Forecasting cesarean deliveries with robust time series models in a tertiary care hospital

Shital Bhandary[1], Binita Pradhan[2]
[1]Associate Professor, School of Public Health, [2]Associate Professor, Department of Obstetrics and Gynecology
Patan Academy of Health Sciences, Lagankhel, Nepal

ABSTRACT

Introduction: Caesarian section rate was 47% at Patan Hospital in 2014 despite the recommendation of World Health Organization for keeping it below 15%. This has become a public health problem and now debated as the human right violation of childbearing women. This study aims to use robust time series model to get valid forecasts of caesarian deliveries for this hospital.

Method: Univariate time series models were used to forecast 3-year caesarean deliveries using 60-month (2010-2014) data of Patan Hospital. A time series model with low mean average percentage error from validation period and without autocorrelation problem was selected as robust model and it was used to forecast the caesarean deliveries for 2015-2017 periods.

Result: Winter's additive model had lowest validation forecasting error and showed decreasing trend of caesarian deliveries but it showed autocorrelation problem. Quadratic regression gave similar results but is also had problem of autocorrelation. Artificial Neural Network – Multilayer Perceptron model gave close forecasts but autocorrelation was not assessed. Best Autoregressive Integrated Moving Average model with (0,1,1),(0,0,0) parameters did not show autocorrelation for 60-month data and taken as robust model for doing 3-year forecasts.

Conclusion: Best ARIMA model with one difference stationary and first order of moving average correction gave valid forecasts for Patan Hospital. Advanced univariate and multivariate time series models with large samples can be used to get more precise forecasts of caesarean deliveries in Nepal.

Keywords: ARIMA, caesarean delivery, forecasts, time series, Nepal

CORRESPONDENCE

Mr. Shital Bhandary
Associate Professor, School of Public Health
Patan Academy of Health Sciences, Lagankhel, Nepal
Email: shitalbhandary@pahs.edu.np

Journal of General Practice and Emergency Medicine of Nepal. Issue 8: 2019. Available at: www.jgpeman.com, eISSN: 2363-1168

3

INTRODUCTION

Caesarean section has been increasing in the developing world in alarming rate of 25% and more[1] in spite of recommendation from World Health Organization (WHO) to make it less than 15%.[2] It is well known that caesarean section rate above 15% does more harm than good.[3,4] Progressive increase in caesarean delivery around the globe is a great concern as it increases the maternal complications[4].

The national caesarean rate increased from 4.5% in 1965 to 32.8% in 2010/2011.[5,6] Caesarean delivery rate increased from 38.4% in 2010 to 46.9% in 2014 at Patan Hospital, Nepal.[7] This warrants forecasts of caesarean deliveries for next few years in order to assess plans and policies placed to reduce caesarean deliveries/caesarean section rate at local level in Nepal.

Thus, this study aims to use Classical Decomposition Models (CDM) to do validation forecasting using 2010-2013 data and forecast the monthly caesarean deliveries for 2014 at Patan Hospital, Lalitpur, Nepal. It also aims to compare best CDM estimates with more advanced univariate time series models to identify the robust model for doing three year (2015-17) forecasting of caesarean deliveries using it.

METHOD

Retrospective data of 60 months (2010-2014) was collected after approval from the IRC-PAHS. Additive and multiplicative CDMs were fitted on the 48-month caesarean section data of 2010-2013 (80% of total data) to fulfill the minimum sample of 30 observations needed to do reliable and valid predications using univariate time series models.[9]

CDM model decomposed the trend, seasonality, cyclical and random components from caesarean delivery time series data and they were used further to do the validation forecasting for 12-month of 2014 (20% of total data). This study followed 80-20 rule where 80% data (48-months) was used as estimation period and 20% data (12-months) was used as validation period before getting the forecasts for the next 36-months[8].

