

ARIMA modeling for sequential pharmaceutical assay monitoring across production lots: Raw and Box-Cox sensitivity analyses

Mostafa Essam Eissa

Independent Researcher and Freelance Consultant, Nasr City, Cairo, Egypt

ARTICLE INFO

Received: 29 June 2026

Revised: 26 July 2026

Accepted: 10 August 2026

eISSN 2224-7157/© 2026 The Author(s).
Published by Bangladesh Council of
Scientific and Industrial Research
(BCSIR).

This is an open access article under the
terms of the Creative Commons Non
Commercial License (CC BY-NC)
(<https://creativecommons.org/licenses/by-nc/4.0/>)

DOI: <https://doi.org/10.3329/bjsir.v61i3.91215>

Abstract

This preliminary short communication evaluates whether chronological production-lot assay results can support short-term next-lot forecasting using non-seasonal ARIMA models. A lot-level assay series was used to evaluate short-term forecasting of 9-deoxy-9alpha-aza-9alpha-methyl-9alpha-homoerythromycin A assay results. The dataset contained 199 valid chronological production-lot observations. The first 179 observations were used for model fitting, and the final 20 real assay results were held out for validation. Calendar aggregation was assessed but was not selected because monthly, quarterly, and semiannual series contained missing intervals, while annual aggregation yielded only 18 observations. The final analysis therefore used chronological production-lot order as the time index and targeted next-lot assay forecasting rather than equal-calendar stability prediction. Non-seasonal ARIMA models were compared using AICc, parameter significance, Ljung-Box residual diagnostics, model parsimony, and holdout validation. Raw ARIMA(0,1,1) with one regular difference and a constant was selected as the primary model after screening using statistical software. Box-Cox transformation with $\lambda = 2$ produced the same ARIMA(0,1,1) structure but did not improve validation accuracy. The primary raw model gave ME = 0.75, RMSE = 1.01, MAE = 0.80, and MAPE = 0.79% on the 20-result validation set. The raw model was therefore retained as the most interpretable and scientifically reasonable model for next-lot assay monitoring.

Keywords: ARIMA modeling; Assay forecasting; Pharmaceutical quality monitoring; Production-lot sequence; Box-Cox transformation; Validation metrics; Predictive bias

Introduction

Routine assay results can contain useful information about lot-to-lot analytical and process behavior. In pharmaceutical quality monitoring, assay methods are used not only to confirm compliance but also to characterize the consistency of bulk pharmaceutical materials (Misiuk, 2010). Scheduled stability studies and trend reviews remain important, but these approaches may not explicitly consider serial dependence among ordered assay observations. Time-series models can therefore provide a complementary approach when the aim is to forecast the next analytical result rather than to replace established stability protocols (Schaffer *et al.* 2021).

ARIMA models are designed for ordered data in which autocorrelation may influence future observations (Box *et al.* 2015; Brockwell and Davis, 2016). The present study does not aim to model true chemical stability over equal calendar intervals. Instead, it examines whether a chronological production-lot sequence can support short-term forecasting of the next assay result. This distinction is important because calendar-seasonal modeling requires complete and regularly spaced calendar observations, whereas the available manufacturer dataset was irregular by calendar month (Chatfield, 2000; Davidescu *et al.* 2021; Tayib *et al.* 2021).

*Corresponding author's e-mail: mostafaessameissa@yahoo.com

Therefore, the specific objective of this study was to determine whether a chronological production-lot assay sequence can provide a statistically adequate and practically interpretable ARIMA model for short-term next-lot assay forecasting. The study further examined whether Box-Cox transformation improved forecast performance compared with modeling the original assay percentage values. The primary evaluation criterion was not transformation-based model fit alone, but forecast accuracy against the final 20 real assay observations on the original assay scale.

Materials and methods

Data source and preprocessing

The raw dataset was obtained from an anonymized pharmaceutical manufacturing source and processed at the

individual lot level. Each 10-digit batch number was used only for chronological ordering. Assay values were accepted when a plausible assay percentage between 94.5% and 103.0% was found in the designated or shifted assay columns. Rows without a valid batch date or assay result were excluded and retained in the audit trail. No assay value was imputed, manually replaced, or removed as an outlier once it satisfied the pre-specified plausibility range. The general outcome summary of dataset processing is shown in Table I.

