Modeling and Forecasting of Armed Robbery Cases in Nigeria using Auto Regressive Integrated Moving Average (ARIMA) Models

Obubu M^{1*} Nwokolo, P.C²

1. Department of Statistics, Nnamdi Azikiwe University, P.M.B 5025, Awka, Nigeria

2. Department of Mathematics, Federal College of Education (Technical) P.M.B 0189 Umunze, Nigeria

Abstract

We have utilized a twenty-nine year crime data in Nigeria pertaining to Armed Robbery, the study proposes crime modeling and forecasting using Autoregressive Integrated Moving Average Models, the best model were selected based on the minimum Akaike information criteria (AIC), Bayesian information criteria(BIC), and Hannan-Quinn criteria (HQC) values and was used to make forecast. Forecasted values suggest that Armed Robbery would slightly be on the increase

Keywords: crime rate, ARIMA, armed robbery, forecasting, ACF/PACF, Akaike information criteria, Bayesian information criteria, Hannan-Quinn criteria

1. Introduction

At the moment, there is no universal definition of crime. This is primarily as a result of continual changes in social, political, psychological and economic conditions of people in respective crime sites. An act may be a crime in one society, but not in another (Danbazau, 2007). For example, prostitution, adultery and homosexuality between consenting adults have been wholly or partially removed from the criminal law in USA (Feldman, 1997) but are considered as crimes in Muslim communities such as Saudi Arabia and Sharia states in Northern Nigeria. The constant changes in time also change the perception of society on crime. Today, it is also becoming crime to pollute the air and water. Pollution causes few problems and receives little attention in colonial days (Usman et al. 2012). Therefore, the perception of an act as a crime varies with time and space. These perceptions are not determined by any objective indicator of the degree of injury or damage but by cultural values and power relations. Therefore, many scholars have defined crime in different views, mostly bordering on ethical and ideological orientation. In a strict legal definition, a crime is a violation of criminal law, which in most societies can be defined broadly as any 'act or omission forbidden by law on pain of punishment' (Carvell and Swinfen, 1970). Tappan (1947) defined crime as an intentional act in violation of the criminal law committed without defense or excuse and penalized by the state. One of the fundamental techniques to combat criminal activities is to better understanding of the dynamics of crime which can be harnessed by understanding the trend of collected crime statistics which in turn can be used for crime forecasting. The classification of crime differs from one country to another. In Nigeria, the Federal Bureau of Investigation tabulates the annual crime data as Uniform Crime Reports (UCR). They classify violations of laws, which derive from common law as part 1 (index) crimes in UCR data, further categorized as violent as property crimes. Part 1 violent crimes include murder and criminal homicide (voluntary manslaughter), forcible rape, aggravated assault, and robbery; while part 1 property crimes include burglary, arson, larceny/theft, and motor vehicle theft. All other crimes count as part II crimes. In Nigeria, the Police classification of crime also depends on what law prescribed. In Nigeria Police Abstract of Statistics (NPACS), offences are categorized into four main categories: Offences against persons include: manslaughter, murder and attempted murder, assault, rape, child stealing, grievous harm and wounding, etc. Offences against property include: armed robbery, house and store breakings, forgery, theft/stealing, etc. Offences against lawful authority include: forgery of current notes, gambling, breach of peace, bribery and corruption, etc. Offences against local act include: traffic offences, liquor offences, etc. Nigeria has one of the highest crime rates in the world. Murder often accompanies minor burglaries. Rich Nigerians live in high secured compounds. Police in some states are empowered to shoot on sight violent criminals. In the 1980s, serious crime grew to nearly epidemic proportions, particularly in Lagos and other urbanized areas characterized by rapid growth and change, by stark economic inequality and deprivation, by social disorganization, and by inadequate government service and law enforcement capabilities. Annual crime rates fluctuated at around 200 per 100,000 populations until the early 1960s. The purpose of this paper is to examine the modeling and forecasting of Armed Robbery rate using ARIMA models.

2. Materials and Method

In this paper, we have used the Crime data on Armed Robbery Cases for past 29 years (1986 -2014). The crime data were sourced from Nigeria Information Resource Centre. We have used GRETL (Gnu Regression, Econometrics and Time-series Library) software for plotting the graphs and analysis of the data set.