CDM-Additive model decomposed original data (Y_t) as additive combination of trend (T_t), seasonality (S_t), cyclical (C_t) and random (R_t) components. It was represented as follows: $Y_t = T_t + S_t + C_t + R_t$. Likewise, CDM-Multiplicative model decomposed the original data as: $Y_t = T_t \times S_t \times C_t \times R_t$. Then, 12-month validation forecasting was done for 2014 using the best CDM model.[9,10]

Trend for this validation forecasting was estimated using simple trend model ($y=a+bt$) where observed deliveries in each month were used as dependent variable (y) and time (1 for Jan 2010 and 48 for December 2013) was used as independent variable (t).[11] Estimated seasonal factors and trends were used to get validation forecasts for each month and total of 2014. Mean absolute percentage error (MAPE) followed by Mean Square Error (MSE) and Mean Absolute Deviation (MAD) errors were used to select the best CDM model.[9,11]

Validation forecasting for 12-month of 2014 was also done using the best Polynomial Regression model, best Exponential Smoothing model, best Autoregressive Integrated Moving Average (ARIMA) model and Artificial Neural Network - Multilayer Perception (ANN-MLP) model using the same 48-month data.[12]

Validation forecasting MAPE and Ljung-Box test were used to select the best (robust) model, that is, model with low MAPE and without autocorrelation. Finally, 3-year forecasting of caesarean deliveries were done using this robust model.

All the model fits were done in IBM SPSS 20.0 software and forecasting errors were computed using Microsoft Excel 2013 spreadsheet program. Ljung-Box test of autocorrelation was obtained from the SPSS software.

RESULT

The CDM-Additive model fitted the 48-month data with MAPE of 6.49%, MSE of 665.11 and MAD of 19.77 whereas CDM-Multiplicative model fitted the same data with MAPE of 6.49%, MSE of 712.44 and MAD of 19.61. Since MAPE were identical for both the CDM models, model with low MSE was used for doing validation forecasting for 12 months of 2014 in this study.

Seasonal factors revealed more caesarean deliveries in the months of March, April, May, October and November with highest in November followed by March and October based on 48-month data. Forecasted caesarean deliveries obtained by adding seasonal factors and estimated trend were also higher for these three months. Estimated trend, on the other hand, showed increasing caesarean deliveries for the 12 months of 2014. Error from CDM-Additive model was highest for the

month of December followed by April, October and May with difference of more than 50 deliveries (Table 1).

The MAD and MSE for validation forecasting of 12-month were 41.83 and 2835.68 respectively which were higher than the MAD of 19.44 and MSE of 665.11 obtained for the 48-months CDM-Additive model. The MAPE of CDM-Additive model for validation forecasting of 12-month was 13.58%, which was higher than the 6.49% MAPE error obtained for 48-month data using the same model.

Table 1: Validation forecasting and errors - CDM-Additive model fit

Year 2014	Observed values (1)	Seasonal factors (2)	Estimated Trend (3)	Forecast* (4=2+3)	Error (5=1-4)	ABS** of Error (6)	Squared error (7=6^2)	Percent error (8 =6/1*100)
January	357	-6.90	354.37	347.47	9.53	9.53	90.86	2.67%
February	360	-19.15	356.28	337.13	22.87	22.87	523.22	6.35%
March	364	17.68	358.19	375.87	-11.87	11.87	140.83	3.26%
April	298	4.88	360.09	364.97	-66.97	66.97	4484.94	22.47%
May	300	-6.18	362.00	355.82	-55.82	55.82	3116.10	18.61%
June	312	-31.21	363.91	332.70	-20.70	20.70	428.58	6.64%
July	319	-3.75	365.82	362.07	-43.07	43.07	1854.90	13.50%
August	357	-16.10	367.73	351.63	5.37	5.37	28.85	1.50%
September	354	17.63	369.63	387.26	-33.26	33.26	1106.19	9.40%
October	345	29.35	371.54	400.89	-55.89	55.89	3123.64	16.20%
November	312	-15.26	373.45	358.19	-46.19	46.19	2133.18	14.80%
December	274	29.02	375.36	404.37	-130.37	130.37	16996.89	47.58%
Average:					-35.53 (Bias)	41.83 (MAD)	2835.68 (MSE)	13.58% (MAPE)

Note: *Additive forecasting assumed cyclical and random components as zero. **Absolute value

Monthly validation forecasting values of caesarean deliveries were fluctuating for CDM-Additive, CDM-Multiplicative and Winter's additive models. On the other hand, quadratic model forecasts showed decreasing monthly trend from January to December 2014 whereas non-seasonal ARIMA with first difference model as well as ANN-MLP model showed increasing monthly trend of caesarean deliveries for the consecutive 12 months starting from January and ending in December of 2014. The total caesarean deliveries obtained from the validation forecasting for 2014 was lowest for the non-seasonal ARIMA with first difference stationary model and highest for CDM-Multiplicative model. Total forecasted deliveries for 2014 for the Quadratic (3949.40) and Winter's additive (3964.80) models were close to the total observed caesarean deliveries (3952) for 2014.

