Center for Child and Family Studies
UNIVERSITY OF SOUTH CAROLINA
College of Social Work

DSS SOUTH CAROLINA DEPARTMENT of SOCIAL SERVICES

APPLYING TIME SERIES MODELING TO FORECAST MONTHLY REPORTS OF ABUSE, NEGLECT AND/OR EXPLOITATION INVOLVING AN ADULT IN THE STATE OF SOUTH CAROLINA

AUTHORS: NELÍS SOTO-RAMÍREZ, CYNTHIA FLYNN, DIANA TESTER

BACKGROUND

Time series analyses have traditionally been used for forecasting techniques in public health and economics. For instance, forecasting techniques are used to predict the incidence of infectious diseases, by using artificial neural network models and auto-regressive integrated moving average (ARIMA) models. One of the advantages of using ARIMA models is that it takes into consideration the periodic variations, changing trends and random disturbances of time series, and use the associations in the sequentially lagged relationships to predict future state. Forecasting techniques used in epidemiological studies could be also applied in social welfare research. Application of time series modeling to predict reports related to maltreatment of vulnerable adults (APS intakes) can be useful for efficient early planning and resource allocation to handle a high volume of investigations.

For the purpose of this study, APS intake data from the South Carolina (SC) Child and Adult Protective Services System (CAPSS) was used to apply ARIMA time series analysis to fit and forecast monthly APS reports for a subset of the population (21 counties). In 2015, all APS intakes for 21 of 46 counties in SC became centralized into intake hubs. As a result, an increase in all APS intakes was observed after 2015. Starting in May 2017, the remainder of the counties were rolled into the hubs with the last county in November 2017. As a consequence, a second wave or spike was detected and it is expected to continue in the first few months of 2018. Hence, for the time series analysis we only included the 21 counties that became centralized into intake hubs in 2015 to predict the expected monthly APS reports from January 2018 to June 2018. The APS data from January 2014 and December 2017 were subjected to ARIMA modeling adjusting for the exogenous variable intake hub implementation. In addition, the impact of the implementation of the intake hubs in 2015 was assessed.

Goal of the study

- To apply ARIMA time series analysis to fit and forecast monthly reports of abuse, neglect and/or exploitation involving an adult (APS intakes) accepted for assessment reported to the SC Department of Social Services.

CONCLUSIONS

- ARIMA time series models are a valuable tool to allow forecasting of future reports of maltreatment of vulnerable adults with high accuracy and predict increasing intakes into the future.
- Policymakers and program administrators at both the state and federal levels need effective forecasts of future intake reports which can then be used to allow appropriate planning and resource allocation to handle a high volume of monthly intake reports.
- More research on the accurate prediction of the future intake reports of maltreatment should be conducted and compared with other forecasting techniques.

METHODS

- Data source: Monthly APS intakes accepted for assessment was obtained from the SC CAPSS database reported between January 2014 and December 2017. Only 21 of 46 counties were considered for the analysis.
- Statistical Methods: Box–Jenkins (1970) approach was used to fit the best ARIMA model for the aggregated monthly number of APS intakes accepted for assessment. Statistical forecasting using the fitted model was used to predict future APS intakes.
  - ✓ The ARIMA (p, d, q) model consists of three terms, where p is the order of autoregression (AR), d is the order of regular differencing (integration), and q is the order of moving average (MA). In terms of y, the general forecasting equation is:

$$\hat{y}_t = \underbrace{\mu}_{\text{Constant}} + \underbrace{\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}}_{\text{AR terms (lagged values of } y)} - \underbrace{\theta_1 e_{t-1} \cdots - \theta_q e_{t-q}}_{\text{MA terms (lagged errors)}}$$