2.1 **Box-Jenkins ARIMA Model**

 $\Delta y_t = Y_t - Y_{t-1} - \dots - (1)$ $\Delta^2 y_t$

$$Y_{t} = \Delta Y_{t} - \Delta Y_{t-1} = Y_{t} - 2Y_{t-1} + Y_{t-2} - \dots - (2)$$

where, Y_t is time series at time t, Y_{t-1} is the proceeding time series of Y_t , Δy_t is the first order difference, $\Delta^2 y_t$ is the second order difference of the current observation, y_t is the current observation and Y_{t-2} is the preceding time series to Y_{t-1} in the same series.

After the appropriate differencing, the expected time series is expected to exhibit features of a stationary time series so that the appropriate ARIMA (p, d, q) process can be used to model the remaining serial correlation in the series.

Where p is the number of auto regressive terms, d is the number of non seasonal differences, q is the number of lagged forecast errors in the prediction equation.

for a time series process Y_t , ARIMA (0,0,1) / AR(1) is the first order auto-regressive process and is given by;

 $y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$ -----(3)

and a first order moving average process ARIMA (0,0,1) / MA(1) and is given by;

 $y_t = \mu - \theta_l \varepsilon_{t-l} + \varepsilon_t$ -----(4)

where ϕ and θ are coefficients of polynomial with order p and q respectively. Alternatively, the model ultimately derived may be a mixture of these processes and of higher orders, in that case, a stationary ARMA (p, q) process is defined by;

 $y_{t} = \mu + \phi_{1}Y_{t-1} + \phi_{2}Y_{t-2} + \ldots + \phi_{p}Y_{t-p} - \theta_{1}\varepsilon_{t-1} - \theta_{2}\varepsilon_{t-2} - \ldots - \theta_{q}\varepsilon_{t-q} + \varepsilon_{t} - \cdots - (5)$ where y_t is the degree of the differencing, ε_t is independently and normally distributed residual with zero mean and constant variance for t = 1, 2, 3, ..., n.

2.2 The Augmented Dickey - Fuller Test

The augmented Dickey-Fuller (ADF) test is most widely used test for checking Stationarity of a series. If d equals 0, the model becomes ARMA, which is linear stationary model. ARIMA (i.e. d > 0) is a linear nonstationary model. If the underlying time series is non-stationary, taking the difference of the series with itself predecessor to determine d makes it stationary, and then ARMA is applied onto the differenced series. A stationary process has a constant mean and variance over the time period. There are various methods available to make a time series stationary. Normally differencing techniques are used to transform a time series from a nonstationary to stationary by subtracting each datum in the series from its predecessor.

2.3 **Model Identification Criteria**

At the identification stage different ARIMA are formulated and tested on the data then their respective Akaike Information Criterion, Schwarz-Bayesian Information Criteria (BIC) and Hannan-Quinn Criteria (HQC) were considered and recorded. In each case, the model with the least AIC, BIC and HQC values were selected and subjected to diagnostic check to ensure that they fit well with the data.

$$AIC = (-2logL + 2m) ------(6)$$
where $m = p + q + P + Q$
and L is the likelihood function
$$Also, -2logL = n (1 + log 2\pi) + n log \sigma^{2} -----(7)$$
where σ^{2} is the mean square error, this implies that;
$$AIC = \{n (1 + log 2\pi) + n log \sigma^{2}) + 2m \} ------(8)$$

$$BIC = log \sigma^{2} + \{(m log n) / n \} ------(9)$$

3. **Results and Discussion**

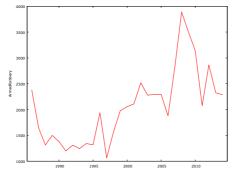


Fig 1: Time Series plot for Armed Robbery data Series

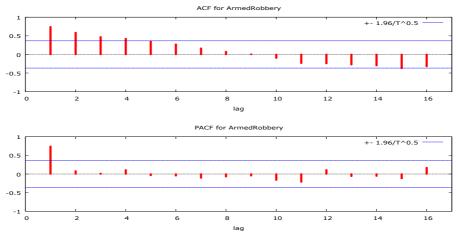


Fig 2: Correlogram for Armed Robbery data series

Table 1: Augmented Dickey-Fuller test for Stationarity of Armed Robbery data series