Time-basis selection and validation design

The cleaned observations were arranged in chronological production/inspection lot order. The first 179 observations were used for model fitting (Table II), while the final 20 observations were kept unchanged as an external holdout

Table I. Summary of the processing of the original dataset

Cleaning item	Result
Valid lot-level assay observations	199
Model-fitting observations	179
Validation observations	20 final real lot results
Time index	Chronological production/inspection lot order
Forecast target	Next production/inspection lot assay result
Missing calendar imputation	None
Excluded/unusable rows	9

Table II. Interval-basis selection screening for dataset

Candidate basis	Observed time points	Missing time points	Main limitation	Decision
Monthly calendar	82	127	Many blank months	Rejected
Quarterly calendar	51	19	Incomplete regular series	Rejected
Semiannual calendar	33	2	Incomplete regular series	Rejected
Annual calendar	18	0	Too small for 20-result validation	Rejected
Chronological lot sequence	199	0	Event-order spacing only	Selected

validation set. No validation observation was used during model fitting. Calendar-based aggregation was evaluated before ARIMA modeling; however, monthly, quarterly, and semiannual aggregation produced missing time intervals, and annual aggregation produced only 18 observations, which was not sufficient for a 20-result validation set. Therefore, chronological lot sequence was selected as the modeling time basis. This analysis should be interpreted as next-lot assay forecasting, not as equal-calendar stability modeling. Forecast evaluation was aligned with the prediction target and the available validation structure (Chatfield, 2000; Cerqueira *et al.* 2025). Prediction intervals were interpreted as uncertainty intervals and not as product acceptance limits (Hazra, 2017).

ARIMA model identification

Non-seasonal ARIMA models were evaluated using Minitab V21.2 statistical software (Minitab, LLC, 2022). Candidate differencing orders $d = 0$, $d = 1$, and $d = 2$ were assessed. For each differencing order, candidate p and q terms were searched up to 5. Model selection was based on corrected Akaike Information Criterion (AICc), AIC, BIC, parameter significance, residual diagnostics, model parsimony, and validation performance. Seasonal ARIMA terms were not included because the final time index represented production-lot order rather than equally spaced calendar time, and no predefined lot-based seasonal cycle was available (Box *et al.* 2015; Brockwell and Davis, 2016; Davidescu *et al.* 2021; Tayib *et al.* 2021).

Box-Cox transformation and validation metrics

Box-Cox transformation was assessed as a sensitivity analysis. The Box-Cox transformation was defined as $y(\lambda) = (y^\lambda - 1)/\lambda$ for λ not equal to 0, and $y(\lambda) = \log(y)$ for λ equal to 0. The primary Box-Cox analysis selected $\lambda = 2$ over the interval $[-1, 2]$, corresponding to Assay_pct squared. An expanded Box-Cox search over $\lambda = [-5, +5]$ was also performed as a sensitivity check and selected $\lambda = 5$ at the boundary. Because transformation changes the response scale, AICc values from raw and transformed models were not compared directly across scales. Instead, raw and transformed models were compared after forecasts were returned to the original assay percentage scale (Guerrero, 1993; Nwakuya and Nwabueze, 2018; Chatfield, 2000).

Forecast errors were calculated as actual assay minus forecast assay. Validation performance was summarized using mean error/bias (ME), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and sample error variance. Final model selection considered residual adequacy, model parsimony, interpretability on the original assay scale, and performance against the final 20 real assay observations.

Results and discussion

Differencing and ARIMA identification: Differencing order was determined within the documented Minitab V21.2 ARIMA workflow. Candidate models with $d = 0$, $d = 1$, and $d = 2$ were evaluated. Model choice was based on AICc, AIC,

Table III. Raw ARIMA model selection

Raw model set	Best model	LogLikelihood	AICc	AIC	BIC	Decision
$d = 0$	ARIMA(1,0,1)	-259.021	526.271	526.041	538.791	Close alternative Selected Rejected: higher AICc and complexity
$d = 1$	ARIMA(0,1,1)	-259.459	525.057	524.919	534.464	
$d = 2$	ARIMA(4,2,1)	-261.375	535.243	534.749	553.806	

Table IV. Best ARIMA(0,1,1) statistical parameter estimates

Parameter	Coefficient	SE Coef	T-value	P-value
MA(1)	0.6508	0.0567	11.48	0.000
Constant	0.0131	0.0272	0.48	0.631

BIC, parameter significance, residual diagnostics, parsimony, and validation performance. The final $d = 1$ model was retained because it gave the most favorable overall combination of fit, residual behavior, and holdout performance.