Table 2: Validation forecasting of caesarean deliveries and MAPE - six time series model fits

Year 2014	Observed values	CDM-Additive Model	CDM-Multiplicative Model	Polynomial regression (Best: Quadratic Model)	Exponential Smoothing (Best: Winter's Additive Model)	ARIMA [Best: (0,1,0), (0,0,0) Model]	ANN-MLP Model (80% training, 20% testing)
January	357	347.47	349.36	334.10	327.09	311.34	335.52
February	360	337.13	345.53	333.52	315.14	312.68	335.79
March	364	375.87	362.19	332.85	343.95	314.02	336.03
April	298	364.97	364.07	332.08	331.26	315.36	336.24
May	300	355.82	342.29	331.20	328.31	316.70	336.43
June	312	332.70	312.86	330.23	315.37	318.04	336.6
July	319	362.07	372.49	329.16	324.93	319.38	336.74
August	357	351.63	366.80	327.99	329.48	320.72	336.88
September	354	387.26	403.61	326.72	348.29	322.06	336.99
October	345	400.89	404.24	325.35	351.10	323.40	337.1
November	312	358.19	350.86	323.88	310.41	324.74	337.19
December	274	404.37	406.01	322.32	339.47	326.09	337.27
Total	3952	4378.36	4380.30	3949.40	3964.80	3824.55	4038.78
MAPE		13.58%	12.88%	8.01%	7.11%	8.47%	8.56%

The 12-month validation forecasting MAPE error for CDM-Additive and CDM-Multiplicative models were 13.58% and 12.88% respectively. MAPE error was 8.56%, 8.47%, 8.01% and 7.11% for ANN-MLP, ARIMA (0,1,0),(0,0,0) Quadratic regression and Winter's Additive models respectively (Table 2).

The R-square and MAPE of Winter's additive model were 0.594 and 6.18% whereas Ljung-Box test was

31.58 with p-value of 0.007 for 48-month data. Table 3 revealed monthly, quarterly and yearly forecasts for 2015-2017 using 60-month data and Winter's additive model. It reported caesarean deliveries as high as 4374, 4388 and 4750 and as low as 3009, 2238 and 1129 for 2015, 2016 and 2017 respectively. Similar results were also observed for all the quarters of 2015-2017

Table 3: Forecasts of caesarean deliveries - Winter's additive model fit

Year, Quarter	Number	Lower Confidence Level (LCL)	Upper Confidence Level (UCL)
2015, Total	**3691**	**3009**	**4374**
2015, Q1	961	804	1117
2015, Q2	902	740	1064
2015, Q3	933	760	1106
2015, Q4	896	705	1087
2016, Total	**3313**	**2238**	**4388**
2016, Q1	866	650	1082
2016, Q2	807	560	1055
2016, Q3	838	553	1123
2016, Q4	801	474	1129
2017, Total	**2936**	**1121**	**4750**
2017, Q1	772	398	1146
2017, Q2	713	288	1138
2017, Q3	744	265	1223
2017, Q4	707	170	1244
Total (2015-17)	**9940**	**6368**	**13512**