d	t-statistic	p-value	A-value
0	-0.6480	0.4274	0.05
1	-6.308	3.24e-07	0.05

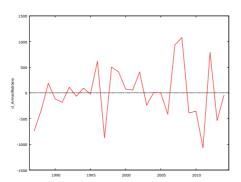


Fig 3: Time Series plot for differenced Armed Robbery data series

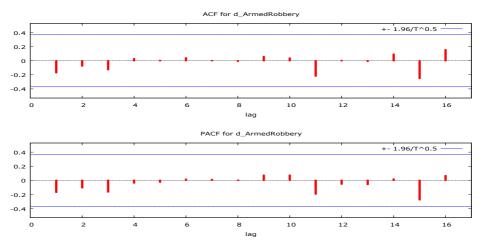


Fig 4: Correlogram for differenced Armed Robbery(d_Armed Robbery) data series

ARIMA	AIC	BIC	HOC
Model	AIC	DIC	HQC
000	464.2668	465.6341	464.6950
001	454.4714	458.5733	455.7561
100	445.1840	449.2858	446.4686
101	447.0676	452.5368	448.7805
111	432.2431	437.5719	433.8722
002	450.3334	455.8026	452.0463
201	449.0168	455.8533	451.1579

Table 2. Identification	of Best ARIMA model for Armed Robbery	
I ubic 2. Iuchiljiculion	of Desi ARIMA model for Armed Robbery	

Akaike information criteria (AIC), Bayesian information criteria(BIC), and Hannan-Quinn criteria (HQC).

3.1 Diagnostic check on the best model for Armed Robbery / Model Verification

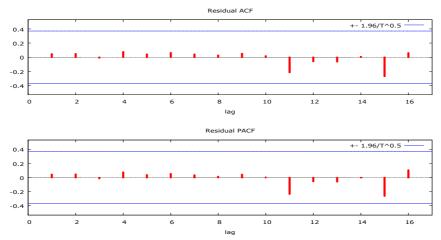
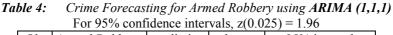


Fig 5: Correlogram of residuals for Armed Robbery

Table 3: Residual autocorrelation function ***, **, * indicate significance at the 1%, 5%, 10% levels using standard error 1/T^0.5

LAG	ACF	PACF	Q-stat.	[p-value]
1	0.0453	0.0453		
2	0.0477	0.0458		
3	-0.0097	-0.0139	0.1406	[0.708]
4	0.0761	0.0753	0.3435	[0.842]
5	0.0420	0.0368	0.4079	[0.939]
6	0.0630	0.0532	0.5596	[0.967]
7	0.0411	0.0354	0.6273	[0.987]
8	0.0268	0.0146	0.6574	[0.995]
9	0.0516	0.0437	0.7753	[0.998]
10	0.0166	0.0029	0.7882	[0.999]
11	-0.2167	-0.2342	3.1074	[0.960]
12	-0.0609	-0.0560	3.3020	[0.973]
13	-0.0649	-0.0590	3.5376	[0.982]
14	0.0090	-0.0001	3.5425	[0.990]
15	-0.2696	-0.2620	8.2399	[0.828]
16	0.0600	0.1049	8.4919	[0.862]

For 95% confidence intervals, $z(0.025) = 1.96$							
Obs	Armed Robbery	prediction	std. error	95% interval			
2016	undefined	2623.85	529.198	(1586.64, 3661.05)			
2017	undefined	2722.48	553.252	(1638.13, 3806.84)			
2018	undefined	2798.19	561.541	(1697.59, 3898.79)			



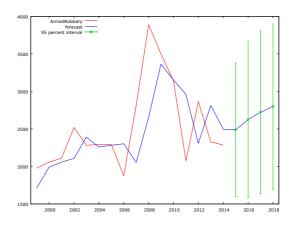


Fig 6: Correlogram of residuals for Armed Robbery

The time series plot and correllogram of the Armed Robbery data series on *fig* 1 and *fig* 2 respectively shows a strong evidence of a non stationary series, the non Stationarity was also confirmed with the help of the augmented dickey-fuller (ADF) test on *table* 1, which tests the null hypothesis that armed robbery data series follows a unit root process which was accepted at 5% alpha level. Thus, implying that the armed robbery data series is non stationary. By using first order differencing transformation, we obtained a t-statistic lesser than what was obtained at d = 0, and a p-value lesser than 5% alpha level. Thus, we select the condition that d = 1 and transform the data using first order differencing to make it stationary as seen on *fig* 3. The Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots of the differenced series is shown in *fig* 4. Seven tentative models were entertained, and the model with the minimum AIC, BIC and HQC, **ARIMA (1,1,1)** defined by