  - ✓ The time series model consist of three iterative steps: identification, estimation, and diagnostic checking.
    i. In the identification stage, the IDENTIFY statement was used to describe the serial correlation of the data and any relations to external factors. Lagged scatter-plots, autocorrelation function (ACF), partial autocorrelation function (PACF) plots, and augmented Dickey-Fuller (ADF) test were used to identify whether or not the time series was stationary.
    ii. In the estimation and diagnosis checking stage, the ESTIMATE statement was used to estimate the parameters of the model selected and to test for significance of the parameter estimates and goodness of fit. To compare different ARIMA models the following measures of overall fit were evaluated: (1) coefficient of determination (R2), (2) Akaike Information Criterion (AIC), (3) Schwartz Bayesian Criterion (SBC), and (4) Mean Square Error (MSE).
    iii. Once the appropriate ARIMA model was fitted, the goodness of fit was examined by means of Ljung-Box Q-test and by plotting the ACF of the residuals of the fitted model. If model is correctly specified, residuals should be uncorrelated (white-noise) and Q should be small. A non-significant value indicates that the chosen model fits well. The maximum likelihood (ML) was used to estimate the parameter estimates.
- The exogenous variable, Intake Hub implementation in 2015, was added to the ARIMA model to evaluate the impact of the Hub implementation.
- In the forecasting stage, the FORECAST statement was used to predict subsequent observations and their corresponding confidence intervals (95% CI).
- The most parsimonious model with the highest accuracy was applied to predict the expected monthly APS intakes from January 2018 to June 2018. The counts of APS intakes were transformed using a quartic root transformation to stabilize the variance. To obtain the forecast values in the original scales, the reverse transformation was calculated.
- SAS Software Version 9.4 (North Carolina State University, Raleigh, NC, USA) was used to develop the ARIMA model.

RESULTS

- In 2015, all APS intakes for 21 of 46 counties became centralized into intake hubs. As a result, an increase of 67.2% in monthly APS intakes was observed in 2015, with a yearly prevalence rate of 8.79 clients per 10,000 in 2014 and 13.24 clients per 10,000 in 2017 (Table & Figure 1).
- Figure 2 shows the trend of monthly APS intakes accepted for assessment between January 2014 and December 2017. Orange bar indicates when the intake Hubs were implemented.
- The monthly APS intakes distribution for the period 2014 to 2017 is shown in Figure 3. The highest average number of APS intake reports was registered from June to August (Median: Jun - 331, Jul – 327, Aug – 341) followed by March (Median: 303).

Identification Stage

- ARIMA models were first-differenced since the Augmented Dickey-Fuller unit root tests showed that the original series were nonstationary (ADF p-value ≥ 0.05). In addition, the counts of APS intakes were transformed using a quartic root transformation to stabilize the variance.

Estimation Stage

- Different combinations of AR and MA orders were tested after evaluating the ACFs and PACFs of the stationary series (Table 2). The ACF function decays for the first lag, then it drop off to zero abruptly. Therefore, a MA(1) was considered. An AR(1) and AR(12) were considered based on the PACF plot.
- Of all the models tested, an ARIMA (12), 1, 1 model (p = (12), d = 1, q = 1) was found to work best for overall APS intakes after evaluating all overall fit measures (AIC = -60.87, SBC = -53.47, R2 = 0.83, MAE = 0.013). The Ljung-Box chi-square statistics and the autocorrelation function of the residuals indicate that the residuals are independent (P Box-Ljung (Q) = 0.69).
- The constructed ARIMA model includes a positive AR component (coefficient = 0.575, $p < 0.0001$) indicating that current month's APS intake reports depends on the APS intakes received twelve months ago (average value plus some fraction of its deviation from this average value a year ago, plus a random error). In addition, a positive MA component with a lag of one (coefficient = 0.573, $p < 0.0001$) indicates that each value of the variable is determined by the current disturbance and the previous disturbance (Table 3).

Impact of the Hub Implementation

- The ARIMA model improves after adding an exogenous dummy variable that takes the value '0' before January 2015 and '1' after that. The impact of the hub implementation variable is significant (coefficient = 0.267, p = 0.003), showing an increase on monthly APS reports after January 2015 (Table 3).