Note: Year and quarter totals may not add due to rounding effec

The R-square value for this best Winter's additive model using 60-month data was 0.550 with MAPE of 6.32%. The Ljung-Box test statistic for this model was 27.57 with p-value of 0.024. Forecasts of this model were lower than the observed caesarean deliveries in the past (Figure 1).
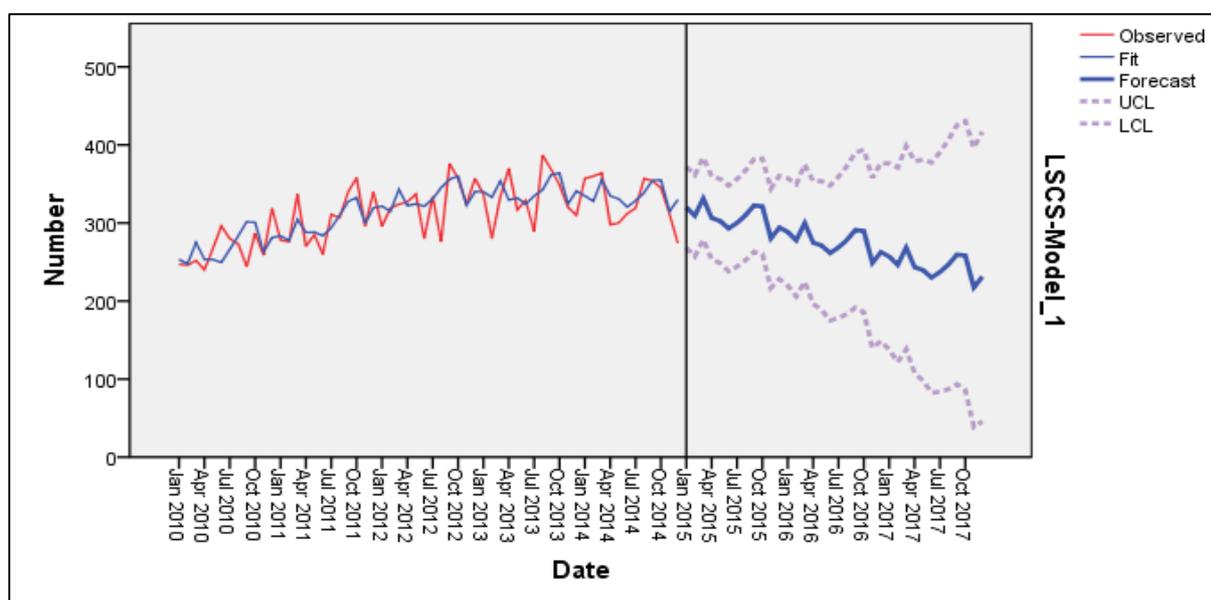

Figure 1: Observed, fitted and forecasted deliveries, 2010-2017-Winter's additive model

Monthly forecasts from quadratic linear regression fit, with next lowest MAPE of validation forecasting, were similar to the Winter's additive model. It reported R-square of 0.498 and standard error of estimate of 27.40 and also showed decreasing trend of caesarean deliveries for 2015-2017 period. Variance inflation factor of this model was 16.52.

As ARIMA model had next low validation forecasting MAPE after quadratic model, 60-month data was used to identify best ARIMA model. The ARIMA (0,1,1),(0,0,0) model was identified as the best ARIMA model for this data with R-square of 0.301 and MAPE of 8.01%. It also reported Ljung-Box test of 22.35 with p-value of 0.172 and

produced a constant forecasts of 320 deliveries for each month of 2015-2017 period.