 $Y_t = 41.9206 + 0.5956Y_{t-1} - Y_{t-1} + e_t$ was chosen as the best model. To verify that the chosen ARIMA (1,1,1) is an appropriate model for armed robbery, a diagnostic check is done using residual ACF/PACF plot at different lags and testing the significance of the correlations up to 16 lags by Q statistic and respective p-values. Fig 5 above clearly shows evidence of random walk as the values are within the bounds and undulate about zero. Hence we uphold the first order differencing. Clearly, non of the correlations is significantly different from zero at a reasonable level. The ACF and PACF of the residuals also indicate good fit of the model and the respective p-values on *table 3* are greater than the α -value which is a desirable result. This proves that the selected ARIMA (1,1,1) is appropriate for modeling Armed Robbery in Nigeria. Finally, from the forecast on *table 4*, we see armed robbery in Nigeria would increase in 2016, and also in subsequent years.

4. Conclusion

ARIMA (1,1,1) has been successfully used to forecast Armed Robbery in Nigeria using a twenty-nine year data series. Armed robbery was found to be on the increase in the forecasted period. Hence, government in various states of the federation should put up structures to positively keep the citizens busy especially our youths. The Nigerian government should review its anti-crime strategies and sanitize the Nigeria Police Force and other security agencies. Government and its agencies should partner with the private sector and encourage them to build industries that will help to gainfully employ her citizens. Parents and school administrators should take moral instruction seriously for their children so as to minimize criminal tendencies.

5. References

-Agyemang, B. (2012). Autoregressive Integrated Moving Average (ARIMA) Intervention Analysis Model for the Major Crimes in Ghana (the Case of Eastern Region), M. Sc Thesis for Masters of Philosophy in Applied Mathematics, Nkrumah University of Science and Technology Ghana.

-Anand Kumar Shrivastav and Ekata A.K.S, 2012. "Applicability of Box Jenkins ARIMA model in crime forecasting: A case study of counterfeiting in Gujarat State." Int. J. Advance Research in Computer Engineering and Technology, 1(4), 494-497.

-Box G E, Jenkinks G M; "Time series analysis: Forecasting and control"; Holden Day, San Francisco.

-Bowerman, B. L., Connell, R. T., and Koehler, A, b; "Forecasting Time Series and Regression: An Applied Approach", Thomson, Belmont, CA 2005.

-Brockwell, P. J. and Davis, R.A. 2012. "Introduction to time series and forecasting second edition." Springer - Verlag New York, Inc.

-Carvell, I.G. and Swinfen, G.E, 1970. Criminal law and procedure. London: Sweet and Maxwell.

-Chamlin, M.B 1998 "Crime arrest: an autoregressive integrated moving average (ARIMA) approach" Journal of quantitative criminology, 4, 3, 247 - 258.

-Chen, P., H. Yuan and X. Shu, 2008. "Forecasting crime using ARIMA model. Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery." Vol. 5, October 18-20, 2008, Jinan, Shandong, Chine, pp: 627-630.

-Danbazau, A.B, 2007. Criminology Justice. 2nd edition. Ibadan: Spectrum Books Limited.

-Feldman, M.P, 1997. The psychology of crime. Cambridge: University Press.

-Gomez, V. and Maravall, A, 1998 "Automatic Modeling Methods for Univariate Series, Banco de Espana Working Paper No. 9808.

-Gorr, W., A. Olligschlaeger and Y. Thompson, 2003. "Short term forecasting of crime." Int. J. Criminol. Sociol. Theory, 4: 648-656.

-Khashei, M., M. Bijari and G.A.R. Ardali, 2009. "Improvement of auto-regressive integrated moving average models using fuzzy logic and Artificial Neutral Networks (ANNs). Neurocomputing" 72: 956-967.

-Tappan, P. W, 1947 Who is the criminal? American Sociological Review, 12, 10, 96 - 102.

68.7420 [0.000]

77.7086 [0.000] 84.8826 [0.000]

-Usman, U. Yakubu, M. and Bello, A. Z, 2012 "An investigation on the rate of crime in Sokoto state using principle component analysis." Nigerian Journal of Basic and applied Science, 20, 2, 152 – 160.

Appendix

1.

