its usefulness in ensuring that the specifications are correct. We note here that the number of units in the input layer is the number of covariates which is 12 in all, January to December, likewise, the number of hidden layers is 1 and number of units in the Hidden layer is 3.

4.4 Comparison of Models performance for forecasting crime rate

In order to examine the fitness of the standard selected ARIMA model and Neural Network model, both of them have been used to make out of sample forecast and the results are reported in the table below.

Table4.5. Error Measures for ARIMA and Neural Network

Sample	Error	ARIMA	NN
In-Sample	MAE	34.64211	3.84614
	MAPE	41.74100	34.4500
	RMSE	42.76374	1.41586
Out-of-Sample	MAE	36.38562	4.43571
	MAPE	60.37820	45.6231
	RMSE	44.62937	2.14356

From the above table, it is observed that Neural Network model has minimum error measures in both the in-sample and out-of-sample sets compared to ARIMA. From the above study it is observed that, Neural Network is good at forecasting crime rate.

5.0 Conclusion

Since ARIMA (0,1,1) has shown the best performance and thus selected for construction of ANN model. The crime rate of both models has been calculated based on the test data which were not used in the calibration phase. The computational results show that the ANN model for forecasting crime rate outperform ARIMA model.

References

Becker, G and Mulligan, C. (1997), The Endogenous Determination of Time Preference, The Quarterly Journal of Economics, 112/3:729-58.

Chatfield, C. 2004. *The analysis of time series: an introduction.* 6th edition, Chapman and Hall/CRC. Boca Raton.

Entorf, H. and Spengler, H. (2000). Socioeconomic and Demographic Factors of Crime in Germany: Evidence From Panel Data of the German States, International. Review of Law and Economics, 20:75-106..

Raphael, S. and Winter-Ebmer, R. (2001), Identifying the Effect of Unemployment on Crime, Journal of Law and Economics, XLIV:259-283.

Saridakis, G. (2004), "Violent Crime in the United States of America: A Time-Series Analysis Between 1960-2000", European Journal of Law and Economics, 18(2), pp 203-221, ISSN 0929-1261.

Witte and Witt (2001), Crime Causation: Economic Theories, Encyclopedia of Crime and Justice.

Usher, D. (1997), Education as a Deterrent to Crime, Canadian Journal of Economics, 30/2:367-84.