Based on the next low validation forecasting MAPE after ARIMA, ANN-MLP model fitted with 80% training and 20% testing partition produced sum of square error of 13.07 and 3.320 for training and testing partition. It reported relative error of 0.54 and 0.58 for training and training partitions respectively. It also produced a constant forecast of 311 caesarean deliveries between January 2015 (61st month) and March 2016 (75th month) and, 312 deliveries between April 2016 (76th month) and December 2017 (96th month).
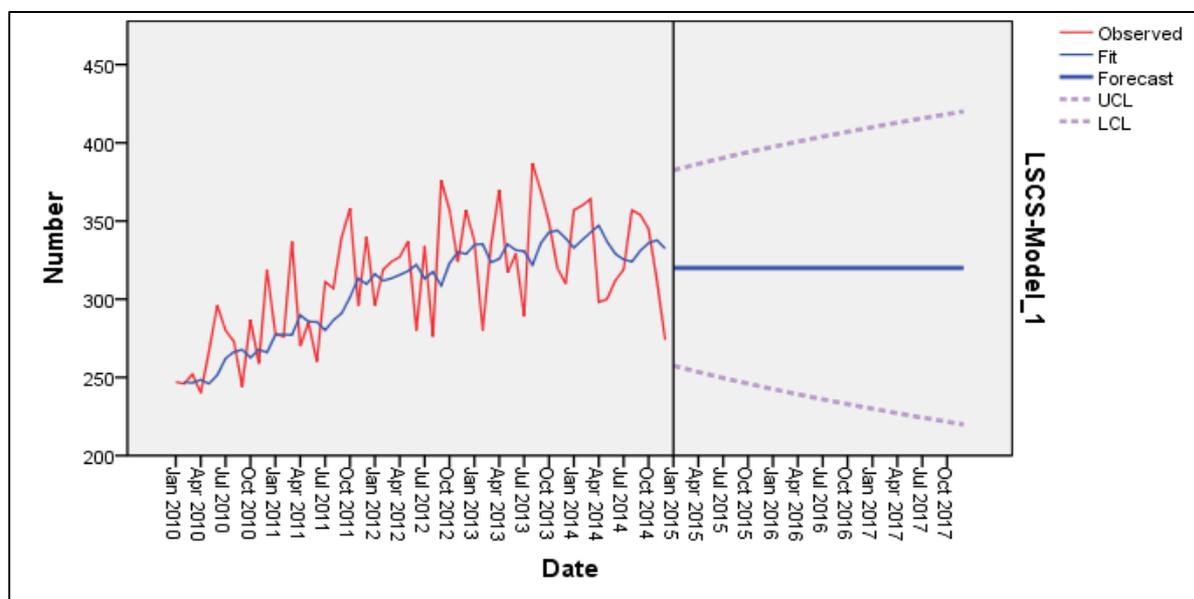

Figure 2: Observed, fitted and forecasted deliveries, 2010-2017-ARIMA (0, 1, 1),(0,0,0) model

DISCUSSION

Classical Decomposition models were found to have highest validation forecasting errors (>10%) in this study, which suggested that they were not useful for predicting caesarean deliveries at Patan Hospital. They are crude methods to do decomposition of time series and not used regularly to do precision forecasting.[9,10] Although quadratic and ARIMA models had low errors compared to CDM models, Winter's additive model had the lowest validation forecasting MAPE. So, it was used as the robust model for forecasting caesarean deliveries of 2015-2017.[11]

Winter's additive model showed decreasing trend of caesarean deliveries at Patan Hospital. Wider confidence intervals from this model, on the other hand, suggested that total caesarean deliveries

could be as high as 4400 in 2015 and 2016 and 4800 in 2017 approximately. This model reported R-square of 0.550, which means 60-month time explained nearly 55% of variation in caesarean deliveries with MAPE of 6.32% indicating good precision of the forecasts. The main limitation of this model fit was statistically significant Ljung-Box test (p-value<0.024) as it suggested problem of autocorrelation among the observed deliveries at various lags of 60-months.[12] This means forecasts from this model are not valid and must be interpreted with caution.

Quadratic linear model fit also showed decreasing trend of caesarean deliveries for 2015-17 but it had problem of multicollinearity (variance inflation factor > 10). This happened as time and time-square variables used in the model were correlated highly

indicating a serious problem of confounding effect.[13] Thus, results from this model are also not reliable as well as valid for this study.

ARIMA (0,1,1),(0,0,0) model was found to be the most promising robust model as it solved the problem of autocorrelation even though it had low R-square value and high MAPE error than the Winter's additive model. The first difference of this model made the data stationary to provide best forecasts. This was not applicable for the exponential smoothing models thus it was the main limitation of Winter's additive model too. This ARIMA model also corrected seasonality using first order of moving average.[14] Thus, forecasts of 320 deliveries for each month of 2015-2017 period was valid than the Winter's additive model. It was confirmed as total caesarean deliveries at Patan Hospital in 2017 was recorded as 3811, which was close to the ARIMA forecast of 3840 than Winter's additive model forecast of 2936 for 2017.

ANN-MLP model[15] also had low MSE values than all the six models and produced forecast of 311 and 312 deliveries for 61-75 and 76-96 months respectively giving total caesarean delivery for 2017 as 3744, which is again better than the Winter's additive model. This means ARIMA (0,1,1),(0,0,0) is the robust univariate forecasting model as it did not show autocorrelation and also had low model error (less than 10% MAPE). Yet, advanced univariate and multivariate time series models can be used to get the more precise forecasts of the caesarian deliveries in Nepal.[16]

Forecasting for 36 months of 2015-2017 from this robust ARIMA model could not be compared with the actual monthly caesarean deliveries occurred at Patan Hospital in 2015-17 periods, which is the main limitation of this study.