Forecasting Stage

- Table 4 shows the monthly forecast of APS intakes according to the model in 2018 with 95%CI for the 21 counties.
- The optimum ARIMA model predicted an average monthly APS intake of 463 between January and June 2018, representing an 18% increase from 2017 (median=392).
- In Figure 4 we display the actual number of APS intakes and the prediction from the ARIMA model with the corresponding 95%CIs.
- Major peaks can be observed during June to August, and again a light peak for March, adequately capturing the pattern in the data (Figure 4).

TABLES & FIGURES

Table 1. Distribution of APS intakes accepted for assessment with Intake Hubs implemented in 2015 (21 counties)

Calendar Year	Number of Intakes	Prevalence (per 10,000 population)	Median	25th Percentile	75th Percentile	Quartile Range
2014	2275	8.790	181.5	173.0	210.0	37.0
2015	3470	13.242	303.5	243.5	325.5	82.0
2016	3474	13.087	295.0	265.5	304.0	38.5
2017	4595	17.310	392.0	340.0	415.5	75.5

Table 2. ARIMA models and selection criteria

ARIMA model	AIC	SBC	R^2	MSE	Ljung-Box p-value
p=(1)(12) q=0 d=1	-60.16	-52.76	0.83	0.013	0.29
p=(1)(12) q=1 d=1	-60.63	-51.38	0.83	0.013	0.78
p=(1,12) q = 0 d=1	-54.57	-47.13	0.80	0.015	0.18
p = (12) q = 1 d=1	***-60.87***	***-53.47***	***0.83***	***0.013***	***0.69***

Table 3. Parameters of the selected ARIMA model for APS intakes

Parameter	Estimate	Standard Error	t Value	p-value	Lag
μ	0.014	0.013	1.12	0.2616	0
MA1,1	0.573	0.121	4.70	<.0001	1
AR1,1	0.575	0.136	4.23	<.0001	12
Hub	0.267	0.092	2.88	0.0039	0

Table 4. Forecasted monthly APS intakes accepted for assessment (95%CIs) for 21 counties, 2018

2018	Predicted APS intakes	Lower 95% CI	Higher 95% CI
January	431.30	344.49	526.86
February	416.90	322.58	522.13
March	487.58	371.30	617.73
April	439.98	324.41	571.24
May	509.58	371.57	666.45
June	507.95	362.45	674.68
Median	463.78	416.90	509.58

Figure 1. Box-plot distribution of APS intakes accepted for assessment by calendar year (2014-2017)

Legend: Line within the box represents median values, border lines represent the first and the third quartile

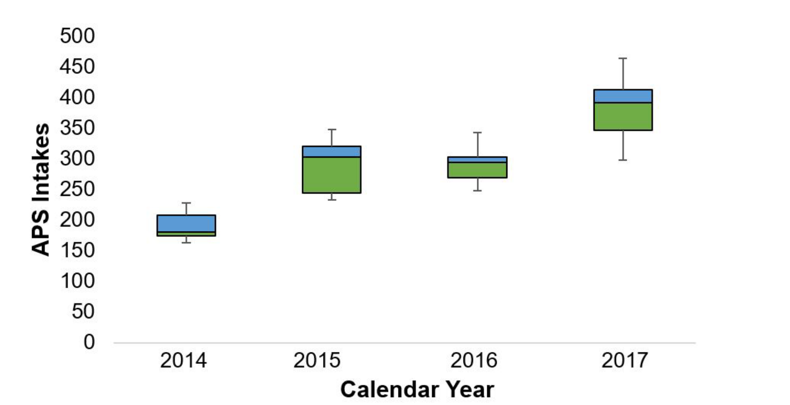


Figure 2. Trend of monthly APS intakes accepted for assessment between January 2014 and December 2017 (21 counties)