The raw assay models were first compared across $d = 0$, $d = 1$, and $d = 2$. Among the retained raw-data candidates, ARIMA(0,1,1) with one regular difference and a constant gave the lowest AICc and was selected for further validation (Tables III and IV). Seasonal terms were not included because the final time index was production-lot order, not equal calendar time, and no predefined lot-based seasonal cycle was available (Box *et al.* 2015; Brockwell and Davis, 2016; Davidescu *et al.* 2021; Tayib *et al.* 2021). For the raw assay scale, ARIMA(0,1,1) with one regular difference and a constant gave the lowest AICc among the retained raw-data candidates (Tables III and IV). The MA(1) coefficient was statistically significant, whereas the constant term was not significant. This supported a parsimonious non-seasonal moving-average structure for next-lot forecasting (Box *et al.* 2015).

Box-Cox transformation was assessed as a sensitivity analysis. The transformation was defined as: $y^{(\lambda)} = (y^\lambda - 1)/\lambda$ for $\lambda \neq 0$, and $y^{(\lambda)} = \ln(y)$ for $\lambda = 0$

The primary Box-Cox analysis selected $\lambda = 2$ over the interval $[-1, 2]$, corresponding to Assay_pct squared. Because transformation changes the response scale, AICc values from raw and transformed models were not compared directly across scales. Raw and transformed models were instead compared on the same 20 real validation observations after returning forecasts to the original assay percentage scale (Guerrero, 1993; Nwakuya and Nwabueze, 2018; Chatfield, 2000).

An expanded Box-Cox search over $\lambda = [-5, +5]$ was performed as a sensitivity check and selected $\lambda = 5$ at the boundary; however, the transformed models did not improve validation performance on the original assay scale and were not considered further in the main results. The Box-Cox $\lambda = 2$ sensitivity analysis selected the same ARIMA(0,1,1) structure as the raw analysis (Tables V and VI). The MA(1) term was significant, but the transformed model did not improve validation accuracy when forecasts were compared on the original assay scale. Therefore, the $\lambda = 2$ transformed model was retained as a sensitivity model rather than as the primary analysis.

Ljung-Box p-values were non-significant at all evaluated lags for the selected raw model and for the $\lambda = 2$ sensitivity model.

Table V. Box-Cox sensitivity results

Box-Cox model set*	Best model	LogLikelihood	AICc	AIC	BIC	Decision
$\lambda = 2$, $d = 0$	ARIMA(1,0,1)	-1204.00	2416.23	2416.00	2428.75	Not selected
$\lambda = 2$, $d = 1$	ARIMA(0,1,1)	-1199.17	2404.49	2404.35	2413.89	Sensitivity model
$\lambda = 2$, $d = 2$	ARIMA(4,2,1)	-1196.03	2404.55	2404.06	2423.12	Rejected: more complex

*The $\lambda = 2$ Box-Cox analysis selected the same ARIMA(0,1,1) structure as the raw analysis. Its MA(1) term was significant, but the transformed model did not improve validation accuracy when forecasts were compared on the original assay scale. The expanded $\lambda = [-5, +5]$ analysis selected $\lambda = 5$ at the search boundary, but this did not change the final model decision and was not adopted as the primary analysis.

Table VI. Box-Cox $\lambda = 2$, ARIMA(0,1,1) statistical parameter estimates

Box-Cox $\lambda = 2$ parameter	Coefficient	SE Coef	T-value	P-value
MA(1)	0.6502	0.0567	11.46	0.000
Constant	2.55	5.34	0.48	0.634

Table VII. Residual diagnostics

Lag	Raw ARIMA(0,1,1) p - value	Box-Cox lambda=2 ARIMA(0,1,1) p -value	Interpretation
12	0.545	0.575	No significant residual autocorrelation
24	0.649	0.672	No significant residual autocorrelation
36	0.732	0.750	No significant residual autocorrelation
48	0.918	0.925	No significant residual autocorrelation