CONCLUSION

Winter's additive model fit, based on lowest validation forecasting MAPE, revealed a decreasing trend of caesarean delivery at Patan Hospital for 2015-2017. However, this model showed problem of autocorrelation making the forecasts from this model unreliable. Quadratic model forecasts were problematic too as it revealed confounding effect. Yet, ANN-MLP model with 80% training and 20% training architecture provided valid forecasts of caesarean deliveries but whether it corrected the autocorrelation in the data was not known. Finally, best ARIMA model (0,1,1),(0,0,0) identified by the IBM SPSSS 20.0 software using 60-month of data

provided a valid forecasts of caesarean deliveries and it must be considered as the robust model for doing similar forecasts at Patan Hospital in the future. This means that caesarean deliveries will remain high at Patan Hospital and counseling for benefits of standard vaginal (normal) delivery must be done to all the couples visiting this hospital. This in turn will help to reduce caesarian delivery further in the future at Patan Hospital to achieve the caesarean section rate of less than 15% as recommended by the World Health Organization.

REFERENCES
1. Geidem AD, Audu BM, Kawuwa BM, Obed JY. Rising trend and indications of caesarean section at the University of Maiduguri Teaching Hospital, Nigeria. Ann Afr Med.2009; 8:127-32.
2. WHO World Health Organisation. Appropriate technology for birth. Lancet 1985;2(8452): 436-437.
3. WHO World Health Organisation. WHO Statement on caesarean section rates. WHO/RHR/15.02, 2015, Geneva, Switzerland:1-8.
4. MacKenzie IZ, Cooke I, Annan B. Indications for caesarean section in a consultant obstetric unit over three decades. J Obstet Gynaecol.2003 May;23(3):233-8.
5. Amatya A, Paudel R, Poudyal A, Wagle RR, Singh M, et al. (2013) Examining stratified cesarean section rates using Robson classification system at Tribhuvan University Teaching Hospital. J Nepal Health Res Counc 11(25): 255-258.
6. Laxmi T, Goma D, Kumariniraula H, Roshnitui T, Binod A. Rising Cesarean Section Rates in Nepal: Question of safety and Integrity on Obstetric Emergency Practice. J Gynecol Women's Health 2017; 7(4): 555716.
7. Pradhan B, Shrestha SD, RC L, Sharma P, Bhandary S. Increasing Trend of Caesarean Section in Patan Hospital. Journal of General Practice and Emergency Medicine of Nepal. 2015 December;6:3-5.
8. Nau R. Three types of forecasts: estimation, validation, and the future in simple forecasting models. Fuqua School of Business. 2019 December. [Available from: http://people.duke.edu/~rnau/three.htm]
9. Arsham H, Shao SP. Seasonal and Cycle Forecasting for Small Firm. American Journal of Small Business. 1985;9(4):36-57.
10. McKenzie SK. Concurrent Seasonal Adjustment with Census X-11. Journal of Business and Economic Statistics. 1984 July;2(3):235-248.

11. Evans JR, Basu A. Forecasting. Statistics, Data Analysis, and Decision Modelling (5e). 2013, Pearson Education, USA:238-271.

12. Bowerman BL, O'Connell RT, Murphree ES. Time Series Forecasting and Index. Business Statistics in Practice (7e), 2014, New York, USA: 630-679.

13. Frank KA. Impact of a confounding variable on a regression coefficient. Sociological Methods and Research 2000 November;29(2):147-194.

14. Coghlan A. A little book of R for Time Series: Release 0.2. 2017, Cambridge, UK: 45-66.

15. IBM Software Group. IBM SPSS Neural Networks 22 – Multilayer Perceptron. 2013. Chicago, USA: 5-13.

16. Nuepane HS, Shrestha CL, Upadhyay TP. Modelling Monthly International Tourist Arrivals and its Risk in Nepal. NRB Economic Review 2012;24(1):28-47.