Autocorrelation function for Armed Robbery

14 -0.3000

16 -0.3221 *

15 -0.3737 ** -0.1196

***, **, * indicate significance at the 1%, 5%, 10% levels using standard error 1/T^0.5

LAG ACF PACF Q-stat. [p-value] 1 0.7445 *** 0.7445 *** 17.7962 [0.000] 29.4015 [0.000] 2 0.5904 *** 0.0810 3 0.4725 ** 0.0188 37.1222 [0.000] 4 0.4267 ** 43.6700 [0.000] 0.1128 5 0.3521 * -0.0378 48.3126 [0.000] 6 0.2737 -0.0445 51.2410 [0.000] 52.3768 [0.000] 7 0.1667 -0.1084 8 0.0787 -0.0714 52.6417 [0.000] 9 0.0086 -0.0473 52.6450 [0.000] 10 -0.0973 -0.1634 53.0926 [0.000] 11 -0.2368 55.8934 [0.000] -0.2127 59.1132 [0.000] 12 -0.2468 0.1139 13 -0.2745 -0.0578 63.3462 [0.000]

-0.0575

0.1744

2. Autocorrelation function for d_Armed Robbery ***, **, * indicate significance at the 1%, 5%, 10% levels using standard error 1/T^0.5

LAG	ACF	PACF	Q-stat. [p-value]
1 -0.1	697	-0.1697	0.8964 [0.344]
2 -0.0	741	-0.1060	1.0739 [0.585]
3 -0.1	266	-0.1651	1.6125 [0.657]
4 0.0	288 -	-0.0378	1.6416 [0.801]
5 0.0	052 ·	-0.0278	1.6426 [0.896]
6 0.0	406	0.0167	1.7055 [0.945]
7 0.0	013	0.0124	1.7055 [0.974]
8 -0.0	093	-0.0004	1.7092 [0.989]
9 0.0	592	0.0736	1.8640 [0.993]
10 0.0)390	0.0730	1.9351 [0.997]
11 -0.2	2181	-0.1959	4.2847 [0.961]
12 0.0	0048	-0.0555	4.2859 [0.978]
13 -0.0)098	-0.0598	4.2913 [0.988]
14 0.0)944	0.0183	4.8258 [0.988]
15 -0.2	2508	-0.2775	8.8887 [0.883]
16 0.1	568	0.0662	10.6105 [0.833]

Augmented Dickey-Fuller test for Armed Robbery including 0 lags of (1-L)Armed Robbery (max was 1, criterion t-statistic) sample size 28 unit-root null hypothesis: a = 1

test without constant model: (1-L)y = (a-1)*y(-1) + eestimated value of (a - 1): -0.0288058 test statistic: tau_nc(1) = -0.647953 p-value 0.4274 1st-order autocorrelation coeff. for e: -0.147

Dickey-Fuller regression OLS, using observations 1987-2014 (T = 28) Dependent variable: d_Armed Robbery

coefficient std. error t-ratio p-value Armed Robbery_1 -0.0288058 0.0444566 -0.6480 0.4274

AIC: 429.441 BIC: 430.773 HQC: 429.849

4.

Augmented Dickey-Fuller test for d Armed Robbery including 0 lags of (1-L)d Armed Robbery (max was 1, criterion t-statistic) sample size 27 unit-root null hypothesis: a = 1test without constant model: (1-L)y = (a-1)*y(-1) + eestimated value of (a - 1): -1.17011 test statistic: tau nc(1) = -6.30817p-value 3.239e-007 1st-order autocorrelation coeff. for e: -0.073 Dickey-Fuller regression OLS, using observations 1988-2014 (T = 27) Dependent variable: d d Armed Robbery

> coefficient std. error t-ratio p-value _____ d_Armed Robbery_1 -1.17011 0.185491 -6.308 3.24e-07 ***

AIC: 412.559 BIC: 413.855 HQC: 412.945

5. Model 000: ARMA, using observations 1986-2014 (T = 29) Dependent variable: Armed Robbery

	Coefficient	Std. Erro	or z	p-value	
const	2051.17	132.301	15.5039	< 0.0001	***
Mean dependent var	2051.17	72	S.D. depende	nt var	712.4608
Mean of innovations	-4.70e-1	14	S.D. of innov	ations	712.4608
Log-likelihood	-231.13	34	Akaike criter	ion	464.2668
Schwarz criterion	465.634	41	Hannan-Quin	n	464.6950

6. Model 001: ARMA, using observations 1986-2014 (T = 29) Estimated using Kalman filter (exact ML) Dependent variable: Armed Robbery Standard errors based on Hessian

> coefficient std. error z p-value

2066.25 157.303 13.14 2.06e-039 *** const theta 1 0.563140 0.121732 4.626 3.73e-06 ***

Mean dependent var 2051.172 S.D. dependent var 712.4608 Mean of innovations -4.491555 S.D. of innovations 548.2605 Log-likelihood –224.2357 Akaike criterion 454.4714 Schwarz criterion 458.5733 Hannan-Quinn 455.7561