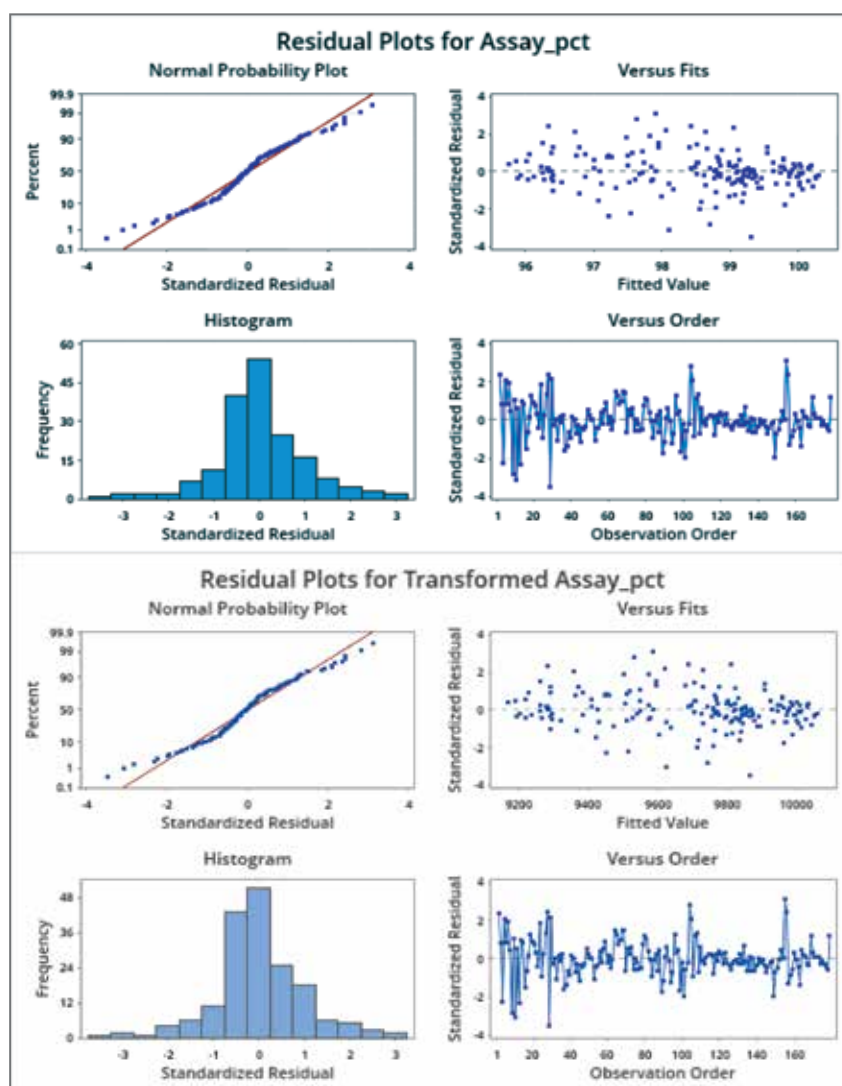


Fig. 1. Residual plots for untransformed and transformed assay% dataset of the API

Table VIII. Forecast validation

Time period	Actual assay	Raw forecast	Box-Cox forecast
180	100.5	99.3196	99.3162
181	99.5	99.3326	99.3284
182	99.4	99.3457	99.3406
183	99.2	99.3588	99.3528
184	99.8	99.3718	99.3650
185	100.8	99.3849	99.3772
186	101.0	99.3979	99.3894
187	99.2	99.4110	99.4016
188	99.7	99.4240	99.4137
189	99.6	99.4371	99.4259
190	99.5	99.4502	99.4381
191	101.2	99.4632	99.4503
192	101.1	99.4763	99.4625
193	100.2	99.4893	99.4746
194	99.4	99.5024	99.4868
195	100.8	99.5155	99.4990
196	100.4	99.5285	99.5112
197	100.8	99.5416	99.5233
198	100.6	99.5546	99.5355
199	101.2	99.5677	99.5477

Model	ME/Bias	RMSE	MAE	MAPE (%)	Sample error variance	Decision
Raw ARIMA(0,1,1)	0.7514	1.0067	0.7986	0.7929	0.4726	Primary model
Box-Cox lambda=2 ARIMA(0,1,1)	0.7630	1.0167	0.8071	0.8013	0.4753	Sensitivity model

Residual ACF/PACF and residual-vs-order plots from the Minitab V21.2 output were reviewed qualitatively (Minitab, LLC, 2022). Residual adequacy was assessed using software-generated Ljung-Box statistics (Table VII) and residual plots, while final model selection also considered parsimony, interpretability, and validation performance against the final 20 real assay observations. Fig.1 demonstrates residual plots for both models.

Errors were calculated as actual minus forecast. The raw model had slightly lower RMSE, MAE, MAPE, absolute bias, and error variance than the Box-Cox sensitivity model (Table VIII). These differences were small, but they favored the simpler and directly interpretable raw assay model. Forecast performance was therefore judged on validation against real holdout observations rather than on AICc values across different response scales (Chatfield, 2000; Cerqueira *et al.* 2025). Time-series graph in Fig. 2 depicts the forecasting of both models with 95% prediction interval. Fig.