*Orange bar indicates when the Hubs were implemented

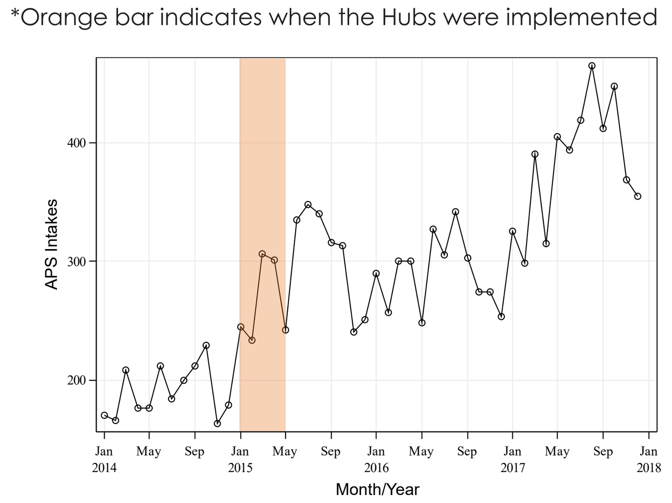


Figure 3. Box-plot distribution of monthly APS intakes accepted for assessment

Legend: Line within the box represents median values, border lines represent the first and the third quartile

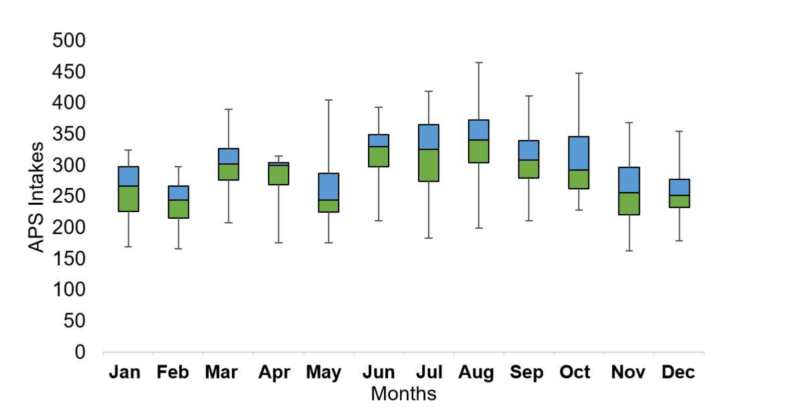
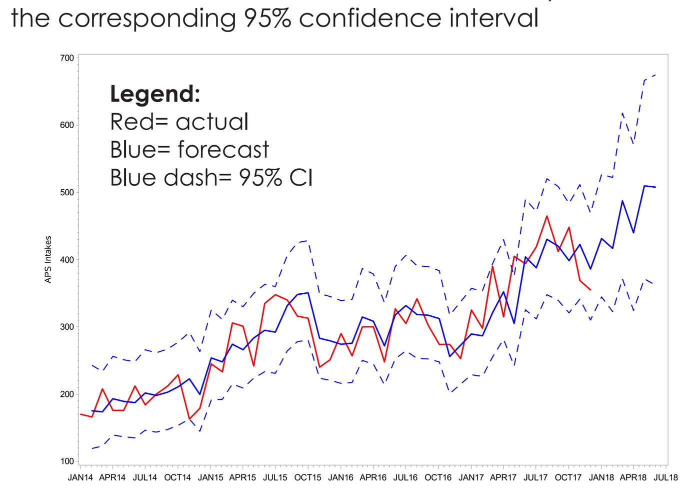


Figure 4. Monthly time series 2014-2017 for observed and fitted APS intakes, and forecast APS intakes for January 2018 – June 2018 with the corresponding 95% confidence interval



REFERENCES

- Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J. Am. Stat. Assoc. 65, 1509–1526
- Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control, 2nd ed. Holden Day, San Francisco
- The ARIMA Procedure SAS Institute Inc. 2014. SAS/ETS® 13.2 User's Guide. Cary, NC: SAS Institute Inc.
- Ljung, G.M., Box, G.E., 1978. On a measure of lack of fit in time series models. Biometrika. http://dx.doi.org/10.1093/biomet/65.2.297
- Soyiri, I.N., Reidpath, D.D., 2013. An overview of health forecasting. Environ Health Prev Med. 18:1-9

ACKNOWLEDGEMENTS

- We gratefully acknowledge the Instructional Design Team of the Center for Child and Family Studies (CCFS) for creating this poster. We acknowledge the CCFS Research and Evaluation Team for their insightful comments.

Visit our website at: http://ccfs.sc.edu/