Real Imaginary Modulus Frequency

MA				
Root 1	-1.7758	0.0000	1.7758	0.5000

www.iiste.org

 Model 100: ARMA, using observations 1986-2014 (T = 29) Estimated using Kalman filter (exact ML) Dependent variable: Armed Robbery Standard errors based on Hessian

	coefficient s	std. error	Z	p-va	lue	
					9.35e-013 ***	
phi_1	0.72958	5 0.1177	50 6	6.196	5.79e-010 ***	

Mean dependent var 2051.172 S.D. dependent var 712.4608 Mean of innovations -7.184895 S.D. of innovations 464.0985 Log-likelihood -219.5920 Akaike criterion 445.1840 Schwarz criterion 449.2858 Hannan-Quinn 446.4686

Real ImaginaryModulus FrequencyAR
Root 11.37060.00001.37060.0000

 Model 101: ARMA, using observations 1986-2014 (T = 29) Estimated using Kalman filter (exact ML) Dependent variable: Armed Robbery Standard errors based on Hessian

	coefficient	std. error	Z	p-value	
phi_1	0.76925	306.907 52 0.1547	31	4.972 6.	64e-07 ***
theta_	1 -0.0922	2184 0.274	4034	-0.3365	0.7365

Mean dependent var 2051.172 S.D. dependent var 712.4608 Mean of innovations -6.566929 S.D. of innovations 463.1898 Log-likelihood -219.5338 Akaike criterion 447.0676 Schwarz criterion 452.5368 Hannan-Quinn 448.7805

	Real Imag	ginary M	lodulus Fr	requency
AR				
Root 1	1.3000	0.0000	1.3000	0.0000
MA				
Root 1	10.8438	0.0000	10.8438	0.0000

Real Imaginary Modulus Frequency

www.iiste.org

9. Model 002: ARMA, using observations 1986-2014 (T = 29) Estimated using Kalman filter (exact ML) Dependent variable: Armed Robbery Standard errors based on Hessian

С	oefficient	std. error	z p-va	lue
theta_1	0.7168	05 0.1759	46 4.074	8.67e-028 *** 4.62e-05 *** 0.0045 ***

Mean dependent var 2051.172 S.D. dependent var 712.4608 Mean of innovations -8.448745 S.D. of innovations 490.7652 Log-likelihood -221.1667 Akaike criterion 450.3334 Schwarz criterion 455.8026 Hannan-Quinn 452.0463

	Real Imaginary		Modulus Frequence	
MA				
Root 1	-0.8715	-1.2931	1.5594	-0.3444
Root 2	-0.8715	1.2931	1.5594	0.3444

 Model 201: ARMA, using observations 1986-2014 (T = 29) Estimated using Kalman filter (exact ML) Dependent variable: Armed Robbery Standard errors based on Hessian

	coefficient	std. error	z p-v	value	
const	2098.11	316.687	6.625	3.47e-011 ***	
phi_1	1.13329	0.97320	6 1.164	0.2442	
phi_2	-0.2638	84 0.7331	92 -0.3	599 0.7189	
theta_	1 -0.4614	41 0.9399	942 -0.4	909 0.6235	

Mean dependent var 2051.172 S.D. dependent var 712.4608 Mean of innovations -4.985078 S.D. of innovations 462.7944 Log-likelihood -219.5084 Akaike criterion 449.0168 Schwarz criterion 455.8533 Hannan-Quinn 451.1579

	Real Imag	jinary M	odulus Fr	equency
AR				
Root 1	1.2410	0.0000	1.2410	0.0000
Root 2	3.0537	0.0000	3.0537	0.0000
MA				
Root 1	2.1671	0.0000	2.1671	0.0000

 Model 111: ARIMA, using observations 1987-2014 (T = 28) Estimated using Kalman filter (exact ML) Dependent variable: (1-L) Armed Robbery Standard errors based on Hessian

coeffi	cient std.	error z	p-value	
phi_1 0.	595752		2.989 0.0	

Mean dependent var -3.428571 S.D. dependent var 512.7648 Mean of innovations -64.06475 S.D. of innovations 454.6336 Log-likelihood -212.1216 Akaike criterion 432.2431 Schwarz criterion 437.5719 Hannan-Quinn 433.8722

Real Imaginary Modulus Frequency

AR				
Root 1	1.6785	0.0000	1.6785	0.0000
MA				
Root 1	1.0000	0.0000	1.0000	0.0000