2 shows time-series ARIMA with forecasting where actual assay values are shown in blue, forecasts in red, and 95% prediction intervals as upper and lower bounds. The x-axis represents chronological lot index.

This analysis narrows the interpretation of the study to the evidence actually supported by the data. The dataset is suitable for next-lot assay monitoring, but not for equal-calendar stability modeling. This distinction reduces overstatement and makes the modeling objective consistent with the available observations. The importance of model specification and time-index choice is well recognized in forecasting applications (Arslan *et al.* 2021; Schaffer *et al.* 2021), and assay monitoring remains a relevant pharmaceutical quality application (Misiuk, 2010).

The raw and lambda = 2 Box-Cox analyses both selected an ARIMA(0,1,1) structure after first differencing. This is consistent with ARIMA practice, where differencing is used

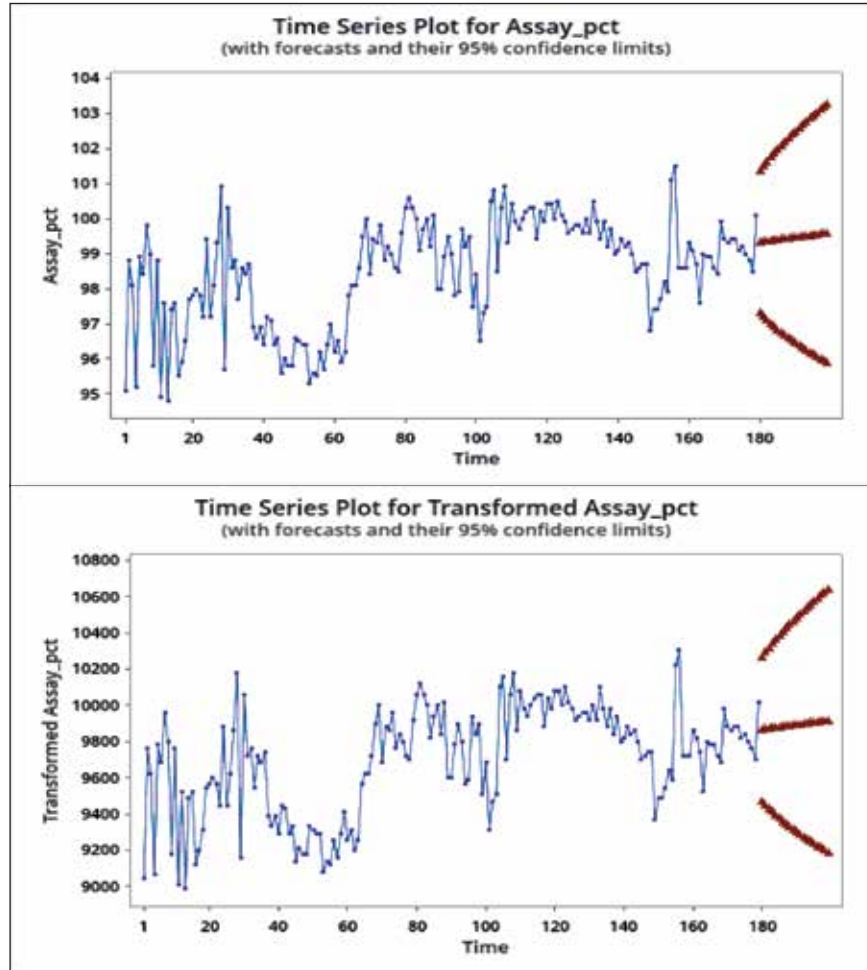


Fig. 2. Time-series plot for ARIMA(0,1,1) on the assay percentage scale

to improve model adequacy and moving-average terms represent short-run innovation effects (Box *et al.* 2015; Brockwell and Davis, 2016). The Box-Cox transformation did not produce a practical validation advantage. Because AICc values are not directly comparable across different response scales, holdout performance on the original assay scale was used to guide the final choice (Chatfield, 2000; Guerrero, 1993).

The expanded Box-Cox search over $\lambda = [-5, +5]$ did not alter the conclusion. Selection of $\lambda = 5$ at the boundary did not provide a more useful forecasting approach. Transformation should be retained only when it improves diagnostics, interpretability, or validation performance, not merely because a mathematical transformation is available (Guerrero, 1993; Nwakuya and Nwabueze, 2018; Cerqueira *et al.* 2025).

The main limitation is the event-order time basis. Consecutive production or inspection lots were treated as sequential observations, but they do not represent equal calendar intervals. Therefore, the model forecasts the next production/inspection lot assay result and should not be interpreted as a true calendar-stability forecast. Calendar seasonality was not modeled because the available data did not provide a complete monthly or quarterly calendar series, and no lot-based seasonal mechanism was documented (Tayib *et al.* 2021). Future prospective validation with additional lots is required before routine operational use (Hazra, 2017).

Conclusion

The chronological production-lot assay sequence supported a statistically adequate short-term forecasting model for

next-lot assay monitoring. Among the evaluated raw ARIMA candidates, ARIMA(0,1,1) with one regular difference and a constant was retained as the primary model because it showed acceptable Ljung-Box residual diagnostics, a significant MA(1) term, direct interpretability on the assay percentage scale, and the best validation performance on the final 20 real assay observations. Box-Cox transformation with $\lambda = 2$ produced the same ARIMA structure but did not improve RMSE, MAE, MAPE, bias, or error variance on the original assay scale. Therefore, the raw ARIMA(0,1,1) model is the preferred model for this dataset. The model should be interpreted as a next-lot assay monitoring tool, not as an equal-calendar chemical stability model, and prospective validation with future lots is recommended before routine use.

References

- Arslan J, Benke KK, Samarasinghe G, Sowmya A, Guymier RH and Baird PN (2021), Model structure uncertainty in the characterization and growth of geographic atrophy, *Transl Vis Sci.* **10**(6): 2. [https:// doi.org/ 10.1167/tvst.10.6.2](https://doi.org/10.1167/tvst.10.6.2)
- Box GEP, Jenkins GM, Reinsel GC and Ljung GM (2015), Time Series Analysis: Forecasting and Control. 5th Ed: (John Wiley & Sons, Hoboken, New Jersey) p 720.
- Brockwell PJ and Davis RA (2016), Introduction to Time Series and Forecasting. 3rd Ed: (Springer International Publishing, Switzerland) p 425.
- Chatfield C (2000), Time-Series Forecasting. 1st Ed: (Chapman & Hall/CRC, Boca Raton) p 280.
- Cerqueira V, Roque L and Soares C (2025), Modelradar: aspect-based forecast evaluation, *Mach Learn.* **114**(10): 229. [https:// doi.org/10. 1007/ s10994-025-06877-z](https://doi.org/10.1007/s10994-025-06877-z)
- Davidescu AA, Apostu SA and Paul A (2021), Comparative analysis of different univariate forecasting methods in modelling and predicting the Romanian unemployment rate for the period 2021–2022, *Entropy.* **23**(3): 325. [https:// doi.org/ 10.3390/ e23030325](https://doi.org/10.3390/e23030325)
- Guerrero VM (1993), Time-series analysis supported by power transformations, *J Forecast.* **12**(1): 37–48. <https://doi.org/10.1002/for.3980120104>
- Hazra A (2017), Using the confidence interval confidently, *J. Thorac. Dis.* **9**(10): 4125. [https:// doi.org/ 10.21037/ jtd.2017.09.14](https://doi.org/10.21037/jtd.2017.09.14)
- Minitab, LLC (2022), Minitab Statistical Software. Minitab, LLC, State College, PA, USA. Available at: <https://www.minitab.com>
- Misiuk W (2010), The role of assay methods in characterizing the quality of bulk pharmaceuticals, *J. Pharm. Bioallied Sci.* **2**(2): 88–92. [https:// doi.org/ 10.4103/ 0975-7406.67007](https://doi.org/10.4103/0975-7406.67007)
- Nwakuya MT and Nwabueze JC (2018), Application of box-cox transformation as a corrective measure to heteroscedasticity using an economic data, *Am J Math Stat.* **8**(1): 8–12. [https:// doi.org/ 10.5923/ j.ajms. 20180801.02](https://doi.org/10.5923/j.ajms.20180801.02)
- Schaffer AL, Dobbins TA and Pearson SA (2021), Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions, *BMC Med Res Methodol.* **21**(1): 58. [https:// doi.org/ 10.1186/ s12874-021-01235-8](https://doi.org/10.1186/s12874-021-01235-8)
- Tayib SA, Nor SR and Norulashikin SM (2021), Forecasting on the crude palm oil production in Malaysia using SARIMA model, *J Phys: Conf Ser.* **1988**(1): 012106. <https://doi.org/10.1088/1742-6596/1988/1/012106>